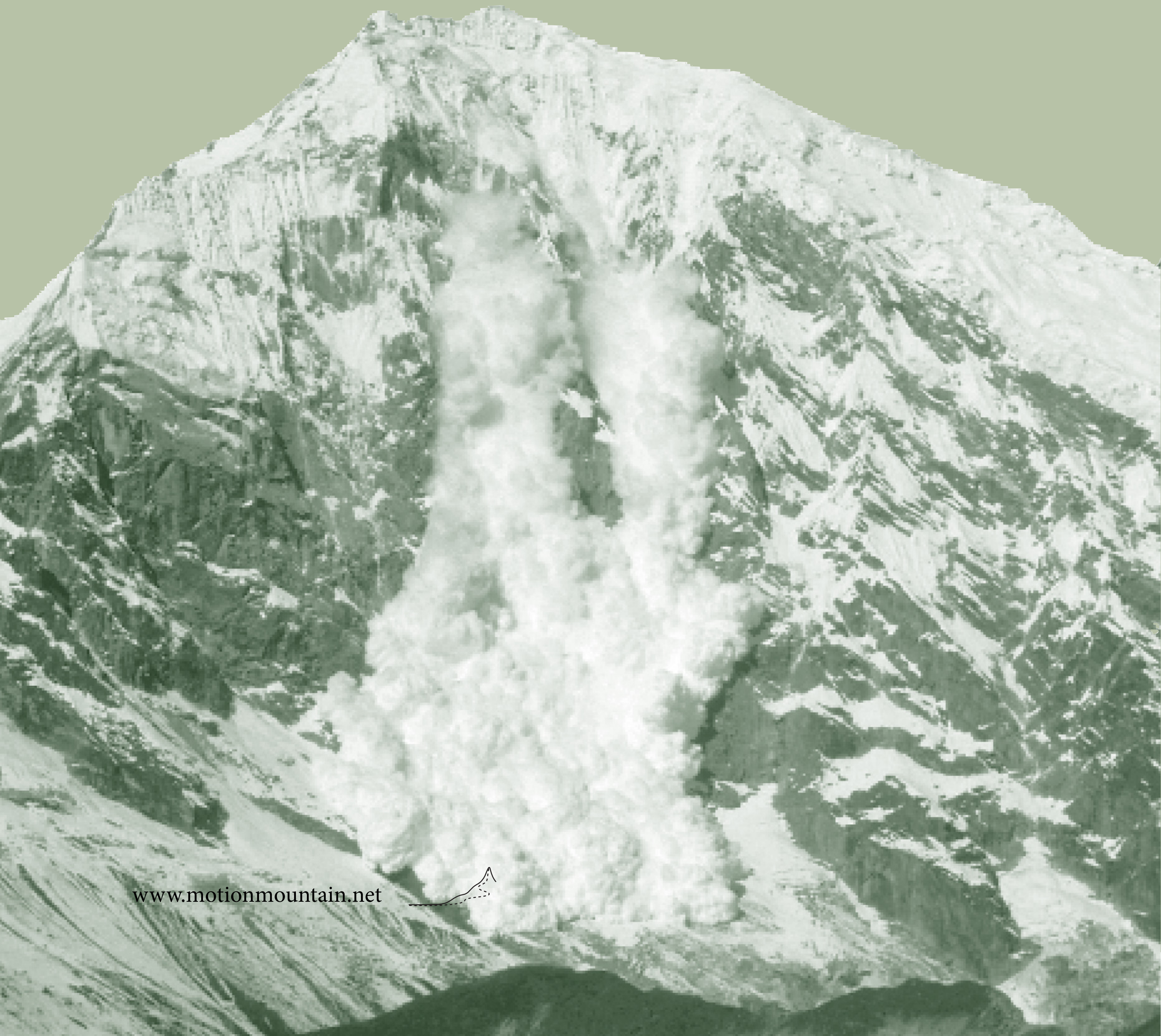
Christoph Schiller

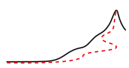
# MOTION MOUNTAIN

## THE ADVENTURE OF PHYSICS – VOL. II

## RELATIVITY



www.motionmountain.net



Christoph Schiller

# Motion Mountain

The Adventure of Physics
Volume II

## Relativity

Edition 24.14, available as free pdf at
www.motionmountain.net

Twenty-fourth edition, ISBN 978-300-021946-7.

To Britta, Esther and Justus Aaron

τῷ ἐμοὶ δαὶμονι

Die Menschen stärken, die Sachen klären.

# PREFACE

> Primum movere, deinde docere.*
>
> Antiquity

This book is written for anybody who is curious about nature and motion. Curiosity about how people, animals, things, images and empty space move leads to many adventures. This volume presents the best of them in the domain of relativity and cosmology.

*Special* relativity is the exploration of the energy speed limit $c$; *general* relativity is the exploration of the force limit $c^4/4G$. The text shows that in both domains, all equations follow from these two limit values. This simple, intuitive and unusual way of learning relativity and cosmology should reward the curiosity of every reader – whether student or professional. In the structure of modern physics, shown in Figure 1, special and general relativity form two important building blocks.

The present volume is the second of a six-volume overview of physics that arose from a threefold aim that I have pursued since 1990: to present motion in a way that is simple, up to date and captivating.

In order to be *simple*, the text focuses on concepts, while keeping mathematics to the necessary minimum. Understanding the concepts of physics is given precedence over using formulae in calculations. The whole text is within the reach of an undergraduate.

In order to be *up to date*, the text is enriched by the many gems – both theoretical and empirical – that are scattered throughout the scientific literature.

In order to be *captivating*, the text tries to startle the reader as much as possible. Reading a book on general physics should be like going to a magic show. We watch, we are astonished, we do not believe our eyes, we think, and finally we understand the trick. When we look at nature, we often have the same experience. Indeed, every page presents at least one surprise or provocation for the reader to think about. Numerous interesting challenges are proposed.

The motto of the text, *die Menschen stärken, die Sachen klären*, a famous statement by Hartmut von Hentig on pedagogy, translates as: 'To fortify people, to clarify things.' Clarifying things requires courage, as changing habits of thought produces fear, often hidden by anger. But by overcoming our fears we grow in strength. And we experience intense and beautiful emotions. All great adventures in life allow this, and exploring motion is one of them.

Munich, 1 January 2011.

* 'First move, then teach.' In modern languages, the mentioned type of *moving* (the heart) is called *motivating*; both terms go back to the same Latin root.

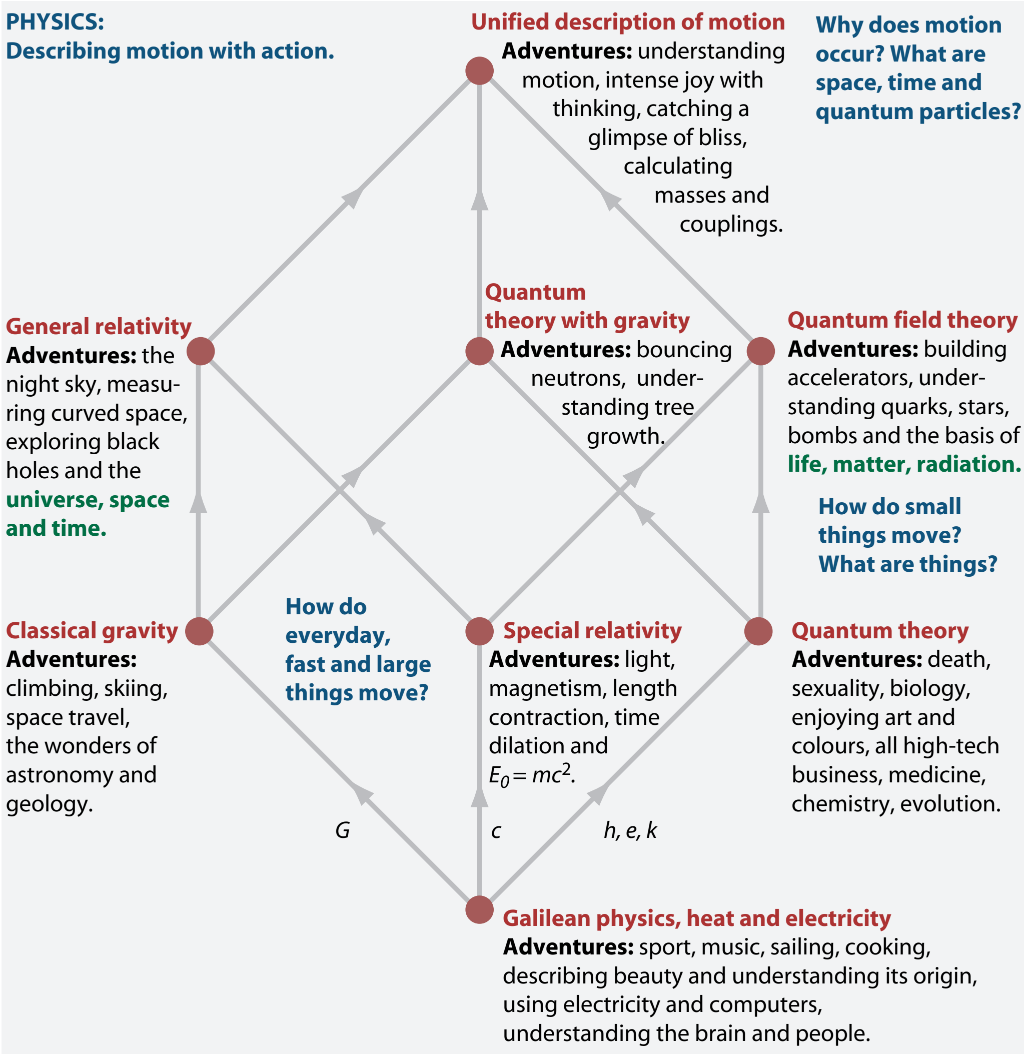



FIGURE 1 A complete map of physics: the connections are defined by the speed of light $c$, the gravitational constant $G$, the Planck constant $h$, the Boltzmann constant $k$ and the elementary charge $e$.

## Advice for learners

In my experience as a teacher, there was one learning method that never failed to transform unsuccessful pupils into successful ones: if you read a book for study, summarize every section you read, *in your own words, aloud*. If you are unable to do so, read the section again. Repeat this until you can clearly summarize what you read in your own words, aloud. You can do this alone in a room, or with friends, or while walking. If you do this with everything you read, you will reduce your learning and reading time significantly. In addition, you will enjoy learning from good texts much more and hate bad texts much less. Masters of the method can use it even while listening to a lecture, in a low voice, thus avoiding to ever take notes.

## Using this book

Text in green, as found in many marginal notes, marks a link that can be clicked in a pdf reader. Such green links are either bibliographic references, footnotes, cross references to other pages, challenge solutions, or pointers to websites.

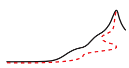
Solutions and hints for *challenges* are given in the appendix. Challenges are classified as research level (r), difficult (d), standard student level (s) and easy (e). Challenges of type r, d or s for which no solution has yet been included in the book are marked (ny).

## A request

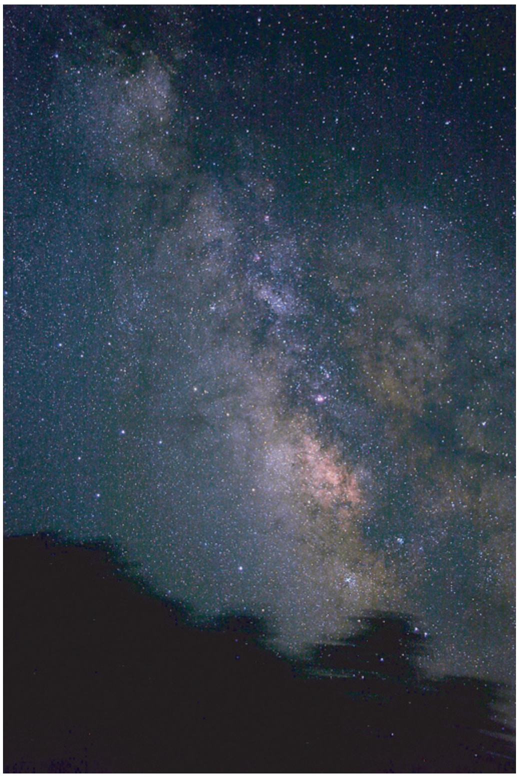
The text is and will remain free to download from the internet. In exchange, I would be delighted to receive an email from you at fb@motionmountain.net, especially on the following issues:

Challenge 1 s
— What was unclear and should be improved?
— What story, topic, riddle, picture or movie did you miss?
— What should be corrected?

Alternatively, you can provide feedback online, on www.motionmountain.net/wiki. The feedback will be used to improve the next edition. On behalf of all readers, thank you in advance for your input. For a particularly useful contribution you will be mentioned – if you want – in the acknowledgements, receive a reward, or both. But above all, enjoy the reading!



# Contents



# Relativity

In our quest to learn how things move,
the experience of hiking and other motion
leads us to discover that there is a maximum speed in nature,
and that two events that happen at the same time for one observer
may not for another.
We discover that empty space can bend, wobble and move,
we find that there is a maximum force in nature,
and we understand why we can see the stars.

CHAPTER 1

# MAXIMUM SPEED, OBSERVERS AT REST, AND MOTION OF LIGHT

> Fama nihil est celerius.*
>
> Antiquity

LIGHT is indispensable for a precise description of motion. To check whether a line or a path of motion is straight, we must look along it. In other words, we use light to define straightness. How do we decide whether a plane is flat? We look across it,** again using light. How do we observe motion? With light. How do we measure length to high precision? With light. How do we measure time to high precision? With light: once it was light from the Sun that was used; nowadays it is light from caesium atoms. Page 266

Light is important because it is the standard for *undisturbed motion*. Physics would have evolved much more rapidly if, at some earlier time, light propagation had been recognized as the ideal example of motion.

But is light really a phenomenon of motion? Yes. This was already known in ancient Greece, from a simple daily phenomenon, the *shadow*. Shadows prove that light is a moving entity, emanating from the light source, and moving in straight lines.*** The Greek Ref. 1 thinker Empedocles (*c.* 490 to *c.* 430 BCE) drew the logical conclusion that light takes a certain amount of time to travel from the source to the surface showing the shadow.

Empedocles thus stated that the speed of light is finite. We can confirm this result with a different, equally simple, but subtle argument. Speed can be measured. And measurement is comparison with a standard. Therefore the *perfect* speed, which is used as the implicit measurement standard, must have a finite value. An infinite velocity standard

---

* 'Nothing is faster than rumour.' This common sentence is a simplified version of Virgil's phrase: *fama, malum qua non aliud velocius ullum.* 'Rumour, the evil faster than all.' From the *Aeneid*, book IV, verses 173 and 174.

** Note that looking along the plane from all sides is not sufficient for this check: a surface that a light beam touches right along its length in *all* directions does not need to be flat. Can you give an example? One needs other methods to check flatness with light. Can you specify one? Challenge 2 s

*** Whenever a source produces shadows, the emitted entities are called *rays* or *radiation*. Apart from light, other examples of radiation discovered through shadows were *infrared rays* and *ultraviolet rays*, which emanate from most light sources together with visible light, and *cathode rays*, which were found to be to the motion of a new particle, the *electron*. Shadows also led to the discovery of *X-rays*, which again turned out to be a version of light, with high frequency. *Channel rays* were also discovered via their shadows; they turn out to be travelling ionized atoms. The three types of radioactivity, namely *$\alpha$-rays* (helium nuclei), *$\beta$-rays* (again electrons), and *$\gamma$-rays* (high-energy X-rays) also produce shadows. All these discoveries were made between 1890 and 1910: those were the 'ray days' of physics.

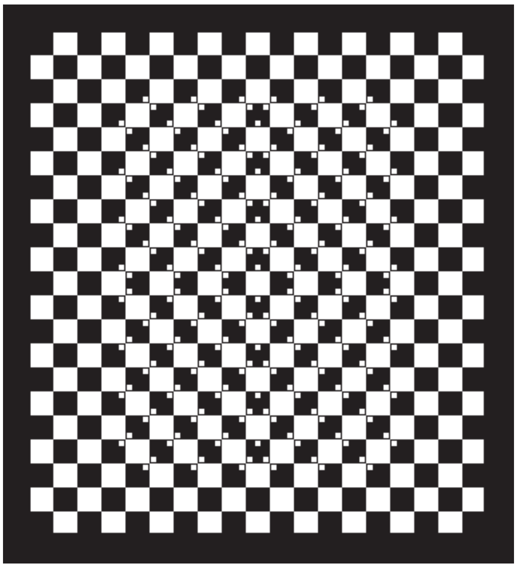


FIGURE 2 How do you check whether the lines are curved or straight?

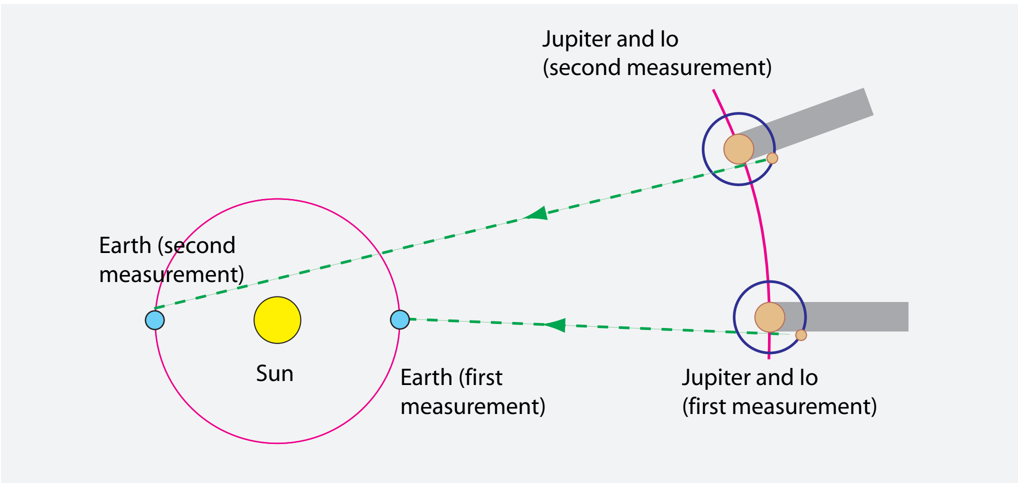



FIGURE 3 Rømer's method of measuring the speed of light

Challenge 3 s

would not allow measurements at all. In nature, lighter entities tend to move with higher speed. Light, which is indeed extremely light, is an obvious candidate for motion with perfect but finite speed. We will confirm this in a minute.

A finite speed of light means that whatever we see is a message from the *past*. When we see the stars,* the Sun or a person we love, we always see an image of the past. In a sense, nature prevents us from enjoying the present – we must therefore learn to enjoy the past.

* The photograph of the night sky and the milky way, on page 14 is copyright Anthony Ayiomamitis and is found on his splendid website www.perseus.gr.

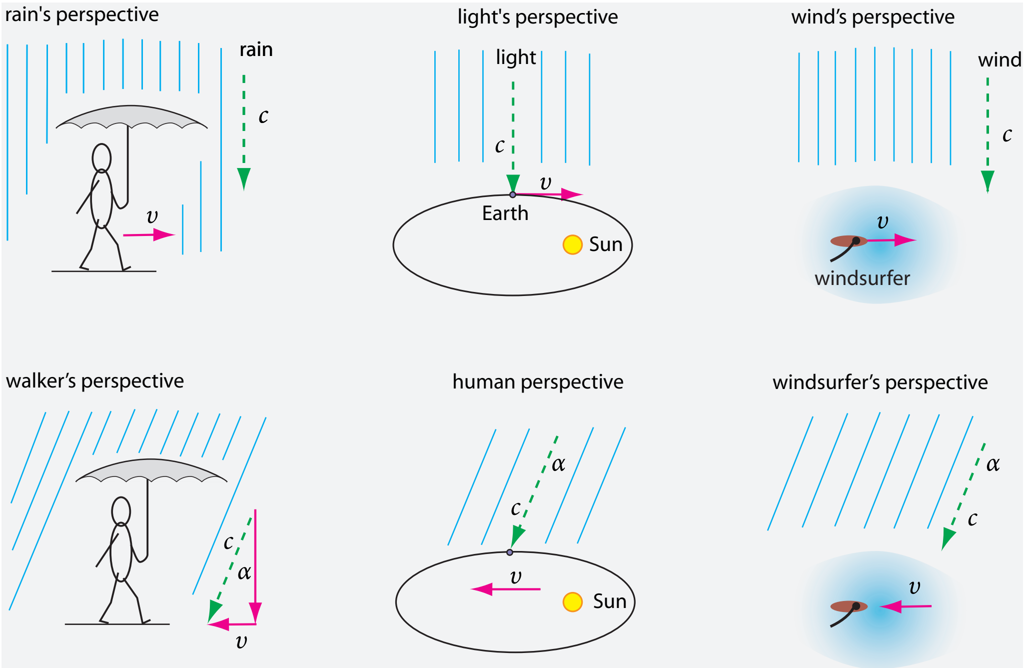



FIGURE 4 The rainwalker's or windsurfer's method of measuring the speed of light

The speed of light is high; therefore it was not measured until the years 1668 to 1676, even though many, including Galileo, had tried to do so earlier. The first measurement method was worked out and published by the Danish astronomer Ole Rømer* when he was studying the orbits of Io and the other Galilean satellites of Jupiter. He did not obtain any specific value for the speed of light because he had no reliable value for the satellite's distance from Earth and because his timing measurements were imprecise. The lack of a numerical result was quickly corrected by his peers, mainly Christiaan Huygens and Edmund Halley. (You might try to deduce Rømer's method from Figure 3.) Since Rømer's time it has been known that light takes a bit more than 8 minutes to travel from the Sun to the Earth. This result was confirmed in a beautiful way fifty years later, in 1726, by the astronomer James Bradley. Being English, Bradley thought of the 'rain method' to measure the speed of light.

Ref. 2 Page 170 Ref. 3 Challenge 4 s Page 127 Ref. 4

How can we measure the speed of falling rain? We walk rapidly with an umbrella, measure the angle $\alpha$ at which the rain appears to fall, and then measure our own velocity $v$. (We can clearly see the angle while walking if we look at the rain to our left or right, if possible against a dark background.) As shown in Figure 4, the speed $c$ of the rain is

* Ole (Olaf) Rømer (1644 Aarhus – 1710 Copenhagen), Danish astronomer. He was the teacher of the Dauphin in Paris, at the time of Louis XIV. The idea of measuring the speed of light in this way was due to the Italian astronomer Giovanni Cassini, whose assistant Rømer had been. Rømer continued his measurements until 1681, when Rømer had to leave France, like all protestants (such as Christiaan Huygens), so that his work was interrupted. Back in Denmark, a fire destroyed all his measurement notes. As a result, he was not able to continue improving the precision of his method. Later he became an important administrator and reformer of the Danish state.

then given by

$$c = v/\tan\alpha\,. \tag{1}$$

In the same way we can measure the speed of wind when on a surfboard or on a ship. The same measurement can be made for light. Figure 4 shows that we just need to measure the angle between the motion of the Earth and the light coming from a star above Earth's orbit. Because the Earth is moving relative to the Sun and thus to the star, the angle is not 90°. This deviation is called the *aberration* of light; the aberration is determined most easily by comparing measurements made six months apart. The value of the aberration angle is $20.5''$. (Nowadays it can be measured with a precision of five decimal digits.) Given that the speed of the Earth around the Sun is $v = 2\pi R/T = 29.7\,\mathrm{km/s}$, the speed of light must therefore be $c = 0.300\,\mathrm{Gm/s}$.* This is an astonishing value, especially when compared with the highest speed ever achieved by a man-made object, namely the Voyager satellites, which travel away from us at 52 Mm/h = 14 km/s, with the growth of children, about 3 nm/s, or with the growth of stalagmites in caves, about 0.3 pm/s. We begin to realize why measurement of the speed of light is a science in its own right.

The first *precise* measurement of the speed of light was made in 1849 by the French physicist Hippolyte Fizeau (1819–1896). His value was only 5 % greater than the modern one. He sent a beam of light towards a distant mirror and measured the time the light took to come back. How did Fizeau measure the time without any electric device? In fact, Page 56 he used the same ideas that are used to measure bullet speeds; part of the answer is given Challenge 9 s in Figure 5. (How far away does the mirror have to be?) A modern reconstruction of Ref. 7 his experiment by Jan Frercks has achieved a precision of 2 %. Today, the experiment is

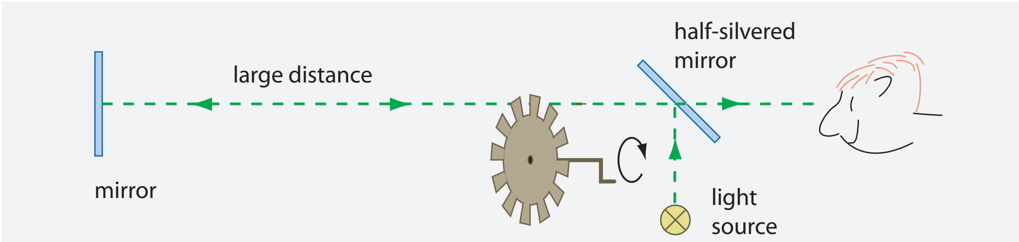
* Umbrellas were not common in Britain in 1726; they became fashionable later, after being introduced from China. The umbrella part of the story is made up. In reality, Bradley had his idea while sailing on the Thames, when he noted that on a moving ship the apparent wind has a different direction from that on land. He had observed 50 stars for many years, notably Gamma Draconis, and during that time he had been puzzled by the *sign* of the aberration, which was *opposite* to the effect he was looking for, namely that of the star parallax. Both the parallax and the aberration for a star above the ecliptic make them describe a Challenge 5 s small ellipse in the course of an Earth year, though with different rotation senses. Can you see why?

By the way, it follows from the invariance of the speed of light that the formula (1) is wrong, and that the Challenge 6 s correct formula is $c = v/\sin\alpha$; can you see why?

To determine the speed of the Earth, we first have to determine its distance from the Sun. The simplest method is the one by the Greek thinker Aristarchus of Samos (*c.* 310 to *c.* 230 bce). We measure the angle between the Moon and the Sun at the moment when the Moon is precisely half full. The cosine of that angle gives the ratio between the distance to the Moon (determined, for example, by the methods of page 143) and Challenge 7 s the distance to the Sun. The explanation is left as a puzzle for the reader.

The angle in question is almost a right angle (which would yield an infinite distance), and good instru- Ref. 5 ments are needed to measure it with precision, as Hipparchus noted in an extensive discussion of the problem around 130 bce. Precise measurement of the angle became possible only in the late seventeenth century, when it was found to be 89.86°, giving a distance ratio of about 400. Today, thanks to radar measurements Page 275 of planets, the distance to the Sun is known with the incredible precision of 30 metres. Moon distance vari- Challenge 8 s ations can even be measured to the nearest centimetre; can you guess how this is achieved?

Ref. 6 Aristarchus also determined the radius of the Sun and of the Moon as multiples of those of the Earth. Aristarchus was a remarkable thinker: he was the first to propose the heliocentric system, and perhaps the first to propose that stars were other, faraway suns. For these ideas, several of his contemporaries proposed that he should be condemned to death for impiety. When the Polish monk and astronomer Nicolaus Copernicus (1473–1543) again proposed the heliocentric system two thousand years later, he did not mention Aristarchus, even though he got the idea from him.



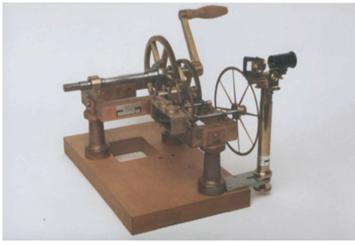
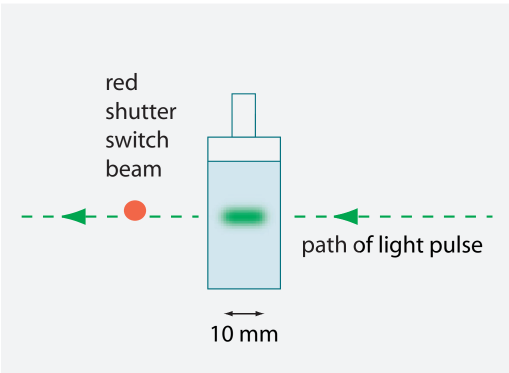



FIGURE 5 Fizeau's set-up to measure the speed of light (photo © AG Didaktik und Geschichte der Physik, Universität Oldenburg)



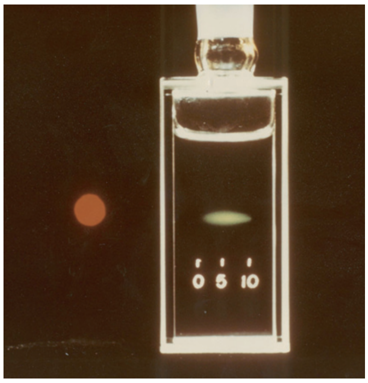




FIGURE 6 A photograph of a light pulse moving from right to left through a bottle with milky water, marked in millimetres (photograph © Tom Mattick)

much simpler; in the chapters on electrodynamics we will discover how to measure the speed of light using two standard UNIX or Linux computers connected by a cable, using the 'ping' command. Vol. III, page 30

The speed of light is so high that it is even difficult to prove that it is *finite*. Perhaps the most beautiful way to prove this is to photograph a light pulse flying across one's field of view, in the same way as one can photograph a car driving by or a bullet flying through the air. Figure 6 shows the first such photograph, produced in 1971 with a standard off-the-shelf reflex camera, a very fast shutter invented by the photographers, and, most noteworthy, not a single piece of electronic equipment. (How fast does such a shutter have to be? How would you build such a shutter? And how would you make sure it opened at the right instant?) Ref. 8 Challenge 10 s

A finite speed of light also implies that a rapidly rotating light beam bends, as shown as in Figure 7. In everyday life, the high speed of light and the slow rotation of lighthouses

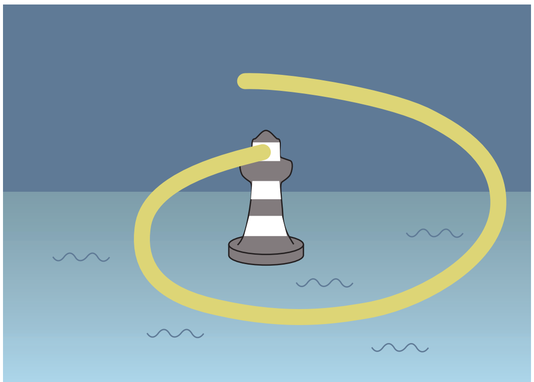


FIGURE 7 A consequence of the finiteness of the speed of light. Watch out for the tricky details – light *does* travel straight from the source, it does *not* move along the drawn curved line; the same occurs for water emitted by a rotating water sprinkler.

TABLE 1 Properties of the motion of light

| Observations about light |
| --- |
| Light can move through vacuum. |
| Light transports energy. |
| Light has momentum: it can hit bodies. |
| Light has angular momentum: it can rotate bodies. |
| Light moves across other light undisturbed. |
| Light in vacuum always moves faster than any material body does. |
| The speed of light, its true signal speed, is the forerunner speed. Page 111 |
| In vacuum, the speed of light is 299 792 458 m/s (or roughly 30 cm/ns). |
| The proper speed of light is infinite. Page 43 |
| Shadows can move without any speed limit. |
| Light moves in a straight line when far from matter. |
| High-intensity light is a wave. |
| Light beams are approximations when the wavelength is neglected. |
| In matter, both the forerunner speed and the energy speed of light are lower than in vacuum. |
| In matter, the group velocity of light pulses can be zero, positive, negative or infinite. |

make the effect barely noticeable.

In short, light moves extremely rapidly. It is much faster than lightning, as you might like to check yourself. A century of increasingly precise measurements of the speed have culminated in the modern value

Challenge 11 s

$$c = 299\,792\,458\ \mathrm{m/s}. \tag{2}$$

In fact, this value has now been fixed *exactly*, by definition, and the metre has been defined in terms of $c$. An approximate value for $c$ is thus 0.3 Gm/s or 30 cm/ns. Table 1 gives a summary of what is known today about the motion of light. Two of the most surprising properties were discovered in the late nineteenth century. They form the basis of what is called the theory of special relativity.

Ref. 9

## Can one play tennis using a laser pulse as the ball and mirrors as rackets?

> Et nihil est celerius annis.*
> Ovid, *Metamorphoses*.

We all know that in order to throw a stone as far as possible, we run as we throw it; we know instinctively that in that case the stone's speed with respect to the ground is higher than if we do not run. However, to the initial astonishment of everybody, experiments show that light emitted from a moving lamp has the same speed as light emitted from a resting one. The simplest way to prove this is to look at the sky. The sky shows many examples of *double stars*: these are two stars that rotate around each other along ellipses. In some of these systems, we see the ellipses (almost) edge-on, so that each star periodically moves towards and away from us. If the speed of light would vary with the speed of the source, we would see bizarre effects, because the light emitted from some positions would catch up the light emitted from other positions. In particular, we would not be able to see the elliptical shape of the orbits. However, bizarre effects are not seen, and the ellipses are observed. Willem de Sitter gave this beautiful argument already in 1913; he confirmed the validity with a large number of double stars. Ref. 10

In other words, light (in vacuum) is never faster than light; all light beams have the same speed. Many specially designed experiments have confirmed this result to high precision. The speed of light can be measured with a precision of better than 1 m/s; but even for lamp speeds of more than 290 000 000 m/s no differences have been found. (Can you guess what lamps were used?) Ref. 11 Challenge 12 s

In everyday life, we also know that a stone arrives more rapidly if we run towards it than in the case that we stand still or even run away from it. But astonishingly again, for light no such effect exists! All experiments clearly show that if we run towards a lamp, we measure the same speed of light as in the case that we stand still or even run away from it. Also these experiments have been performed to the highest precision possible. Ref. 12

All experiments thus show that the velocity of light has the *same value* for all observers, even if they are moving with respect to each other or with respect to the light source. The speed of light is indeed the ideal, perfect measurement standard.**

Ref. 15 There is also a second set of experimental evidence for the constancy, or better, the invariance of the speed of light. Every electromagnetic device, such as an electric vacuum cleaner, shows that the speed of light is *invariant*. We will discover that magnetic fields would not result from electric currents, as they do every day in every electric motor Vol. III, page 46

* 'Nothing is faster than the years.' Book X, verse 520.

** An equivalent alternative term for the speed of light is 'radar speed' or 'radio speed'; we will see below why this is the case. Page 93

The speed of light is also not far from the speed of neutrinos. This was shown most spectacularly by the observation of a supernova in 1987, when the light flash and the neutrino pulse arrived on Earth only 12 seconds apart. (It is not known whether the difference is due to speed differences or to a different starting point of the two flashes.) What would be the first digit for which the two speed values could differ, knowing that the supernova was $1.7 \cdot 10^5$ light years away, and assuming the same starting point? Challenge 13 s

Experiments also show that the speed of light is the same in all directions of space, to at least 21 digits of precision. Other data, taken from gamma ray bursts, show that the speed of light is independent of frequency to at least 20 digits of precision. Ref. 13 Ref. 14


FIGURE 8 All devices based on electric motors prove that the speed of light is invariant (© Miele, EasyGlide)

FIGURE 9 Albert Einstein (1879–1955)

and in every loudspeaker, if the speed of light were not invariant. This was actually how the invariance was first deduced, by several researchers. Only *after* these results did the German–Swiss physicist Albert Einstein show that the invariance of the speed of light is also in agreement with the observed motion of bodies. We will check this agreement in
Ref. 16 this chapter. The connection between relativity and electric vacuum cleaners, as well as
Vol. III, page 46 other machines, will be explored in the chapters on electrodynamics.

The main connection between light and motion of bodies can be stated in a few words. If the speed of light were not invariant, observers would be able to move at the speed of light. Why? Since light is a wave, an observer moving at the same speed would as the wave would see a *standing* wave. However, electromagnetism forbids such a phenomenon. Therefore, observers cannot reach the speed of light. The speed of light is thus a limit speed. Observers and bodies thus always move *slower* than light. Therefore, light is also an invariant speed. In other words, tennis with light is not fun: the speed of light is always the same.

## Albert Einstein

Albert Einstein (b. 1879 Ulm, d. 1955 Princeton) was one of the greatest physicists and of the greatest thinkers ever. (The 's' in his name is pronounced 'sh'.) In 1905, he published

three important papers: one about Brownian motion, one about special relativity, and one about the idea of light quanta. The first paper showed definitely that matter is made of molecules and atoms; the second showed the invariance of the speed of light; and the third paper was one of the starting points of quantum theory. Each paper was worth a Nobel Prize, but he was awarded the prize only for the last one. Also in 1905, he proved the famous formula $E_0 = mc^2$ (published in early 1906), after a few others also had pro-
Page 68
posed it. Although Einstein was one of the founders of quantum theory, he later turned against it. His famous discussions with his friend Niels Bohr nevertheless helped to clarify the field in its most counter-intuitive aspects. He also explained the Einstein–de Haas effect which proves that magnetism is due to motion inside materials. After many other discoveries, in 1915 and 1916 he published his highest achievement: the general theory of
Page 122
relativity, one of the most beautiful and remarkable works of science.

Being Jewish and famous, Einstein was a favourite target of attacks and discrimination by the National Socialist movement; therefore, in 1933 he emigrated from Germany to the USA; since that time, he stopped contact with Germans, except for a few friends, among them Max Planck. Until his death, Einstein kept his Swiss passport. He was not only a
Ref. 17
great physicist, but also a great thinker; his collection of thoughts about topics outside physics are well worth reading. His family life was disastrous, and he made each of his family members unhappy.

Anyone interested in emulating Einstein should know first of all that he published *many* papers. He was ambitious and hard-working. Moreover, many of his papers were wrong; he would then correct them in subsequent papers, and then do so again. This happened so frequently that he made fun of himself about it. Einstein indeed realized the well-known definition of a genius as a person who makes the largest possible number of mistakes in the shortest possible time.

## An invariant limit speed and its consequences

Experiments and theory show that observers cannot reach the speed of light. Equivalently, no object can reach the speed of light. In other words, not only is light the *standard* of speed; it is also the *maximum* speed in nature. More precisely, the velocity $v$ of any physical system in nature (i.e., any localized mass or energy) is bound by

$$v \leqslant c \,. \tag{3}$$

This relation is the basis of special relativity; in fact, the complete theory of special relativity is contained in it.

An invariant limit speed is not as surprising at we might think. We need such an
Page 93
invariant in order be able to *measure* speeds. Nevertheless, an invariant maximum speed implies many fascinating results: it leads to observer-varying time and length intervals, to an intimate relation between mass and energy, and to the existence of event horizons, as we will see.

Already in 1895, Henri Poincaré* called the discussion of viewpoint invariance the *theory of relativity*, and the name was common in 1905. Einstein regretted that the theory

* Henri Poincaré (1854–1912), important French mathematician and physicist. Poincaré was one of the most productive men of his time, advancing relativity, quantum theory, and many parts of mathematics.

TABLE 2 How to convince yourself and others that there is a maximum speed $c$ in nature. Compare this table with the table about maximum force, on page 97 below, and with the table about a smallest action, on page 17 in volume IV.

| Issue | Test Method |
|---|---|
| The energy speed value $c$ is observer-invariant | check all observations |
| Local energy speed values $> c$ are not observed | check all observations |
| Observed speed values $> c$ are either non-local or not due to energy transport | check all observations |
| Local energy speed values $> c$ cannot be produced | check all attempts |
| Local energy speed values $> c$ cannot be imagined | solve all paradoxes |
| A maximum local energy speed value $c$ is consistent | 1 – check that all consequences, however weird, are confirmed by observation |
| | 2 – deduce the theory of special relativity from it and check it |

was called this way; he would have preferred the name 'Invarianztheorie' or 'theory of invariance', but was not able to change the name any more. Thus he called the description of motion *without* gravity the theory of *special* relativity, and the description of motion *with* gravity the theory of *general* relativity. Both fields are full of fascinating and counter-intuitive results.*

Ref. 18 Ref. 15

Can an invariant limit speed exist in nature? Table 2 shows that we need to explore three points to accept the idea. We need to show that first, no higher speed is *observed*, secondly, that no higher energy speed *can* ever be observed, and thirdly, that *all* consequences of the invariance of the speed of light, however weird they may be, apply to nature. In fact, this programme defines the theory of special relativity; thus it is all we do in the remaining of this chapter.

The invariance of the speed of light is in complete contrast with Galilean mechanics, which describes the behaviour of stones, and proves that Galilean mechanics is *wrong* at high velocities. At low velocities the Galilean description remains good, because the error is small. But if we want a description valid at *all* velocities, we have to discard Galilean mechanics. For example, when we play tennis, by hitting the ball in the right way, we can increase or decrease its speed. But with light this is impossible. Even if we mount a

* The most beautiful and simple introduction to relativity is still that given by Albert Einstein himself, for example in *Über die spezielle und allgemeine Relativitätstheorie*, Vieweg, 1917 and 1997, or in *The Meaning of Relativity*, Methuen, 1951. It has taken a century for books almost as beautiful to appear, such as the texts by Schwinger or by Taylor and Wheeler.

Ref. 19, Ref. 20

mirror on an aeroplane and reflect a light beam with it, the light still moves away with the same speed. All experiments confirm this weird behaviour of light.

If we accelerate a bus we are driving, the cars on the other side of the road pass by with higher and higher speeds. For light, experiment shows that this is *not* so: light always passes by with the *same* speed.* Light does not behave like cars or any other matter object. Again, all experiments confirm this weird behaviour.

Why exactly is the invariance of the speed of light almost unbelievable, even though the measurements show it unambiguously? Take two observers O and Ω (pronounced 'omega') moving with relative velocity $v$, such as two cars on opposite sides of the street. Imagine that at the moment they pass each other, a light flash is emitted by a lamp in O. The light flash moves through positions $x(t)$ for observer O and through positions $\xi(\tau)$ (pronounced 'xi of tau') for Ω. Since the speed of light is the same for both, we have

$$\frac{x}{t} = c = \frac{\xi}{\tau} \,. \tag{4}$$

However, in the situation described, we obviously have $x \neq \xi$. In other words, the invariance of the speed of light implies that $t \neq \tau$, i.e., that *time is different for observers moving*
Challenge 14 e
*relative to each other*. Time is thus not unique. This surprising result, which has been con-
Ref. 21
firmed by many experiments, was first stated clearly in 1905 by Albert Einstein. Though many others knew about the invariance of $c$, only the young Einstein had the courage to say that time is observer-dependent, and to explore and face the consequences. Let us do so as well.

One remark is in order. The speed of light is a limit speed. What is meant with this statement is that the speed of light *in vacuum* is a limit speed. Indeed, particles can move faster than the speed of light *in matter*, as long as they move slower than the speed of light *in vacuum*. This situation is regularly observed.

In solid or liquid matter, the speed of light is regularly two or three times lower than the speed of light in vacuum. For special materials, the speed of light can be even lower: in the centre of the Sun, the speed of light is estimated to be only around 10 km/year = 0.3 mm/s, and even in the laboratory, for some materials, the speed of light has been
Ref. 22, Ref. 23
found to be as low as 0.3 m/s.

Page 258
When an aeroplane moves faster than the speed of sound in air, it creates a cone-shaped shock wave behind it. When a charged particle moves faster that the speed of light in matter, it emits a cone of radiation, so-called *Vavilov–Čerenkov radiation*. Vavilov–Čerenkov radiation is regularly observed; for example, it is the cause of the blue glow of the water in nuclear reactors and it appears in transparent plastic crossed by fast particles, a connection used in detectors for accelerator experiments.

In this and the following chapters, when we use the term 'speed of light', we mean the speed of light in vacuum. In fact, the speed of light in air is smaller than that in vacuum only by a fraction of one per cent, so that in most cases, the difference between air and vacuum can be neglected.

* Indeed, even with the current measurement precision of $2 \cdot 10^{-13}$, we cannot discern any changes of the
Ref. 13
speed of light for different speeds of the observer.

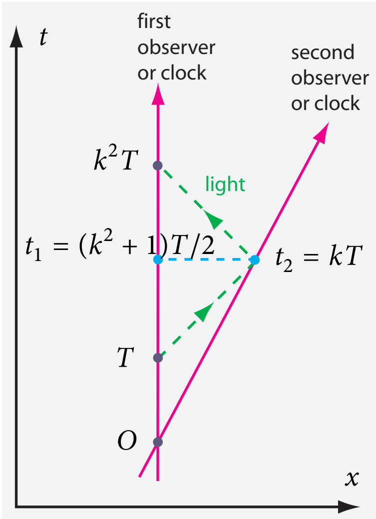



FIGURE 10 A drawing containing most of special relativity, including the expressions for time dilation and for the Lorentz transformation



## Special relativity with a few lines

The speed of light is invariant and constant for all observers. We can thus deduce all relations between what two different observers measure with the help of Figure 10. It shows two observers moving with constant speed against each other, drawn in space-time. The first is sending a light flash to the second, from where it is reflected back to the first. Since the speed of light is invariant, light is the only way to compare time and space coordinates for two distant observers. Also two distant clocks (like two distant metre bars) can only be compared, or synchronized, using light or radio flashes. Since light speed is invariant, all light paths in the same direction are parallel in such diagrams.

Ref. 24

A constant relative speed between two observers implies that a constant factor $k$ relates the time coordinates of events. (Why is the relation linear?) If a flash starts at a time $T$ as measured for the first observer, it arrives at the second at time $kT$, and then back again at the first at time $k^2T$. The drawing shows that

Challenge 15 s
Challenge 16 s

$$k = \sqrt{\frac{c+v}{c-v}} \quad \text{or} \quad \frac{v}{c} = \frac{k^2-1}{k^2+1} \; . \tag{5}$$

This factor will appear again in the Doppler effect.*

Page 28

Figure 10 also shows that the first observer measures a time $t_1$ for the event when the light is reflected; however, the second observer measures a different time $t_2$ for the *same* event. Time is indeed *different* for two observers in relative motion. Time is relative. Figure 11 illustrates the result.

The *time dilation factor* between the two observers is found from Figure 10 by com-

* The explanation of relativity using the factor $k$ is often called *k-calculus*.

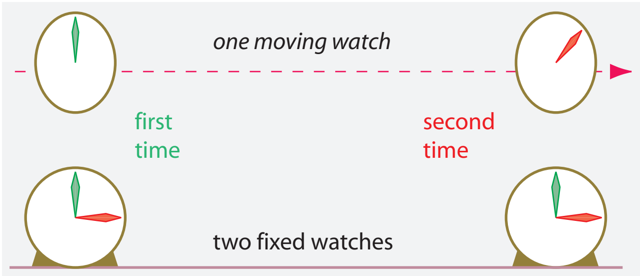



FIGURE 11 Moving clocks go slow: moving clocks mark time more slowly than do stationary clocks

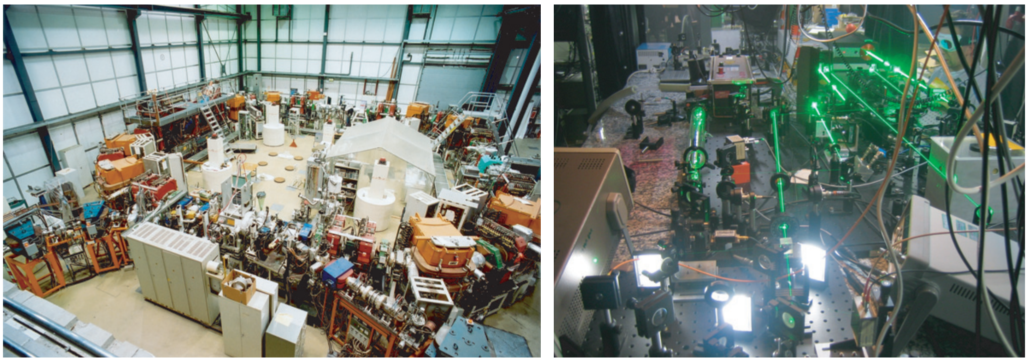


FIGURE 12 Moving clocks go slow: moving lithium atoms in a storage ring (left) read out with lasers (right) confirm the prediction to highest precision (© Max Planck Gesellschaft, TSR relativity team)

paring the values $t_1$ and $t_2$; it is given by

$$\frac{t_1}{t_2} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \gamma(v) \ . \tag{6}$$

Time intervals for a moving observer are *shorter* by this factor $\gamma$; the time dilation factor is always larger than 1. In other words, *moving clocks go slower*. For everyday speeds the effect is tiny. That is why we do not detect time differences in everyday life. Nevertheless, Galilean physics is not correct for speeds near that of light; the correct expression (6) has been tested to a precision better than one part in 10 million, with an experiment shown in Figure 12. The same factor $\gamma$ also appears in the formula $E = \gamma mc^2$ for the equivalence of mass and energy, which we will deduce below. Expressions (5) or (6) are the only pieces of mathematics needed in special relativity: all other results derive from it.

Challenge 17 e

Ref. 25

If a light flash is sent forward starting from the second observer to the first and reflected back, the second observer will make a similar statement: for him, the first clock is moving, and also for him, the moving clock marks time more slowly. *Each of the observers observes that the other clock marks time more slowly.* The situation is similar to that of two men comparing the number of steps between two identical ladders that are not parallel. A man on either ladder will always observe that the steps of the *other* ladder are shorter, as shown in Figure 13. There is nothing deeper than this observation at the basis of time dilation and length contraction.

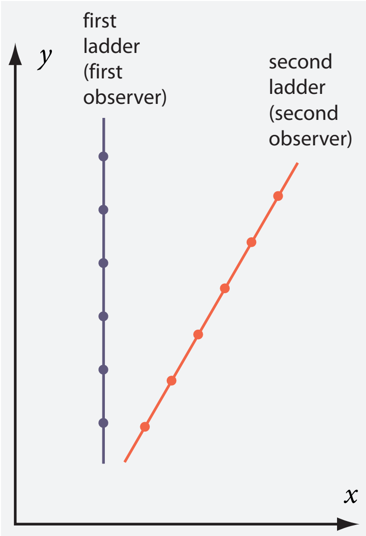



FIGURE 13 A man on each ladder claims that the other ladder is shorter

Naturally, many people have tried to find arguments to avoid the strange conclusion that time differs from observer to observer. But none have succeeded, and all experimental results confirm that conclusion: time is indeed relative. Let us have a look at some of the experiments.

## Acceleration of light and the Doppler effect

Can light in vacuum be accelerated? It depends what you mean. Most physicist are snobbish and say that every mirror accelerates light, because it changes its direction. We will see in the chapter on electromagnetism that matter also has the power to *bend* light, and Vol. III, page 123 thus to accelerate it. However, it will turn out that all these methods only change the *direction* of propagation; none has the power to change the *speed* of light in a vacuum. In particular, light is an example of a motion that cannot be stopped. There are only a few Challenge 18 s other such examples. Can you name one?

What would happen if we could accelerate light to higher speeds? For this to be possible, light would have to be made of massive particles. If light had mass, it would be necessary to distinguish the 'massless energy speed' $c$ from the speed of light $c_L$, which would be lower and would depend on the kinetic energy of those massive light particles. The speed of light would not be invariant, but the massless energy speed would still be so. Massive light particles could be captured, stopped and stored in a box. Such boxes would make electric illumination unnecessary; it would be sufficient to store some daylight in them and release the light, slowly, during the following night, maybe after giving it a push to speed it up.*

Physicists have tested the possibility of massive light in quite some detail. Observa- Ref. 26, Ref. 12 tions now put any possible mass of light (particles) at less than $1.3 \cdot 10^{-52}$ kg from terres-

* Incidentally, massive light would also have *longitudinal* polarization modes. This is in contrast to observations, which show that light is polarized exclusively *transversally* to the propagation direction.

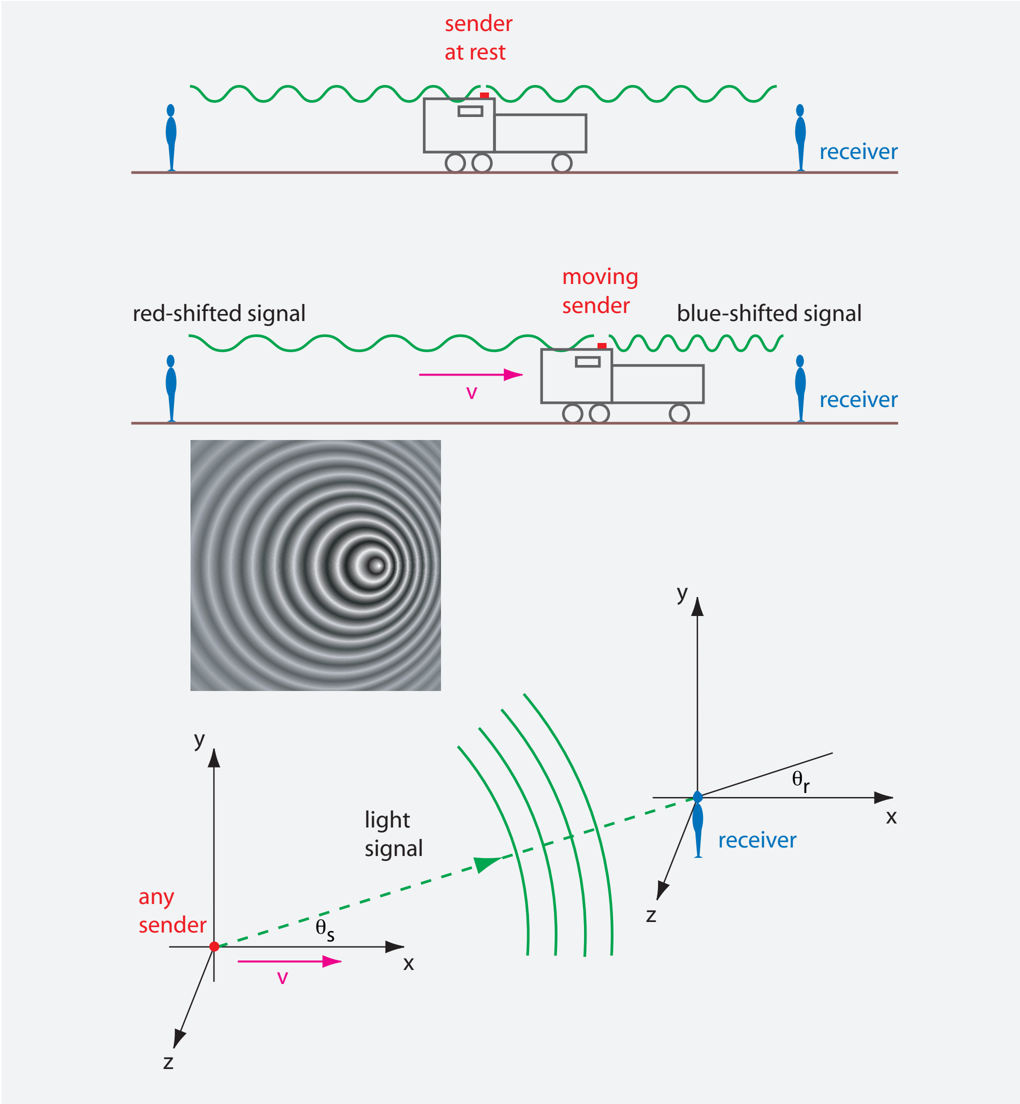



FIGURE 14 The set-up for the observation of the Doppler effect in one and three dimensions: waves emitted by an approaching source arrive with higher frequency and shorter wavelength, in contrast to waves emitted by a departing source (shadow waves courtesy Pbroks13/Wikimedia)

trial experiments, and at less than $4 \cdot 10^{-62}$ kg from astrophysical arguments (which are slightly less compelling). In other words, light is not heavy, light is light.

But what happens when light hits a *moving* mirror? The situation is akin to that of a light source moving with respect to the receiver: the receiver will observe a *different colour* from that observed by the sender. This frequency shift is called the *Doppler effect*. Christian Doppler* was the first to study the frequency shift in the case of sound waves. We all know the change in whistle tone between approaching and departing trains: that

* Christian Andreas Doppler (b. 1803 Salzburg, d. 1853 Venezia), Austrian physicist. Doppler studied the effect named after him for sound and light. Already in 1842 he predicted (correctly) that one day we would

is the Doppler effect for sound. We can determine the speed of the train in this way. Bats, dolphins, and wales use the acoustical Doppler effect to measure the speed of prey, and it is used to measure blood flow and heart beat in ultrasound systems (despite being extremely loud to babies), as shown in Figure 15. Page 246

Doppler was also the first to extend the concept to the case of light waves. As we will see, light is (also) a wave, and its colour is determined by its frequency, or equivalently, by its wavelength $\lambda$. Like the tone change for moving trains, Doppler realized that a moving light source produces a colour at the receiver that is different from the colour at the source. Simple geometry, and the conservation of the number of maxima and minima, leads to the result Vol. III, page 93 Challenge 19 e

$$\frac{\lambda_r}{\lambda_s} = \frac{1}{\sqrt{1 - v^2/c^2}}(1 - \frac{v}{c}\cos\theta_r) = \gamma\,(1 - \frac{v}{c}\cos\theta_r)\ . \tag{7}$$

The variables $v$ and $\theta_r$ in this expression are defined in Figure 14. Light from an approaching source is thus blue-shifted, whereas light from a departing source is red-shifted.

The first observation of the Doppler effect for light was made by Johannes Stark* in 1905, who studied the light emitted by moving atoms. All subsequent experiments confirmed the calculated colour shift within measurement errors; the latest checks have found agreement to within two parts per million. Ref. 28

In contrast to sound waves, a colour change is also found when the motion is *transverse* to the light signal. Thus, a yellow rod in rapid motion across the field of view will have a blue leading edge and a red trailing edge prior to the closest approach to the observer. The colours result from a combination of the longitudinal (first-order) Doppler shift and the transverse (second-order) Doppler shift. At a particular angle $\theta_{unshifted}$ the colour will stay the same. (How does the wavelength change in the purely transverse case? What is the expression for $\theta_{unshifted}$ in terms of the speed $v$?) Challenge 20 s

The colour or frequency shift explored by Doppler is used in many applications. Almost all solid bodies are mirrors for radio waves. Many buildings have doors that open automatically when one approaches. A little sensor above the door detects the approaching person. It usually does this by measuring the Doppler effect of radio waves emitted by the sensor and reflected by the approaching person. (We will see later that radio waves and light are manifestations of the same phenomenon.) So the doors open whenever something moves towards them. Police radar also uses the Doppler effect, this time to measure the speed of cars.** Page 93

be able to use the effect to measure the motion of distant stars by looking at their colours. For his discovery of the effect – and despite its experimental confirmation in 1845 and 1846 – Doppler was expelled from the Imperial Academy of Science in 1852. His health degraded and he died shortly afterwards. Ref. 27

* Johannes Stark (1874–1957), discovered in 1905 the optical Doppler effect in channel rays, and in 1913 the splitting of spectral lines in electrical fields, nowadays called the *Stark effect*. For these two discoveries he received the 1919 Nobel Prize for physics. He left his professorship in 1922 and later turned into a full-blown National Socialist. A member of the NSDAP from 1930 onwards, he became known for aggressively criticizing other people's statements about nature purely for ideological reasons; he became rightly despised by the academic community all over the world.

** At what speed does a red traffic light appear green? Challenge 21 s

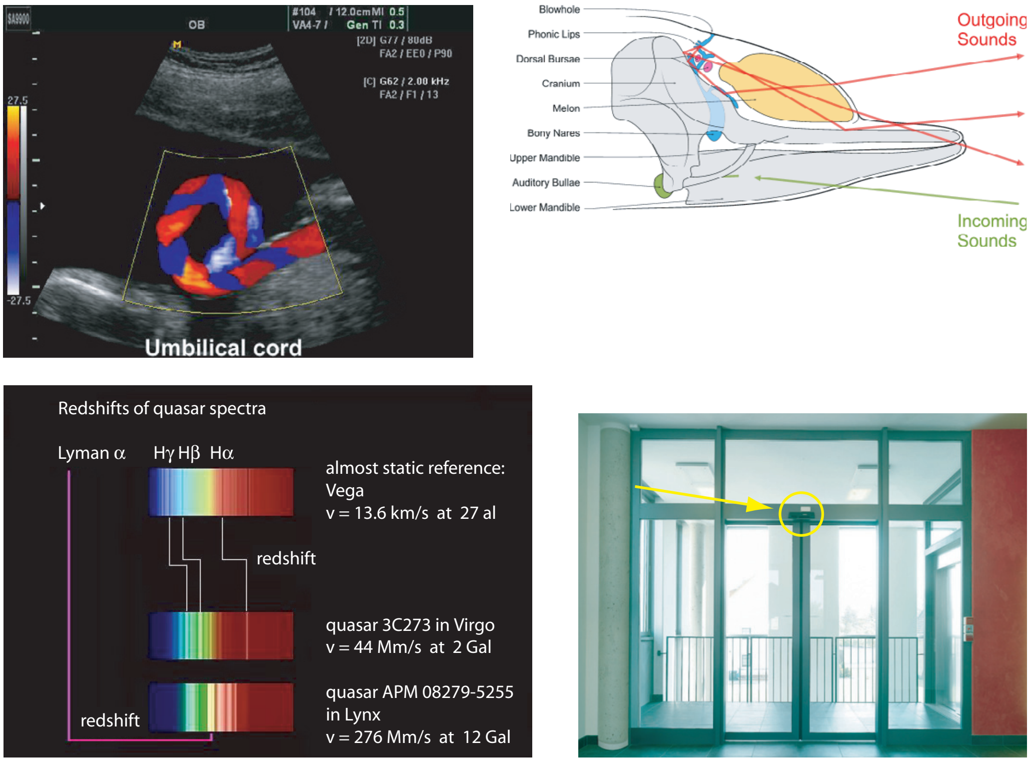



FIGURE 15 A Doppler sonography to detect blood flow (coloured) in a foetus, the Doppler sonar system of dolphins, the Doppler effect for light from two quasars and the Doppler effect system in a sliding door opener (© Medison, Wikimedia, Maurice Gavin, Hörmann AG)

The Doppler effect is regularly used to measure the speed of distant stars. In these cases, the Doppler shift is often characterized by the *red-shift number* $z$, defined with the help of wavelength $\lambda$ or frequency $f$ by

$$z = \frac{\Delta\lambda}{\lambda} = \frac{f_S}{f_R} - 1 = \sqrt{\frac{c+v}{c-v}} - 1 \, . \tag{8}$$

Challenge 22 s Can you imagine how the number $z$ is determined? Typical values for $z$ for light sources in the sky range from $-0.1$ to 3.5, but higher values, up to more than 10, have also been found. Can you determine the corresponding speeds? How can they be so high? Challenge 23 s

Because of the rotation of the Sun and the Doppler effect, one edge of the Sun is blue-shifted, and the other is red-shifted. It is possible to determine the rotation speed of the Sun in this way. (Ref. 29) The time of a rotation lies between 27 and 33 days, depending of the latitude. The Doppler effect also showed that the surface of the Sun oscillates with periods of the order of 5 minutes. Also the rotation of our galaxy was discovered using the Doppler effect of its stars; the Sun takes about 220 million years for a rotation around the centre of the galaxy.

In summary, whenever we try to change the *speed* of light, we only manage to change its *colour*. That is the Doppler effect. In short, acceleration of light leads to colour change.

Page 164 This connection leads to a puzzle: we know from classical physics that when light passes Challenge 24 s a large mass, such as a star, it is deflected. Does this deflection lead to a Doppler shift?

## The difference between light and sound

The Doppler effect for light is much more fundamental than the Doppler effect for sound. Even if the speed of light were not yet known to be invariant, the Doppler effect alone would *prove* that time is different for observers moving relative to each other. Why? Time is what we read from our watch. In order to determine whether another watch is synchronized with our own one, we *look* at both watches. In short, we need to use light signals Ref. 30 to synchronize clocks. Now, any change in the colour of light moving from one observer to another necessarily implies that their watches run differently, and thus that time is *different* for the two of them. To see this, note that also a light source is a clock – 'ticking' very rapidly. So if two observers see different colours from the same source, they measure different numbers of oscillations for the same clock. In other words, time is different for observers moving against each other. Indeed, equation (5) for the Doppler effect implies the whole of special relativity, including the invariance of the speed of light. (Can you confirm that the connection between observer-dependent frequencies and observer- Challenge 25 s dependent time breaks down in the case of the Doppler effect for *sound*?)

Why does the behaviour of light imply special relativity, while that of sound in air does not? The answer is that light is a limit for the motion of energy. Experience shows that there are supersonic aeroplanes, but there are no superluminal rockets. In other words, the limit $v \leqslant c$ is valid only if $c$ is the speed of light, not if $c$ is the speed of sound in air.

However, there is at least one system in nature where the speed of sound is indeed a limit speed for energy: the speed of sound is the limit speed for the motion of *disloca-* Page 221 *tions* in crystalline solids. (We discuss this in detail later on.) As a result, the theory of special relativity is also valid for dislocations, provided that the speed of light is replaced everywhere by the speed of sound! Indeed, dislocations obey the Lorentz transforma- Ref. 31 tions, show length contraction, and obey the famous energy formula $E = \gamma mc^2$. In all these effects the speed of sound $c$ plays the same role for dislocations as the speed of light plays for general physical systems.

Given special relativity is based on the statement that nothing can move faster than light, we need to check this statement carefully.

## Can one shoot faster than one's shadow?

> “Quid celerius umbra?*”
>
> Antiquity

For Lucky Luke to achieve the feat shown in Figure 16, his bullet has to move faster than Challenge 26 e the speed of light. (What about his hand?) In order to emulate Lucky Luke, we could take the largest practical amount of energy available, taking it directly from an electrical power station, and accelerate the lightest 'bullets' that can be handled, namely electrons. This experiment is carried out daily in particle accelerators such as the Large Electron Positron ring, the LEP, of 27 km circumference, located partly in France and partly in

* 'What is faster than the shadow?' A motto often found on sundials.

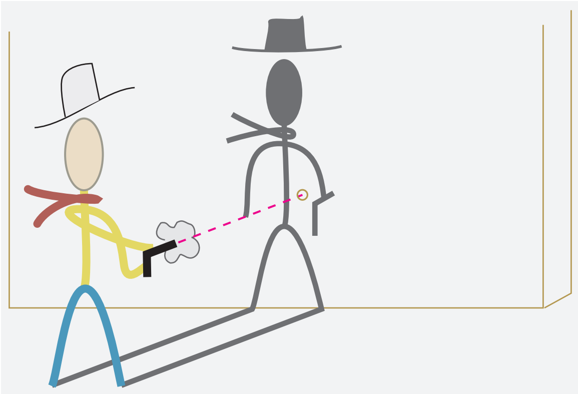


FIGURE 16 Lucky Luke

Switzerland, near Geneva. There, 40 MW of electrical power (the same amount used by a small city) were used to accelerate electrons and positrons to record energies of over 16 nJ (104.5 GeV) each, and their speed was measured. The result is shown in Figure 17: even with these impressive means it is impossible to make electrons move more rapidly than light. (Can you imagine a way to measure kinetic energy and speed separately?)

Ref. 32
Challenge 27 e

The speed–energy relation of Figure 17 is a consequence of the maximum speed, and its precise details are deduced below. These and many similar observations thus show that there is a *limit* to the velocity of objects and radiation. Bodies and radiation cannot move at velocities higher that the speed of light.* The accuracy of Galilean mechanics was taken for granted for more than two centuries, so that nobody ever thought of checking it; but when this was finally done, as in Figure 17, it was found to be wrong.

Page 62
Ref. 35

The same result appears when we consider momentum instead of energy. Particle accelerators show that momentum is *not* proportional to speed: at high speeds, doubling the momentum does *not* lead to a doubling of speed. In short, experiments show that neither increasing the energy nor increasing the momentum of even the lightest particles allows to reach the speed of light.

The people most unhappy with this speed limit are computer engineers: if the speed limit were higher, it would be possible to build faster microprocessors and thus faster computers; this would allow, for example, more rapid progress towards the construction of computers that understand and use language.

The existence of a limit speed runs counter to Galilean mechanics. In fact, it means that for velocities near that of light, say about 15 000 km/s or more, the expression $mv^2/2$ is *not* equal to the kinetic energy $T$ of the particle. In fact, such high speeds are rather common: many families have an example in their home. Just calculate the speed of electrons inside a television, given that the transformer inside produces 30 kV.

Challenge 28 s

* There are still people who refuse to accept this result, as well as the ensuing theory of relativity. Every reader should enjoy the experience, at least once in his life, of conversing with one of these men. (Strangely, no woman has yet been reported as belonging to this group of people. Despite this conspicuous effect, studying the influences of *sex* on physics is almost a complete waste of time.)

Crackpots can be found, for example, via the internet, in the sci.physics.relativity newsgroup. See also the www.crank.net website. Crackpots are a mildly fascinating lot, especially since they teach the importance of *precision* in language and in reasoning, which they all, without exception, neglect.

Ref. 33
Ref. 34

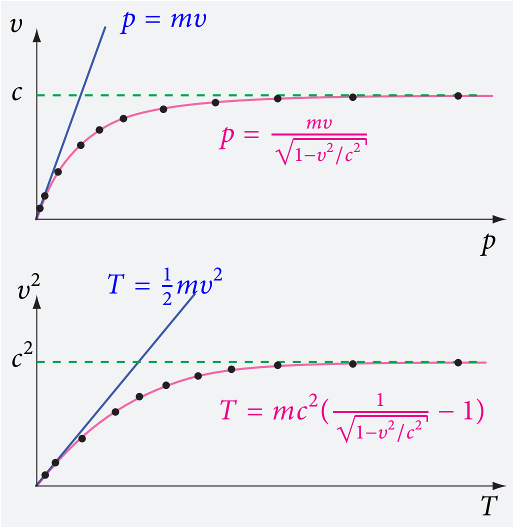



FIGURE 17 Experimental values (black dots) for the electron velocity $v$ as function of their kinetic energy $T$ and of their momentum $p$, compared with the prediction of Galilean physics (blue) and that of special relativity (red)



The speed of light is a *limit* speed for objects. This property is easily seen to be a consequence of its *invariance*. Bodies that can be at rest in one frame of reference obviously move more slowly than light in that frame. Now, if something moves more slowly than something else for *one* observer, it does so for all other observers as well. (Trying to imagine a world in which this would not be so is interesting: bizarre phenomena would occur, such as things interpenetrating each other.) Since the speed of light is the same for all observers, no object can move faster than light, for every observer.

Challenge 29 d

We conclude that the maximum speed is the speed of *massless* entities. Electromagnetic waves, including light, are the only known entities that can travel at the maximum speed. Gravitational waves are also predicted to achieve maximum speed, but this has not yet been observed. Though the speed of neutrinos cannot be distinguished experimentally from the maximum speed, recent experiments showed that they do have a tiny mass.

Ref. 36

Conversely, if a phenomenon exists whose speed is the limit speed for one observer, then this limit speed must necessarily be the same for all observers. Is the connection between limit property and observer invariance generally valid in nature?

Challenge 30 e
Challenge 31 r

## The composition of velocities

If the speed of light is a limit, no attempt to exceed it can succeed. This implies that when two velocities are composed, as when one throws a stone while running or travelling, the values cannot simply be added. Imagine a train that is travelling at velocity $v_{te}$ relative to the Earth, and a passenger throws a stone inside it, in the same direction, with velocity $v_{st}$ relative to the train. It is usually assumed as evident that the velocity of the stone relative to the Earth is given by $v_{se} = v_{st} + v_{te}$. In fact, both reasoning and measurement show a different result.

Page 24

The existence of a maximum speed, together with Figure 18, implies that the $k$-factors

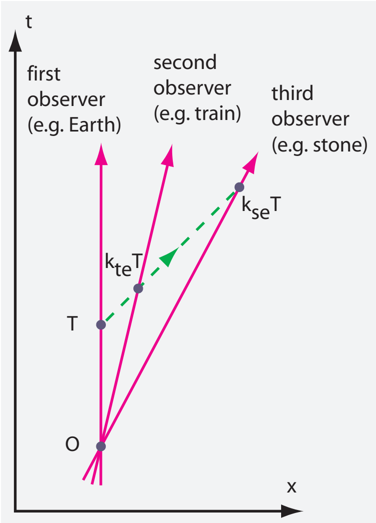



FIGURE 18 How to deduce the composition of velocities

must satisfy $k_{se} = k_{st}k_{te}$.* Then we only need to insert the relation (5) between each
Challenge 32 e $k$-factor and the respective speed to get

$$v_{se} = \frac{v_{st} + v_{te}}{1 + v_{st}v_{te}/c^2} . \tag{9}$$

Challenge 33 e This is called the *velocity composition formula*. The result is never larger than $c$ and is always smaller than the naive sum of the velocities.** Expression (9) has been confirmed
Page 60 by each of the millions of cases for which it has been checked. You may check that it
Ref. 12 simplifies with high precision to the naive sum for everyday life speed values.

## Observers and the principle of special relativity

Special relativity is built on a simple principle:

▷ *The local maximum speed of energy transport is the same for all observers.*

Ref. 38 Or, as Hendrik Lorentz*** liked to say, the equivalent:

* By taking the (natural) logarithm of this equation, one can define a quantity, the *rapidity*, that quantifies the speed and is additive.

Ref. 37 ** One can also deduce the Lorentz transformation directly from this expression.

*** Hendrik Antoon Lorentz (b. 1853 Arnhem, d. 1928 Haarlem) was, together with Boltzmann and Kelvin, one of the most important physicists of his time. He deduced the so-called Lorentz transformation and the Lorentz contraction from Maxwell's equations for the electromagnetic field. He was the first to understand, long before quantum theory confirmed the idea, that Maxwell's equations for the vacuum also describe matter and all its properties, as long as moving charged point particles – the electrons – are included. He showed this in particular for the dispersion of light, for the Zeeman effect, for the Hall effect and for the Faraday effect. He also gave the correct description of the Lorentz force. In 1902, he received the physics Nobel Prize together with Pieter Zeeman. Outside physics, he was active in the internationalization of scientific collaborations. He was also instrumental in the creation of the largest human-made structures on Earth: the polders of the Zuiderzee.

▷ *The speed $v$ of a physical system is bound by*

$$v \leqslant c \tag{10}$$

*for all observers, where $c$ is the speed of light.*

This invariance of the speed of light was known since the 1850s, because the expression
Vol. III, page 93 $c = 1/\sqrt{\varepsilon_0 \mu_0}$, known to people in the field of electricity, does not depend on the speed of the observer or of the light source, nor on their orientation or position. The invariance, including the speed independence, was found by optical experiments that used moving prisms, moving water, moving bodies with double refraction, interfering light beams travelling in different directions, interfering circulating light beams or light from moving stars. The invariance was also found by electromagnetic experiments that used moving insulators in electric and magnetic fields.* All experiments showed without exception that the speed of light in vacuum is invariant, whether performed *before* and *after* special relativity was formulated. The experiment performed by Albert Michelson, and the high-precision version to date, by Stephan Schiller and his team, are illustrated in Figure 19. All such experiments found no change of the speed of light with the motion
Ref. 41 of the Earth within measurement precision, which is around 2 parts in $10^{-17}$ at present. You can also confirm the invariance of the speed of light yourself at home; the way to do
Vol. III, page 46 this is explained in the section on electrodynamics.

The existence of an invariant limit speed has several interesting consequences. To explore them, let us keep the rest of Galilean physics intact.** The limit property and the invariance of the speed of light imply:

— In a closed free-floating ('inertial') room, there is no way to tell the speed of the room. Or, as Galileo writes in his *Dialogo*: il moto [ ...] niente opera ed è come s' e' non fusse. 'Motion [ ...] has no effect and behaves as if it did not exist'. Sometimes this statement is shortened to: motion is like nothing.
— There is no notion of absolute rest: rest is an observer-dependent, or *relative* concept.***
— Length and space depend on the observer; length and space are not absolute, but relative.
— Time depends on the observer; time is not absolute, but relative.
— Mass and energy are equivalent.

* All these experiments, which Einstein did not bother to cite in his 1905 paper, were performed by the
Ref. 39 complete who's who of 19th century physics, such as Wilhelm Röntgen, Alexander Eichenwald, François
Ref. 40 Arago, Augustin Fresnel, Armand Fizeau , Martin Hoek, Harold Wilson, Albert Abraham Michelson, (the first US-American to receive, in 1907, the Nobel Prize in Physics) Edward Morley, Oliver Lodge, John Strutt Rayleigh, Dewitt Brace, Georges Sagnac and Willem de Sitter among others.

** This point is essential. For example, Galilean physics states that only *relative* motion is observable.
Page 130 Galilean physics also excludes various mathematically possible ways to realize an invariant light speed that would contradict everyday life.

Einstein's original 1905 paper starts from two principles: the *invariance* of the speed of light and the equivalence, or *relativity*, of all inertial observers. The latter principle had already been stated in 1632 by Galileo; only the invariance of the speed of light was new. Despite this fact, the new theory was named – by
Ref. 18 Poincaré – after the old principle, instead of calling it 'invariance theory', as Einstein would have preferred.
Challenge 34 s *** Can you give the precise argument leading to this deduction?

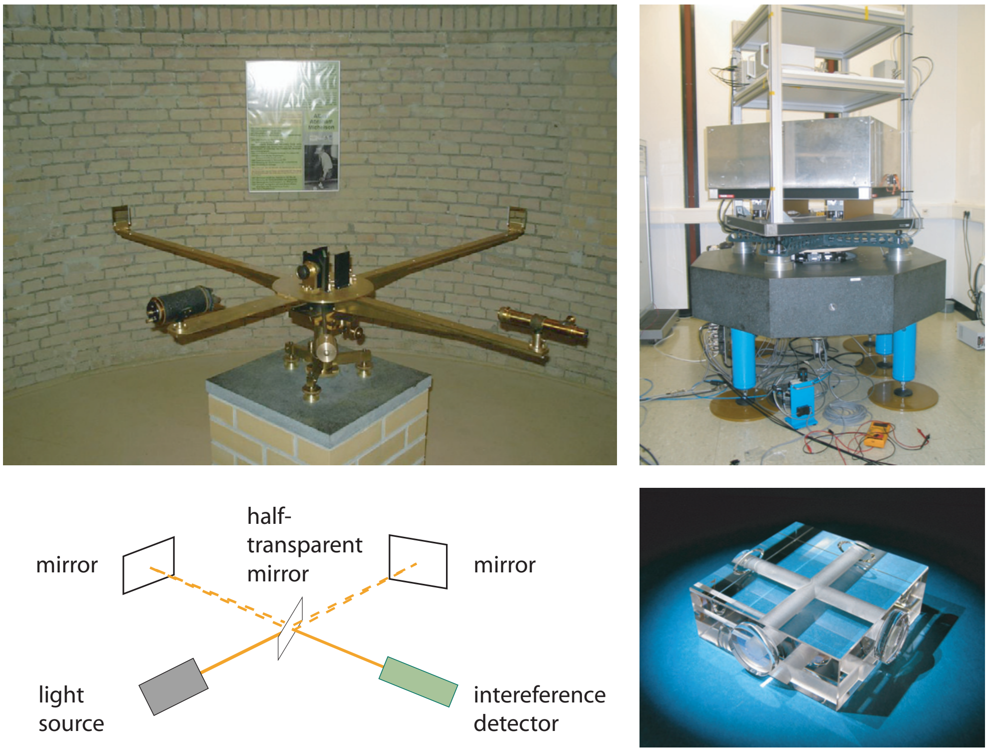



FIGURE 19 Testing the invariance of the speed of light on the motion of the observer: the reconstructed set-up of the first experiment by Albert Michelson in Potsdam, performed in 1881, and a modern high-precision, laser-based set-up that keeps the mirror distances constant to less than a proton radius and constantly rotates the whole experiment around a vertical axis (© Astrophysikalisches Institut Potsdam, Stephan Schiller)

We can draw more specific conclusions when two additional conditions are realised. First, we study situations where gravitation can be neglected. (If this not the case, we need *general* relativity to describe the system.) Secondly, we also assume that the data about the bodies under study – their speed, their position, etc. – can be gathered without disturbing them. (If this not the case, we need *quantum theory* to describe the system.)

How *exactly* differ the time intervals and lengths measured by two observers? To answer, we only need a pencil and a ruler. To start, we explore situations where no interaction plays a role. In other words, we star with *relativistic kinematics*: all bodies move without disturbance.

If an undisturbed body is observed to travel along a straight line with a constant velocity (or to stay at rest), one calls the observer *inertial*, and the coordinates used by the observer an *inertial frame of reference*. Every inertial observer is itself in undisturbed motion. Examples of inertial observers (or frames) thus include – in *two* dimensions – those moving on a frictionless ice surface or on the floor inside a smoothly running train or ship. For a full example – in all *three* spatial dimensions – we can take a cosmonaut travelling in a space-ship as long as the engine is switched off or a person falling in vacuum. Inertial observers in three dimensions can also be called *free-floating* observers, where 'free' stands again for 'undisturbed'. They are thus much rarer than non-inertial

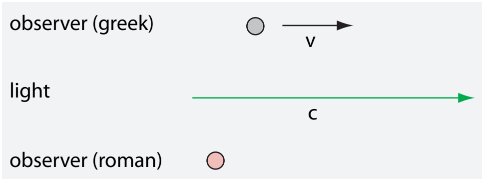



FIGURE 20 Two inertial observers and a beam of light

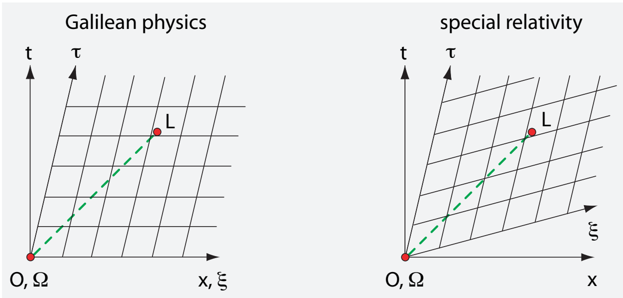



FIGURE 21 Space-time diagrams for light seen from two inertial observers, using coordinates $(t, x)$ and $(\tau, \xi)$

Challenge 35 e observers. Can you confirm this? Nevertheless, inertial observers are the most simple ones, and they form a special set:

— Any two inertial observers move with *constant velocity* relative to each other (as long as gravity and interactions play no role, as assumed above).
— All inertial observers are *equivalent*: they describe the world with the same equations. This statement, due to Galileo, was called the *principle of relativity* by Henri Poincaré.

To see how exactly the measured length and space intervals change from one inertial observer to the other, we assume a Roman one, using coordinates $x$, $y$, $z$ and $t$, and a Greek one, using coordinates $\xi$, $\upsilon$, $\zeta$ and $\tau$,* that move with velocity $\boldsymbol{v}$ relative to each other. The invariance of the speed of light in any direction for any two observers means
Challenge 36 e that the coordinate differences found by two observers are related by

$$(c\mathrm{d}t)^2 - (\mathrm{d}x)^2 - (\mathrm{d}y)^2 - (\mathrm{d}z)^2 = (c\mathrm{d}\tau)^2 - (\mathrm{d}\xi)^2 - (\mathrm{d}\upsilon)^2 - (\mathrm{d}\zeta)^2 . \qquad (11)$$

We now chose the axes in such a way that the velocity points in the $x$ and $\xi$-direction. Then we have

$$(c\mathrm{d}t)^2 - (\mathrm{d}x)^2 = (c\mathrm{d}\tau)^2 - (\mathrm{d}\xi)^2 . \qquad (12)$$

Assume that a flash lamp is at rest at the origin for the Greek observer, thus with $\xi = 0$, and produces two flashes separated by a time interval $\mathrm{d}\tau$. For the Roman observer, the flash lamp moves with speed $v$, so that $\mathrm{d}x = v\mathrm{d}t$. Inserting this into the previous

* They are read as 'xi', 'upsilon', 'zeta' and 'tau'. The names, correspondences and pronunciations of all Greek letters are explained in Appendix A.

Challenge 37 e expression, we deduce

$$dt = \frac{d\tau}{\sqrt{1 - v^2/c^2}} = \gamma d\tau . \tag{13}$$

This expression thus relates clock intervals measured by one observer to the clock intervals measured by another. At relative speeds $v$ that are *small* compared to the velocity of light $c$, such as occur in everyday life, the *stretch factor*, *relativistic correction* or *relativistic contraction* $\gamma$ is equal to 1 for all practical purposes. In these cases, the time intervals found by the two observers are essentially equal: time is then the same for all. However, for velocities *near* that of light the value of $\gamma$ increases. The largest value humans have ever achieved is about $2 \cdot 10^5$; the largest observed value in nature is about $10^{12}$. Can you imagine where they occur? Challenge 38 s

For a relativistic correction $\gamma$ larger than 1, the time measurements of the two observers give different values: moving observers observe *time dilation*. Time differs from one observer to another.

But that is not all. Once we know how clocks behave, we can easily deduce how *coordinates* change. Figures 20 and 21 show that the $x$ coordinate of an event L is the sum of two intervals: the $\xi$ coordinate and the length of the distance between the two origins. In other words, we have

$$\xi = \gamma(x - vt) . \tag{14}$$

Using the invariance of the space-time interval, we get

$$\tau = \gamma(t - xv/c^2) . \tag{15}$$

Henri Poincaré called these two relations the *Lorentz transformations of space and time* after their discoverer, the Dutch physicist Hendrik Antoon Lorentz.* In one of the most Ref. 42 beautiful discoveries of physics, in 1892 and 1904, Lorentz deduced these relations from Page 64 the equations of electrodynamics, where they had been lying, waiting to be discovered, since 1865.** In that year James Clerk Maxwell had published the equations that describe everything electric, magnetic and optical. However, it was Einstein who first understood that $t$ and $\tau$, as well as $x$ and $\xi$, are *equally valid* descriptions of space and time.

The Lorentz transformation describes the change of viewpoint from one inertial frame to a second, moving one. This change of viewpoint is called a (Lorentz) *boost*. The formulae (14) and (15) for the boost are central to the theories of relativity, both special and general. In fact, the mathematics of special relativity will not get more difficult than that: if you know what a square root is, you can study special relativity in all its beauty.

The Lorentz transformations (14) and (15) contain many curious results. Again they show that time depends on the observer. They also show that length depends on the Challenge 39 e observer: in fact, moving observers observe *length contraction*. Space and time are thus indeed relative.

* For information about Hendrik Antoon Lorentz, see page 35.
** The same discovery had been published first in 1887 by the German physicist Woldemar Voigt (1850–1919); Voigt – pronounced 'Fohgt' – was also the discoverer of the Voigt effect and the Voigt tensor. Independently, in 1889, the Irishman George F. Fitzgerald also found the result.

The Lorentz transformations (14) and (15) are also strange in another respect. When two observers look at each other, each of them claims to measure shorter intervals than
Challenge 40 s the other. In other words, special relativity shows that the grass on the other side of the fence is always *shorter* – if we ride along beside the fence on a bicycle and if the grass is inclined. We explore this bizarre result in more detail shortly.

Many alternative formulae for Lorentz boosts have been explored, such as expressions in which the relative acceleration of the two observers is included, as well as the relative
Ref. 43 velocity. However, all alternatives to be discarded after comparing their predictions with experimental results. Before we have a look at such experiments, we continue with a few logical deductions from the boost relations.

## What is space-time?

> “Von Stund' an sollen Raum für sich und Zeit für sich völlig zu Schatten herabsinken und nur noch eine Art Union der beiden soll Selbstständigkeit bewahren.*”
> Hermann Minkowski.

The Lorentz transformations tell us something important: space and time are two aspects of the same basic entity. They 'mix' in different ways for different observers. The mixing is commonly expressed by stating that time is the *fourth dimension*. This makes sense because the common basic entity – called *space-time* – can be defined as the set of all events, events being described by four coordinates in time and space, and because the
Challenge 41 s set of all events has the properties of a manifold.** (Can you confirm this?) Complete space-time is observer-invariant and absolute; space-time remains unchanged by boosts. Only its split into time and space depends on the viewpoint.

In other words, the existence of a maximum speed in nature forces us to introduce the invariant space-time manifold, made of all possible events, for the description of nature. In the absence of gravitation, i.e., in the theory of special relativity, the space-time manifold is characterized by a simple property: the *space-time interval* d$i$ between
Ref. 44 two events, defined as

$$\mathrm{d}i^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = c^2\mathrm{d}t^2\left(1 - \frac{v^2}{c^2}\right) , \qquad (16)$$

is independent of the (inertial) observer: it is an invariant. Space-time is also called *Minkowski space-time*, after Hermann Minkowski,*** the teacher of Albert Einstein; he was the first, in 1904, to define the concept of space-time and to understand its usefulness and importance. We will discover that later that when gravitation is present, the whole of

* 'Henceforth space by itself and time by itself shall completely fade into shadows and only a kind of union of the two shall preserve autonomy.' This famous statement was the starting sentence of Minkowski's 1908 talk at the meeting of the Gesellschaft für Naturforscher und Ärzte.

Vol. V, page 286 ** The term 'manifold' is defined in all mathematical details later in our walk.

*** Hermann Minkowski (1864–1909), German mathematician. He had developed similar ideas to Einstein, but the latter was faster. Minkowski then developed the concept of space-time. Minkowski died suddenly at the age of 44.

space-time *bends*; such bent space-times, called *Riemannian space-times*, will be essential in general relativity.

The space-time interval $\mathrm{d}i$ of equation (16) has a simple physical meaning. It is the time measured by an observer moving from event $(t, x)$ to event $(t + \mathrm{d}t, x + \mathrm{d}x)$, the so-called *proper time*, multiplied by $c$. If we neglect the factor $c$, we can also call the interval the *wristwatch time*.

In short, we can say that we *live in* space-time. Space-time exists independently of all things; it is a container, a background for everything that happens. And even though coordinate systems differ from observer to observer, the underlying entity, space-time, is the same and *unique*, even though space and time by themselves are not. (All this applies also in the presence of gravitation, in general relativity.)

How does Minkowski space-time differ from Galilean space-time, the combination of everyday space and time? Both space-times are manifolds, i.e., continuum sets of points, both have one temporal and three spatial dimensions, and both manifolds have the topology of the punctured sphere. (Can you confirm this?) Both manifolds are flat, i.e., free of curvature. In both cases, space is what is measured with a metre rule or with a light ray, and time is what is read from a clock. In both cases, space-time is fundamental, unique and absolute; it is and remains the *background* and the *container* of things and events.

Challenge 42 s

The central difference, in fact the only one, is that Minkowski space-time, in contrast to the Galilean case, *mixes* space and time. The mixing is different for observers with different speeds, as shown in Figure 21. The mixing is the reason that time and space are observer-dependent, or relative, concepts.

Mathematically, time is a fourth dimension; it expands space to space-time. Calling time the *fourth dimension* is thus only a statement on how relativity calculates – we will do that below – and has *no* deeper meaning.

The maximum speed in nature thus forces us to describe motion with space-time. That is interesting, because in space-time, speaking in simple terms, *motion does not exist*. Motion exists only in space. In space-time, nothing moves. For each point particle, space-time contains a *world-line*. (See Figure 22.) In other words, instead of asking *why* motion exists, we can equivalently ask why space-time is criss-crossed by world-lines. But at this point of our adventure we are still far from answering either question. What we can do is to explore *how* motion takes place.

## Can we travel to the past? – Time and causality

We know that time is different for different observers. Does time nevertheless order events in sequences? The answer given by relativity is a clear 'yes and no'. Certain sets of events are not naturally ordered by time; others sets are. This is best seen in a space-time diagram, such as Figure 22.

Clearly, two events can be placed in a time sequence only if one event is or could be the *cause* of the other. But this connection can only apply if the first event could send energy, e.g. through a signal, to the second. In other words, a temporal sequence between two events implies that the signal speed connecting the two events must not be larger than the speed of light. Figure 22 shows that event E at the origin of the coordinate system can only be influenced by events in quadrant IV (the *past light cone*, when all space dimensions are included), and can itself influence only events in quadrant II, the *future light cone*.

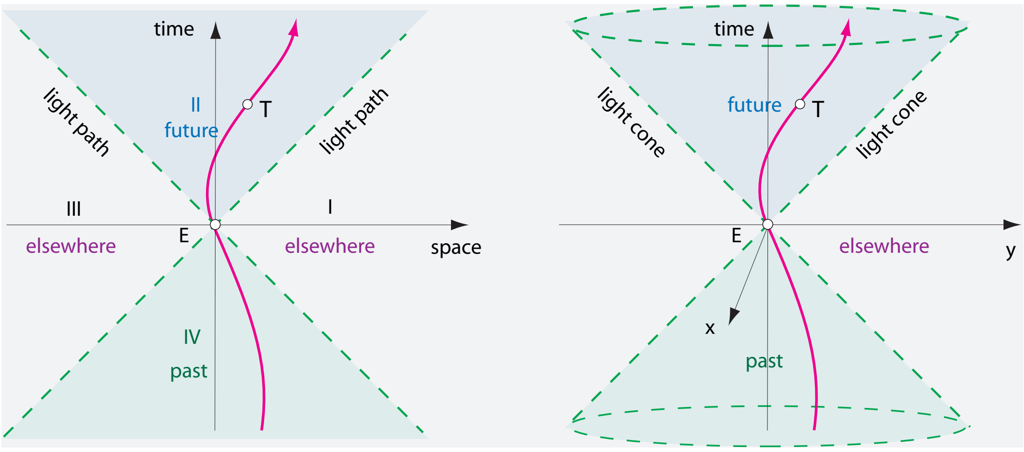



FIGURE 22 A space-time diagram for a moving object T seen from an inertial observer O in the case of one and two spatial dimensions; the slope of the world-line at a point is the speed at that point, and thus is never steeper than that of light

Events in quadrants I and III neither influence nor are influenced by event E: signal speed above that of light would be necessary to achieve that. Thus the full light cone defines the boundary between events that *can* be ordered with respect to event E – namely those inside the cone – and those that *cannot* – those outside the cone, which happen *elsewhere* for all observers. (Some people call all the events happening elsewhere the *present*.)

The past light cone gives the complete set of events that can influence what happens at E, the coordinate origin. One says that E is *causally connected* to events in the past light cone. Note that causal connection is an invariant concept: all observers agree on whether or not it applies to two given events. Can you confirm this? Challenge 43 s

In short, time orders events only *partially*. In particular, for two events that are not causally connected, their temporal order (or their simultaneity) depends on the observer! Challenge 44 e

A vector inside the light cone is called *timelike*; one on the light cone is called *lightlike* or *null*; and one outside the cone is called *spacelike*. For example, the *world-line* of an observer, i.e., the set of all events that make up its past and future history, consists of timelike events only.

Special relativity thus teaches us that causality and time can be defined *only* because light cones exist. If transport of energy at speeds faster than that of light did exist, time could not be defined. Causality, i.e., the possibility of (partially) ordering events for all observers, is due to the existence of a maximal speed.

If the speed of light could be surpassed, we could always win the lottery. Can you see why? In other words, if the speed of light could be surpassed in some way, the future could influence the past. Can you confirm this? In such situations, one would observe *acausal* effects. However, there is an everyday phenomenon which tells that the speed of light is indeed maximal: our memory. If the future could influence the past, we would also be able to *remember* the future. To put it in another way, if the future could influence the past, the second principle of thermodynamics would not be valid.* No known data from everyday life or from experiments provide any evidence that the future can

Challenge 45 e
Challenge 46 s

* Another related result is slowly becoming common knowledge. Even if space-time had a non-trivial shape,

influence the past. In other words, *time travel to the past is impossible*. How the situation changes in quantum theory will be revealed later on. Interestingly, time travel to the future *is* possible, as we will see shortly.

## Curiosities about special relativity

Special relativity is full of curious effects. Let us start with a puzzle that helps to sharpen our thinking. Seen by an observer on an island, two lightning strokes hit simultaneously: one hits the island, and another, many kilometers away, the open sea. A second observer is a pilot in a relativistic aeroplane and happens to be just above the island when the
Challenge 47 e lightning hits the island. Which lightning hits first for the pilot?

For the pilot, the distant lightning, hitting the sea, hits first. But this is a trick question: despite being the one that hits first, the distant lightning is observed by the pilot to hit *after* the one on the island, because light from the distant hit needs time to reach him. However, the pilot can compensate for the propagation time and can deduce that the
Challenge 48 e distant lightning hit first.

Let us explore a few additional consequences of special relativity.

## Faster than light: how far can we travel?

How far away from Earth can we travel, given that the trip should not last more than a lifetime, say 80 years, and given that we are allowed to use a rocket whose speed can approach the speed of light as closely as desired? Given the time $t$ we are prepared to spend in a rocket, given the speed $v$ of the rocket, and assuming optimistically that it can accelerate and decelerate in a negligible amount of time, the distance $d$ we can move
Challenge 49 e away is given by

$$d = \frac{vt}{\sqrt{1 - v^2/c^2}} \; . \tag{17}$$

The distance $d$ is larger than $ct$ already for $v > 0.72c$, and, if $v$ is chosen large enough, it increases beyond all bounds! In other words, light speed does *not* limit the distance we can travel in a lifetime or in any other time interval. We could, in principle, roam the
Page 46 entire universe in less than a second. (The fuel issue is discussed below.)

For rocket trips it makes sense to introduce the concept of *proper velocity* $w$, defined as

$$w = \frac{d}{t} = \frac{v}{\sqrt{1 - v^2/c^2}} = \gamma\, v \; . \tag{18}$$

As we have just seen, proper velocity is *not* limited by the speed of light; in fact the proper velocity of light itself is infinite.*

---

such as a cylindrical topology with closed time-like curves, one still would not be able to travel into the
Ref. 45 past, in contrast to what many science fiction novels suggest. This is made clear by Stephen Blau in a recent pedagogical paper.

* Using proper velocity, the relation given in equation (9) for the composition of two velocities $\mathbf{w}_a = \gamma_a \mathbf{v}_a$
Challenge 50 e and $\mathbf{w}_b = \gamma_b \mathbf{v}_b$ simplifies to

$$w_{s\parallel} = \gamma_a \gamma_b (v_a + v_{b\parallel}) \quad \text{and} \quad w_{s\perp} = w_{b\perp} \; , \tag{19}$$

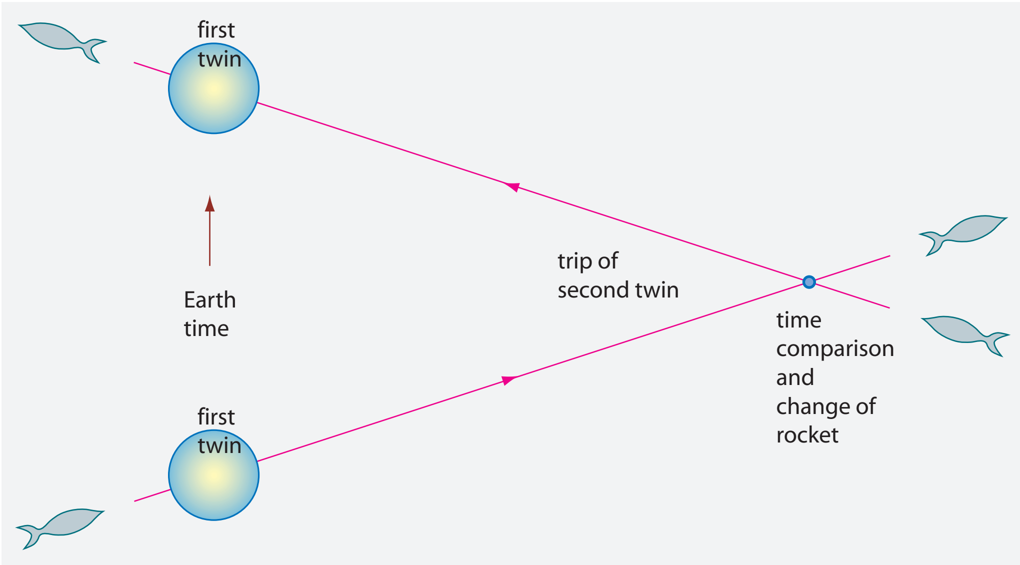



FIGURE 23 The twin paradox

## Synchronization and time travel – can a mother stay younger than her own daughter?

The maximum speed in nature implies that time is different for different observers moving relative to each other. So we have to be careful about how we synchronize clocks that are far apart, even if they are at rest with respect to each other in an inertial reference frame. For example, if we have two similar watches showing the same time, and if we carry one of them for a walk and back, they will show different times afterwards. This experiment has actually been performed several times and has fully confirmed the prediction of special relativity. The time difference for a person or a watch in an aeroplane travelling around the Earth once, at about 900 km/h, is of the order of 100 ns – not very noticeable in everyday life. This is sometimes called the *clock paradox*. In fact, the delay is easily calculated from the expression

Ref. 47, Ref. 48

$$\frac{t}{t'} = \gamma \ . \tag{20}$$

Also human bodies are clocks; they show the elapsed time, usually called *age*, by various changes in their shape, weight, hair colour, etc. If a person goes on a long and fast trip, on her return she will have aged *less* and thus stayed younger than a second person who stayed at her (inertial) home. Travellers stay younger.

The most extreme illustration of this is the famous *twin paradox*. An adventurous twin jumps on a relativistic rocket that leaves Earth and travels for many years. Far from Earth, he jumps on another relativistic rocket going the other way and returns to Earth. The trip is illustrated in Figure 23. At his arrival, he notes that his twin brother on Earth

where the signs ∥ and ⊥ designate the component in the direction of and the component perpendicular to $\mathbf{v}_a$, respectively. One can in fact express all of special relativity in terms of 'proper' quantities.

Ref. 46

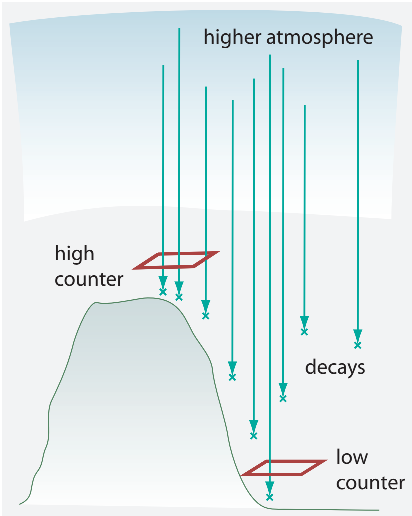



FIGURE 24 More muons than expected arrive at the ground because fast travel keeps them young

Ref. 49 is much older than himself. This result has also been confirmed in many experiments. Can you explain the result, especially the asymmetry between the two brothers?

Special relativity thus confirms, in a surprising fashion, the well-known observation that those who travel a lot remain younger. On the other hand, the human traveller with the largest measured youth effect so far was the cosmonaut Sergei Krikalyov, who has spent 803 days in orbit, and nevertheless aged only a few milliseconds less than people on Earth.

The twin paradox is also the confirmation of the possibility of time travel to the future. With the help of a fast rocket that comes back to its starting point, we can arrive at local times that we would never have reached within our lifetime by staying home. Alas, we can *never* return to the past to talk about it.*

One of the simplest experiments confirming the prolonged youth of *really fast* travellers involves the counting of muons. Muons are particles that are continuously formed Page 119 in the upper atmosphere by cosmic radiation and then fly to the ground. Muons *at rest* (with respect to the measuring clock) have a finite half-life of 2.2 μs (or, at the speed of light, 660 m). After this amount of time, half of the muons have decayed. This half-life can be measured using simple muon counters. In addition, there exist more special counters that only count muons travelling within a certain speed range, say from $0.9950c$ to $0.9954c$. One can put one of these special counters on top of a mountain and put another in the valley below, as shown in Figure 24. The first time this experiment was performed, Ref. 51 the height difference was 1.9 km. Flying 1.9 km through the atmosphere at the mentioned speed takes about 6.4 μs. With the half-life just given, a naive calculation finds that only Challenge 51 s about 13% of the muons observed at the top should arrive at the lower site in the valley. However, it is observed that about 82% of the muons arrive below. The reason for this result is the relativistic time dilation. Indeed, at the mentioned speed, muons experience a

Ref. 50 * There are even special books on time travel, such as the well-researched text by Nahin. Note that the concept of time travel has to be clearly defined; otherwise one has no answer to the clerk who calls his office chair a time machine, as sitting on it allows him to get to the future.

proper time difference of only 0.62 μs during the travel from the mountain top to the valley. This time is much shorter than that observed by the human observers. The shortened muon time yields a much lower number of lost muons than would be the case without time dilation; moreover, the measured percentage confirms the value of the predicted time dilation factor $\gamma$ within experimental errors, as you may want to check. The same effect is observed when relativistic muons are made to run in circles at high speed inside a so-called storage ring. The faster the muons turn, the longer they live.

Challenge 52 s
Ref. 52

Half-life dilation has also been found for many other decaying systems, such as pions, hydrogen atoms, neon atoms and various nuclei, always confirming the predictions of special relativity. Since all bodies in nature are made of particles, the 'youth effect' of high speeds – usually called *time dilation* – applies to bodies of all sizes; indeed, it has not only been observed for particles, but also for lasers, radio transmitters and clocks.

Ref. 12

If motion leads to time dilation, a clock on the Equator, constantly running around the Earth, should go slower than one at the poles. However, this prediction, which was made by Einstein himself, is incorrect. The centrifugal acceleration leads to a reduction in gravitational acceleration whose time dilation exactly cancels that due to the velocity. This story serves as a reminder to be careful when applying special relativity in situations involving gravity: special relativity is only applicable when space-time is flat, i.e., when gravity is *not* present.

Ref. 54

In summary, a mother *can* stay younger than her daughter. The mother's wish to remain younger than her daughter is not easy to fulfil, however. Let us imagine that a mother is accelerated in a spaceship away from Earth at $10\,\mathrm{m/s^2}$ for ten years, then decelerates at $10\,\mathrm{m/s^2}$ for another ten years, then accelerates for ten additional years towards the Earth, and finally decelerates for ten final years in order to land safely back on our planet. The mother has taken 40 years for the trip. She got as far as 22 000 light years from Earth. At her return on Earth, 44 000 years have passed. All this seems fine, until we realize that the necessary amount of fuel, even for the most efficient engine imaginable, is so large that the mass returning from the trip is only one part in $2 \cdot 10^{19}$ of the mass that started. The necessary amount of fuel does not exist on Earth. The same problem appears for shorter trips.

Challenge 53 e
Ref. 53

We also found that we cannot (simply) synchronize clocks at rest with respect to each other simply by walking, clock in hand, from one place to another. The correct way to do so is to exchange light signals. Can you describe how? The precise definition of synchronization is necessary, because we often need to call two distant events *simultaneous*, for example when we define coordinates. Obviously, a maximum speed implies that simultaneity depends on the observer. Indeed, this dependence has been confirmed by all experiments.

Challenge 54 s

## Length contraction

The length of an object measured by an observer attached to the object is called its *proper length*. The length measured by an inertial observer passing by is always smaller than the proper length. This result follows directly from the Lorentz transformations.

Challenge 55 e

For a Ferrari driving at 300 km/h or 83 m/s, the length is contracted by 0.15 pm: less than the diameter of a proton. Seen from the Sun, the Earth moves at 30 km/s; this gives

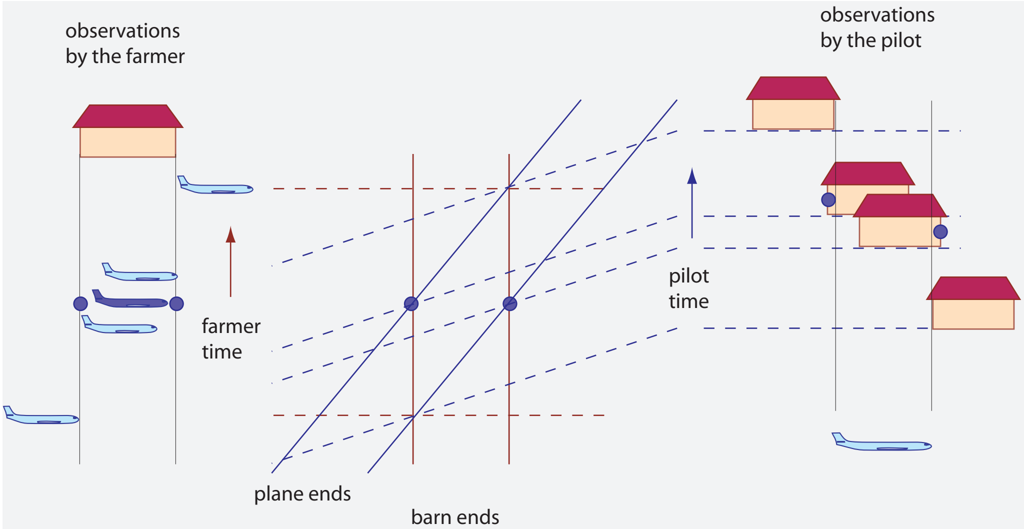



FIGURE 25 The observations of the pilot and the barn owner

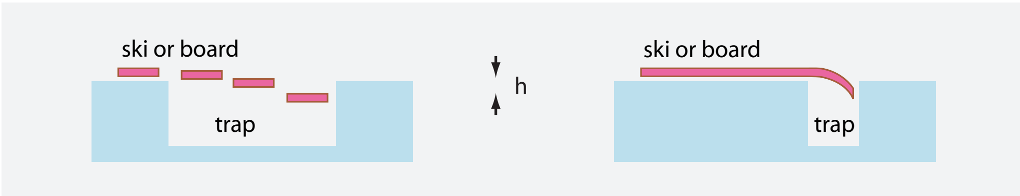



FIGURE 26 The observations of the trap digger and of the snowboarder, as (misleadingly) published in the literature

a length contraction of 6 cm. Neither of these effects has ever been measured.* But larger effects could be. Let us explore the consequences.

Imagine a pilot flying with his plane through a barn with two doors, one at each end. The plane is slightly longer than the barn, but moves so rapidly that its relativistically contracted length is shorter than the length of the barn. Can the farmer close the barn (at least for a short time) with the plane completely inside? The answer is positive. But why can the pilot not say the following: relative to him, the barn is contracted; therefore the plane does not fit inside the barn? The answer is shown in Figure 25. For the farmer, the doors close (and reopen) at the same time. For the pilot, they do not. For the farmer, the pilot is in the dark for a short time; for the pilot, the barn is never dark. (That is not
Challenge 57 s completely true: can you work out the details?)

We now explore some variations of the general case. Can a rapid snowboarder fall into a hole that is a bit shorter than his board? Imagine him boarding so (unrealistically) fast that the length contraction factor $\gamma$ is 4. For an observer on the ground, the snowboard is four times shorter, and when it passes over the hole, it will fall into it. However, for the boarder, it is the hole which is four times shorter; it seems that the snowboard cannot fall into it.

Challenge 56 s * Is the Earth contraction value measurable measurable at all?

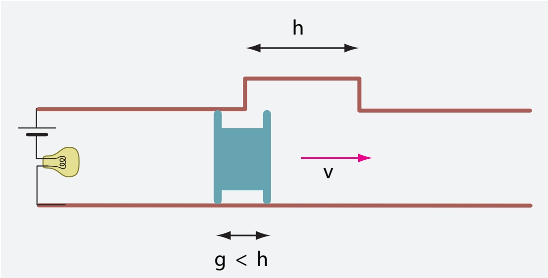



FIGURE 27 Does the conducting glider keep the lamp lit at large speeds?

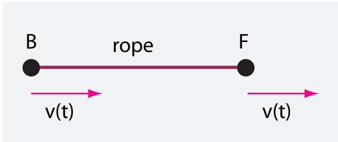



FIGURE 28 What happens to the rope?

Ref. 55 | Challenge 58 e | Challenge 59 s

A first careful analysis shows that, in contrast to the observation of the hole digger, the snowboarder does not experience the board's shape as fixed: while passing over the hole, the boarder observes that the board takes on a parabolic shape and falls into the hole, as shown in Figure 26. Can you confirm this? In other words, shape is not an observer-invariant concept. (However, rigidity *is* observer-invariant, if defined properly; can you confirm this?)

Ref. 56 | Challenge 60 e

This explanation however, though published, is not correct, as Harald van Lintel and Christian Gruber have pointed out. One should not forget to estimate the size of the effect. At relativistic speeds the time required for the hole to affect the full thickness of the board cannot be neglected. The snowboarder only sees his board take on a parabolic shape if it is extremely thin and flexible. For usual boards moving at relativistic speeds, the snowboard has no time to fall any appreciable height $h$ or to bend into the hole before passing it. Figure 26 is so exaggerated that it is incorrect. The snowboarder would simply speed over the hole.

Ref. 57 | Challenge 61 s

The paradoxes around length contraction become even more interesting in the case of a conductive glider that makes electrical contact between two rails, as shown in Figure 27. The two rails are parallel, but one rail has a gap that is longer than the glider. Can you work out whether a lamp connected in series stays lit when the glider moves along the rails with relativistic speed? (Make the simplifying and not fully realistic assumption that electrical current flows as long and as soon as the glider touches the rails.) Do you get the same result for all observers? And what happens when the glider is longer than the detour? Or when it approaches the lamp from the other side of the detour? (Warning: this problem gives rise to *heated* debates!) What is unrealistic in this experiment?

Ref. 58 | Challenge 62 s

Another example of length contraction appears when two objects, say two cars, are connected over a distance $d$ by a straight rope, as shown in Figure 28. Imagine that both are at rest at time $t = 0$ and are accelerated together in exactly the same way. The observer at rest will maintain that the two cars always remain the same distance apart. On the other hand, the rope needs to span a distance $d' = d/\sqrt{1 - v^2/c^2}$, and thus has to expand when the two cars are accelerating. In other words, the rope will break. You can check by yourself that this prediction is confirmed by all observers, in the cars and on Earth.

Ref. 59

A funny – but quite unrealistic – example of length contraction is that of a submarine moving horizontally. Imagine that before moving, the resting submarine has tuned its weight to float in water without any tendency to sink or to rise. Now the submarine moves (possibly with relativistic speed) in horizontal direction. The captain observes the water outside to be Lorentz contracted; thus the water is denser and he concludes that the submarine will rise. A nearby fish sees the submarine to be contracted, thus denser than water, and concludes that the submarine will sink. Who is wrong, and what is the buoyancy force? Alternatively, answer the following question: why is it impossible for a submarine to move at relativistic speed?

Challenge 63 s
Challenge 64 s

In summary, for macroscopic bodies, length contraction will probably never be observed. However, it does play an important role for *images*.

## Relativistic films – aberration and Doppler effect

In our adventure so far, we have encountered several ways in which the observed surroundings change when we move at relativistic speed. We now put them all together. First of all, Lorentz contraction and aberration lead to *distorted* images. Secondly, aberration increases the viewing angle beyond the roughly 180 degrees that we are used to in everyday life. At relativistic speeds, when we look in the direction of motion, we see light that is invisible for an observer at rest, because for the latter, it comes from behind. Thirdly, the Doppler effect produces *colour-shifted* images. Fourthly, our rapid motion changes the *brightness* and *contrast* of the image: the so-called *searchlight effect*. Each of these changes depends on the direction of sight; they are shown in Figure 30.

Modern computers enable us to simulate the observations made by rapid observers with photographic quality, and even to produce simulated films and computer games.* The images of Figure 29 are particularly helpful in allowing us to understand image distortion. They show the viewing angle, the circle which distinguish objects in front of the observer from those behind the observer, the coordinates of the observer's feet and the point on the horizon toward which the observer is moving. Adding these markers in your head when watching other pictures or films may help you to understand more clearly what they show.

We note that the image seen by a moving observer is a *distorted* version of that seen by one at rest at the same point. A moving observer never sees different things than a resting one at the same point. Indeed, light cones are independent of observer motion.

Studying the images with care shows another effect. Even though the Lorentz contraction is measurable, it *cannot* be photographed. This surprising result was discovered only in 1959. Measuring implies simultaneity at the object's position; in contrast, photographing implies simultaneity at the observer's position. On a photograph or in a film, the Lorentz contraction is modified by the effects due to different light travel times from the different parts of an object; the result is a change in shape that is reminiscent of, but not exactly the same as, a rotation. This is shown in Figure 32. The total deformation is the result of the angle-dependent aberration. We discussed the aberration of star positions at the beginning of this chapter. In complete images, aberration transforms circles

Ref. 60
Page 18

* See for example the many excellent images and films at www.anu.edu.au/Physics/Searle by Anthony Searle and www.anu.edu.au/Physics/vrproject by Craig Savage and his team; you can even do interactive motion steering with the free program downloadable at realtimerelativity.org. There is also beautiful material at www.tat.physik.uni-tuebingen.de/~weiskopf/gallery/index.html by Daniel Weiskopf, at www.itp.uni-hannover.de/~dragon/stonehenge/stone1.htm by Norbert Dragon and Nicolai Mokros, and at www.tempolimit-lichtgeschwindigkeit.de by Ute Kraus, once at Hanns Ruder's group.

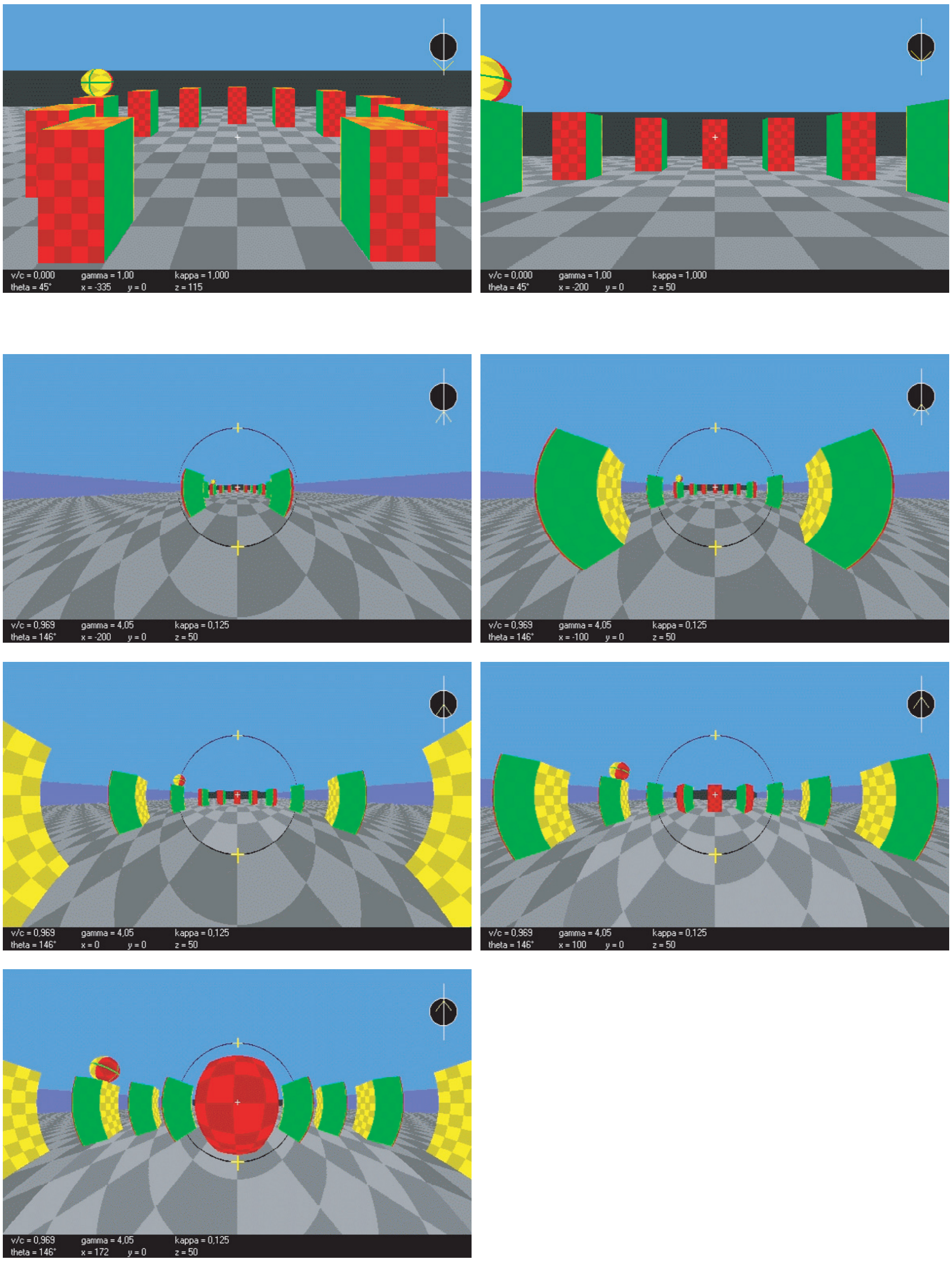



FIGURE 29 Flying through twelve vertical columns (shown in the two uppermost images) with 0.9 times the speed of light as visualized by Nicolai Mokros and Norbert Dragon, showing the effect of speed and position on distortions (© Nicolai Mokros)

FIGURE 30 Flying through three straight and vertical columns with 0.9 times the speed of light as visualized by Daniel Weiskopf: on the left with the original colours; in the middle including the Doppler effect; and on the right including brightness effects, thus showing what an observer would actually see (© Daniel Weiskopf)

FIGURE 31 What a researcher standing and one running rapidly through a corridor observe (ignoring colour and brightness effects) (© Daniel Weiskopf)



into circles: such transformations are called *conformal*. As a result, a sphere is seen as a sphere even at relativistic speeds; in a sense, the aberration compensates the Lorentz contraction.

Aberration leads to the *pearl necklace paradox*. If the relativistic motion transforms spheres into spheres, and rods into shorter rods, what happens to a pearl necklace moving along its own long axis? Does it get shorter or not?

Challenge 65 s

A further puzzle: imagine that a sphere that moves and rotates at high speed. Can all the mentioned effects lead to an apparent, observer-dependent sense of rotation?

Challenge 66 r

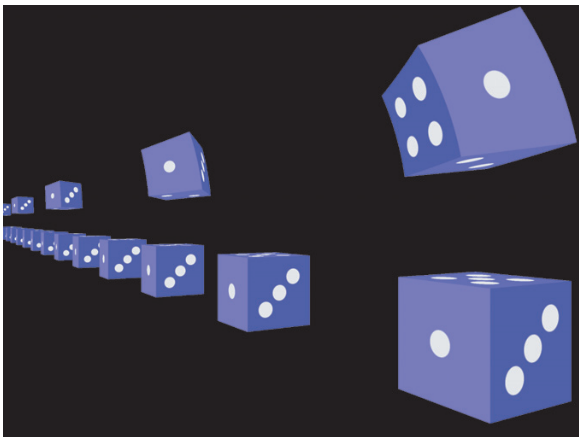


FIGURE 32 A stationary row of dice (below), and the same row, flying above it at relativistic speed towards the observer, though with Doppler effect switched off. (Mpg film © Ute Kraus at www.tempolimit-lichtgeschwindigkeit.de.)

## Which is the best seat in a bus?

Ref. 58 Let us explore another surprise of special relativity. Imagine two twins inside two identically accelerated cars, one in front of the other, starting from standstill at time $t = 0$, as described by an observer at rest with respect to both of them. (There is no connecting rope now.) Both cars contain the same amount of fuel. We easily deduce that the acceleration of the two twins stops, when the fuel runs out, at the same time in the frame of the outside observer. In addition, the distance between the cars has remained the same all along for the outside observer, and the two cars continue rolling with an identical constant velocity $v$, as long as friction is negligible. If we call the events at which the front car and back car engines switch off f and b, their time coordinates in the outside frame at rest are related simply by $t_f = t_b$. By using the Lorentz transformations you can deduce for the frame of the freely rolling twins the relation

Challenge 67 e

Challenge 68 e

$$t'_b = \gamma \Delta x \, v/c^2 + t'_f \, , \tag{21}$$

which means that the front twin has aged *more* than the back twin! Thus, in accelerated systems, ageing is position-dependent.

For choosing a seat in a bus, though, this result does not help. It is true that the best seat in an accelerating bus is the back one, but in a decelerating bus it is the front one. At the end of a trip, the choice of seat does not matter.

Is it correct to deduce from the above that people on high mountains age faster than people in valleys, so that living in a valley helps postponing grey hair?

Challenge 69 s

## How fast can one walk?

In contrast to running, walking means to move the feet in such a way that at least one of them is on the ground at any time. This is one of the rules athletes have to follow in Olympic walking competitions; they are disqualified if they break it. A student athlete was thinking about the theoretical maximum speed he could achieve in the Olympic

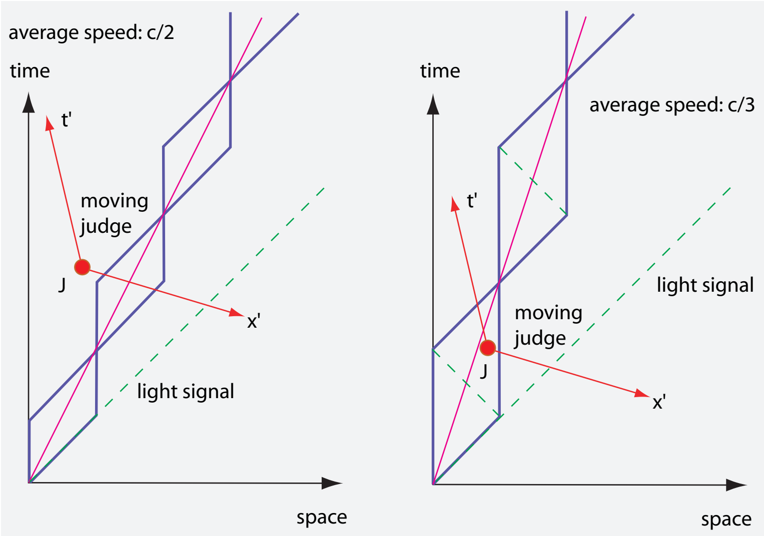



FIGURE 33 For the athlete on the left, the judge moving in the opposite direction sees both feet off the ground at certain times, but not for the athlete on the right

Games. The ideal would be that each foot accelerates instantly to (almost) the speed of light. The highest walking speed is then achieved by taking the second foot off the ground at exactly the same instant at which the first is put down. By 'same instant', the student originally meant 'as seen by a competition judge at rest with respect to Earth'. The motion of the feet is shown in the left diagram of Figure 33; it gives a limit speed for walking of half the speed of light.

But then the student noticed that a *moving* judge will regularly see both feet off the ground and thus disqualify the athlete for running. To avoid disqualification by *any* judge, the rising foot has to wait for a light signal from the lowered one. The limit speed for Olympic walking then turns out to be only one third of the speed of light. Ref. 61

## Is the speed of shadow greater than the speed of light?

Actually, motion faster than light does exist and is even rather common. Nature only constrains the motion of mass and energy. However, non-material points or non-energy-transporting features and images *can* move faster than light. There are several simple examples. To be clear, we are not talking about *proper* velocity, which in these cases cannot be defined anyway. (Why?) The following examples show speeds that are genuinely higher than the speed of light in vacuum. Page 43 Challenge 70 s

As first example, consider the point at which scissors cut paper, marked X in Figure 34. If the scissors are closed rapidly enough, the point moves faster than light. Similar examples can also be found in every window frame, and in fact in any device that has twisting parts.

Another example of superluminal motion is a music record – an old-fashioned LP – disappearing into its sleeve, as shown in Figure 35. The point where the border of the record meets the border of the sleeve can travel faster than light.

Another example suggests itself when we remember that we live on a spherical planet. Imagine you lie on the floor and stand up. Can you show that the initial speed with which

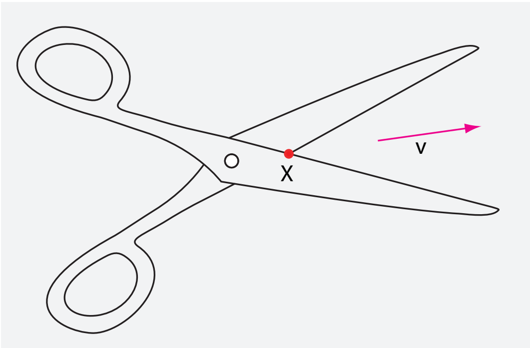



FIGURE 34 A simple example of motion that can be faster than light

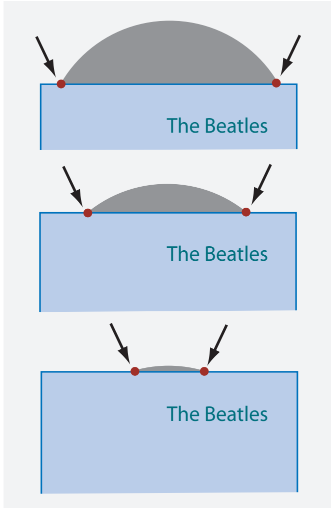



FIGURE 35 Another example of faster-than-light motion

Challenge 71 s the horizon moves away from you can be larger than that of light?

Finally, a standard example is the motion of a spot of light produced by shining a laser beam onto the Moon. If the laser is moved, the spot can easily move faster than light. The same applies to the light spot on the screen of an oscilloscope when a signal of sufficiently high frequency is fed to the input.

All these are typical examples of the *speed of shadows*, sometimes also called the *speed of darkness*. Both shadows and darkness can indeed move faster than light. In fact, there is no limit to their speed. Can you find another example? Challenge 72 s

In addition, there is an ever-increasing number of experimental set-ups in which the phase velocity or even the group velocity of light is higher than $c$. They regularly make headlines in the newspapers, usually along the lines of 'light moves faster than light'. We will discuss this surprising phenomenon in more detail later on. Page 107 In fact, these cases can also be seen – with some abstraction – as special cases of the 'speed of shadow' phenomenon.

For a different example, imagine that we are standing at the exit of a straight tunnel of length $l$. We see a car, whose speed we know to be $v$, entering the other end of the tunnel and driving towards us. We know that it entered the tunnel because the car is no longer in the Sun or because its headlights were switched on at that moment. At what time $t$, after we see it entering the tunnel, does it drive past us? Simple reasoning shows that $t$ is given by

$$t = l/v - l/c \; . \qquad (22)$$

In other words, the approaching car seems to have a velocity $v_{appr}$ of

$$v_{appr} = \frac{l}{t} = \frac{vc}{c - v} \; , \qquad (23)$$

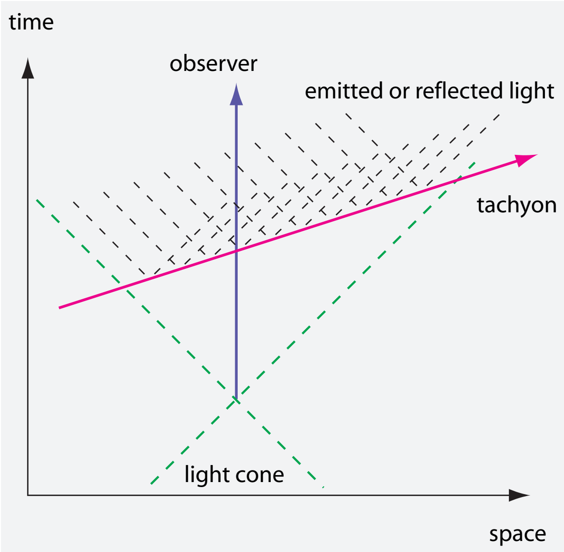



FIGURE 36 Hypothetical space-time diagram for tachyon observation

which is higher than $c$ for any car velocity $v$ higher than $c/2$. For cars this does not happen too often, but astronomers know a type of bright object in the sky called a *quasar* (a contraction of 'quasi-stellar object'), which sometimes emits high-speed gas jets. If the emission is in or near the direction of the Earth, its apparent speed – even the purely transverse component – is higher than $c$. Such situations are now regularly observed with telescopes. Ref. 62

Note that to a second observer at the *entrance* of the tunnel, the apparent speed of the car *moving away* is given by

$$v_{\text{leav}} = \frac{vc}{c+v} , \tag{24}$$

which is *never* higher than $c/2$. In other words, objects are never seen departing with more than half the speed of light.

The story has a final twist. We have just seen that motion faster than light can be observed in several ways. But could an *object* moving faster than light be observed at all? Surprisingly, it could be observed only in rather unusual ways. First of all, since such an imaginary object, usually called a *tachyon*, moves faster than light, we can never see it approaching. If it can be seen at all, a tachyon can only be seen departing. Seeing a tachyon would be similar to hearing a supersonic jet. Only *after* a tachyon has passed nearby, assuming that it is visible in daylight, could we notice it. We would first see a flash of light, corresponding to the bang of a plane passing with supersonic speed. Then we would see *two* images of the tachyon, appearing somewhere in space and departing in opposite directions, as can be deduced from Figure 36. Even if one of the two images were approaching us, it would be getting fainter and smaller. This is, to say the least, rather unusual behaviour. Moreover, if you wanted to look at a tachyon at night, illuminating it with a torch, you would have to turn your head in the direction opposite to the arm with the torch! This requirement also follows from the space-time diagram: can you see why? Challenge 73 e

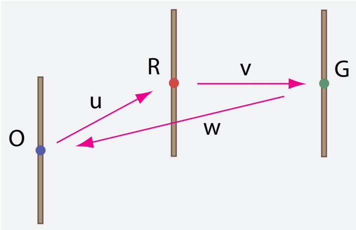



FIGURE 37 If O's stick is parallel to R's and R's is parallel to G's, then O's stick and G's stick are not

Nobody has ever seen such phenomena.

Ref. 63 Tachyons, if they existed, would be strange objects: they would accelerate when they Page 66 lose energy, a zero-energy tachyon would be the fastest of all, with infinite speed, and the direction of motion of a tachyon depends on the motion of the observer. No object with these properties has ever been observed. Worse, as we just saw, tachyons would seem to appear from nothing, defying laws of conservation; and note that, just as tachyons cannot be seen in the usual sense, they cannot be touched either, since both processes are due to electromagnetic interactions, as we will see later in our ascent of Motion Mountain. Tachyons therefore cannot be objects in the usual sense. In the quantum part of our adventure we will show that quantum theory actually *rules out* the existence of (real) tachyons. However, quantum theory also *requires* the existence of 'virtual' tachyons, as we will discover.

## Parallel to parallel is not parallel – Thomas rotation

The limit speed has many strange consequences. Any two observers can keep a stick parallel to the other's, even if they are in motion with respect to each other. But strangely, given a chain of three or more sticks for which any two adjacent ones are parallel, the first and the last sticks will *not* generally be parallel. In particular, they *never* will be if the motions of the various observers are in different directions, as is the case when the velocity vectors form a loop.

The simplest set-up is shown in Figure 37. In special relativity, a general concatenation Ref. 64 of pure boosts does not give a pure boost, but a boost plus a rotation. As a result, the first and last stick in a chain of parallel sticks are usually not parallel.

An example of this effect appears in rotating motion. Imagine that we walk in a circle with relativistic speed holding a stick. We always keep the stick parallel to the direction it had just before. At the end of the turn, the stick will have an angle with respect to the direction at the start. Similarly, the *axis* of a rotating body circling a second body will *not* be pointing in the same direction after one turn. This effect is called *Thomas precession*, after Llewellyn Thomas, who discovered it in 1925, a full 20 years after the birth of special relativity. It had escaped the attention of dozens of other famous physicists. Thomas precession is important for the orbit of electrons inside atoms, where the stick is the spin axis of the rapidly orbiting electron. All these surprising phenomena are purely relativistic, and are thus measurable *only* in the case of speeds comparable to that of light.

## A never-ending story – temperature and relativity

What temperature is measured by an observer who moves with respect to a heat bath? The literature on the topic is confusing. Max Planck, Albert Einstein and Wolfgang Pauli agreed on the following result: the temperature $T$ seen by an observer moving with speed $v$ is related to the temperature $T_0$ measured by the observer at rest with respect to the heat bath via

$$T = T_0 \sqrt{1 - v^2/c^2} \; . \tag{25}$$

A moving observer thus always measures lower temperature values than a resting one.

In 1908, Max Planck used this expression, together with the corresponding transformation for thermal energy, to deduce that the entropy is invariant under Lorentz transformations. Being the discoverer of the Boltzmann constant $k$, Planck proved in this way
Ref. 65 that the Boltzmann constant is a relativistic invariant.

Not all researchers agree on the expression for the transformation of energy, however. (They do agree on the invaraince of $k$, though.) Others maintain that $T$ and $T_0$ should be interchanged in the formula. Also, powers other than the simple square root have been
Ref. 66 proposed. The origin of these discrepancies is simple: temperature is only defined for equilibrium situations, i.e., for baths. But a bath for one observer is not a bath for the other. For low speeds, a moving observer sees a situation that is *almost* a heat bath; but at higher speeds the issue becomes tricky. Temperature is deduced from the speed of matter particles, such as atoms or molecules. For rapidly moving observers, there is no good way to measure temperature, because the distribution is not in equilibrium. Any naively measured temperature value depends on the energy range of matter particles that is used! In short, thermal equilibrium is not an observer-invariant concept. Therefore, *no* temperature transformation formula is correct for high speeds. (Only with certain additional assumptions, Planck's expression holds.) In fact, there are not even any experimental observations that would allow such a formula to be checked. Realizing such a measurement is a challenge for future experimenters – but not for relativity itself.

# Chapter 2

# RELATIVISTIC MECHANICS

The speed of light is an invariant quantity and a limit value. Therefore, we need to rethink all observables that we defined with the help of velocity – thus all of them! The most basic observables are mass, momentum and energy. In other words, we need to recreate mechanics based on the invariant limit speed: we need to build *relativistic mechanics*.

## Mass in relativity

Page 88 In Galilean physics, the mass ratio between two bodies was defined using collisions; it was given by the negative inverse of the velocity change ratio

$$\frac{m_2}{m_1} = -\frac{\Delta v_1}{\Delta v_2} . \tag{26}$$

However, experiments show that this expression is wrong for speeds near that of light and must be changed. In fact, experiments are not needed: thinking alone can show this. Challenge 74 s Can you do so?

There is only one solution to this problem. Indeed, experiments confirm that the two Ref. 67 Galilean conservation theorems for momentum and for mass have to be changed into

$$\sum_i \gamma_i m_i \boldsymbol{v}_i = \text{const} \tag{27}$$

and

$$\sum_i \gamma_i m_i = \text{const} . \tag{28}$$

These expressions are the (relativistic) conservation of momentum and the (relativistic) conservation of mass–energy. They will remain valid throughout the rest of our ascent of Motion Mountain.

The conservation of momentum and energy implies, among other things, that telepor- Challenge 75 s tation is *not* possible in nature, in contrast to science fiction. Can you confirm this?

Obviously, in order to recover Galilean physics, the relativistic correction (factors) $\gamma_i$ have to be almost equal to 1 for everyday velocities, that is, for velocities nowhere near the speed of light. That is indeed the case. In fact, even if we did not know the expression of the relativistic correction factor, we can deduce it from the collision shown in Figure 38.

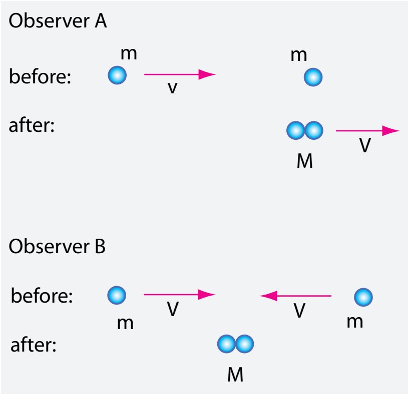



FIGURE 38 An inelastic collision of two identical particles seen from two different inertial frames of reference

In the first frame of reference (A) we have $\gamma_v mv = \gamma_V MV$ and $\gamma_v m + m = \gamma_V M$. From the observations of the second frame of reference (B) we deduce that $V$ composed with

Challenge 76 e

$V$ gives $v$, in other words, that

$$v = \frac{2V}{1 + V^2/c^2} . \tag{29}$$

When these equations are combined, the relativistic correction $\gamma$ is found to depend on the magnitude of the velocity $v$ through

$$\gamma_v = \frac{1}{\sqrt{1 - v^2/c^2}} . \tag{30}$$

With this expression the *mass* ratio between two colliding particles is defined as the ratio

$$\frac{m_1}{m_2} = -\frac{\Delta(\gamma_2 v_2)}{\Delta(\gamma_1 v_1)} . \tag{31}$$

This is the generalization of the definition of mass ratio from Galilean physics. (In the

Page 91

chapter on Galilean mechanics we also used a generalized mass definition based on acceleration ratios. We do not explore its relativistic generalization because it contains some subtleties which we will encounter shortly.) The correction factors $\gamma_i$ ensure that the mass defined by this equation is the same as the one defined in Galilean mechanics, and that it is the same for all types of collision a body may have.* In this way, mass remains a quantity characterizing the difficulty of accelerating a body, and it can still be used for *systems* of bodies as well.

Following the example of Galilean physics, we call the quantity

$$\boldsymbol{p} = \gamma m \boldsymbol{v} \tag{32}$$

the (*linear*) *relativistic* (*three-*) *momentum* of a particle. Total momentum is a *conserved*

Challenge 77 e

* The results below also show that $\gamma = 1 + T/mc^2$, where $T$ is the kinetic energy of a particle.

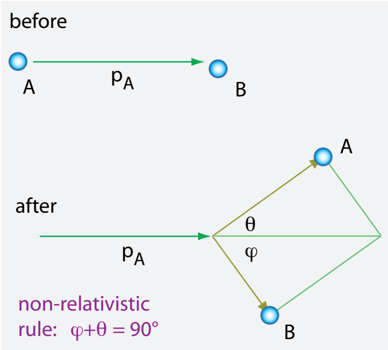



FIGURE 39 A useful rule for playing non-relativistic snooker

quantity for any system not subjected to external influences, and this conservation is a direct consequence of the way mass is defined.

For low speeds, or $\gamma \approx 1$, relativistic momentum is the same as Galilean momentum, and is then proportional to velocity. But for high speeds, momentum increases faster than velocity, tending to infinity when approaching light speed. The result is confirmed Page 34 by experimental data, as shown in Figure 17.

## Why relativistic snooker is more difficult

There is a well-known property of collisions between a moving sphere or particle and a resting one of the *same mass* that is important when playing snooker, pool or billiards. After such a collision, the two spheres will depart at a *right angle* from each other, as shown in Figure 39.

However, experiments show that the right angle rule does *not* apply to relativistic collisions. Indeed, using the conservation of momentum and a bit of dexterity you can Challenge 78 e calculate that

$$\tan\theta \tan\varphi = \frac{2}{\gamma+1} , \tag{33}$$

where the angles are defined in Figure 41. It follows that the sum $\varphi + \theta$ is *smaller* than a right angle in the relativistic case. Relativistic speeds thus completely change the game of snooker. Indeed, every accelerator physicist knows this: for electrons or protons, these angles can easily be deduced from photographs taken in cloud or bubble chambers, which show the tracks left by particles when they move through them, as shown in Figure 40. Ref. 12 All such photographs confirm the above expression. In fact, the *shapes* of detectors are chosen according to expression (33), as sketched in Figure 41. If the formula – and relativity – were wrong, most of these detectors would not work, as they would miss most of the particles after the collision. If relativity were wrong, such detectors would have to be much larger. In fact, these experiments also prove the formula for the composition of Challenge 79 ny velocities. Can you show this?

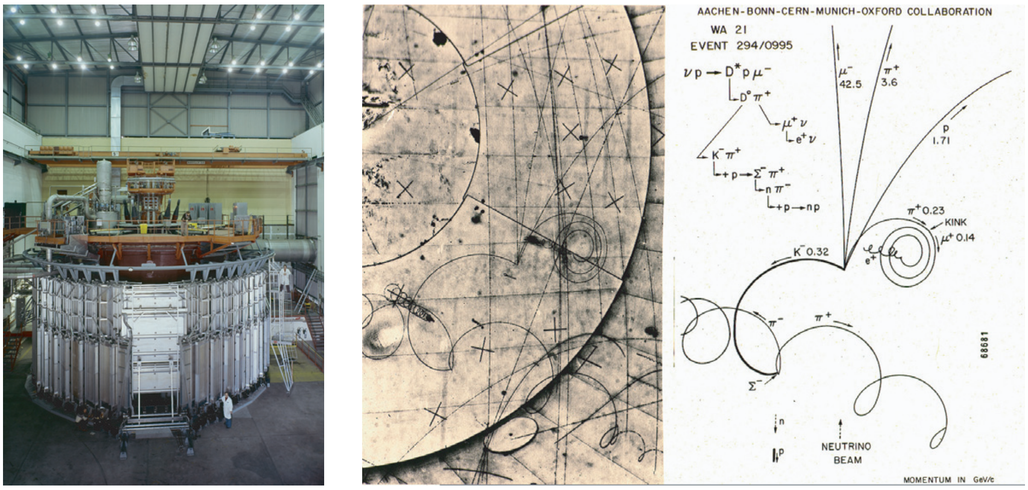



FIGURE 40 The 'Big European Bubble Chamber' and an example of tracks of relativistic particles it produced, with the momentum values deduced from the photograph (© CERN)

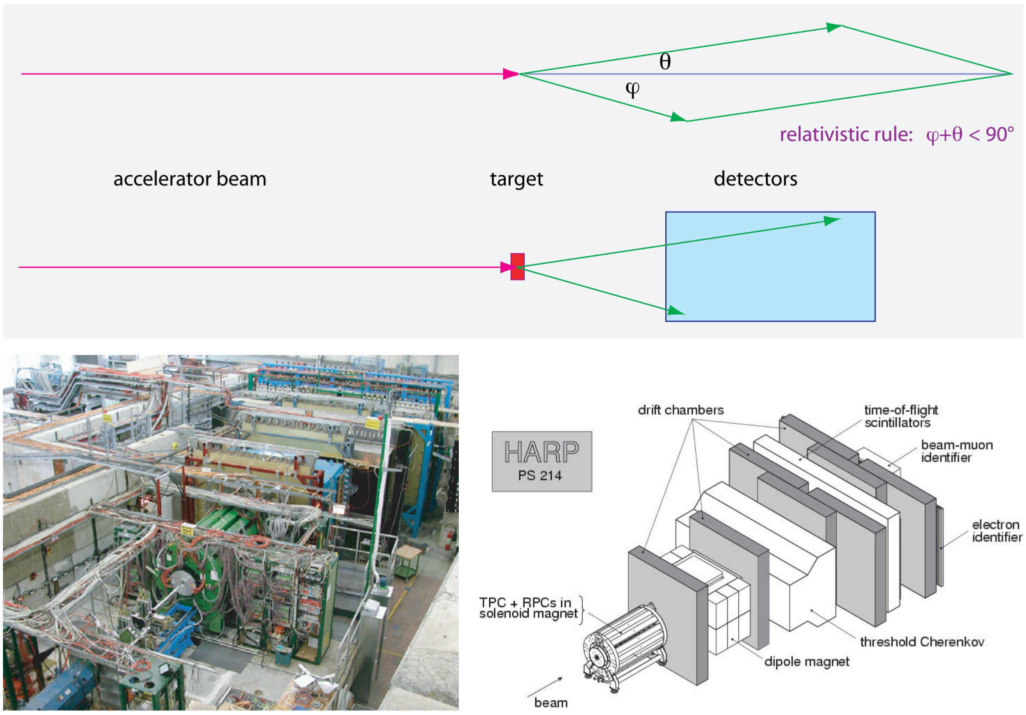



FIGURE 41 The dimensions of detectors for particle accelerators with single beams are based on the relativistic snooker angle rule – as an example, the HARP experiment at CERN (© CERN)

## Mass and energy are equivalent

Let us go back to the collinear and inelastic collision of Figure 38. What is the mass $M$ of the final system? Calculation shows that
Challenge 80 s

$$M/m = \sqrt{2(1+\gamma_v)} > 2 \ . \tag{34}$$

In other words, the mass of the final system is *larger* than the sum $2m$ of the two original masses. In contrast to Galilean mechanics, the sum of all masses in a system is *not* a conserved quantity. Only the sum $\sum_i \gamma_i m_i$ of the *corrected* masses is conserved.

Relativity provides the solution to this puzzle. Everything falls into place if, for the *energy* $E$ of an object of mass $m$ and velocity $v$, we use the expression

$$E = \gamma m c^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} \ , \tag{35}$$

applying it both to the total system and to each component. The conservation of the corrected mass can then be read as the conservation of energy, simply without the factor $c^2$. In the example of the two identical masses sticking to each other, the two parts are thus each described by mass and energy, and the resulting system has an energy $E$ given by the sum of the energies of the two parts. (We recall that the uncorrected masses do *not* add up.) In particular, it follows that the energy $E_0$ of a body *at rest* and its mass $m$ are related by

$$E_0 = mc^2 \ . \tag{36}$$

The expression $E = \gamma m c^2$ is perhaps the most beautiful and famous discovery of modern physics. In other words, the existence of a maximum speed implies that every mass has energy, and that energy has mass. Mass and energy are two terms for the same basic concept: they are *equivalent*.

Since mass and energy are equivalent, energy has all properties of mass. In particular, energy has inertia and weight. For example, a full battery is more massive and heavier than an empty one, and a warm glass of water is heavier than a cold one. Radio waves and light have weight. Conversely, mass has all properties of energy. For example, one can use mass to make engines run. But this is no news, as it is realized in every engine! Muscles, car engines or nuclear ships work by losing a tiny bit of mass and use the corresponding energy to overcome friction and move the person, car or ship.

Since $c^2$ is so large, we can also say that *mass is concentrated energy*. Increasing the energy of a system increases its mass a little bit, and decreasing the energy content decreases the mass a little bit. If a bomb explodes inside a closed box, the mass, weight and momentum of the box are the same before and after the explosion, but the combined mass of the debris inside the box will be a little bit *smaller* than before. All bombs – not only nuclear ones – thus take their power of destruction from a reduction in mass. In fact, every activity of a system – such as a caress, a smile or a look – takes its energy from a reduction in mass.

The *kinetic energy* $T$ is thus given by the difference between total energy and rest energy. This gives

$$T = \gamma mc^2 - mc^2 = \frac{1}{2}mv^2 + \frac{1\cdot 3}{2\cdot 4}m\frac{v^4}{c^2} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}m\frac{v^6}{c^4} + ... \quad (37)$$

Challenge 81 e
(using the binomial theorem). The expression reduces to the well-known Galilean value $T_{\text{Galilean}} = \frac{1}{2}mv^2$ only for low, everyday speeds.

The mass–energy equivalence $E = \gamma mc^2$ implies that extracting *any* energy from a material system results in a mass decrease. When a person plays the piano, thinks or runs, its mass decreases. When a cup of tea cools down or when a star shines, its mass decreases. When somebody uses somebody else's electric power, he is taking away some mass: electric power theft is thus mass theft! The mass–energy equivalence pervades all of nature.

There is just one known way to transform the *full* mass of a body into kinetic, in this case electromagnetic, energy: we annihilate it with the same amount of antimatter. Fortunately, there is almost no antimatter in the universe, so that the process does not occur in everyday life, because the energy content of even a speck of dust is already substantial.

Challenge 82 e

The mass–energy equivalence $E = \gamma mc^2$ means the death of many science fiction fantasies. It implies that there are *no* undiscovered sources of energy on or near Earth. If such sources existed, they would be measurable through their mass. Many experiments have looked for, and are still looking for, such effects with a negative result. There is no freely available energy in nature.*

## Weighing light

Challenge 83 e

The mass–energy equivalence $E = \gamma mc^2$ also implies that one needs about 90 thousand million kJ (or 21 thousand million kcal) to increase one's weight by one single gram. Of course, dieticians have slightly different opinions on this matter! As mentioned, humans do get their everyday energy from the material they eat, drink and breathe by reducing its combined mass before expelling it again; however, this *chemical mass defect* cannot yet be measured by weighing the materials before and after the reaction: the difference is too small, because of the large conversion factor $c^2$. Indeed, for any chemical reaction, bond energies are about 1 aJ (6 eV) per bond; this gives a weight change of the order of one part in $10^{10}$, too small to be measured by weighing people or determining mass differences between food and excrement. Therefore, for everyday chemical reactions mass can be taken to be constant, in accordance with Galilean physics.

Ref. 68

The mass–energy equivalence $E = \gamma mc^2$ has been confirmed by all experiments performed so far. The measurement is simplest for the nuclear mass defect. The most precise experiment, from 2005, compared the masses difference of nuclei before and after neutron capture on one hand, and emitted gamma ray energy on the other hand. The mass–energy relation was confirmed to a precision of more than 6 digits.

Page 197

* Two extremely diluted, yet somewhat mysterious forms of energy, called *dark matter* and (confusingly) *dark energy*, are distributed throughout the universe, with a density of about 1 nJ/m$^3$. Their existence is deduced from quite delicate measurements that detected their mass, but their nature has not yet been fully resolved.

Modern methods of mass measurement of single molecules have even made it possible to measure the *chemical* mass defect: it is now possible to compare the mass of a single molecule with that of its constituent atoms. David Pritchard's group has developed so-called *Penning traps*, which allow masses to be determined from the measurement of frequencies; the attainable precision of these cyclotron resonance experiments is sufficient to confirm $\Delta E_0 = \Delta mc^2$ for chemical bonds. In the future, bond energies will be determined in this way with high precision. Since binding energy is often radiated as light, we can also say that these modern techniques make it possible to *weigh* light.

Ref. 69

Thinking about light and its mass was the basis for Einstein's derivation of the mass–energy relation. When an object of mass $m$ emits two equal light beams of total energy $E$ in opposite directions, its own energy decreases by the emitted amount. Let us look at what happens to its mass. Since the two light beams are equal in energy and momentum, the body does not move, and we cannot deduce anything about its mass change. But we can deduce something if we describe the same situation when moving with the nonrelativistic velocity $v$ along the beams. We know that due to the Doppler effect one beam is red-shifted and the other blue-shifted, by the factors $1 + v/c$ and $1 - v/c$. The blue-shifted beam therefore acquires an extra momentum $vE/2c^2$ and the red-shifted beam loses momentum by the same amount. In nature, momentum is conserved. Therefore, after emission, we find that the body has a momentum $p = mv - vE/c^2 = v(m - E/c^2)$. We thus conclude that a body that loses an energy $E$ reduces its mass by $E/c^2$. This is the equivalence of mass and energy.

Challenge 84 e

In short, we find that the *rest energy* $E_0$ of an object, the maximum energy that can be extracted from a mass $m$, is

$$E_0 = mc^2 . \tag{38}$$

We saw above that the Doppler effect is a consequence of the invariance of the speed of light. Whenever the invariance of the speed of light is combined with momentum and energy conservation we find the equivalence of mass and energy.

How are momentum and energy related? The definitions of momentum (32) and energy (35) lead to two basic relations. First of all, their *magnitudes* are related by

Challenge 85 e

$$m^2c^4 = E^2 - p^2c^2 \tag{39}$$

for all relativistic systems, be they objects or, as we will see below, radiation. For the momentum *vector* we get the other important relation

$$\boldsymbol{p} = \frac{E}{c^2}\boldsymbol{v} , \tag{40}$$

Challenge 86 e

which is equally valid for *any* type of moving energy, be it an object or a beam or pulse of radiation.* We will use both relations often in the rest of our ascent of Motion Mountain, including the following discussion.

* Using 4-vector notation, we can write $\boldsymbol{v}/c = \boldsymbol{p}/P_0$, where $P_0 = E/c$.

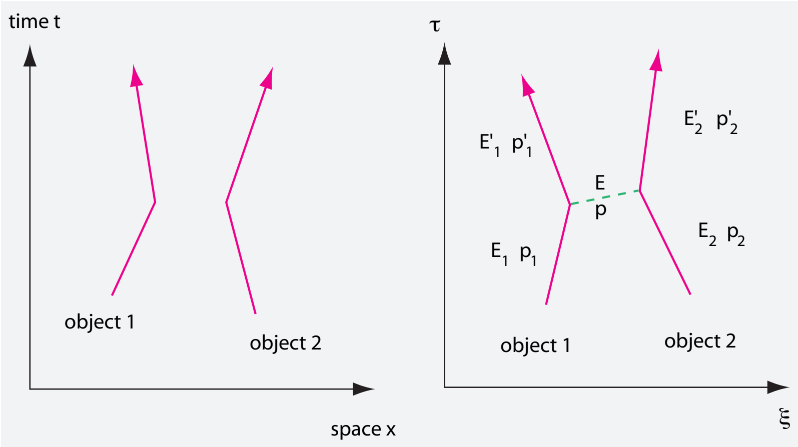



FIGURE 42 Space-time diagrams of the same collision for two different observers

## Collisions, virtual objects and tachyons

We have just seen that in relativistic collisions the conservation of total energy and momentum are intrinsic consequences of the definition of mass. Let us now have a look at collisions in more detail. A *collision* is a process, i.e., a series of events, for which

— the total momentum before the interaction and after the interaction is the same;
— the momentum is exchanged in a small region of space-time;
— for small velocities, the Galilean description is valid.

In everyday life an *impact* is the event at which both objects change momentum. But the two colliding objects are located at *different* points when this happens. A collision is
Ref. 70 therefore described by a space-time diagram such as the left-hand one in Figure 42; it is reminiscent of the Orion constellation. It is easy to check that the process described by
Challenge 87 e such a diagram is a collision according to the above definition.

The right-hand side of Figure 42 shows the same process seen from another, Greek, frame of reference. The Greek observer says that the first object has changed its momentum *before* the second one. That would mean that there is a short interval when momentum and energy are *not* conserved!

The only way to make sense of the situation is to assume that there is an exchange of a third object, drawn with a dotted line. Let us find out what the properties of this object are. We give numerical subscripts to the masses, energies and momenta of the two bodies,
Challenge 88 e and give them a prime after the collision. Then the unknown mass $m$ obeys

$$m^2c^4 = (E_1 - E_1')^2 - (p_1 - p_1')^2c^2 = 2m_1^2c^4 - 2E_1E_1'\left(\frac{1 - v_1v_1'}{c^2}\right) < 0\,. \qquad (41)$$

This is a strange result, because it means that the unknown mass is an *imaginary* number!* On top of that, we also see directly from the second graph that the exchanged object moves faster than light. It is a *tachyon*, from the Greek ταχύς 'rapid'. In other words, collisions involve motion that is faster than light! We will see later that collisions are indeed the *only* processes where tachyons play a role in nature. Since the exchanged objects appear only during collisions, never on their own, they are called *virtual* objects, to distinguish them from the usual, *real* objects, which we observe everyday.** We will study the properties of virtual particle later on, when we come to discuss quantum theory. Page 55, page 164

In nature, a tachyon is always a virtual object. Real objects are always *bradyons* – from the Greek βραδύς 'slow' – or objects moving slower than light. Note that tachyons, despite their high velocity, do not allow the transport of energy faster than light; and that they do not violate causality if and only if they are emitted or absorbed with equal probability. Can you confirm all this? Challenge 89 ny

When we will study quantum theory, we will also discover that a general contact interaction between objects is described not by the exchange of a *single* virtual object, but by a continuous *stream* of virtual particles. For standard collisions of everyday objects, the interaction turns out to be electromagnetic. In this case, the exchanged particles are virtual photons. In other words, when one hand touches another, when it pushes a stone, or when a mountain supports the trees on it, streams of virtual photons are continuously exchanged. Vol. IV, page 54

There is an additional secret hidden in collisions. In the right-hand side of Figure 42, the tachyon is emitted by the first object and absorbed by the second one. However, it is easy to imagine an observer for which the opposite happens. Challenge 90 s In short, the direction of travel of a tachyon depends on the observer! In fact, this is a hint about *antimatter*. In space-time diagrams, matter and antimatter travel in opposite directions. The connection between relativity and antimatter will become more apparent in quantum theory. Vol. IV, page 163

## Systems of particles – no centre of mass

Relativity also forces us to eliminate the cherished concept of *centre of mass*. We can see this already in the simplest example possible: that of two equal objects colliding.

Figure 43 shows that from the viewpoint in which one of two colliding particles is at rest, there are at least three different ways to define the centre of mass. In other words, the centre of mass is not an observer-invariant concept. Ref. 71 We can deduce from the figure that the concept only makes sense for systems whose components move with *small* velocities relative to each other. An atom is an example. For more general systems, centre of mass is not uniquely definable. Will this hinder us in our ascent? No. We are more interested in the motion of single particles than that of composite objects or systems.

* It is usual to change the mass–energy and mass–momentum relation of tachyons to $E = \pm mc^2/\sqrt{v^2/c^2 - 1}$ and $p = \pm mv/\sqrt{v^2/c^2 - 1}$; this amounts to a redefinition of $m$. After the redefinition, tachyons have *real* mass. The energy and momentum relations show that tachyons lose energy and momentum when they get faster. (Provocatively, a single tachyon in a box could provide humanity with all the energy we need.) Both signs for the energy and momentum relations must be retained, because otherwise the equivalence of all inertial observers would not be generated. Tachyons thus do not have a minimum energy or a minimum momentum.

** More precisely, a virtual particle does not obey the relation $m^2c^4 = E^2 - p^2c^2$, valid for real particles.

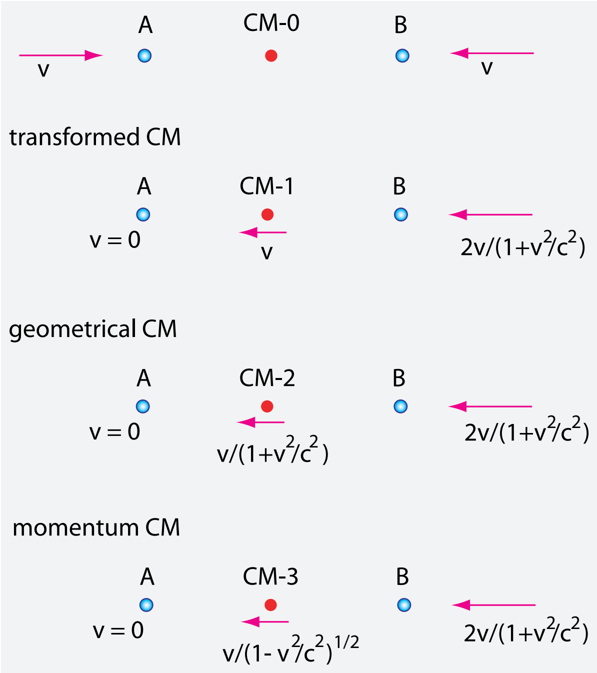



FIGURE 43 There is no consistent way to define a relativistic centre of mass

## Why is most motion so slow?

For most everyday systems, dilation factors $\gamma$ are very near to 1; noticeable departures from 1, thus speeds of more than a few per cent of the speed of light, are uncommon. Most such situations are microscopic. We have already mentioned the electrons inside a television tube or inside a particle accelerator. The particles making up cosmic radiation are another example; it is important, because their high energy has produced many of the mutations that are the basis of evolution of animals and plants on this planet. Later we will discover that the particles involved in radioactivity are also relativistic.

But why don't we observe any relativistic *macroscopic* bodies? Because the universe exists since as long time. Bodies that collide with relativistic velocities undergo processes not found in everyday life: when they collide, part of their kinetic energy is converted into new matter via $E = \gamma mc^2$. In the history of the universe this has happened so many times that practically all macroscopic bodies move with low speed with respect to their environment, and practically all of the bodies still in relativistic motion are microscopic particles.

A second reason for the disappearance of rapid relative motion is radiation damping. Can you imagine what happens to relativistic charges during collisions, or in a bath of light? Radiation damping also slows down microscopic particles. Challenge 91 s

In short, almost all matter in the universe moves with small velocity relative to other matter. The few known counter-examples are either very old, such as the quasar jets mentioned above, or stop after a short time. For example, the huge energies necessary for macroscopic relativistic motion are available in supernova explosions, but the relativistic motion ceases to exist after a few weeks. In summary, the universe is mainly filled

Page 206 with slow motion because it is *old*. We will determine its age shortly.

### The history of the mass–energy equivalence formula

Albert Einstein took several months after his first paper on special relativity to deduce the expression

$$E = \gamma m c^2 \tag{42}$$

which is often called the most famous formula of physics. He published it in a second, Ref. 15 separate paper towards the end of 1905. Arguably, the formula could have been discovered thirty years earlier, from the theory of electromagnetism.

In fact, several persons deduced similar results before Einstein. In 1903 and 1904, Ref. 73 *before* Einstein's first relativity paper, Olinto De Pretto, a little-known Italian engineer, calculated, discussed and published the formula $E = mc^2$.It might well be that Einstein got the idea for the formula from De Pretto,* possibly through Einstein's friend Michele Besso or other Italian-speaking friends he met when he visited his parents, who were living in Italy at the time. Of course, the value of Einstein's efforts is not diminished by this.

Ref. 73 In fact, a similar formula had also been deduced in 1904 by Friedrich Hasenöhrl and published again in Annalen der Physik in 1905, before Einstein, though with an incorrect numerical factor, due to a calculation mistake. The formula $E = mc^2$ is also part of several expressions in two publications in 1900 by Henri Poincaré. Also Paul Langevin knew the formula, and Einstein said of him that he would surely have discovered the theory of special relativity had it not been done before. The real hero in the story might well be Tolver Preston, who discussed the equivalence of mass and energy already in 1875, in his book *Physics of the Ether*. The mass–energy equivalence was thus indeed floating in the air, waiting to be understood and put into the correct context.

Page 103 In the 1970s, a similar story occurred: a simple relation between the acceleration and the temperature of the vacuum was discovered. The result had been waiting to be discovered for over 50 years. Indeed, a number of similar, anterior results were found in the Challenge 92 s libraries. Could other simple relations be hidden in modern physics waiting to be found?

### 4-vectors

How can we describe motion consistently for *all* observers? We have to introduce a simple idea: 4-vectors. We already know that the motion of a particle can be seen as a sequence of *events*. Events are points in space-time. To describe events with precision, we introduce event coordinates, also called *4-coordinates*. These are written as

$$\boldsymbol{X} = (ct, \boldsymbol{x}) = (ct, x, y, z) = X^i \ . \tag{43}$$

In this way, an event is a point in four-dimensional space-time, and is described by four coordinates. The four coordinates are called the *zeroth*, namely time $X^0 = ct$, the *first*, usually called $X^1 = x$, the *second*, $X^2 = y$, and the *third*, $X^3 = z$. In fact, $\boldsymbol{X}$ is the simplest

* Umberto Bartocci, mathematics professor of the University of Perugia in Italy, published the details of Ref. 72 this surprising story in several papers and in a book.

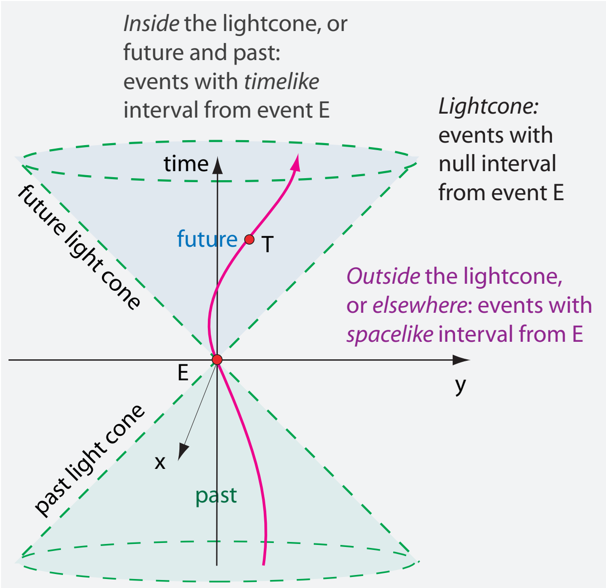



FIGURE 44 The space-time diagram of a moving object T, with one spatial dimension missing

example of a *4-vector*. The old vectors $\boldsymbol{x}$ of Galilean physics are also called *3-vectors*. We see that time is treated like the zeroth of four dimensions.

We can now define a *space-time distance* or *space-time interval* between two events as the length of the difference vector $\boldsymbol{X}$. In fact, we usually use the square of the length, the *magnitude*, to avoid those unwieldy square roots. In special relativity, the magnitude $\boldsymbol{X}^2$ of any 4-vector $\boldsymbol{X}$ is defined as

$$\boldsymbol{X}^2 = X_0{}^2 - X_1{}^2 - X_2{}^2 - X_3{}^2 = ct^2 - x^2 - y^2 - z^2 = X_a X^a = \eta_{ab} X^a X^b = \eta^{ab} X_a X_b \,. \quad (44)$$

Page 38 The squared space-time interval is thus the squared time interval *minus* the squared length interval. We have seen above that this minus sign results from the invariance of the speed of light. In contrast to a squared space interval, a squared space-time interval can be positive, negative or even zero.

How can we imagine the space-time interval? The magnitude of the *space-time interval* is the square of $c$ times the proper time. The *proper time* is the time shown by a clock moving in a straight line and with constant velocity between two events in space-time. For example, if the start and end events in space-time require motion with the speed of light, the proper time and the space-time interval vanish. This situation defines the so-called null vectors or *lightlike* intervals. We call the set of all null vector end points the light cone; it is shown in Figure 44. If the motion between two events is *slower* than the speed of light, the squared proper time is positive and the space-time interval is called timelike. For negative space-time intervals the interval is called spacelike. In this last case, the negative of the magnitude, which then is a positive number, is called the squared *proper distance*. The proper distance is the length measured by an odometer as the object moves along. Page 42

We note that the definition of the light cone, its interior and its exterior, are *observer-*

Challenge 93 e *invariant*. We therefore use these concepts regularly.

In the definition for the space-time interval we have introduced for the first time two notations that are useful in relativity. First of all, we automatically sum over repeated indices. Thus, $X_a X^a$ means the sum of all products $X_a X^a$ as $a$ ranges over all indices. Secondly, for every 4-vector $\boldsymbol{X}$ we distinguish two ways to write the coordinates, namely coordinates with superscripts and coordinates with subscripts. (For 3-vectors, we only use subscripts.) They are related by the following general relation

$$\begin{aligned} X^b &= (ct, x, y, z) \\ X_a &= (ct, -x, -y, -z) = \eta_{ab} X^b \,, \end{aligned} \tag{45}$$

where we have introduced the so-called *metric* $\eta^{ab}$, an abbreviation of the matrix*

$$\eta^{ab} = \eta_{ab} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{46}$$

Don't panic: this is all, and it won't get more difficult! (A generalization of this matrix is used later on, in general relativity.) We now go back to physics; in particular, we are now ready to describe motion in space-time.

## 4-velocity

We now define velocity of an body in a way that is useful for all observers. We cannot define the velocity as the derivative of its coordinates with respect to time, since time and temporal sequences depend on the observer. The solution is to define all observables with respect to the just-mentioned *proper time* $\tau$, which is defined as the time shown by a clock attached to the body. In relativity, motion and change are always measured with respect to clocks attached to the moving system.

Therefore the *relativistic velocity* or *4-velocity* $\boldsymbol{U}$ of an body is defined as the rate of change of its 4-coordinates $\boldsymbol{X} = (ct, \boldsymbol{x})$ with respect to proper time, i.e., as

$$\boldsymbol{U} = \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}\tau} . \tag{47}$$

The coordinates $\boldsymbol{X}$ are measured in the coordinate system defined by the chosen inertial observer. The value of the 4-velocity $\boldsymbol{U}$ depends on the observer or coordinate system used, as does usual velocity in everyday life. Using $\mathrm{d}t = \gamma\, \mathrm{d}\tau$ and thus

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma \frac{\mathrm{d}x}{\mathrm{d}t} \quad \text{, where as usual} \quad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}} , \tag{48}$$

* This is the so-called *timelike convention*, used in about 70 % of all physics texts. Note that 30 % of all physics textbooks use the negative of $\eta$ as the metric, the so-called *spacelike convention*, and thus have opposite signs in this definition.

we get the relation of 4-velocity with the 3-velocity $\boldsymbol{v} = \mathrm{d}\boldsymbol{x}/\mathrm{d}t$:

$$U^0 = \gamma c \;,\; U^i = \gamma v_i \quad \text{or} \quad \boldsymbol{U} = (\gamma c, \gamma \boldsymbol{v}) \;. \tag{49}$$

For small velocities we have $\gamma \approx 1$, and then the last three components of the 4-velocity are those of the usual, Galilean 3-velocity. For the magnitude of the 4-velocity $\boldsymbol{U}$ we find $\boldsymbol{UU} = U_a U^a = \eta_{ab} U^a U^b = c^2$, which is therefore independent of the magnitude of the 3-velocity $\boldsymbol{v}$ and makes it a timelike vector, i.e., a vector *inside* the light cone.

In general, a *4-vector* is defined as a quantity $(H^0, H^1, H^2, H^3)$ that transforms under boosts as

$$\begin{aligned} H^0_V &= \gamma_V (H^0 - H^1 V/c) \\ H^1_V &= \gamma_V (H^1 - H^0 V/c) \\ H^2_V &= H^2 \\ H^3_V &= H^3 \end{aligned} \tag{50}$$

when changing from one inertial observer to another moving with a relative velocity $V$ in the $x$ direction; the corresponding generalizations for the other coordinates are understood. This relation allows us to deduce the relativistic transformation laws for any 3-vector. Can you deduce the 3-velocity composition formula (9) from this definition?¥¥¥«««««««« Challenge 94 s

We know that the magnitude of a 4-vector can be zero even though all its components are different from zero. Such a vector is called *null*. Which motions have a null velocity vector? Challenge 95 s

## 4-acceleration and proper acceleration

Similarly, the *4-acceleration* $\boldsymbol{B}$ of a body is defined as

$$\boldsymbol{B} = \frac{\mathrm{d}\boldsymbol{U}}{\mathrm{d}\tau} = \frac{\mathrm{d}^2\boldsymbol{X}}{\mathrm{d}\tau^2} \;. \tag{51}$$

Using $\mathrm{d}\gamma/\mathrm{d}\tau = \gamma \mathrm{d}\gamma/\mathrm{d}t = \gamma^4 \boldsymbol{va}/c^2$, we get the following relations between the four components of $\boldsymbol{B}$ and the 3-acceleration $\boldsymbol{a} = \mathrm{d}\boldsymbol{v}/\mathrm{d}t$: Ref. 74

$$B^0 = \gamma^4 \frac{\boldsymbol{va}}{c} \quad , \quad B^i = \gamma^2 a_i + \gamma^4 \frac{(\boldsymbol{va}) v_i}{c^2} \;. \tag{52}$$

Challenge 96 e The magnitude $B$ of the 4-acceleration is easily found via $\boldsymbol{BB} = \eta_{cd} B^c B^d = -\gamma^4 (a^2 + \gamma^2 (\boldsymbol{va})^2/c^2) = -\gamma^6 (a^2 - (\boldsymbol{v} \times \boldsymbol{a})^2/c^2)$. Note that the magnitude does depend on the value of the 3-acceleration $\boldsymbol{a}$. We see that a body that is accelerated for one inertial observer is also accelerated for all other inertial observers. We also see directly that 3-accelerations are *not* Lorentz invariant, unless the velocities are small compared to the speed of light. *Different inertial observers measure different 3-accelerations.* This is in contrast to our everyday experience and to Galilean physics, where accelerations are independent of the speed of the observer.

We note that 4-acceleration lies *outside* the light cone, i.e., that it is a spacelike vector. We also note that $\boldsymbol{BU} = \eta_{cd}B^c U^d = 0$, which means that the 4-acceleration is always perpendicular to the 4-velocity.*

When the 3-acceleration $\boldsymbol{a}$ is parallel to the 3-velocity $\boldsymbol{v}$, we get $B = \gamma^3 a$; when $\boldsymbol{a}$ is perpendicular to $\boldsymbol{v}$, as in circular motion, we get $B = \gamma^2 a$. We will use this result shortly.

How does the 3-acceleration change from one inertial observer to another? To simplify the discussion, we introduce the so-called *comoving observer*, the observer for which a particle is at rest. We call the magnitude of the 3-acceleration for the *comoving observer* the comoving or *proper acceleration*; in this case $\boldsymbol{B} = (0, \boldsymbol{a})$ and $\boldsymbol{B}^2 = -a^2$. Proper acceleration describes what the comoving observer *feels*: proper acceleration describes the experience of being pushed into the back of the accelerating seat. Proper acceleration is the most important and useful concept when studying accelerated motion in relativity.

Proper acceleration is an important quantity, because no observer, whatever his speed relative to the moving body, ever measures a 3-acceleration that is higher than the proper acceleration, as we will see now.

We can calculate how the value of 3-acceleration $\boldsymbol{a}$ measured by a general *inertial* observer is related to the proper acceleration $\boldsymbol{a}_c$ measured by the comoving observer using expressions (52) and (50). In this case $\boldsymbol{v}$ is both the relative speed of the two observers and the speed of the accelerated particle. We get

Ref. 80

$$a^2 = \frac{1}{\gamma_v^4}\left(a_c^2 - \frac{(\boldsymbol{a}_c\boldsymbol{v})^2}{c^2}\right) , \qquad (55)$$

Page 71

which we know already in a slightly different form. It shows (again) that the comoving or proper 3-acceleration is always *larger* than the 3-acceleration measured by any other inertial observer. The faster an inertial observer is moving relative to the accelerated system, the smaller the 3-acceleration he observes. The expression also confirms that whenever the speed is perpendicular to the acceleration, a boost yields a factor $\gamma_v^2$, whereas a speed parallel to the acceleration gives the already mentioned factor $\gamma_v^3$.

Challenge 98 e

The maximum property of proper acceleration implies that accelerations, in contrast to velocities, *cannot* be called relativistic. In other words, accelerations require relativistic treatment only when the involved velocities are relativistic. If the velocities involved are low, even the highest accelerations can be treated with Galilean physics.

* Similarly, the *4-jerk* $\boldsymbol{J}$ of a body is defined as

$$\boldsymbol{J} = \mathrm{d}\boldsymbol{B}/\mathrm{d}\tau = \mathrm{d}^2\boldsymbol{U}/\mathrm{d}\tau^2 . \qquad (53)$$

Challenge 97 e

For the relation with the 3-jerk $\boldsymbol{j} = \mathrm{d}\boldsymbol{a}/\mathrm{d}t$ we then get

$$\boldsymbol{J} = (J^0, J^i) = \left( \frac{\gamma^5}{c}(\boldsymbol{jv} + a^2 + 4\gamma^2\frac{(\boldsymbol{va})^2}{c^2}) , \gamma^3 j_i + \frac{\gamma^5}{c^2}((\boldsymbol{jv})v_i + a^2 v_i + 4\gamma^2\frac{(\boldsymbol{va})^2 v_i}{c^2} + 3(\boldsymbol{va})a_i) \right) \qquad (54)$$

Page 84

which we will use later on. Surprisingly, $\boldsymbol{J}$ does not vanish when $\boldsymbol{j}$ vanishes. Why not?

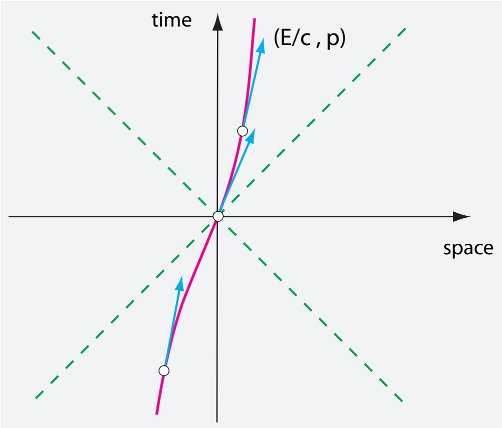



FIGURE 45 Energy–momentum is tangent to the world line

## 4-MOMENTUM OR ENERGY–MOMENTUM OR MOMENERGY

To describe motion, we need the concept of momentum. The *4-momentum* is defined as

$$\boldsymbol{P} = m\boldsymbol{U} \tag{56}$$

and is therefore related to the 3-momentum $\boldsymbol{p}$ by

$$\boldsymbol{P} = (\gamma mc, \gamma m\boldsymbol{v}) = (E/c, \boldsymbol{p}) . \tag{57}$$

For this reason 4-momentum is also called the *energy–momentum* 4-vector. In short, *the 4-momentum of a body is given by the mass times 4-displacement per proper time*. This is the simplest possible definition of momentum and energy. The concept was introduced by Max Planck in 1906.

The energy–momentum 4-vector, sometimes also called *momenergy*, is, like the 4-velocity, *tangent* to the world line of a particle. This connection, shown in Figure 45, follows directly from the definition, since

$$(E/c, \boldsymbol{p}) = (\gamma mc, \gamma m\boldsymbol{v}) = m(\gamma c, \gamma \boldsymbol{v}) = m(c\mathrm{d}t/\mathrm{d}\tau, \mathrm{d}\boldsymbol{x}/\mathrm{d}\tau) . \tag{58}$$

The (square of the) length of momenergy, namely $\boldsymbol{PP} = \eta_{ab}P^aP^b$, is, like any squared length of a 4-vector, the same for all inertial observers; it is found to be

$$E^2/c^2 - p^2 = m^2c^2 , \tag{59}$$

thus confirming a result given above. We have already mentioned that energies or situations are called *relativistic* if the kinetic energy $T = E - E_0$ is not negligible when compared to the rest energy $E_0 = mc^2$. A particle whose kinetic energy is *much* higher than its rest mass is called *ultrarelativistic*. Particles in accelerators or in cosmic rays fall into this category. What is their energy–momentum relation?

Challenge 99 s

The conservation of energy, momentum and mass of Galilean mechanics thus merge, in special relativity, into the conservation of momenergy. In short, in nature *momenergy is conserved.* In particular, mass is not a conserved quantity any more.

In contrast to Galilean mechanics, relativity implies an absolute zero for the energy. We cannot extract more energy than $mc^2$ from a system of mass $m$. In particular, a zero value for potential energy is fixed in this way. In short, relativity shows that energy is bounded from below. There is no infinite amount of energy available in nature.

Not all Galilean energy contributes to mass: potential energy in an outside field does not. Relativity forces us into precise energy bookkeeping. We keep in mind for later that 'potential energy' in relativity is an abbreviation for 'energy reduction of the outside field'.

Can you show that for two particles with 4-momenta $P_1$ and $P_2$, one has $P_1 P_2 = m_1 E_2 = m_2 E_1 = c^2 \gamma_{12} m_1 m_2$, where $\gamma_{12}$ is the Lorentz factor due to their relative velocity $v_{12}$? Challenge 100 s

Note that by the term 'mass' $m$ we always mean what is sometimes called the *rest mass*. This name derives from the bad habit of many science fiction and secondary-school books of calling the product $\gamma m$ the *relativistic mass*. Workers in the field usually (but not unanimously) reject this concept, as did Einstein himself, and they also reject the often-heard expression that '(relativistic) mass increases with velocity'. Relativistic mass and energy would then be two words for the same concept: this way to talk is at the level of the tabloid press. Ref. 75

## 4-force

The 4-force $\boldsymbol{K}$ is defined with 4-momentum $\boldsymbol{P}$ as

$$\boldsymbol{K} = \mathrm{d}\boldsymbol{P}/\mathrm{d}\tau = m\boldsymbol{B} \ . \tag{60}$$

Therefore force remains equal to mass times acceleration in relativity. From the definition of $\boldsymbol{K}$ we deduce the relation with 3-force $\boldsymbol{F} = \mathrm{d}\boldsymbol{p}/\mathrm{d}t = m\mathrm{d}(\gamma\boldsymbol{v})/\mathrm{d}t$, namely* Ref. 74, Ref. 76

$$\boldsymbol{K} = (K^0, K^i) = \left(\gamma^4 \frac{m\boldsymbol{va}}{c}, \gamma^2 m a_i + \gamma^4 v_i \frac{m\boldsymbol{va}}{c^2}\right) = \left(\frac{\gamma}{c}\frac{\mathrm{d}E}{\mathrm{d}t}, \gamma\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}\right) = \left(\gamma\frac{\boldsymbol{Fv}}{c}, \gamma\boldsymbol{F}\right) \ . \tag{61}$$

The 4-force, like the 4-acceleration, is orthogonal to the 4-velocity. The meaning of the zeroth component of the 4-force can easily be discerned: it is the *power* required to accelerate the object. Indeed, we have $\boldsymbol{KU} = c^2 \mathrm{d}m/\mathrm{d}\tau = \gamma^2(\mathrm{d}E/\mathrm{d}t - \boldsymbol{Fv})$: this is the proper rate at which the internal energy of a system increases. The product $\boldsymbol{KU}$ vanishes only for rest-mass-conserving forces. Many particle collisions lead to reactions and thus do not belong to this class of forces; such collisions and forces do not conserve rest mass. In everyday life however, the rest mass is preserved, and then we get the Galilean expression for power given by $\boldsymbol{Fv} = \mathrm{d}E/\mathrm{d}t$. Challenge 101 e

For rest-mass-preserving forces we get $\boldsymbol{F} = \gamma m\boldsymbol{a} + (\boldsymbol{Fv})\boldsymbol{v}/c^2$. In other words, in the general case, 3-force and 3-acceleration are neither parallel nor proportional to each other. In contrast, 3-momentum is parallel, but not proportional to 3-velocity. Challenge 102 s

We note that 3- force has the largest possible value, the *proper force*, in the comoving frame. A boost keeps the component of the force in the direction of the boost unchanged,

* Some authors define 3-force as $\mathrm{d}\boldsymbol{p}/\mathrm{d}\tau$; then $\boldsymbol{K}$ looks slightly different.

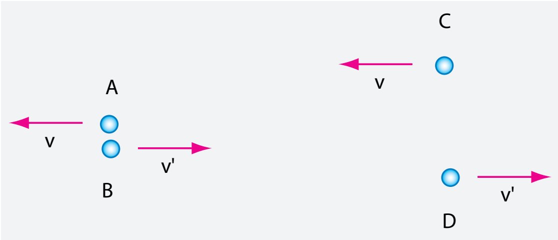



FIGURE 46 On the definition of relative velocity

Challenge 103 ny and reduces the compenents in the perpendicular directions. In particular, boost cannot be used to increase 3-force values beyond all bounds. The situation is somewhat recalls Page 72 the situation for 3-acceleration, though the transformation behaviour differs.

The 4-force can thus also be called the *power–force* 4-vector. In Galilean mechanics, when we defined force, we also explored potentials. However, we cannot do this easily in special relativity. In contrast to Galilean mechanics, where interactions and potentials can have almost any desired behaviour, special relativity has strict requirements for them. There is no way to define potentials and interactions in a way that makes sense Ref. 77 for all observers – except if the potentials are related to fields that can carry energy and momentum. In other terms, *relativity only allows potentials related to radiation.* In fact, only two type of potentials are allowed by relativity in everyday life: those due to electromagnetism and those due to gravity. (In the microscopic domain, also the two nuclear interactions are possible.) In particular, this result implies that when two everyday objects collide, the collision is either due to gravitational or to electric effects. To put it even more bluntly: relativity forbids 'purely mechanical' interactions. Mechanics is not a fundamental part of nature. Indeed, in the volume on quantum theory we will confirm that everything that we call *mechanical* in everyday life is, without exception, *electromagnetic*. Every caress and every kiss is an electromagnetic process. To put it in another way, and using the fact that light is an electromagnetic process, we can say: if we bang two objects hard enough onto each other, we will inevitably produce light.

The inclusion of gravity into relativity yields the theory of general relativity. In general relativity, the just defined power–force vector will play an important role. It will turn out that in nature, the 3-force $\boldsymbol{F}$ and the 3-power $\boldsymbol{F v}$ are limited in magnitude. Can you guess Challenge 104 d how?

## Rotation in relativity

If at night we turn around our own axis while looking at the sky, the stars move with a velocity much higher than that of light. Most stars are masses, not images. Their speed should be limited by that of light. How does this fit with special relativity?

This example helps to clarify in another way what the limit velocity actually is. Physically speaking, a rotating sky does *not* allow superluminal energy transport, and thus does not contradict the concept of a limit speed. Mathematically speaking, the speed of light limits relative velocities only between objects that come *near* to each other, as shown on the left of Figure 46. To compare velocities of *distant* objects, like between ourselves and the stars, is only possible if all velocities involved are constant in time; this is not the case if we turn. The differential version of the Lorentz transformations make this point

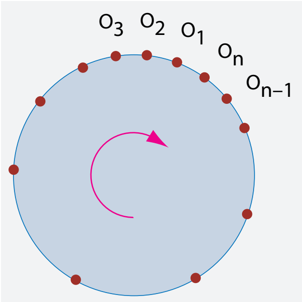



FIGURE 47 Observers on a rotating object

particularly clear. Indeed, the relative velocities of *distant* objects are frequently higher than the speed of light. We encountered one example earlier, when discussing the car in the tunnel, and we will encounter a more examples shortly.

Page 54 Page 89

With this clarification, we can now briefly consider *rotation* in relativity. The first question is how lengths and times change in a rotating frame of reference. You may want to check that an observer in a rotating frame agrees with a non-rotating colleague on the radius of a rotating body; however, both find that the rotating body, even if it is rigid, has a circumference *different* from the one it had before it started rotating. Sloppily speaking, the value of π *changes* for rotating observers! For the rotating observer, the ratio between the circumference $c$ and the radius $r$ turns out to be $c/r = 2\pi\gamma$: the ratio increases with rotation speed. This counter-intuitive result is often called *Ehrenfest's paradox*. It shows that space-time for a rotating observer is *not* the flat Minkowski space-time of special relativity. The paradox also shows that rigid bodies do not exist.

Challenge 105 e Challenge 106 e Ref. 78

Rotating bodies behave strangely in many ways. For example, we get into trouble when we try to synchronize clocks mounted on a rotating circle, as shown in Figure 47. If we start synchronizing the clock at position $O_2$ with that at $O_1$, and so on, continuing up to last clock $O_n$, we find that the last clock is *not* synchronized with the first. This result reflects the change in circumference just mentioned. In fact, a careful study shows that the measurements of length and time intervals lead all observers $O_k$ to conclude that they live in a rotating space-time, one that is not flat. Rotating discs can thus be used as an introduction to general relativity, where spatial curvature and its effects form the central topic. More about this in the next chapter.

Ref. 20 In relativity, rotation and translation combine in strange ways. Imagine a cylinder in uniform rotation along its axis, as seen by an observer at rest. As Max von Laue has discussed, the cylinder will appear *twisted* to an observer moving along the rotation axis. Can you confirm this?

Challenge 107 e

For train lovers, here is a well-known puzzle. A train travels on a circular train track. The train is as long as the track, so that is forms a circle. What happens if the same train runs at relativistic speeds: does the train fall out of the track, remain on the track or fall inside the track?

Challenge 108 ny

Is angular velocity limited? Yes: the tangential speed in an inertial frame of reference cannot reach that of light. The limit on angular velocity thus depends on the *size* of the body in question. That leads to a neat puzzle: can we *see* an object that rotates very rapidly?

Challenge 109 s

We mention that *4-angular momentum* is defined naturally as

$$l^{ab} = x^a p^b - x^b p^a \,. \tag{62}$$

The two indices imply that the 4-angular momentum is a *tensor*, not a vector. Angular
Challenge 110 ny momentum is conserved, also in special relativity. The moment of inertia is naturally defined as the proportionality factor between angular velocity and angular momentum. By the way, how would you determine whether a microscopic particle, too small to be
Challenge 111 ny seen, is rotating?

For a rotating particle, the rotational energy is part of the rest mass. You may want to
Challenge 112 ny calculate the fraction for the Earth and the Sun. It is not large.

Here is a last puzzle about rotation. We know that velocity is relative: its measured
Challenge 113 ny value depends on the observer. Is this the case also for angular velocity?

## Wave motion

Page 236 We saw in Galilean physics that a harmonic or sine wave is described, among others, by an angular frequency $\omega = 2\pi\nu$ and by a wave vector $\boldsymbol{k}$, with $k = 2\pi/\lambda$. In special relativity, the two quantities are combined in the *wave 4-vector* $\boldsymbol{L}$ that is given by

$$L^a = \left(\frac{\omega}{c}, \boldsymbol{k}\right) \,. \tag{63}$$

As usual, the phase velocity of a harmonic wave is $\omega/k = \lambda\nu$. The wave 4-vector *for light* has magnitude 0, it is a null vector. For slower waves, such as sound waves, the wave
Challenge 114 e 4-vector is timelike.

The *phase* $\varphi$ of a wave can now be defined as

$$\varphi = L_a x^a = L^a x_a \,. \tag{64}$$

Being a scalar, as expected, the phase of any wave, be it light, sound or any other type, is
Challenge 115 e the same for all observers; the phase is a relativistic invariant.*

Suppose an observer with 4-velocity $\boldsymbol{U}$ finds that a wave with wave 4-vector $\boldsymbol{L}$ has frequency $\nu$. Show that

$$\nu = \boldsymbol{L}\boldsymbol{U} \tag{65}$$

Challenge 116 s must be obeyed.

Interestingly, the wave phase 4-velocity $\omega/k$ transforms in a different way than particle
Ref. 19 velocity, except in the case $\omega/k = c$. Also the aberration formula for wave motion differs
Challenge 117 ny from that for particle motion, except in the case $\omega/k = c$. Can you find the two relations?

## The action of a free particle – how do things move?

If we want to describe relativistic motion of a free particle in terms of the least action
Vol. I, page 199 principle, we need a definition of the action. We already know that physical action is a

* In component notation, the important relations are $(\omega/c, \mathbf{k})(ct, \mathbf{x}) = \varphi$, then $(\omega/c, \mathbf{k})(c, \mathbf{v}_{\text{phase}}) = 0$ and finally $(\mathrm{d}\omega/c, \mathrm{d}\mathbf{k})(c, \mathbf{v}_{\text{group}}) = 0$.

measure of the change occurring in a system. For an inertially moving or free particle, the only change is the ticking of its proper clock. As a result, the action of a free particle will be proportional to the elapsed proper time. In order to get the standard unit of energy times time, or Js, for the action, the obvious guess for the action of a free particle is

$$S = -mc^2 \int_{\tau_1}^{\tau_2} \mathrm{d}\tau \; , \tag{66}$$

where $\tau$ is the proper time along its path. This is indeed the correct expression. It implies conservation of (relativistic) energy and momentum, as the change in proper time is maximal, and the action minimal, for straight-line motion with constant velocity. Can you confirm this?

Challenge 118 ny

Indeed, in nature, all particles move in such a way that the elapsed proper time is maximal. In other words, we again find that in nature things change *as little as possible*. Nature is like a wise old man: its motions are as slow as possible – it does as little as possible. If you prefer, every change in nature is maximally effective. As we mentioned before, Bertrand Russell called the this the *'law' of cosmic laziness*.

Page 203

The expression (66) for the action is due to Max Planck. In 1906, by exploring it in detail, he found that the quantum of action $\hbar$, which he had discovered together with the Boltzmann constant, is a relativistic invariant (like the Boltzmann constant $k$). Can you imagine how he did this?

Challenge 119 ny

The action can also be written in more complex, seemingly more frightening ways. These equivalent ways to write it are particularly appropriate to prepare us for general relativity:

$$S = \int L \,\mathrm{d}t = -mc^2 \int_{t_1}^{t_2} \frac{1}{\gamma}\,\mathrm{d}t = -mc \int_{\tau_1}^{\tau_2} \sqrt{u_a u^a}\,\mathrm{d}\tau = -mc \int_{s_1}^{s_2} \sqrt{\eta^{ab} \frac{\mathrm{d}x_a}{\mathrm{d}s}\frac{\mathrm{d}x_b}{\mathrm{d}s}}\,\mathrm{d}s \; , \tag{67}$$

where $s$ is some arbitrary, but monotonically increasing, function of $\tau$, such as $\tau$ itself. As usual, the *metric* $\eta^{\alpha\beta}$ of special relativity is

$$\eta^{ab} = \eta_{ab} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{68}$$

You can easily confirm the form of the action (67) by deducing the equation of motion in the usual way.

Challenge 120 ny

In short, nature is not in a hurry: every object moves in a such way that its own clock shows the *longest* delay possible, compared with any alternative motion nearby. This general principle is also valid for particles under the influence of gravity, as we will see in the section on general relativity, and for particles under the influence of electric or magnetic interactions. In fact, the principle of maximum proper time, i.e., the least action principle, is valid in *all* cases of motion found in nature, as we will discover step by step. For the moment, we just note that the longest proper time is realized when the average

Challenge 121 ny difference between kinetic and potential energy is minimal. (Can you confirm this?) We thus recover the principle of least action in its everyday formulation.

Page 199 Earlier on, we saw that the action measures the change going on in a system. Special relativity shows that nature minimizes change by maximizing proper time. In nature, *proper time is always maximal*. In other words, things move along paths defined by the *principle of maximal ageing*. Can you explain why 'maximal ageing' and 'cosmic laziness' are equivalent? Challenge 122 e

When you throw a stone, the stone follows more or less a parabolic path. Had it flown higher, it would have to move faster, which slows down its aging. Had it flown lower, it would also age more slowly, because at lower height stay younger, as we will see. Page 134 The actual path is thus indeed the path of maximum aging.

We thus again find that nature is the opposite of a Hollywood film: nature changes in the most economical way possible – all motion realizes the smallest possible amount of action. Exploring the deeper meaning of this result is left to you: enjoy it!

## Conformal transformations – why is the speed of light invariant?

The distinction between space and time in special relativity depends on the inertial observer. On the other hand, all inertial observers agree on the position, shape and orientation of the light cone at a point. Thus, in the theory of relativity, the light cones are the basic physical 'objects'. For any expert of relativity, space-time is a large collection of light cones. Given the importance of light cones, we might ask if inertial observers are the only ones that observe the same light cones. Interestingly, it turns out that *additional* observers do as well.

The first category of additional observers that keep light cones invariant are those using units of measurement in which all time and length intervals are multiplied by a *scale factor* $\lambda$. The transformations among these observers or points of view are given by

$$x_a \mapsto \lambda x_a \tag{69}$$

and are called *dilations* or *scaling transformations*.

A second category of additional observers are found by applying the so-called *special conformal transformations*. These are compositions of an *inversion*

$$x_a \mapsto \frac{x_a}{x^2} \tag{70}$$

with a *translation* by a 4-vector $b_a$, namely

$$x_a \mapsto x_a + b_a \,, \tag{71}$$

Challenge 123 e and a second inversion. Therefore the special conformal transformations are

$$x_a \mapsto \frac{x_a + b_a x^2}{1 + 2b_a x^a + b^2 x^2} \,. \tag{72}$$

These transformations are called *conformal* because they do not change angles of (infinitesimally) small shapes, as you may want to check. The transformations therefore leave the *form* (of infinitesimally small objects) unchanged. For example, they transform infinitesimal circles into infinitesimal circles, and infinitesimal (hyper-)spheres into infinitesimal (hyper-)spheres. The transformations are called *special* because the *full* conformal group includes the dilations and the inhomogeneous Lorentz transformations as well.*

Challenge 124 ny

Note that the way in which special conformal transformations leave light cones invariant is rather subtle.

Challenge 126 ny

Since dilations do not commute with time translations, there is no conserved quantity associated with this symmetry. (The same is true of Lorentz boosts.) In contrast, rotations and spatial translations do commute with time translations and thus do lead to conserved quantities.

In summary, vacuum is conformally invariant – in the special sense just mentioned – and thus also dilation invariant. This is another way to say that vacuum alone is not sufficient to define lengths, as it does not fix a scale factor. As we would expect, matter is necessary to do so. Indeed, (special) conformal transformations are not symmetries of situations containing matter. Vacuum is conformally invariant; nature as a whole is not.**

However, conformal invariance, or the invariance of light cones, is sufficient to allow velocity measurements. Conformal invariance is also *necessary* for velocity measurements, as you might want to check.

Challenge 128 ny

We have seen that conformal invariance implies inversion symmetry: that is, that the large and small scales of a vacuum are related. This suggests that the invariance of the speed of light is related to the existence of inversion symmetry. This mysterious connection gives us a glimpse of the adventures we will encounter in the final part of our ascent of Motion Mountain. There, conformal invariance turns out to be an important property that will lead to some surprising insights.

Challenge 125 e

* The set of all *special* conformal transformations forms a group with four parameters; adding dilations and the inhomogeneous Lorentz transformations one gets fifteen parameters for the *full* conformal group. Mathematically speaking, the conformal group is locally isomorphic to SU(2,2) and to the simple group SO(4,2). These concepts are explained later on. Note that all this is true only for *four* space-time dimensions. In *two* dimensions – the other important case – the conformal group is isomorphic to the group of arbitrary analytic coordinate transformations, and is thus infinite-dimensional.

Vol. V, page 277

** A field that has mass cannot be conformally invariant; therefore conformal invariance is not an exact symmetry of all of nature. Can you confirm that a mass term $m\varphi^2$ in a Lagrangian density is not conformally invariant?

Challenge 127 ny

We note that the conformal group does not appear only in the kinematics of special relativity and thus is not only a symmetry of the vacuum: the conformal group is also the symmetry group of physical interactions, such as electromagnetism, as long as the involved radiation bosons have zero mass, as is the case for the photon. In simple words, both the vacuum and all those radiation fields that are made of massless particles are conformally invariant. Fields due to massive particles are not.

We can go even further. All elementary particles observed up to now have masses that are many orders of magnitude smaller than the Planck mass $\sqrt{\hbar c/G}$. Thus it can be said that they have *almost* vanishing mass; conformal symmetry can then be seen as an *approximate* symmetry of nature. In this view, all massive particles can be seen as small corrections, or perturbations, of massless, i.e., conformally invariant, fields. Therefore, for the construction of a fundamental theory, conformally invariant Lagrangians are often assumed to provide a good starting approximation.


FIGURE 48 The animation shows an observer accelerating down the road in a desert, until he reaches relativistic speeds. The inset shows the position along the road. Note how things seem to recede, despite the advancing motion. (Quicktime film © Anthony Searle and Australian National University, from www.anu.edu.au/Physics/Savage/TEE.)

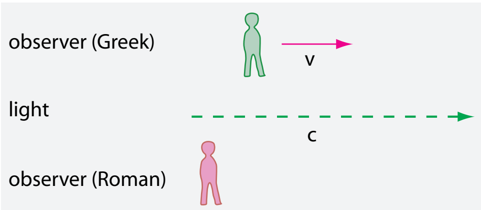



FIGURE 49 The simplest situation for an inertial and an accelerated observer

## Accelerating observers

So far, we have only studied what inertial, or free-flying, observers say to each other when they talk about the same observation. For example, we saw that moving clocks always run slow. The story gets even more interesting when one or both of the observers are accelerating.

One sometimes hears that special relativity cannot be used to describe accelerating observers. That is wrong, just as it is wrong to say that Galilean physics cannot be used for accelerating observers. Special relativity's only limitation is that it cannot be used in non-flat, i.e., curved, space-time. Accelerating bodies do exist in flat space-time, and therefore they can be discussed in special relativity.

As an appetizer, let us see what an accelerating, Greek, observer says about the clock Ref. 79 of an inertial, Roman, one, and vice versa. We assume that the Greek observer, shown in Figure 49, moves along the path $\boldsymbol{x}(t)$, as observed by the inertial Roman one. In general, the Greek–Roman clock rate ratio is given by $\Delta\tau/\Delta t = (\tau_2 - \tau_1)/(t_2 - t_1)$. Here the Greek coordinates are constructed with a simple procedure: take the two sets of events defined by $t = t_1$ and $t = t_2$, and let $\tau_1$ and $\tau_2$ be the points where these sets intersect the time axis of the Greek observer.*

We first briefly assume that the Greek observer is also inertial and moving with velocity $v$ as observed by the Roman one. The clock ratio of a Greek observer is then given

* These sets form what mathematicians call *hypersurfaces*.

by

$$\frac{\Delta\tau}{\Delta t} = \frac{d\tau}{dt} = \sqrt{1 - v^2/c^2} = \frac{1}{\gamma_v} , \qquad (73)$$

Challenge 129 ny a formula we are now used to. We find again that inertially moving clocks run slow.

For accelerated motions of the Greek observer, the differential version of the above Ref. 79 reasoning is necessary. The Greek/Roman clock rate ratio is $d\tau/dt$, and $\tau$ and $\tau + d\tau$ are calculated in the same way from the times $t$ and $t + dt$. To do this, we assume again that the Greek observer moves along the path $\boldsymbol{x}(t)$, as measured by the Roman one. We find directly that

$$\frac{\tau}{\gamma_v} = t - \boldsymbol{x}(t)\boldsymbol{v}(t)/c^2 \qquad (74)$$

and thus

$$\frac{\tau + d\tau}{\gamma_v} = (t + dt) - [\boldsymbol{x}(t) + dt\boldsymbol{v}(t)][\boldsymbol{v}(t) + dt\boldsymbol{a}(t)]/c^2 . \qquad (75)$$

Together, and to first order, these equations yield

$$\text{‘}d\tau/dt\text{’} = \gamma_v(1 - \boldsymbol{v}\boldsymbol{v}/c^2 - \boldsymbol{x}\boldsymbol{a}/c^2) . \qquad (76)$$

This result shows that accelerated clocks can run *fast or slow*, depending on their position $\boldsymbol{x}$ and the sign of their acceleration $\boldsymbol{a}$. There are quotes in the above equation because we can see directly that the Greek observer notes

$$\text{‘}dt/d\tau\text{’} = \gamma_v , \qquad (77)$$

which is *not* the inverse of equation (76). This difference becomes most apparent in the simple case of two clocks with the same velocity, one of which has a constant acceleration Ref. 79 $g$ towards the origin, whereas the other moves inertially. We then have

$$\text{‘}d\tau/dt\text{’} = 1 + gx/c^2 \qquad (78)$$

and

$$\text{‘}dt/d\tau\text{’} = 1 . \qquad (79)$$

Page 88 We will discuss this situation in more detail shortly. But first we must clarify the concept of acceleration.

## Accelerating frames of reference

How do we check whether we live in an inertial frame of reference? Let us first define the term. An *inertial frame* (*of reference*) has two defining properties. First, lengths and distances measured with a ruler are described by Euclidean geometry. In other words, rulers behave as they do in daily life. In particular, distances found by counting how many rulers (rods) have to be laid down end to end to reach from one point to another –

FIGURE 50 An observer accelerating down a road in a city. The film shows the 360° view around the observer; the borders thus show the situation behind his back, where the houses, located near the event horizon, remain at constant size and distance. (Mpg film © Anthony Searle and Australian National University.)



the so-called *rod distances* – behave as in everyday life. For example, rod distances obey Pythagoras' theorem in the case of right-angled triangles. Secondly, in inertial frames, the speed of light is invariant. In other words, any two observers in that frame, independent of their time and of the position, make the following observation: the ratio $c$ between twice the rod distance between two points and the time taken by light to travel from one point to the other and back is always the same.

Equivalently, an inertial frame is one for which all clocks always remain synchronized and whose geometry is Euclidean. In particular, in an inertial frame all observers at fixed coordinates always remain *at rest* with respect to each other. This last condition is, however, a more general one. There are other, non-inertial, situations where this is still the case.

Non-inertial frames, or *accelerating frames*, are a useful concept in special relativity. In fact, we all live in such a frame. And we can use special relativity to describe motion in such a accelerating frame, in the same way that we used Galilean physics to describe it at the beginning of our journey.

A general *frame of reference* is a continuous set of observers remaining at rest with
Ref. 82 respect to each other. Here, 'at rest with respect to each other' means that the time for a light signal to go from one observer to another and back again is constant over time, or equivalently, that the rod distance between the two observers is constant. Any frame of reference can therefore also be called a *rigid* collection of observers. We therefore note that a general frame of reference is *not* the same as a general set of coordinates; the latter is usually *not* rigid. But if all the rigidly connected observers have constant coordinate values, we speak of a *rigid coordinate system*. Obviously, these are the most useful when it comes to describing accelerating frames of reference.*

Ref. 81 * There are essentially only two other types of rigid coordinate frames, apart from the inertial frames:

— The frame $\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 - c^2\mathrm{d}t^2(1 + g_k x_k/c^2)^2$ with arbitrary, but constant, acceleration of the origin. The acceleration is $\boldsymbol{a} = -\boldsymbol{g}(1 + \boldsymbol{g}x/c^2)$.
— The uniformly rotating frame $\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + 2\omega(-y\,\mathrm{d}x + x\,\mathrm{d}y)\mathrm{d}t - (1 - r^2\omega^2/c^2)\mathrm{d}t$. Here the $z$-axis is the rotation axis, and $r^2 = x^2 + y^2$.

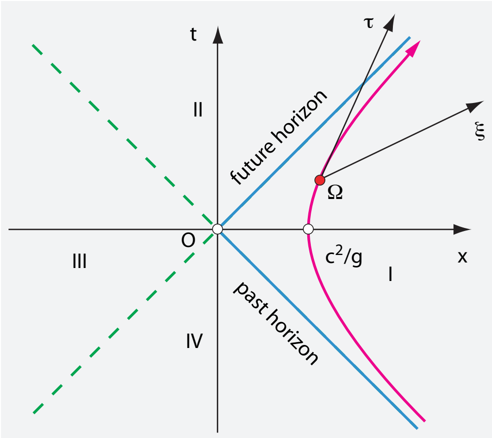



FIGURE 51 The hyperbolic motion of an rectilinearly, uniformly accelerating observer Ω

Ref. 82 Note that if two observers both move with a velocity $\boldsymbol{v}$, as measured in some *inertial* frame, they observe that they are at rest with respect to each other *only* if this velocity is *constant*. Again we find, as above, that two people tied to each other by a rope, and at a distance such that the rope is under tension, will see the rope break (or hang loose) if they accelerate together to (or decelerate from) relativistic speeds in precisely the same way. Acceleration in relativity requires careful thinking. Challenge 130 ny Page 48

Can you state how the acceleration ratio enters into the definition of mass in special relativity? Page 59 Challenge 131 ny

## Constant acceleration

Acceleration is a tricky topic. An observer who always *feels* the *same* force on his body is called *uniformly* accelerating. His proper acceleration is constant. More precisely, a uniformly accelerating observer is an observer whose acceleration at every moment, measured by the inertial frame with respect to which the observer is at rest *at that moment*, always has the same value $\boldsymbol{B}$. It is important to note that uniform acceleration is *not* uniformly accelerating when always observed from the *same* inertial frame. This is an important difference from the Galilean case.

For uniformly accelerated motion in the sense just defined, 4-jerk is zero, and we need

$$\boldsymbol{B} \cdot \boldsymbol{B} = -g^2 \ , \tag{80}$$

Ref. 83 where $g$ is a constant independent of $t$. The simplest case is uniformly accelerating motion that is also *rectilinear*, i.e., for which the acceleration $\boldsymbol{a}$ is parallel to $\boldsymbol{v}$ at one instant of time and (therefore) for all other times as well. In this case we can write, using 3-vectors, Challenge 132 e

$$\gamma^3 \boldsymbol{a} = \boldsymbol{g} \quad \text{or} \quad \frac{\mathrm{d}\gamma\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{g} \ . \tag{81}$$

Challenge 133 e

Taking the direction we are talking about to be the $x$-axis, and solving for $v(t)$, we get

$$v = \frac{gt}{\sqrt{1 + \frac{g^2 t^2}{c^2}}} , \tag{82}$$

where it was assumed that $v(0) = 0$. We note that for small times we get $v = gt$ and for large times $v = c$, both as expected. The momentum of the accelerated observer increases linearly with time, again as expected. Integrating, we find that the accelerated observer moves along the path

Challenge 134 e

$$x(t) = \frac{c^2}{g}\sqrt{1 + \frac{g^2 t^2}{c^2}} , \tag{83}$$

where we assumed that $x(0) = c^2/g$, in order to keep the expression simple. Because of this result, visualized in Figure 51, a rectilinearly and uniformly accelerating observer is said to undergo *hyperbolic* motion. For small times, the world-line reduces to the usual $x = gt^2/2 + x_0$, whereas for large times it is $x = ct$, as expected. The motion is thus uniformly accelerated only for the moving body itself, but *not* for an outside observer, again as expected.

The proper time $\tau$ of the accelerated observer is related to the time $t$ of the inertial frame in the usual way by $dt = \gamma d\tau$. Using the expression for the velocity $v(t)$ of equation (82) we get*

Ref. 83, Ref. 84

$$t = \frac{c}{g} \sinh \frac{g\tau}{c} \quad \text{and} \quad x = \frac{c^2}{g} \cosh \frac{g\tau}{c} \tag{84}$$

for the relationship between proper time $\tau$ and the time $t$ and position $x$ measured by the external, inertial Roman observer. We will encounter this relation again during our study of black holes.

Does the last formula sound boring? Just imagine accelerating on your motorbike at $g = 10\,\text{m/s}^2$ for the proper time $\tau$ of 25 years. That would bring you beyond the end of the known universe! Isn't that worth a try? Unfortunately, neither motorbikes nor missiles that accelerate like this exist, as their fuel tanks would have to be enormous. Can you confirm this?

Challenge 135 s

For uniform rectilinear acceleration, the coordinates transform as

$$\begin{aligned} t &= \left(\frac{c}{g} + \frac{\xi}{c}\right) \sinh \frac{g\tau}{c} \\ x &= \left(\frac{c^2}{g} + \xi\right) \cosh \frac{g\tau}{c} \\ y &= \upsilon \\ z &= \zeta , \end{aligned} \tag{85}$$

Ref. 85

* Use your favourite mathematical formula collection – every person should have one – to deduce this. The *hyperbolic sine* and the *hyperbolic cosine* are defined by $\sinh y = (e^y - e^{-y})/2$ and $\cosh y = (e^y + e^{-y})/2$. They imply that $\int dy/\sqrt{y^2 + a^2} = \text{arsinh}\, y/a = \text{Arsh}\, y/a = \ln(y + \sqrt{y^2 + a^2})$.

where $\tau$ now is the time coordinate in the Greek, accelerated frame. We note also that the space-time interval $\mathrm{d}\sigma$ satisfies

$$\mathrm{d}\sigma^2 = (1 + g\xi/c^2)^2 c^2 \mathrm{d}\tau^2 - \mathrm{d}\xi^2 - \mathrm{d}\upsilon^2 - \mathrm{d}\zeta^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \,, \qquad (86)$$

and since for $\mathrm{d}\tau = 0$ distances are given by Pythagoras' theorem, the Greek, accelerated reference frame is indeed rigid. Ref. 86

After this forest of formulae, let's tackle a simple question, shown in Figure 51. The inertial, Roman observer O sees the Greek observer $\Omega$ departing under continuous acceleration, moving further and further away, following equation (83). What does the Greek observer say about his Roman colleague? With all the knowledge we have now, that is easy to answer. At each point of his trajectory $\Omega$ sees that O has the coordinate $\tau = 0$ (can you confirm this?), which means that the distance to the Roman observer, as seen by the Greek one, is the same as the space-time interval O$\Omega$. Using expression (83), we see that this is

Challenge 136 e

Ref. 87

$$d_{\mathrm{O}\Omega} = \sqrt{\xi^2} = \sqrt{x^2 - c^2t^2} = c^2/g \,, \qquad (87)$$

which, surprisingly enough, is constant in time! In other words, the Greek observer will observe that he stays at a constant distance from the Roman one, in complete contrast to what the Roman observer says. Take your time to check this strange result in some other way. We will need it again later on, to explain why the Earth does not explode. (Can you guess how that is related to this result?)

Challenge 137 s

## Event horizons

We now explore one of the most surprising consequences of accelerated motion, one that is intimately connected with the result just deduced. We explore the trajectory, in the coordinates $\xi$ and $\tau$ of the rigidly accelerated frame, of an object located at the departure point $x = x_0 = c^2/g$ at all times $t$. We get the two relations*

Challenge 138 ny

$$\begin{aligned} \xi &= -\frac{c^2}{g}\left(1 - \operatorname{sech}\frac{g\tau}{c}\right) \\ \mathrm{d}\xi/\mathrm{d}\tau &= -c \operatorname{sech}\frac{g\tau}{c} \tanh\frac{g\tau}{c} \,. \end{aligned} \qquad (89)$$

These equations are strange. For large times $\tau$ the coordinate $\xi$ approaches the limit value $-c^2/g$ and $\mathrm{d}\xi/\mathrm{d}\tau$ approaches zero. The situation is similar to that of riding a car accelerating away from a woman standing on a long road. For the car driver, the woman moves away; however, after a while, the only thing the driver notices is that she is slowly approaching the horizon. In everyday life, both the car driver and the woman on the road

* The functions appearing above, the *hyperbolic secant* and the *hyperbolic tangent*, are defined using the expressions from the footnote on page 85:

$$\operatorname{sech} y = \frac{1}{\cosh y} \quad \text{and} \quad \tanh y = \frac{\sinh y}{\cosh y} \,. \qquad (88)$$

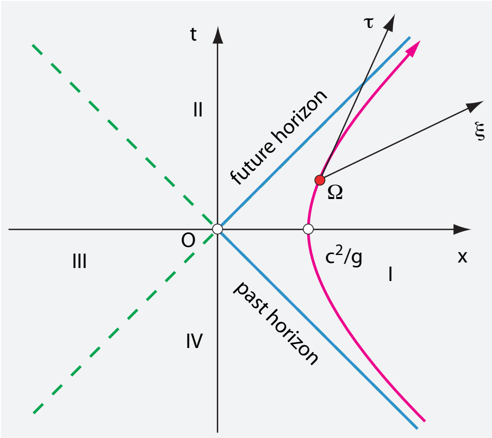



FIGURE 52 Hyperbolic motion and event horizons

see the other person approaching their respective horizon; in special relativity, only the accelerated observer makes a observation of this type.

A graph of the situation helps to clarify the result. In Figure 52 we can see that light emitted from any event in regions II and III cannot reach the Greek observer. Those events are hidden from him and cannot be observed. The boundary between the part of space-time that can be observed and the part that cannot is called the *event horizon*. Strangely enough, however, light from the Greek observer *can* reach region II. Event horizons thus act like *one-way gates* for light and other signals. For completeness, the graph also shows the past event horizon. We note that an event horizon is a *surface*. It is thus a different phenomenon than the everyday horizon, which is a *line*. Can you confirm that event horizons are *black*? Challenge 139 ny

So, not all events observed in an inertial frame of reference can be observed in a uniformly accelerating frame of reference. Accelerated observers are limited. Uniformly accelerating frames of reference produce event horizons at a distance $-c^2/g$. For example, a person who is standing can never see further than this distance below his feet.

By the way, is it true that a light beam cannot catch up with a massive observer in hyperbolic motion, if the observer has a sufficient head start? Challenge 140 s

Here is a more advanced challenge, which prepares us for general relativity. What is the two-dimensional *shape* of the horizon seen by a uniformly accelerated observer? Challenge 141 s

Another challenge: what horizon is seen by an observer on a carousel? Challenge 142 s

## The importance of horizons

In special relativity, horizons might seem to play a secondary role. But this impression is wrong, for two reasons. First, in general relativity, horizons become frequent: the dark night sky is an example of a horizon, and so is the surface of a black hole. And there a billions of black holes in the universe. But the second reason for the interest of horizons is even more important.

Two and a half thousand years ago, Leucippus of Elea (*c.* 490 to *c.* 430 bce) and Democritus of Abdera (*c.* 460 to *c.* 356 or 370 bce) founded atomic theory. In particular, they made the statement that everything found in nature is – in modern words – *particles* and *empty space*. For many centuries, modern physics corroborated this statement. For example, all matter turned out to be made of particles. Also light and all other types of radiation are made of particles. But then came relativity and the discovery of horizons.

Horizons show that *atomism is wrong*: we will discover soon that horizons have colours, and that they can have mass, spin and charge. But horizons are extended, not localized. In short, we will discover that horizons are neither space nor particles. Horizons are something new.

Only in the last two volumes of our adventure will we discover that horizons are effectively a *mixture* of space and particles. But we will need some time to find out what this means exactly. So far, special relativity only tells us that horizons are a new phenomenon of nature, an unexpected addition to particles and space-time.

## Acceleration changes colours

Page 28 We saw above that a moving receiver sees different colours than the sender. So far, we discussed this colour shift, or Doppler effect, for inertial motion only. For accelerating frames the situation is even stranger: sender and receiver do not agree on colours even Ref. 83, Ref. 88 if they are at *rest* with respect to each other. Indeed, if light is emitted in the direction of the acceleration, the formula for the space-time interval gives

$$\mathrm{d}\sigma^2 = \left(1 + \frac{g_0 x}{c^2}\right)^2 c^2 \mathrm{d}t^2 \qquad (90)$$

in which $g_0$ is the proper acceleration of an observer located at $x = 0$. We can deduce in Challenge 143 ny a straightforward way that

$$\frac{f_\mathrm{r}}{f_\mathrm{s}} = 1 - \frac{g_\mathrm{r} h}{c^2} = \frac{1}{1 + \frac{g_\mathrm{s} h}{c^2}} \qquad (91)$$

where $h$ is the rod distance between the source and the receiver, and where $g_\mathrm{s} = g_0/(1 + g_0 x_\mathrm{s}/c^2)$ and $g_\mathrm{r} = g_0/(1 + g_\mathrm{o} x_\mathrm{r}/c^2)$ are the proper accelerations measured at the source and at the detector. In short, the frequency of light decreases when light moves in the direction of acceleration. By the way, does this have an effect on the colour of trees along Challenge 144 s their vertical extension?

The formula usually given, namely

$$\frac{f_\mathrm{r}}{f_\mathrm{s}} = 1 - \frac{g h}{c^2} , \qquad (92)$$

is only correct to a first approximation. In accelerated frames of reference, we have to be careful about the meaning of every quantity. For everyday accelerations, however, the Challenge 145 e differences between the two formulae are negligible. Can you confirm this?

## Can light move faster than $c$?

What speed of light does an accelerating observer measure? Using expression (92) above, an accelerated observer deduces that

$$v_{\text{light}} = c\left(1 + \frac{gh}{c^2}\right) \qquad (93)$$

which is higher than $c$ for light moving in front of or 'above' him, and lower than $c$ for light moving behind or 'below' him. This strange result follows from a basic property of any accelerating frame of reference: in such a frame, even though all observers are at rest with respect to each other, clocks do *not* remain synchronized. This predicted change of the speed of light has also been confirmed by experiment: the propagation delays to be Page 161 discussed in general relativity can be seen as confirmations of this effect.

In short, the speed of light is only invariant when it is defined as $c = \mathrm{d}x/\mathrm{d}t$, *and* if $\mathrm{d}x$ is measured with a ruler located at a point *inside* the interval $\mathrm{d}x$, *and* if $\mathrm{d}t$ is measured with a clock read off *during* the interval $\mathrm{d}t$. In other words, the speed of light is only invariant if measured locally.

If, however, the speed of light is defined as $\Delta x/\Delta t$, or if the ruler measuring distances or the clock measuring times is located *away* from the propagating light, the speed of light is different from $c$ for accelerating observers! This is the same effect you can experience when you turn around your vertical axis at night: the star velocities you observe are much higher than the speed of light. In short, *$c$ is the speed of light only relative to* nearby *matter.*

Note that this result does not imply that signals or energy can be moved faster than $c$. Challenge 146 s You may want to check this for yourself.

In fact, all these effects are negligible for distances $l$ that are much less than $c^2/a$. For an acceleration of $9.5\,\mathrm{m/s^2}$ (about that of free fall), distances would have to be of the order of one light year, or $9.5 \cdot 10^{12}$ km, in order for any sizeable effects to be observed.

By the way, everyday gravity is equivalent to a constant acceleration. So, why then do Challenge 147 s distant objects, such as stars, not move faster than light, following expression (93)?

## The composition of accelerations

To get a better feeling for acceleration, we explore another topic: the composition theorem for accelerations. This situation is more complex than for velocities, and is often Ref. 89 avoided. However, a good explanation of this was published by Mishra.

If we call $a_{nm}$ the acceleration of system $n$ by observer $m$, we are seeking to express the object acceleration $a_{01}$ as function of the value $a_{02}$ measured by the other observer, the relative acceleration $a_{12}$, and the proper acceleration $a_{22}$ of the other observer: see Figure 53. Here we will only study one-dimensional situations, where all observers and all objects move along one axis. (For clarity, we also write $v_{12} = v$ and $v_{02} = u$.)

Challenge 148 e In Galilean physics we have the general connection

$$a_{01} = a_{02} - a_{12} + a_{22} \qquad (94)$$

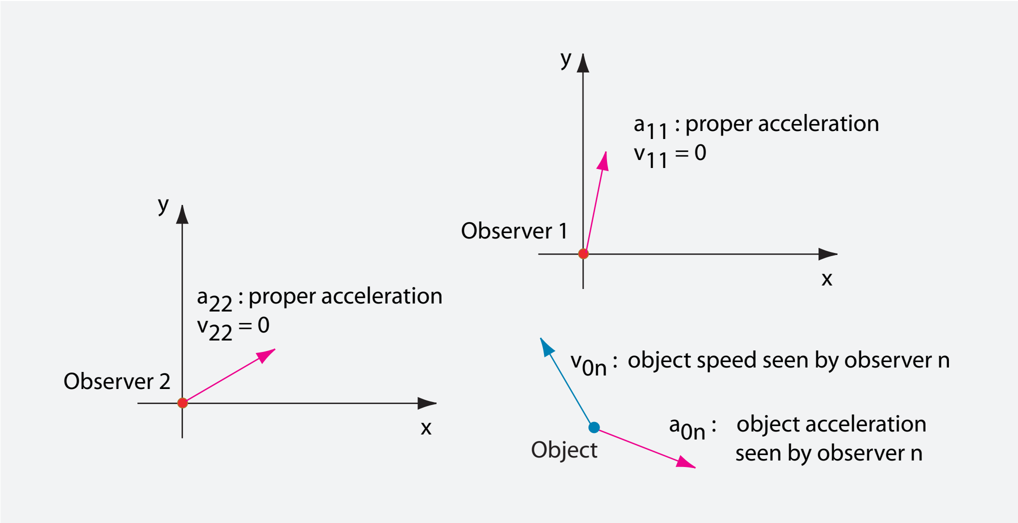



FIGURE 53 The definitions necessary to deduce the composition behaviour of accelerations

because accelerations behave simply. In special relativity, we get

$$a_{01} = a_{02}\frac{(1-v^2/c^2)^{3/2}}{(1-uv/c^2)^3} - a_{12}\frac{(1-u^2/c^2)(1-v^2/c^2)^{-1/2}}{(1-uv/c^2)^2} + a_{22}\frac{(1-u^2/c^2)(1-v^2/c^2)^{3/2}}{(1-uv/c^2)^3} \qquad (95)$$

Challenge 149 ny
and you might enjoy checking the expression.

## A curiosity: what is the one-way speed of light?

We have seen that the speed of light, as usually defined, is given by $c$ only if either the observer is inertial or the observer measures the speed of light passing nearby (rather than light passing at a distance). In short, the speed of light has to be measured locally. But this condition does not eliminate one last subtlety.

Usually, length is measured by the time it takes light to travel. In this case the speed of light will *obviously* be invariant. So how can we check the invariance? We need to eliminate length measurements. The simplest way to do this is to reflect light from a mirror, as shown in Figure 54. The invariance of the speed of light implies that if light goes up and down a short straight line, then the clocks at the two ends measure times given by

$$t_3 - t_1 = 2\,(t_2 - t_1)\,. \qquad (96)$$

Here it is assumed that the clocks have been synchronised according to the prescription on page 46. If the factor were not exactly two, the speed of light would not be invariant. In fact, all experiments so far have yielded a factor of two, within measurement errors.

Ref. 90, Ref. 91
Challenge 150 s
But these experiments instil us with a doubt: it seems that the *one-way* velocity of light cannot be measured. Do you agree? Is the issue important?

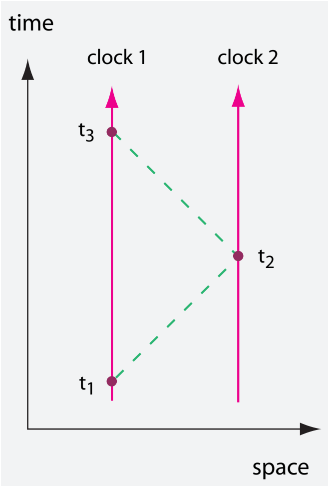



FIGURE 54 Clocks and the measurement of the speed of light as two-way velocity

## LIMITS ON THE LENGTH OF SOLID BODIES

An everyday solid object breaks when some part of it moves with respect to some nearby part with more than the speed of sound $c$ of the material.* For example, when an object hits the floor and its front end is stopped within a distance $d$, the object breaks at the latest when

$$\frac{v^2}{c^2} \geqslant \frac{2d}{l} \ . \qquad (97)$$

In this way, we see that we can avoid the breaking of fragile objects by packing them into foam rubber – which increases the stopping distance. This may explain why boxes containing presents are usually so much larger than their contents.

The fracture limit can also be written in a different way. To avoid breaking, the acceleration $a$ of a solid body with length $l$ must obey

$$la < c^2 \ , \qquad (98)$$

where $c$ is the speed of sound, which is the speed limit for the material parts of solids. Let us now repeat the argument in relativity, using the speed of light instead of that of sound. Imagine accelerating the front of a *solid* body with some *proper* acceleration $a$. The back end cannot move with an acceleration $\alpha$ equal or larger than infinity, or if one prefers, it cannot move with more than the speed of light. A quick check shows that therefore the length $l$ of a solid body must obey

Ref. 92

Challenge 151 s

$$la < c^2 \ , \qquad (99)$$

where $c$ is now the speed of light. The speed of light thus limits the size of solid bodies. For example, for 9.8 m/s$^2$, the acceleration of good motorbike, this expression gives a length

* The (longitudinal) speed of sound is about 5.9 km/s for glass, iron or steel; about 4.5 km/s for gold; and about 2 km/s for lead. Other sound speeds are given on page 237.

limit of 9.2 Pm, about a light year. Not a big restriction: most motorbikes are shorter.

However, there are other, more interesting situations. Today, high accelerations are produced in particle accelerators. Atomic nuclei have a size of a few femtometres. Can you deduce at which energies they break when smashed together in an accelerator? Challenge 152 ny In fact, inside a nucleus, the nucleons move with accelerations of the order of $v^2/r \approx \hbar^2/m^2r^3 \approx 10^{31}$ m/s$^2$; this is one of the highest values found in nature. Is the length limit also obeyed by nuclei? Challenge 153 s

We find that Galilean physics and relativity produce similar conclusions: a limiting speed, be it that of sound or that of light, makes it impossible for solid bodies to be *rigid*. When we push one end of a body, the other end always can move only a little bit later.

A puzzle: does the speed limit imply a relativistic 'indeterminacy relation'

$$\Delta l \; \Delta a \leqslant c^2 \tag{100}$$

for the length and acceleration indeterminacies? Challenge 154 s

What does all this mean for the size of elementary particles? Take two electrons a distance $d$ apart, and call their size $l$. The acceleration due to electrostatic repulsion then leads to an upper limit for their size given by Challenge 155 ny

$$l < \frac{4\pi\varepsilon_0 c^2 d^2 m}{e^2} \; . \tag{101}$$

The nearer electrons can get, the smaller they must be. The present experimental limit gives a size smaller than $10^{-19}$ m. Can electrons be exactly point-like? We will come back to this question during our study of general relativity and quantum theory.

Chapter 3

# SPECIAL RELATIVITY IN FOUR SENTENCES

The results encountered in our ascent of Motion Mountain can be quickly summarized.

— All (free floating and nearby) observers observe that there is a unique, maximal and invariant energy speed in nature, the 'perfect' speed $c = 0.3\,\mathrm{Gm/s}$, which is realized by massless radiation such as light or radio signals, but cannot be achieved by material systems.
— Therefore, even though space-time is the same for every observer, times and lengths vary from one observer to another, as described by the Lorentz transformations (14) and (15), and as confirmed by experiment. Page 39
— Collisions show that the maximum energy speed implies that mass is equivalent to energy, and that the total energy of a body is given by $E = \gamma mc^2$, as again confirmed by experiment.
— Applied to accelerated objects, these results lead to numerous counter-intuitive consequences, such as the twin paradox, the appearance of event horizons and the appearance of short-lived, i.e., virtual, tachyons in collisions.

Special relativity shows that all motion of radiation and matter is limited in speed and is defined and measured using the propagation of light. The other properties of everyday motion remain. In particular, the six basic properties of everyday motion that follow from its predictability are still valid: also relativistic motion is continuous, conserves energy–momentum and angular momentum, is relative, is reversible, is mirror-invariant (except for the weak interaction, where a different way to predict mirror-inverse motion holds) and is lazy, i.e., it minimizes action. Vol. I, page 27 Vol. V, page 178

## Could the speed of light vary?

The speed of massless light is the limit speed of energy in nature. Could the limit speed change from place to place, or as time goes by? This tricky question still makes a fool out of many physicists. The first answer is usually a loud: 'Yes, of course! Just look at what happens when the value of $c$ is changed in formulae.' (Several such 'variable speed of light' conjectures have even been explored.) However, this often-heard answer is wrong. Ref. 93

Since the speed of light enters into our definition of time and space, 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 the value actually varies. No imaginable experiment could detect a variation of the limit speed, as the limit speed is the basis for all measurements. 'That is intellectual cruelty!', you might say. 'All Challenge 156 s

experiments show that the speed of light is invariant; we had to swallow one counter-intuitive result after another to accept the invariance of the speed of light, and now we are even supposed to admit that there is no other choice?' Yes, we are. That is the irony of progress in physics. The observer-invariance of the speed of light is counter-intuitive and astonishing when compared to the observer-dependence of everyday, Galilean speeds. But had we taken into account that every speed measurement is – whether we like it or not – a comparison with the speed of light, we would not have been astonished by the invariance of the speed of light at all; rather, we would have been astonished by the strange properties of *small* speeds.

There is, in principle, *no* way to check the invariance of a measurement standard. To put it another way, the truly surprising aspect of relativity is not the invariance of $c$; it is the disappearance of $c$ from the formulae of everyday motion.

## Where does special relativity break down?

As we approach the speed of light, the quantities in the Lorentz transformation diverge. However, this is only half the story. In nature, no observable actually reaches arbitrary large values. Indeed, approaching the speed of light as nearly as possible, even special relativity breaks down.

At extremely large Lorentz contractions, there is no way to ignore the curvature of space-time that the moving matter or radiation creates; *gravitation* has to be taken into account. At extremely large Lorentz contractions, there is also no way to ignore the fluctuations of speed and position of the moving particles; *quantum theory* has to be taken into account. The exploration of these two limitations define the next two stages of our ascent of Motion Mountain. We start with the first.

CHAPTER 4

# SIMPLE GENERAL RELATIVITY: GRAVITATION, MAXIMUM SPEED AND MAXIMUM FORCE

GENERAL relativity is easy! Nowadays, it can be made as intuitive as universal ravity and its inverse square law – by using the right approach. The main ideas of eneral relativity, like those of special relativity, are accessible to secondary-school students. In particular, black holes, gravitational waves, space-time curvature and the limits of the universe can then be understood as easily as the Doppler effect or the twins paradox.

In the following pages we will discover that, just as special relativity is based on a maximum speed $c$, general relativity is based on a maximum force $c^4/4G$ or on a maximum power $c^5/4G$. We first show that all known experimental data are consistent with these limits. Then we find that the maximum force and the maximum power are achieved only on insurmountable limit surfaces.

▷ The surfaces that realize maximum force (momentum change) or maximum power are called ***horizons***.

Page 86

Horizons are simple generalizations of those horizons that we encountered in special relativity. Horizons play the role in general relativity that is played by light beams in special relativity: they are the systems that *realize* the limit. A horizon is the reason that the sky is dark at night and that the universe is of finite size. Horizons tell us that in general, space-time is curved. And horizons will allow us to deduce the field equations of general relativity.

We also discuss the main counter-arguments and paradoxes arising from the force and power limits. The resolutions of the paradoxes clarify why the limits have remained dormant for so long, both in experiments and in teaching.

Challenge 157 s

After this introduction, we will study the effects of relativistic gravity in detail. We will explore the consequences of space-time curvature for the motions of bodies and of light in our everyday environment. For example, the inverse square law will be modified. (Can you explain why this is necessary in view of what we have learned so far?) Most fascinating of all, we will discover how to move and bend the vacuum. Then we will study the universe at large. Finally, we will explore the most extreme form of gravity: black holes.

FIGURE 55 Effects of gravity: a dripping stalactite (© Richard Cindric) and the rings of Saturn, photographed when the Sun is hidden behind the planet (courtesy CICLOPS, JPL, ESA, NASA)

## Maximum force – general relativity in one statement

> “One of the principal objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity.”
> Willard Gibbs

Ref. 94

We just saw that the theory of *special* relativity appears when we recognize the speed limit $c$ in nature and take this limit as a basic principle. At the turn of the twenty-first century it was shown that *general* relativity can be approached by using a similar basic principle: Ref. 95, Ref. 96

▷ *There is in nature a maximum force:*

$$F \leqslant \frac{c^4}{4G} = 3.0 \cdot 10^{43}\,\mathrm{N}\,. \tag{102}$$

In nature, no force in any muscle, machine or system can exceed this value. For the curious, the value of the force limit is the energy of a (Schwarzschild) black hole divided by twice its radius. The force limit can be understood intuitively by noting that (Schwarzschild) black holes are the densest bodies possible for a given mass. Since there is a limit to how much a body can be compressed, forces – whether gravitational, electric, centripetal or of any other type – cannot be arbitrary large.

Alternatively, it is possible to use another, equivalent statement as a basic principle:

▷ *There is a maximum power in nature:*

$$P \leqslant \frac{c^5}{4G} = 9.1 \cdot 10^{51}\,\mathrm{W}\,. \tag{103}$$

No power of any lamp, engine or explosion can exceed this value. The maximum power is realized when a (Schwarzschild) black hole is radiated away in the time that light takes to travel along a length corresponding to its diameter. We will see below precisely what black holes are and why they are connected to these limits.

The existence of a maximum force or power implies the full theory of general rela-

TABLE 3 How to convince yourself and others that there is a maximum force $c^4/4G$ or a maximum power $c^5/4G$ in nature. Compare this table with the table about maximum speed, on page 24 above, and with the table about a smallest action, on page 17 in volume IV.

| Issue | Method |
|---|---|
| The force value $c^4/4G$ is observer-invariant | check all observations |
| Force values $> c^4/4G$ are not observed | check all observations |
| Force values $> c^4/4G$ are either non-local or not due to energy transport | check all observations |
| Force values $> c^4/4G$ cannot be produced | check all attempts |
| Force values $> c^4/4G$ cannot be imagined | solve all paradoxes |
| A maximum force value $c^4/4G$ is consistent | 1 – show that all consequences, however weird, are confirmed by observation |
| | 2 – deduce the theory of general relativity from it and check it |

tivity. In order to prove the correctness and usefulness of this approach, a sequence of arguments is required. The sequence is the same as the one we used for the establishment of the limit speed in special relativity; it is shown in Table 3. The basis is to recognize that the force value is *invariant*. This follows from the invariance of $c$ and $G$. For the first argumrnt, we need to gather all observational evidence for the claimed limit. Secondly, we have to show that the limit applies in all possible and imaginable situations; any apparent paradoxes will need to be resolved. Finally, in order to establish the limit as a principle of nature, we have to show that general relativity follows from it.

Page 24

These three steps structure this introduction to general relativity. We start the story by explaining the origin of the idea of a limiting value.

## The force and power limits

In the nineteenth and twentieth centuries many physicists took pains to avoid the concept of force. Heinrich Hertz made this a guiding principle of his work, and wrote an influential textbook on classical mechanics without ever using the concept. The fathers of quantum theory, who all knew this text, then dropped the term ‘force’ completely from the vocabulary of microscopic physics. Meanwhile, the concept of ‘gravitational force’ was eliminated from general relativity by reducing it to a ‘pseudo-force’. Force fell out of fashion.

Nevertheless, the maximum force principle does make sense, provided that we visualize it by means of the definition of force: *force is the flow of momentum per unit time*. In nature, momentum cannot be created or destroyed. We use the term 'flow' to remind us that momentum, being a conserved quantity, can only change by inflow or outflow. In other words, change of momentum, and thus force, always takes place through some boundary surface. This fact is of central importance. Whenever we think about force at a point, we *really* mean the momentum 'flowing' through a surface at that point. General relativity states this idea usually as follows: forces keep bodies from following geodesics. (A *geodesic* is a path followed by a freely falling particle.) The mechanism underlying a measured force is not important; in order to have a concrete example to guide the discussion it can be helpful to imagine force as electromagnetic in origin. However, any type of force is possible.

We also stress that the force limit concerns 3-force, or what we call force in everyday life, and that the power limit concerns what we call power in everyday life. In other words, in nature, both 3-velocity and 3-force are limited.

The maximum force principle boils down to the following statement: if we imagine any physical surface (and cover it with observers), the integral of momentum flow through the surface (measured by all those observers) never exceeds the limit value $c^4/4G$. It does not matter how the surface is chosen, as long as it is physical, i.e., as long as we can fix observers* onto it.

The principle of maximum force imposes a limit on muscles, the effect of hammers, the flow of material, the acceleration of massive bodies, and much more. No system can create, measure or experience a force above the limit. No particle, no galaxy and no bulldozer can exceed it.

The existence of a force limit has an appealing consequence. In nature, forces can be measured. Every measurement is a comparison with a standard. The force limit provides a *natural* unit of force that fits into the system of natural units** that Max Planck derived from $c$, $G$ and $h$ (or $\hbar$). The maximum force thus provides a standard of force valid in every place and at every instant of time.

The limit value of $c^4/4G$ differs from Planck's proposed unit in two ways. First, the numerical factor is different (Planck had in mind the value $c^4/G$). Secondly, the force unit is a *limiting* value. In this respect, the maximum force plays the same role as the maximum speed. As we will see later on, this limit property is valid for all other Planck units as well, once the numerical factors have been properly corrected. The factor 1/4 has no deeper meaning: it is just the value that leads to the correct form of the field equations of general relativity. The factor 1/4 in the limit is also required to recover, in everyday situations, the inverse square law of universal gravitation. When the factor is properly taken into account, the maximum force (or power) is simply given by the (corrected) Planck energy divided by the (corrected) Planck length or Planck time.

Ref. 97
Vol. VI, page 24
Page 116

The expression for the maximum force involves the speed of light $c$ and the gravitational constant $G$; it thus qualifies as a statement on relativistic gravitation. The fundamental principle of special relativity states that speed $v$ obeys $v \leqslant c$ for all observers.

* Observers in general relativity, like in special relativity, are massive physical systems that are small enough so that their influence on the system under observation is negligible.

** When Planck discovered the quantum of action, he noticed at once the possibility to define natural units. On a walk with his seven-year-old son in the forest around Berlin, he told him that he had made a discovery as important as the discovery of universal gravity.

Vol. IV, page 18

Analogously, the basic principle of general relativity states that in all cases force $F$ and power $P$ obey $F \leqslant c^4/4G$ and $P \leqslant c^5/4G$. It does not matter whether the observer measures the force or power while moving with high velocity relative to the system under observation, during free fall, or while being strongly accelerated. However, we will see that it is essential that the observer records values measured *at his own location* and that the observer is *realistic*, i.e., made of matter and not separated from the system by a horizon. These conditions are the same that must be obeyed by observers measuring velocity in special relativity.

Since physical power is force times speed, and since nature provides a speed limit, the force bound and the power bound are equivalent. We have already seen that force and power appear together in the definition of 4-force. The statement of a maximum 3-force is valid for every component of the 3-force, as well as for its magnitude. (As we will see below, a boost to an observer with high $\gamma$ value cannot be used to overcome the force or power limits.) The power bound limits the output of car and motorcycle engines, lamps, lasers, stars, gravitational radiation sources and galaxies. It is equivalent to $1.2 \cdot 10^{49}$ horsepower. The maximum power principle states that there is no way to move or get rid of energy more quickly than that.

Page 74
Page 109

The power limit can be understood intuitively by noting that every engine produces *exhausts*, i.e., some matter or energy that is left behind. For a lamp, a star or an evaporating black hole, the exhausts are the emitted radiation; for a car or jet engine they are hot gases; for a water turbine the exhaust is the slowly moving water leaving the turbine; for a rocket it is the matter ejected at its back end; for a photon rocket or an electric motor it is electromagnetic energy. Whenever the power of an engine gets close to the limit value, the exhausts increase dramatically in mass–energy. For extremely high exhaust masses, the gravitational attraction from these exhausts – even if they are only radiation – prevents further acceleration of the engine with respect to them. The maximum power principle thus expresses that there is a built-in braking mechanism in nature; this braking mechanism is gravity.

Yet another, equivalent limit appears when the maximum power is divided by $c^2$.

▷ *There is a maximum rate of mass change in nature:*

$$\frac{\mathrm{d}m}{\mathrm{d}t} \leqslant \frac{c^3}{4G} = 1.0 \cdot 10^{35}\ \mathrm{kg/s}\ . \tag{104}$$

This bound imposes a limit on pumps, jet engines and fast eaters. Indeed, the rate of flow of water or any other material through tubes is limited. The mass flow limit is obviously equivalent to either the force or the power limit.

The claim of a maximum force, power or mass change in nature seems almost too fantastic to be true. Our first task is therefore to check it empirically as thoroughly as we can.

## The experimental evidence

Like the maximum speed principle, the maximum force principle must first of all be checked experimentally. Michelson spent a large part of his research life looking for possible changes in the value of the speed of light. No one has yet dedicated so much effort to testing the maximum force or power. However, it is straightforward to confirm that no experiment, whether microscopic, macroscopic or astronomical, has ever measured force values larger than the stated limit. Many people have claimed to have produced speeds larger than that of light. So far, nobody has ever claimed to have produced or observed a force larger than the limit value. Challenge 158 s

The large accelerations that particles undergo in collisions inside the Sun, in the most powerful accelerators or in reactions due to cosmic rays correspond to force values much smaller than the force limit. The same is true for neutrons in neutron stars, for quarks inside protons, and for all matter that has been observed to fall towards black holes. Furthermore, the search for space-time singularities, which would allow forces to achieve or exceed the force limit, has been fruitless.

In the astronomical domain, all forces between stars or galaxies are below the limit value, as are the forces in their interior. Not even the interactions between any two halves of the universe exceed the limit, whatever physically sensible division between the two halves is taken. (The meaning of 'physically sensible division' will be defined below; for divisions that are *not* sensible, exceptions to the maximum force claim *can* be constructed. You might enjoy searching for such an exception.) Page 114 Challenge 159 s

Astronomers have also failed to find any region of space-time whose curvature (a concept to be introduced below) is large enough to allow forces to exceed the force limit. Indeed, none of the numerous recent observations of black holes has brought to light forces larger than the limit value or objects smaller than the corresponding black hole radii.

The power limit can also be checked experimentally. It turns out that the power – or luminosity – of stars, quasars, binary pulsars, gamma ray bursters, galaxies or galaxy clusters can indeed be a sizeable fraction of the power limit. However, no violation of the limit has ever been observed. In fact, the sum of all light output from all stars in the universe does not exceed the limit. Similarly, even the brightest sources of gravitational waves, merging black holes, do not exceed the power limit. Only the brightness of evaporating black holes in their final phase could equal the limit. But so far, none has ever been observed. (The fact that single sources in the universe can approach the power limit, while the universe also has to obey it, suggests the so-called power paradox. More about it below.) Ref. 98 Page 114

Similarly, all observed mass flow rates are orders of magnitude below the corresponding limit. Even physical systems that are mathematical analogues of black holes – for example, silent acoustical black holes or optical black holes – do not invalidate the force and power limits that hold in the corresponding systems.

In summary, the experimental situation is somewhat disappointing. Experiments do not contradict the limit values. But neither do the data do much to confirm them. The reason is the lack of horizons in everyday life and in experimentally accessible systems. The maximum speed at the basis of special relativity is found almost everywhere; maximum force and maximum power are found almost nowhere. Below we will propose some dedicated tests of the limits that could be performed in the future. Page 119

| | | | | |
|---|---|---|---|---|
| Maximum force $c^4/4G$,<br>Maximum power $c^5/4G$,<br>Maximum mass rate $c^3/4G$ | are equivalent to | First law of horizon mechanics<br>(horizon equation) | is equivalent to | Field equations of general relativity |

FIGURE 56 Showing the equivalence of the maximum force or power with the field equations of general relativity

## Deducing general relativity*

In order to establish the maximum force and power limits as fundamental physical principles, it is not sufficient to show that they are consistent with what we observe in nature. It is necessary to show that they imply the complete theory of general relativity. (This section is only for readers who already know the field equations of general relativity. Other Page 105 readers may skip to the next section.)

In order to derive the theory of relativity we need to study those systems that *realize* the limit under scrutiny. In the case of the special theory of relativity, the main system that realizes the limit speed is light. For this reason, light is central to the exploration of special relativity. In the case of general relativity, the systems that realize the limit are less obvious. We note first that a maximum force (or power) cannot be realized throughout a *volume* of space. If this were possible, a simple boost** could transform the force (or power) to a higher value. Therefore, nature can realize maximum force and power only on surfaces, not volumes. In addition, these surfaces must be unattainable. Ref. 97 These unattainable surfaces are basic to general relativity; they are called *horizons*. Maximum force and power only appear on horizons. We have encountered horizons in special Page 87 relativity, where they were defined as surfaces that impose limits to observation. (Note the contrast with everyday life, where a horizon is only a line, not a surface.) The present definition of a horizon as a surface of maximum force (or power) is equivalent to the definition as a surface beyond which no signal may be received. In both cases, a horizon is a surface beyond which interaction is impossible.

The connection between horizons and the maximum force is a central point of relativistic gravity. It is as important as the connection between light and the maximum speed in special relativity. In special relativity, we showed that the fact that light speed is the maximum speed in nature implies the Lorentz transformations. In general relativity, we will now prove that the maximum force in nature, which we can call the *horizon force*, implies the field equations of general relativity. To achieve this aim, we start with the realization that all horizons have an energy flow across them. The flow depends on the horizon curvature, as we will see. This connection implies that horizons cannot be planes, as an infinitely extended plane would imply an infinite energy flow.

The deduction of the equations of general relativity has only two steps, as shown in Figure 56. In the first step, we show that the maximum force or power principle implies the first 'law' of horizon mechanics. In the second step, we show that the first 'law' implies

* This section can be skipped at first reading. (The mentioned proof dates from December 2003.)

** A *boost* was defined in special relativity as a change of viewpoint to a second observer *moving* in relation to the first.

the field equations of general relativity.

The simplest finite horizon is a static sphere, corresponding to a Schwarzschild black hole. A spherical horizon is characterized by its radius of curvature $R$, or equivalently, by its surface gravity $a$; the two quantities are related by $2aR = c^2$. Now, the energy flowing through any horizon is always finite in extension, when measured along the propagation direction. We can thus speak more specifically of an energy pulse. Any energy pulse through a horizon is thus characterized by an energy $E$ and a proper length $L$. When the energy pulse flows perpendicularly through a horizon, the rate of momentum change, or force, for an observer at the horizon is

$$F = \frac{E}{L} . \tag{105}$$

Our goal is to show that the existence of a maximum force implies general relativity. Now, maximum force is realized on horizons. We thus need to insert the maximum possible values on both sides of equation (105) and to show that general relativity follows.

Using the maximum force value and the area $4\pi R^2$ for a spherical horizon we get

$$\frac{c^4}{4G} = \frac{E}{LA} 4\pi R^2 . \tag{106}$$

The fraction $E/A$ is the energy per area flowing through any area $A$ that is part of a horizon. The insertion of the maximum values is complete when we note that the length $L$ of the energy pulse is limited by the radius $R$. The limit $L \leqslant R$ follows from geometrical considerations: seen from the concave side of the horizon, the pulse must be shorter than the radius of curvature. An independent argument is the following. The length $L$ of an
Ref. 99 object accelerated by $a$ is limited, by special relativity, by $L \leqslant c^2/2a$. Special relativity already shows that this limit is related to the appearance of a horizon. Together with relation (106), the statement that horizons are surfaces of maximum force leads to the following important relation for static, spherical horizons:

$$E = \frac{c^2}{8\pi G} a A . \tag{107}$$

This *horizon equation* relates the energy flow $E$ through an area $A$ of a spherical horizon with surface gravity $a$. It states that the energy flowing through a horizon is limited, that this energy is proportional to the area of the horizon, and that the energy flow is proportional to the surface gravity. (The horizon equation is also called the *first law of black hole*
Ref. 100 *mechanics* or the *first law of horizon mechanics*.)

The above derivation also yields the intermediate result

$$E \leqslant \frac{c^4}{16\pi G} \frac{A}{L} . \tag{108}$$

This form of the horizon equation states more clearly that no surface other than a horizon can achieve the maximum energy flow, when the area and pulse length (or surface gravity) are given. No other domain of physics makes comparable statements: they are intrinsic to the theory of gravitation.

An alternative derivation of the horizon equation starts with the emphasis on power instead of on force, using $P = E/T$ as the initial equation.

It is important to stress that the horizon equations (107) and (108) follow from only two assumptions: first, there is a maximum speed in nature, and secondly, there is a maximum force (or power) in nature. No specific theory of gravitation is assumed. The horizon equation might even be testable experimentally, as argued below. (We also note that the horizon equation – or, equivalently, the force or power limit – implies a maximum mass change rate in nature given by $\mathrm{d}m/\mathrm{d}t \leqslant c^3/4G$.)

Next, we have to generalize the horizon equation from static and spherical horizons to general horizons. Since the maximum force is assumed to be valid for *all* observers, whether inertial or accelerating, the generalization is straightforward. For a horizon that is irregularly curved or time-varying the horizon equation becomes

$$\delta E = \frac{c^2}{8\pi G}\, a\, \delta A\ . \tag{109}$$

This differential relation – it might be called the *general horizon equation* – is valid for any horizon. It can be applied separately for every piece $\delta A$ of a dynamic or spatially changing horizon. The general horizon equation (109) has been known to be equivalent to general relativity at least since 1995, when this equivalence was (implicitly) shown

Ref. 101 by Jacobson. We will show that the differential horizon equation has the same role for general relativity as the equation $\mathrm{d}x = c\ \mathrm{d}t$ has for special relativity. From now on, when we speak of the horizon equation, we mean the general, differential form (109) of the relation.

It is instructive to restate the behaviour of energy pulses of length $L$ in a way that holds for any surface, even one that is not a horizon. Repeating the above derivation, we get

$$\frac{\delta E}{\delta A} \leqslant \frac{c^4}{16\pi G}\,\frac{1}{L}\ . \tag{110}$$

Equality is only realized when the surface $A$ is a horizon. In other words, whenever the value $\delta E/\delta A$ in a physical system approaches the right-hand side, a horizon starts to form. This connection will be essential in our discussion of apparent counter-examples to the limit principles.

If we keep in mind that on a horizon the pulse length $L$ obeys $L \leqslant c^2/2a$, it becomes clear that the general horizon equation is a consequence of the maximum force $c^4/4G$ or the maximum power $c^5/4G$. In addition, the horizon equation takes also into account maximum speed, which is at the origin of the relation $L \leqslant c^2/2a$. The horizon equation thus follows purely from these two limits of nature.

The remaining, second step of the argument is the derivation of general relativity

Ref. 101 from the general horizon equation. This derivation was provided by Jacobson, and the essential points are given in the following paragraphs. To see the connection between the general horizon equation (109) and the field equations, we only need to generalize the general horizon equation to general coordinate systems and to general directions of energy–momentum flow. This is achieved by introducing tensor notation that is adapted to curved space-time.

To generalize the general horizon equation, we introduce the general surface element $d\Sigma$ and the local boost Killing vector field $k$ that generates the horizon (with suitable norm). Jacobson uses these two quantities to rewrite the left-hand side of the general horizon equation (109) as

$$\delta E = \int T_{ab} k^a d\Sigma^b , \tag{111}$$

where $T_{ab}$ is the energy–momentum tensor. This expression obviously gives the energy at the horizon for arbitrary coordinate systems and arbitrary energy flow directions.

Jacobson's main result is that the factor $a\,\delta A$ in the right hand side of the general horizon equation (109) can be rewritten, making use of the (purely geometric) Raychaudhuri equation, as

$$a\,\delta A = c^2 \int R_{ab} k^a d\Sigma^b , \tag{112}$$

where $R_{ab}$ is the Ricci tensor describing space-time curvature. This relation describes how the local properties of the horizon depend on the local curvature.

Combining these two steps, the general horizon equation (109) becomes

$$\int T_{ab} k^a d\Sigma^b = \frac{c^4}{8\pi G} \int R_{ab} k^a d\Sigma^b . \tag{113}$$

Jacobson then shows that this equation, together with local conservation of energy (i.e., vanishing divergence of the energy–momentum tensor) can only be satisfied if

$$T_{ab} = \frac{c^4}{8\pi G} \left( R_{ab} - (\frac{R}{2} + \Lambda) g_{ab} \right) , \tag{114}$$

where $R$ is the Ricci scalar and $\Lambda$ is a constant of integration the value of which is not determined by the problem. The above equations are the full field equations of general relativity, including the cosmological constant $\Lambda$. The field equations thus follow from the horizon equation. They are therefore shown to be valid at horizons.

Since it is possible, by choosing a suitable coordinate transformation, to position a horizon at any desired space-time point, the field equations must be valid over the whole of space-time. This observation completes Jacobson's argument. Since the field equations follow, via the horizon equation, from the maximum force principle, we have also shown that at every space-time point in nature the same maximum force holds: the value of the maximum force is an invariant and a constant of nature.

In other words, the field equations of general relativity are a direct consequence of the limit on energy flow at horizons, which in turn is due to the existence of a maximum force (or power). In fact, Jacobson's derivation shows that the argument works in both directions. Maximum force (or power), the horizon equation, and general relativity are equivalent.

In short, *the maximum force principle is a simple way to state that, on horizons, energy* *flow is proportional to area and surface gravity*. This connection makes it possible to deduce the full theory of general relativity. In particular, a maximum force value is sufficient to tell space-time how to curve. We will explore the details of this relation shortly. Note that if no force limit existed in nature, it would be possible to 'pump' any desired amount of energy through a given surface, including any horizon. In this case, the energy flow would not be proportional to area, horizons would not have the properties they have, and general relativity would not hold. We thus get an idea how the maximum flow of energy, the maximum flow of momentum and the maximum flow of mass are all connected to horizons. The connection is most obvious for black holes, where the energy, momentum or mass are those falling into the black hole.

Page 237

We note that the deduction of general relativity's field equations from the maximum power of force is correct only under the assumption that gravity is purely geometric. This is the essential statement of general relativity. If the mechanism of gravity would be based on other fields, such as hitherto unknown particles, the equivalence between gravity and a maximum force would not be given.

Since the derivation of general relativity from the maximum force principle or from the maximum power principle is now established, we can rightly call these limits *horizon force* and *horizon power*. Every experimental or theoretical confirmation of the field equations indirectly confirms their existence.

## Space-time is curved

Imagine two observers who start moving parallel to each other and who both continue straight ahead. If after a while they discover that they are not moving parallel to each other any more, then they can deduce that they have moved on a curved surface (try it!) or in a curved space. In particular, this happens near localized energy, such as masses. The existence of a maximum force implies that space-time is curved near masses.

Challenge 160 s

A horizon so strongly curved that it forms a closed boundary, like the surface of a sphere, is called a black hole. We will study black holes in detail below. The main property of a black hole, like that of any horizon, is that it is impossible to detect what is 'behind' the boundary.*

Page 235

The analogy between special and general relativity can thus be carried further. In special relativity, maximum speed implies $\mathrm{d}x = c\ \mathrm{d}t$, and the change of time depends on the observer. In general relativity, maximum force (or power) implies the horizon equation $\delta E = \frac{c^2}{8\pi G}\, a\, \delta A$ and the observation that space-time is curved.

The maximum force (or power) thus has the same double role in general relativity as the maximum speed has in special relativity. In special relativity, the speed of light is the maximum speed; it is also the proportionality constant that connects space and time, as the equation $\mathrm{d}x = c\ \mathrm{d}t$ makes apparent. In general relativity, the horizon force is the maximum force; it also appears (with a factor $2\pi$) in the field equations as the proportionality constant connecting energy and curvature. The maximum force thus describes both the elasticity of space-time and – if we use the simple image of space-time as a medium – the maximum tension to which space-time can be subjected. This double role of a material constant as proportionality factor and as limit value is well known in materials science.

* Analogously, in special relativity it is impossible to detect what moves faster than the light barrier.

Why is the maximum force also the proportionality factor between curvature and energy? Imagine space-time as an elastic material.* The elasticity of a material is described by a numerical material constant. The simplest definition of this material constant is the ratio of stress (force per area) to strain (the proportional change of length). An exact definition has to take into account the geometry of the situation. For example, the shear modulus $G$ (or $\mu$) describes how difficult it is to move two parallel surfaces of a material against each other. If the force $F$ is needed to move two parallel surfaces of area $A$ and length $l$ against each other by a distance $\Delta l$, we define the shear modulus $G$ by

$$\frac{F}{A} = G\frac{\Delta l}{l} . \qquad (115)$$

The shear modulus for metals and alloys ranges between 25 and 80 GPa. The continuum theory of solids shows that for any crystalline solid without any defect (a 'perfect' solid) there is a so-called theoretical shear stress: when stresses higher than this value are applied, the material breaks. The theoretical shear stress, in other words, the maximum stress in a material, is given by

$$G_{\text{tss}} = \frac{G}{2\pi} . \qquad (116)$$

The maximum stress is thus essentially given by the shear modulus. This connection is similar to the one we found for the vacuum. Indeed, imagining the vacuum as a material
Ref. 102
that can be bent is a helpful way to understand general relativity. We will use it regularly in the following.

What happens when the vacuum is stressed with the maximum force? Is it also torn apart like a solid? Almost: in fact, when vacuum is torn apart, particles appear. We will find out more about this connection later on: since particles are quantum entities, we need to study quantum theory first, before we can describe the effect in the last part of
Vol. VI, page 263
our mountain ascent.

## Conditions of validity of the force and power limits

The maximum force value is valid only under certain conditions. To clarify this point, we can compare the situation to the maximum speed. There are three conditions for the valdity of maximum speed.

First of all, the speed of light (in vacuum) is an upper limit for motion of systems with *momentum* or *energy* only. It can, however, be exceeded for motions of non-material points. Indeed, the cutting point of a pair of scissors, a laser light spot on the Moon, shadows, or the group velocity or phase velocity of wave groups can exceed the speed of
Page 53
light.

Secondly, the speed of light is a limit only if measured *near* the moving mass or energy: the Moon moves faster than light if one turns around one's axis in a second; distant points in a Friedmann universe move apart from each other with speeds larger than the speed of light.

* Does this analogy make you think about aether? Do not worry: physics has no need for the concept of
Vol. III, page 112
aether, because it is indistinguishable from vacuum. General relativity does describe the vacuum as a sort of material that can be deformed and move – but it does not need nor introduce the aether.

Thirdly, the observer measuring speeds must be *realistic*: the observer must be made of matter and energy, thus must move more slowly than light, and must be able to observe
Ref. 103
the system. No system moving at or above the speed of light can be an observer.

The same three conditions apply in general relativity for the validity of maximum force and power. The third point is especially important. In particular, relativistic gravity forbids point-like observers and test masses: they are not realistic. Surfaces moving faster than light are also not realistic. In such cases, counter-examples to the maximum force
Challenge 161 s
claim can be found. Try and find one – many are possible, and all are fascinating. We now explore some of the most important ones.

## Gedanken experiments and paradoxes about the force limit

> “Wenn eine Idee am Horizonte eben aufgeht, ist gewöhnlich die Temperatur der Seele dabei sehr kalt. Erst allmählich entwickelt die Idee ihre Wärme, und am heissesten ist diese (das heisst sie tut ihre grössten Wirkungen), wenn der Glaube an die Idee schon wieder im Sinken ist.”
> Friedrich Nietzsche*

The last, but central, step in our discussion of the force limit is the same as in the discussion of the speed limit. We saw that no real experiment has ever led to a force value large thna the force limit. But we also need to show that no *imaginable* experiment can overcome the force limit. Following a tradition dating back to the early twentieth century, such an imagined experiment is called a *Gedanken experiment*, from the German Gedankenexperiment, meaning ‘thought experiment’.

In order to dismiss all imaginable attempts to exceed the maximum speed, it was sufficient to study the properties of velocity addition and the divergence of kinetic energy near the speed of light. In the case of maximum force, the task is more involved. Indeed, stating a maximum force, a maximum power and a maximum mass change easily provokes numerous attempts to contradict them.

* *

The brute force approach. The simplest attempt to exceed the force limit is to try to accelerate an object with a force larger than the maximum value. Now, acceleration implies the transfer of energy. This transfer is limited by the horizon equation (109) or the limit (110). For any attempt to exceed the force limit, the flowing energy results in the appearance of a horizon. But a horizon prevents the force from exceeding the limit, because it imposes a limit on interaction.

Page 91
We can explore this limit directly. In special relativity we found that the acceleration of an object is limited by its length. Indeed, at a distance given by $c^2/2a$ in the direction opposite to the acceleration $a$, a *horizon* appears. In other words, an accelerated body

* ‘When an idea is just rising on the horizon, the soul’s temperature with respect to it is usually very cold. Only gradually does the idea develop its warmth, and it is hottest (which is to say, exerting its greatest influence) when belief in the idea is already once again in decline.’ Friedrich Nietzsche (1844–1900), German philosopher and scholar. This is aphorism 207 – *Sonnenbahn der Idee* – from his text *Menschliches Allzumenschliches – Der Wanderer und sein Schatten.*

breaks, at the latest, at that point. The force $F$ on a body of mass $M$ and radius $R$ is thus limited by

$$F \leqslant \frac{M}{2R} c^2 . \qquad (117)$$

It is straightforward to add the (usually small) effects of gravity. To be observable, an accelerated body must remain *larger* than a black hole; inserting the corresponding radius $R = 2GM/c^2$ we get the force limit (102). *Dynamic* attempts to exceed the force limit thus fail.

* *

The rope attempt. We can also try to generate a higher force in a *static* situation, for example by pulling two ends of a rope in opposite directions. We assume for simplicity that an unbreakable rope exists. Any rope works because the potential energy between its atoms can produce high forces between them. To produce a rope force exceeding the limit value, we need to store large (elastic) energy in the rope. This energy must enter from the ends. When we increase the tension in the rope to higher and higher values, more and more (elastic) energy must be stored in smaller and smaller distances. To exceed the force limit, we would need to add more energy per distance and area than is allowed by the horizon equation. A horizon thus inevitably appears. But there is no way to stretch a rope across a horizon, even if it is unbreakable. A horizon leads either to the breaking of the rope or to its detachment from the pulling system. Horizons thus make it impossible to generate forces larger than the force limit. In fact, the assumption of infinite wire strength is unnecessary: the force limit cannot be exceeded even if the strength of the wire is only finite.

We note that it is not important whether an applied force pulls – as for ropes or wires – or pushes. In the case of *pushing* two objects against each other, an attempt to increase the force value without end will equally lead to the formation of a horizon, due to the limit provided by the horizon equation. By definition, this happens precisely at the force limit. As there is no way to use a horizon to push (or pull) on something, the attempt to achieve a higher force ends once a horizon is formed. Static forces cannot exceed the limit value.

* *

The braking attempt. A force limit provides a maximum momentum change per time. We can thus search for a way to *stop* a moving physical system so abruptly that the maximum force might be exceeded. The non-existence of rigid bodies in nature, already known from special relativity, makes a completely sudden stop impossible; but special relativity on its own provides no lower limit to the stopping time. However, the inclusion of gravity does. Stopping a moving system implies a transfer of energy. The energy flow per area cannot exceed the value given by the horizon equation. Thus we cannot exceed the force limit by stopping an object.

Page 91

Similarly, if a rapid system is *reflected* instead of stopped, a certain amount of energy needs to be transferred and stored for a short time. For example, when a tennis ball is reflected from a large wall its momentum changes and a force is applied. If many such balls are reflected at the same time, surely a force larger than the limit can be realized?

It turns out that this is impossible. If we attempted it, the energy flow at the wall would reach the limit given by the horizon equation and thus create a horizon. In that case, no reflection is possible any more. So the limit cannot be exceeded.

* *

The classical radiation attempt. Instead of systems that pull, push, stop or reflect *matter*, we can explore systems where *radiation* is involved. However, the arguments hold in exactly the same way, whether photons, gravitons or other particles are involved. In particular, mirrors, like walls, are limited in their capabilities.

It is also impossible to create a force larger than the maximum force by concentrating a large amount of light onto a surface. The same situation as for tennis balls arises: when the limit value $E/A$ given by the horizon equation (110) is reached, a horizon appears that prevents the limit from being broken.

* *

The brick attempt. The force and power limits can also be tested with more concrete Gedanken experiments. We can try to exceed the force limit by stacking weight. But even building an infinitely high brick tower does not generate a sufficiently strong force on its foundations: integrating the weight, taking into account its decrease with height, yields a finite value that cannot reach the force limit. If we continually increase the mass density of the bricks, we need to take into account that the tower and the Earth will change into a black hole. And black holes, as mentioned above, do not allow the force limit to be exceeded.

* *

The boost attempt. A boost can apparently be chosen in such a way that a 3-force value $F$ in one frame is transformed into any desired value $F'$ in another frame. This turns out to be wrong. In relativity, 3-force cannot be increased beyond all bounds using boosts. In all reference frames, the meaured 3-force can never exceed the proper force, i.e., the 3-force value measured in the comoving frame. (The situation can be compared to 3-velocity, where a boost cannot be used to exceed the value $c$, whatever boost we may choose; however, there is no strict equivalence, as the transformation behaviour of 3-force and of 3-velocity differ markedly.)

Ref. 104
Page 74

* *

The divergence attempt. The force on a test mass $m$ at a radial distance $d$ from a Schwarzschild black hole (for $\Lambda = 0$) is given by

Ref. 98

$$F = \frac{GMm}{d^2\sqrt{1 - \frac{2GM}{dc^2}}} . \tag{118}$$

In addition, the inverse square law of universal gravitation states that the force between two masses $m$ and $M$ is

$$F = \frac{GMm}{d^2} . \tag{119}$$

Both expressions can take any value; this suggest that no maximum force limit exists.

A detailed investigation shows that the maximum force still holds. Indeed, the force in the two situations diverges only for non-physical point-like masses. However, the maximum force implies a minimum approach distance to a mass $m$ given by

$$d_{\text{min}} = \frac{2Gm}{c^2} \,. \tag{120}$$

The minimum approach distance – in simple terms, this would be the corresponding black hole radius – makes it impossible to achieve zero distance between two masses or between a horizon and a mass. This implies that a mass can never be point-like, and that there is a (real) minimum approach distance, proportional to the mass. If this minimum approach distance is introduced in equations (118) and (119), we get

$$F = \frac{c^4}{4G} \frac{Mm}{(M+m)^2} \frac{1}{\sqrt{1 - \frac{M}{M+m}}} \leqslant \frac{c^4}{4G} \tag{121}$$

and

$$F = \frac{c^4}{4G} \frac{Mm}{(M+m)^2} \leqslant \frac{c^4}{4G} . \tag{122}$$

The maximum force value is thus never exceeded, as long as we take into account the size of observers.

* *

The consistency problem. If observers cannot be point-like, we might question whether it is still correct to apply the original definition of momentum change or energy change as the integral of values measured by observers attached to a given surface. In general relativity, observers cannot be point-like, but they can be as small as desired. The original definition thus remains applicable when taken as a limit procedure for ever-decreasing observer size. Obviously, if quantum theory is taken into account, this limit procedure comes to an end at the Planck length. This is not an issue for general relativity, as long as the typical dimensions in the situation are much larger than this value.

* *

The quantum problem. If quantum effects are neglected, it is possible to construct surfaces with sharp angles or even fractal shapes that overcome the force limit. However, such surfaces are not physical, as they assume that lengths smaller than the Planck length can be realized or measured. The condition that a surface be physical implies that it must have an intrinsic indeterminacy given by the Planck length. A detailed study shows that quantum effects do not allow the horizon force to be exceeded.

Challenge 162 e

Ref. 97

* *

The relativistically extreme observer attempt. Any extreme observer, whether in rapid inertial or in accelerated motion, has no chance to beat the limit. In classical physics we are used to thinking that the interaction necessary for a measurement can be made as small as desired. This statement, however, is not valid for all observers; in particular, extreme observers cannot fulfil it. For them, the measurement interaction is large. As a result, a horizon forms that prevents the limit from being exceeded.

* *

The microscopic attempt. We can attempt to exceed the force limit by accelerating a small particle as strongly as possible or by colliding it with other particles. High forces do indeed appear when two high energy particles are smashed against each other. However, if the combined energy of the two particles became high enough to challenge the force limit, a horizon would appear before they could get sufficiently close.

Ref. 105 In fact, quantum theory gives exactly the same result. Quantum theory by itself already provides a limit to acceleration. For a particle of mass $m$ it is given by

$$a \leqslant \frac{2mc^3}{\hbar} . \qquad (123)$$

Here, $\hbar = 1.1 \cdot 10^{-34}$ Js is the quantum of action, a fundamental constant of nature. In particular, this acceleration limit is satisfied in particle accelerators, in particle collisions and in pair creation. For example, the spontaneous generation of electron–positron pairs in intense electromagnetic fields or near black hole horizons does respect the limit (123). Inserting the maximum possible mass for an elementary particle, namely the (corrected) Page 36 Planck mass, we find that equation (123) then states that the horizon force is the upper bound for elementary particles.

* *

The compaction attempt. Are black holes really the most dense form of matter or energy? The study of black hole thermodynamics shows that mass concentrations with higher Ref. 98 density than black holes would contradict the principles of thermodynamics. In black hole thermodynamics, surface and entropy are related: reversible processes that reduce entropy could be realized if physical systems could be compressed to smaller values than the black hole radius. As a result, the size of a black hole is the limit size for a mass in nature. Equivalently, the force limit cannot be exceeded in nature.

* *

The force addition attempt. In special relativity, composing velocities by a simple vector addition is not possible. Similarly, in the case of forces such a naive sum is incorrect; any attempt to add forces in this way would generate a horizon. If textbooks on relativity had explored the behaviour of force vectors under addition with the same care with which they explored that of velocity vectors, the force bound would have appeared much earlier in the literature. (Obviously, general relativity is required for a proper treatment.)

* *

In special relativity, a body moving more slowly than light in one frame does so in all frames. Can you show that a force smaller than the invariant limit $c^4/4G$ in one frame

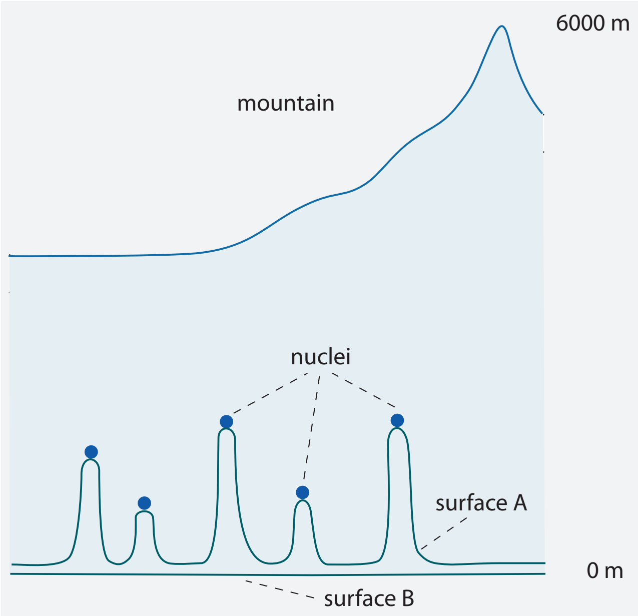



FIGURE 57 The mountain attempt to exceed the maximum mass flow value

Challenge 163 s of reference is also smaller in any other frame?

* *

Can you propose and then resolve an additional attempt to exceed the force or power limit? Challenge 164 r

## Gedanken experiments with the power limit and the mass flow limit

Like the force bound, the power bound must be valid for all *imaginable* systems. Here are some attempts to refute it.

* *

The cable-car attempt. Imagine an engine that accelerates a mass with an unbreakable and massless wire (assuming that such a wire could exist). As soon as the engine reached the power bound, either the engine or the exhausts would reach the horizon equation. When a horizon appears, the engine cannot continue to pull the wire, as a wire, even an infinitely strong one, cannot pass a horizon. The power limit thus holds whether the engine is mounted inside the accelerating body or outside, at the end of the wire pulling it.

* *

The mountain attempt. It is possible to define a surface that is so strangely bent that it passes *just below* every nucleus of every atom of a mountain, like the surface A in Figure 57. All atoms of the mountain above sea level are then *just above* the surface, barely touching it. In addition, imagine that this surface is moving *upwards* with almost the speed of light. It is not difficult to show that the mass flow through this surface is higher than the mass flow limit. Indeed, the mass flow limit $c^3/4G$ has a value of about $10^{35}$ kg/s. In a time of $10^{-22}$ s, the diameter of a nucleus divided by the speed of light, only $10^{13}$ kg need to flow through the surface: that is the mass of a mountain.

This surface seems to provide a counter-example to the limit. However, a closer look shows that this is not the case. The problem is the expression 'just below'. Nuclei are quantum particles and have an indeterminacy in their position; this indeterminacy is essentially the nucleus–nucleus distance. As a result, in order to be sure that the surface of interest has all atoms *above* it, the shape cannot be that of surface A in Figure 57. It must be a flat plane that remains below the whole mountain, like surface B in the figure. However, a flat surface beneath a mountain does not allow the mass change limit to be exceeded.

* *

The multiple atom attempt. We can imagine a number of atoms equal to the number of the atoms of a mountain that all lie with large spacing (roughly) in a single plane. Again, the plane is moving upwards with the speed of light. But also in this case the indeterminacy in the atomic positions makes it impossible to say that the mass flow limit has been exceeded.

* *

The multiple black hole attempt. Black holes are typically large and the indeterminacy in their position is thus negligible. The mass limit $c^3/4G$, or power limit $c^5/4G$, corresponds to the flow of a single black hole moving through a plane at the speed of light. Several black holes crossing a plane together at just under the speed of light thus seem to beat the limit. However, the surface has to be physical: an observer must be possible on each of its points. But no observer can cross a black hole. A black hole thus effectively punctures the plane surface. No black hole can ever be said to cross a plane surface; even less so a multiplicity of black holes. The limit remains valid.

* *

The multiple neutron star attempt. The mass limit seems to be in reach when several neutron stars (which are slightly less dense than a black hole of the same mass) cross a plane surface at the same time, at high speed. However, when the speed approaches the speed of light, the crossing time for points far from the neutron stars and for those that actually cross the stars differ by large amounts. Neutron stars that are almost black holes cannot be crossed in a short time in units of a coordinate clock that is located far from the stars. Again, the limit is not exceeded.

* *

The luminosity attempt. The existence of a maximum luminosity bound has been discussed by astrophysicists. In its full generality, the maximum bound on power, i.e., on energy per time, is valid for any energy flow *through any physical surface whatsoever*. The physical surface may even run across the whole universe. However, not even bringing together all lamps, all stars and all galaxies of the universe yields a surface which has a larger power output than the proposed limit.

Ref. 98

The surface must be *physical*.* A surface is *physical* if an observer can be placed on each of its points. In particular, a physical surface may not cross a horizon, or have local detail finer than a certain minimum length. This minimum length will be introduced later on; it is given by the corrected Planck length. If a surface is not physical, it may provide a counter-example to the power or force limits. However, these counter-examples make no statements about nature. (*Ex falso quodlibet.***)

Vol. VI, page 60
Challenge 165 s

* *

The many lamp attempt, or *power paradox*. An absolute power limit imposes a limit on the rate of energy transport through any imaginable surface. At first sight, it may seem that the combined power emitted by two radiation sources that each emit 3/4 of the maximum value should emit 3/2 times the maximum value, and thus allow to overcome the power limit. However, two such lamps would be so massive that they would form a horizon around them – a black hole would form. No amount of radiation that exceeds the power limit can leave. Again, since the horizon limit (110) is achieved, a horizon appears that swallows the light and prevents the force or power limit from being exceeded.

* *

The light concentration attempt. Another approach is to shine a powerful, short and spherical flash of light onto a spherical mass. At first sight it seems that the force and power limits can be exceeded, because light energy can be concentrated into small volumes. However, a high concentration of light energy forms a black hole or induces the mass to form one. There is no way to pump energy into a mass at a faster rate than that dictated by the power limit. In fact, it is impossible to group light sources in such a way that their total output is larger than the power limit. Every time the force limit is approached, a horizon appears that prevents the limit from being exceeded.

* *

The black hole attempt. One possible system in nature that actually *achieves* the power limit is the final stage of black hole evaporation. However, even in this case the power limit is not exceeded, but only equalled.

* *

The saturation attempt. If the universe already saturates the power limit, a new source would break it, or at least imply that another elsewhere must close down. Can you find the fallacy in this argument?

Challenge 166 ny

* *

The water flow attempt. We could try to pump water as rapidly as possible through a large tube of cross-section $A$. However, when a tube of length $L$ filled with water flowing at speed $v$ gets near to the mass flow limit, the gravity of the water *waiting* to be pumped through the area $A$ will slow down the water that is being pumped through the area. The limit is again reached when the cross-section $A$ turns into a horizon.

* It can also be called *physically sensible*.
** 'Anything can be deduced from a falsehood.'

Checking that no system – from microscopic to astrophysical – ever exceeds the maximum power or maximum mass flow is a further test of general relativity. It may seem easy to find a counter-example, as the surface may run across the whole universe or envelop any number of elementary particle reactions. However, no such attempt succeeds.

In summary, in all situations where the force, power or mass-flow limit is challenged, whenever the energy flow reaches the black hole mass–energy density in space or the corresponding momentum flow in time, an event horizon appears; this horizon makes it impossible to exceed the limits. All three limits are confirmed both in observation and in theory. Values exceeding the limits can neither be generated nor measured. Gedanken experiments also show that the three bounds are the tightest ones possible. Obviously, all three limits are open to future tests and to further Gedanken experiments. (If you can
Challenge 167 r
think of a good one, let me know.)

## Why maximum force has remained undiscovered for so long

The *first* reason why the maximum force principle remained undiscovered for so long is the absence of horizons in everyday life. Due to this lack, experiments in everyday life do not highlight the force or power limits. It took many decades before physicists realized that the dark night sky is not something unique, but only one example of an observation that is common in nature: nature is full of horizons. But in everyday life, horizons do not play an important role – fortunately – because the nearest one is located at the centre of the Milky Way.

The *second* reason why the principle of maximum force remained hidden is the erroneous belief that point particles exist. This is a theoretical prejudice due to a common idealization used in Galilean physics. For a complete understanding of general relativity it is essential to remember regularly that point particles, point masses and point-like observers do not exist. They are approximations that are only applicable in Galilean physics, in special relativity or in quantum theory. In general relativity, horizons prevent the existence of point-like systems. The incorrect habit of believing that the size of a system can be made as small as desired while keeping its mass constant prevents the force or power limit from being noticed.

The *third* reason why the principle of maximum force remained hidden are prejudices against the concept of force. In general relativity, gravitational force is hard to define. Even in Galilean physics it is rarely stressed that force is the flow of momentum through a surface. The teaching of the concept of force is incomplete since centuries – with rare
Ref. 106
notable exceptions – and thus the concept is often avoided.

In short, the principle of maximum force – or of maximum power – has thus remained undiscovered for so long because a 'conspiracy' of nature and of thinking habits hid it from most experimental and theoretical physicists.

## An intuitive understanding of general relativity

> Wir leben zwar alle unter dem gleichen Himmel, aber wir haben nicht alle den gleichen Horizont.*
>
> Konrad Adenauer

The concepts of horizon force and horizon power can be used as the basis for a direct, intuitive approach to general relativity.

* *

What is gravity? Of the many possible answers we will encounter, we now have the first: gravity is the 'shadow' of the maximum force. Whenever we experience gravity as weak, we can remember that a different observer at the same point and time would experience the maximum force. Searching for the precise properties of that observer is a good exercise. Another way to put it: if there were no maximum force, gravity would not exist.

* *

The maximum force implies universal gravity. To see this, we study a simple planetary system, i.e., one with small velocities and small forces. A simple planetary system of size $L$ consists of a (small) satellite circling a central mass $M$ at a radial distance $R = L/2$. Let $a$ be the acceleration of the object. Small velocity implies the condition $aL \ll c^2$, deduced from special relativity; small force implies $\sqrt{4GMa} \ll c^2$, deduced from the force limit. These conditions are valid for the system as a whole and for all its components. Both expressions have the dimensions of speed squared. Since the system has only one characteristic speed, the two expressions $aL = 2aR$ and $\sqrt{4GMa}$ must be proportional, yielding

$$a = f \frac{GM}{R^2} , \tag{124}$$

where the numerical factor $f$ must still be determined. To determine it, we study the escape velocity necessary to leave the central body. The escape velocity must be smaller than the speed of light for any body larger than a black hole. The escape velocity, derived from expression (124), from a body of mass $M$ and radius $R$ is given by $v_{esc}^2 = 2fGM/R$. The minimum radius $R$ of objects, given by $R = 2GM/c^2$, then implies that $f = 1$. Therefore, for low speeds and low forces, the inverse square law describes the orbit of a satellite around a central mass.

* *

If empty space-time is elastic, like a piece of metal, it must also be able to oscillate. Any physical system can show oscillations when a deformation brings about a restoring force. We saw above that there is such a force in the vacuum: it is called gravitation. In other words, vacuum must be able to oscillate, and since it is extended, it must also be able to sustain waves. Indeed, gravitational waves are predicted by general relativity, as we will see below. Page 151

* 'We all live under the same sky, but we do not have the same horizon.' Konrad Adenauer (1876–1967), West German Chancellor.

* *

If curvature and energy are linked, the maximum speed must also hold for gravitational energy. Indeed, we will find that gravity has a finite speed of propagation. The inverse square law of everyday life cannot be correct, as it is inconsistent with any speed limit. More about the corrections induced by the maximum speed will become clear shortly. In addition, since gravitational waves are waves of massless energy, we would expect the
Page 151
maximum speed to be their propagation speed. This is indeed the case, as we will see.

* *

A body cannot be denser than a (non-rotating) black hole of the same mass. The maximum force and power limits that apply to horizons make it impossible to squeeze mass into smaller horizons. The maximum force limit can therefore be rewritten as a limit for the size $L$ of physical systems of mass $m$:

$$L \geqslant \frac{4Gm}{c^2} . \qquad (125)$$

If we call twice the radius of a black hole its 'size', we can state that no physical system of mass $m$ is smaller than this value.* The size limit plays an important role in general relativity. The opposite inequality, $m \geqslant \sqrt{A/16\pi}\, c^2/G$, which describes the maximum 'size' of black holes, is called the *Penrose inequality* and has been proven for many physically realistic situations. The Penrose inequality can be seen to imply the maximum force
Ref. 107, Ref. 108, Ref. 109
limit, and vice versa. The maximum force principle, or the equivalent minimum size of matter–energy systems, thus prevents the formation of naked singularities. (Physicists call the lack of naked singularities the so-called *cosmic censorship*. conjecture.)

* *

There is a power limit for all energy sources. In particular, the value $c^5/4G$ limits the luminosity of all gravitational sources. Indeed, all formulae for gravitational wave emission
Ref. 98
imply this value as an upper limit. Furthermore, numerical relativity simulations never exceed it: for example, the power emitted during the simulated merger of two black holes is below the limit.

* *

Perfectly plane waves do not exist in nature. Plane waves are of infinite extension. But neither electrodynamic nor gravitational waves can be infinite, since such waves would carry more momentum per time through a plane surface than is allowed by the force limit. The non-existence of plane gravitational waves also precludes the production of singularities when two such waves collide.

* *

In nature, there are no infinite forces. There are thus no naked singularities in nature. Horizons prevent the appearance of naked singularities. In particular, the big bang was

* The maximum value for the mass to size limit is obviously equivalent to the maximum mass change given above.

*not* a singularity. The mathematical theorems by Penrose and Hawking that seem to imply the existence of singularities tacitly assume the existence of point masses – often in the form of 'dust' – in contrast to what general relativity implies. Careful re-evaluation of each such proof is necessary.

* *

The force limit means that space-time has a limited stability. The limit suggests that space-time can be torn into pieces. This is indeed the case. However, the way that this happens is not described by general relativity. We will study it in the last part of this text.

* *

The maximum force is the standard of force. This implies that the gravitational constant $G$ is constant in space and time – or at least, that its variations across space and time cannot be detected. Present data support this claim to a high degree of precision. Ref. 110

* *

The maximum force principle implies that gravitational energy – as long as it can be defined – *falls* in gravitational fields in the same way as other type of energy. As a result, the maximum force principle predicts that the Nordtvedt effect vanishes. Ref. 98 The Nordtvedt effect is a hypothetical periodical change in the orbit of the Moon that would appear if the gravitational energy of the Earth–Moon system did not fall, like other mass–energy, in the gravitational field of the Sun. Lunar range measurements have confirmed the absence of this effect.

* *

If horizons are surfaces, we can ask what their colour is. This question will be explored later on. Page 235

* *

Page 34 Challenge 168 e Later on we will find that quantum effects cannot be used to exceed the force or power limit. (Can you guess why?) Quantum theory also provides a limit to motion, namely a lower limit to action; however, this limit is independent of the force or power limit. (A dimensional analysis already shows this: there is no way to define an action by combinations of $c$ and $G$.) Therefore, even the combination of quantum theory and general relativity does not help in overcoming the force or power limits.

## An intuitive understanding of cosmology

Page 216 A maximum power is the simplest possible explanation of Olbers' paradox. Power and luminosity are two names for the same observable. The sum of all luminosities in the universe is finite; the light and all other energy emitted by all stars, taken together, is finite. If we assume that the universe is homogeneous and isotropic, the power limit $P \leqslant c^5/4G$ must be valid across any plane that divides the universe into two halves. The part of the universe's luminosity that arrives on Earth is then so small that the sky is dark at night. In fact, the actually measured luminosity is still smaller than this calculation, as a large part of the power is not visible to the human eye (since most of it is matter anyway). In other words, the night is dark because of nature's power limit. This explanation is *not* in contrast to the usual one, which uses the finite lifetime of stars, their finite density, their finite size, and the finite age and the expansion of the universe. In fact, the combination of all these usual arguments simply implies and repeats in more complex words that the power limit cannot be exceeded. However, this more simple explanation seems to be absent in the literature.

The existence of a maximum force in nature, together with homogeneity and isotropy, implies that the visible universe is of *finite size*. The opposite case would be an infinitely large, homogeneous and isotropic universe. But in that case, any two halves of the universe would attract each other with a force above the limit (provided the universe were sufficiently old). This result can be made quantitative by imagining a sphere whose centre lies at the Earth, which encompasses all the universe, and whose radius decreases with time almost as rapidly as the speed of light. The mass flow $\mathrm{d}m/\mathrm{d}t = \rho A v$ is predicted to reach the mass flow limit $c^3/4G$; thus we have

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \rho_0 4\pi R_0^2 c = \frac{c^3}{4G} \quad , \tag{126}$$

Ref. 111 a relation also predicted by the Friedmann models. The precision measurements of the cosmic background radiation by the WMAP satellite confirm that the present-day total energy density $\rho_0$ (including dark matter and dark energy) and the horizon radius $R_0$ just reach the limit value. The maximum force limit thus predicts the observed size of the universe.

A finite power limit also suggests that a finite age for the universe can be deduced.
Challenge 169 s Can you find an argument?

## Experimental challenges for the third millennium

The lack of direct tests of the horizon force, power or mass flow is obviously due to the lack of horizons in the environment of all experiments performed so far. Despite the difficulties in reaching the limits, their values are observable and falsifiable.

In fact, the force limit might be tested with high-precision measurements in binary pulsars or binary black holes. Such systems allow precise determination of the positions of the two stars. The maximum force principle implies a relation between the position
Ref. 97 error $\Delta x$ and the energy error $\Delta E$. For all systems we have

$$\frac{\Delta E}{\Delta x} \leqslant \frac{c^4}{4G} \ . \tag{127}$$

For example, a position error of 1 mm gives a mass error of below $3 \cdot 10^{23}$ kg. In everyday life, all measurements comply with this relation. Indeed, the left side is so much smaller than the right side that the relation is rarely mentioned. For a direct check, only systems which might achieve direct equality are interesting. Dual black holes or dual pulsars are such systems.

It might be that one day the amount of matter falling into some black hole, such as the one at the centre of the Milky Way, might be measured. The limit $\mathrm{d}m/\mathrm{d}t \leqslant c^3/4G$

could then be tested directly.

The power limit implies that the highest luminosities are only achieved when systems emit energy at the speed of light. Indeed, the maximum emitted power is only achieved when all matter is radiated away as rapidly as possible: the emitted power $P = Mc^2/(R/v)$ cannot reach the maximum value if the body radius $R$ is larger than that of a black hole (the densest body of a given mass) or the emission speed $v$ is lower than that of light. The sources with highest luminosity must therefore be of maximum density and emit entities without rest mass, such as gravitational waves, electromagnetic waves or (maybe) gluons. Candidates to detect the limit are black holes in formation, in evaporation or undergoing mergers.

A candidate surface that reaches the limit is the night sky. The night sky is a horizon. Provided that light, neutrino, particle and gravitational wave flows are added together, the limit $c^5/4G$ is predicted to be reached. If the measured power is smaller than the limit (as it seems to be at present), this might even give a hint about new particles yet to be discovered. If the limit were exceeded or not reached, general relativity would be shown to be incorrect. This might be an interesting future experimental test.

The power limit implies that a wave whose integrated intensity approaches the force limit cannot be plane. The power limit thus implies a limit on the product of intensity $I$ (given as energy per unit time and unit area) and the size (curvature radius) $R$ of the front of a wave moving with the speed of light $c$:

$$4\pi R^2 I \leqslant \frac{c^5}{4G} . \tag{128}$$

Obviously, this statement is difficult to check experimentally, whatever the frequency and type of wave might be, because the value appearing on the right-hand side is extremely large. Possibly, future experiments with gravitational wave detectors, X-ray detectors, gamma ray detectors, radio receivers or particle detectors might allow us to test relation (128) with precision. (You might want to predict which of these experiments will confirm the limit first.)

Challenge 170 e

The lack of direct experimental tests of the force and power limits implies that *indirect tests* become particularly important. All such tests study the motion of matter or energy and compare it with a famous consequence of the force and power limits: the field equations of general relativity. This will be our next topic.

## A summary of general relativity

There is a simple axiomatic formulation of general relativity: the horizon force $c^4/4G$ and the horizon power $c^5/4G$ are the highest possible force and power values. No contradicting observation is known. No counter-example has been imagined. General relativity follows from these limits. Moreover, the limits imply the darkness of the night and the finiteness of the size of the universe.

The principle of maximum force has obvious applications for the teaching of general relativity. The principle brings general relativity to the level of first-year university, and possibly to well-prepared secondary school, students: only the concepts of maximum force and horizon are necessary. Space-time curvature is a consequence of horizon curvature.

The concept of a maximum force points to an additional aspect of gravitation. The cosmological constant $\Lambda$ is not fixed by the maximum force principle. (However, the principle does fix its sign to be positive.) Present measurements give the result $\Lambda \approx 10^{-52}\,/\mathrm{m}^2$. A positive cosmological constant implies the existence of a negative energy volume density $-\Lambda c^4/G$. This value corresponds to a negative pressure, as pressure and energy density have the same dimensions. Multiplication by the (numerically corrected) Planck area $2G\hbar/c^3$, the smallest area in nature, gives a force value

Challenge 171 ny
Page 217
Page 33

$$F = 2\Lambda\hbar c = 0.60 \cdot 10^{-77}\,\mathrm{N}\,. \tag{129}$$

This is also the gravitational force between two (numerically corrected) Planck masses $\sqrt{\hbar c/8G}$ located at the cosmological distance $1/4\sqrt{\Lambda}$. If we make the somewhat wishful assumption that expression (129) is the smallest possible force in nature (the numerical factors are not yet verified), we get the fascinating conjecture that the full theory of general relativity, including the cosmological constant, may be defined by the combination of a *maximum* and a *minimum* force in nature. (Can you find a smaller force?)

Challenge 172 d

Proving the minimum force conjecture is more involved than for the case of the maximum force. So far, only some hints are possible. Like the maximum force, the minimum force must be compatible with gravitation, must not be contradicted by any experiment, and must withstand any Gedanken experiment. A quick check shows that the minimum force, as we have just argued, allows us to deduce gravitation, is an invariant, and is not contradicted by any experiment. There are also hints that there may be no way to generate or measure a smaller value. For example, the minimum force corresponds to the energy per length contained by a photon with a wavelength of the size of the universe. It is hard – but maybe not impossible – to imagine the production of a still smaller force.

We have seen that the maximum force principle and general relativity fail to fix the value of the cosmological constant. Only a unified theory can do so. We thus get two requirements for such a theory. First, any unified theory must predict the same upper limit to force. Secondly, a unified theory must fix the cosmological constant. The appearance of $\hbar$ in the conjectured expression for the minimum force suggests that the minimum force is determined by a combination of general relativity and quantum theory. The proof of this suggestion and the direct measurement of the minimum force are two important challenges for our ascent beyond general relativity.

We are now ready to explore the consequences of general relativity and its field equations in more detail. We start by focusing on the concept of space-time curvature in everyday life, and in particular, on its consequences for the observation of motion.

Chapter 5

# HOW MAXIMUM SPEED CHANGES SPACE, TIME AND GRAVITY

> “Sapere aude.*”
>
> Horace *Epistulae*, 1, 2, 40.

Observation shows that gravitational influences do transport energy.** Our description of gravity must therefore include the speed limit. Only a description that takes into account that the limit speed for energy transport can be a precise description of gravity. Henri Poincaré stated this requirement for a precise description of gravitation as early as 1905. But universal gravity, with its relation $a = GM/r^2$, allows speeds higher than that of light. For example, the speed of a mass in orbit is not limited. In universal gravity it is also unclear how the values of $a$ and $r$ depend on the observer. In short, universal gravity cannot be correct. In order to reach the correct description, called *general relativity* by Albert Einstein, we have to throw quite a few preconceptions overboard.

Ref. 112, Ref. 113

The results of combining maximum speed with gravity will be fascinating: we will find that empty space can bend and move, that the universe has a finite age and that objects can be in permanent free fall. We will discover that even though empty space can be bent, it is much stiffer than steel. Despite the strangeness of these and other consequences, they have all been confirmed by all experiments performed so far.

## Rest and free fall

The opposite of motion in daily life is a body at rest, such as a child sleeping or a rock defying the waves. A body is at rest whenever it is not disturbed by other bodies. In the everyday description of the world, rest is the *absence of velocity*. With Galilean and special relativity, rest became *inertial motion*, since no inertially moving observer can distinguish its own motion from rest: nothing disturbs him. Both the rock in the waves and the rapid protons crossing the galaxy as cosmic rays are at rest. With the inclusion of gravity, we are led to an even more general definition of rest.

▷ *Every observer and every body in free fall can rightly claim to be at rest.*

Challenge 173 e

If any body moving inertially is to be considered at rest, then any body in free fall must also be. Nobody knows this better than Joseph Kittinger, the man who in August 1960

* ‘Venture to be wise.’ Horace is Quintus Horatius Flaccus, (65–8 BCE), the great Roman poet.
** The details of this statement are far from simple. They are discussed on page 151 and page 185.

Ref. 114 stepped out of a balloon capsule at the record height of 31.3 km. At that altitude, the air is so thin that during the first minute of his free fall he felt completely at rest, as if he were floating. Although an experienced parachutist, he was so surprised that he had to turn upwards in order to convince himself that he was indeed moving away from his balloon! Despite his lack of any sensation of movement, he was falling at up to 274 m/s or 988 km/h with respect to the Earth's surface. He only started feeling something when he encountered the first substantial layers of air. That was when his free fall started to be disturbed. Later, after four and a half minutes of fall, his special parachute opened; and nine minutes later he landed in New Mexico.

Kittinger and all other observers in free fall, such as the cosmonauts circling the Earth or the passengers in parabolic aeroplane flights,* make the same observation: it is impossible to distinguish anything happening in free fall from what would happen at rest. This impossibility is called the *principle of equivalence*; it is one of the starting points of general relativity. It leads to the most precise – and final – definition of rest that we will encounter: *rest is free fall*. Rest is lack of disturbance; so is free fall.

The set of all free-falling observers at a point in space-time generalizes the special-relativistic notion of the set of the inertial observers at a point. This means that we must describe motion in such a way that not only all inertial but also all freely falling observers can talk to each other. In addition, a full description of motion must be able to describe gravitation and the motion it produces, and it must be able to describe motion for any observer imaginable. General relativity realizes this aim.

As a first step, we put the result in simple words: true *motion is the opposite of free fall*. This statement immediately rises a number of questions: Most trees or mountains are not Challenge 174 s in free fall, thus they are not at rest. What motion are they undergoing? And if free fall is rest, what is weight? And what then is gravity anyway? Let us start with the last question.

## What clocks tell us about gravity

Page 116 Above, we described gravity as the shadow of the maximum force. But there is a second way to describe it, more closely related to everyday life. As William Unruh likes to Ref. 115 explain, the constancy of the speed of light for all observers implies a simple conclusion:

▷ *Gravity is the uneven running of clocks at different places.***

Challenge 176 e Of course, this seemingly absurd definition needs to be checked. The definition does not talk about a single situation seen by different observers, as we often did in special relativity. The definition depends of the fact that neighbouring, identical clocks, fixed against each other, run differently in the presence of a gravitational field when watched by the *same* observer; moreover, this difference is directly related to what we usually call gravity. There are two ways to check this connection: by experiment and by reasoning. Let us start with the latter method, as it is cheaper, faster and more fun.

* Nowadays it is possible to book such flights in specialized travel agents.

** Gravity is also the uneven length of metre bars at different places, as we will see below. Both effects are needed to describe it completely; but for daily life on Earth, the clock effect is sufficient, since it is much larger than the length effect, which can usually be neglected. Can you see why? Challenge 175 s

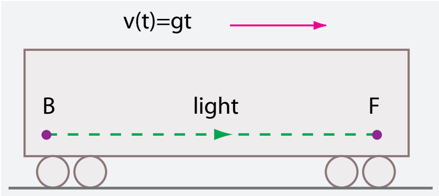



FIGURE 58 Inside an accelerating train or bus

An observer feels no difference between gravity and constant acceleration. We can thus study constant acceleration and use a way of reasoning we have encountered already in the chapter on special relativity. We assume light is emitted at the back end of a train of length $\Delta h$ that is accelerating forward with acceleration $g$, as shown in Figure 58. The light arrives at the front of the train after a time $t = \Delta h/c$. However, during this time the accelerating train has picked up some additional velocity, namely $\Delta v = gt = g\Delta h/c$. As a result, because of the Doppler effect we encountered in our discussion of special relativity, the frequency $f$ of the light arriving at the front has changed. Using the expression of the Doppler effect, we thus get*

Page 52
Challenge 177 e

$$\frac{\Delta f}{f} = \frac{g\Delta h}{c^2} \ . \qquad (130)$$

The sign of the frequency change depends on whether the light motion and the train acceleration are in the same or in opposite directions. For actual trains or buses, the frequency change is quite small; nevertheless, it is measurable. Acceleration induces frequency changes in light. Let us compare this first effect of acceleration with the effects of gravity.

Challenge 179 s

To measure time and space, we use light. What happens to light when gravity is involved? The simplest experiment is to let light fall or rise. In order to deduce what must happen, we add a few details. Imagine a conveyor belt carrying masses around two wheels, a low and a high one, as shown in Figure 59. The descending, grey masses are slightly larger. Whenever such a larger mass is near the bottom, some mechanism – not shown in the figure – converts the mass surplus to light, in accordance with the equation $E = mc^2$, and sends the light up towards the top.** At the top, one of the lighter, white masses passing by absorbs the light and, because of its added weight, turns the conveyor belt until it reaches the bottom. Then the process repeats.***

Ref. 116

As the grey masses on the descending side are always heavier, the belt would turn for ever and this system could continuously *generate* energy. However, since energy conservation is at the basis of our definition of time, as we saw in the beginning of our walk, the whole process must be impossible. We have to conclude that the light changes its energy when climbing. The only possibility is that the light arrives at the top with a frequency

Vol. I, page 226

* The expression $v = gt$ is valid only for non-relativistic speeds; nevertheless, the conclusion of this section is not affected by this approximation.

Challenge 178 e

** As in special relativity, here and in the rest of our mountain ascent, the term 'mass' always refers to rest mass.

*** Can this process be performed with 100% efficiency?

Challenge 180 s

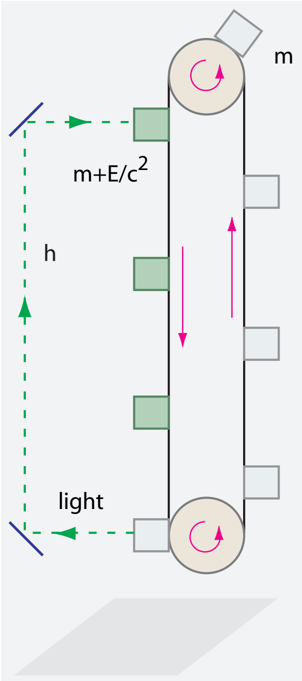



FIGURE 59 The necessity of blue- and red-shift of light: why trees are greener at the bottom

*different* from the one at which it is emitted from the bottom.*

In short, it turns out that rising light is gravitationally red-shifted. Similarly, the light descending from the top of a tree down to an observer is *blue-shifted*; this gives a darker colour to the top in comparison with the bottom of the tree. The combination of light speed invariance and gravitation thus imply that trees have different shades of green along their height.** How big is the effect? The result deduced from the drawing is again the one of formula (130). That is what we would, as light moving in an accelerating train and light moving in gravity are equivalent situations, as you might want to check yourself. The formula gives a relative change of frequency of only $1.1 \cdot 10^{-16}$ /m near the surface of the Earth. For trees, this so-called *gravitational red-shift* or *gravitational Doppler effect* is far too small to be observable, at least using normal light.

Challenge 182 e
Challenge 183 s

In 1911, Einstein proposed an experiment to check the change of frequency with height by measuring the red-shift of light emitted by the Sun, using the famous Fraunhofer lines as colour markers. The results of the first experiments, by Schwarzschild and others, were unclear or even negative, due to a number of other effects that induce colour changes at high temperatures. But in 1920 and 1921, Leonhard Grebe and Albert Bachem, and independently Alfred Perot, confirmed the gravitational red-shift with careful experiments. In later years, technological advances made the measurements much easier, until it was even possible to measure the effect on Earth. In 1960, in a classic experiment using the

Ref. 117
Page 154
Ref. 118

* The precise relation between energy and frequency of light is described and explained in our discussion on quantum theory, on page 39. But we know already from classical electrodynamics that the energy of light depends on its intensity and on its frequency.

** How does this argument change if you include the illumination by the Sun?

Challenge 181 ny

Mössbauer effect, Pound and Rebka confirmed the gravitational red-shift in their university tower using $\gamma$ radiation.

Ref. 119

But our two thought experiments tell us much more. Let us use the same argument as in the case of special relativity: a colour change implies that clocks run differently at different heights, just as they run differently in the front and in the back of a train. The time difference $\Delta\tau$ is predicted to depend on the height difference $\Delta h$ and the acceleration of gravity $g$ according to

$$\frac{\Delta\tau}{\tau} = \frac{\Delta f}{f} = \frac{g\Delta h}{c^2} \; . \tag{131}$$

Therefore, in gravity, *time is height-dependent*. That was exactly what we claimed above. In fact, *height makes old*. Can you confirm this conclusion?

Challenge 184 e

In 1972, by flying four precise clocks in an aeroplane while keeping an identical one on the ground, Hafele and Keating found that clocks indeed run differently at different altitudes according to expression (131). Subsequently, in 1976, the team of Vessot shot a precision clock based on a maser – a precise microwave generator and oscillator – upwards on a missile. The team compared the maser inside the missile with an identical maser on the ground and again confirmed the above expression. In 1977, Briatore and Leschiutta showed that a clock in Torino indeed ticks more slowly than one on the top of the Monte Rosa. They confirmed the prediction that on Earth, for every 100 m of height gained, people age more rapidly by about 1 ns per day. This effect has been confirmed for all systems for which experiments have been performed, such as several planets, the Sun and numerous other stars.

Ref. 120
Ref. 121
Ref. 122
Challenge 185 e

Do these experiments show that time changes or are they simply due to clocks that function badly? Take some time and try to settle this question. We will give one argument only: gravity does change the colour of light, and thus really does change time. Clock precision is not an issue here.

Challenge 186 e

In summary, gravity is indeed the uneven running of clocks at different heights. Note that an observer at the lower position and another observer at the higher position *agree* on the result: both find that the upper clock goes faster. In other words, when gravity is present, space-time is *not* described by the Minkowski geometry of special relativity, but by some more general geometry. To put it mathematically, whenever gravity is present, the 4-distance $\mathrm{d}s^2$ between events is different from the expression without gravity:

$$\mathrm{d}s^2 \neq c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \; . \tag{132}$$

We will give the correct expression shortly.

Is this view of gravity as height-dependent time really reasonable? No. It turns out that it is not yet strange enough! Since the speed of light is the same for all observers, we can say more. If time changes with height, length must also do so! More precisely, if clocks run differently at different heights, the length of metre bars must also change with height. Can you confirm this for the case of horizontal bars at different heights?

Challenge 187 s

If length changes with height, the circumference of a circle around the Earth *cannot* be given by $2\pi r$. An analogous discrepancy is also found by an ant measuring the radius and circumference of a circle traced on the surface of a basketball. Indeed, gravity implies that

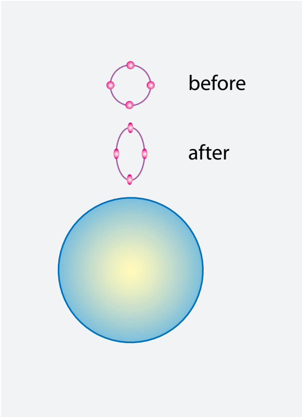



FIGURE 60 Tidal effects: the only effect bodies feel when falling

humans are in a situation analogous to that of ants on a basketball, the only difference being that the circumstances are translated from two to three dimensions. We conclude that wherever gravity plays a role, space is *curved*.

## What tides tell us about gravity

During his free fall, Kittinger was able to specify an inertial frame for himself. Indeed, he felt completely at rest. Does this mean that it is impossible to distinguish acceleration from gravitation? No: distinction *is* possible. We only have to compare *two* (or more) falling observers.

Kittinger could not have found a frame which is also inertial for a colleague falling on the opposite side of the Earth. Such a common frame does not exist. In general, it is impossible to find a *single* inertial reference frame describing different observers freely falling near a mass. In fact, it is impossible to find a common inertial frame even for *nearby* observers in a gravitational field. Two nearby observers observe that during their fall, their relative distance changes. (Why?) The same happens to orbiting observers.

Challenge 188 e

Challenge 189 s

In a closed room in orbit around the Earth, a person or a mass at the centre of the room would not feel any force, and in particular no gravity. But if several particles are located in the room, they will behave differently depending on their exact positions in the room. Only if two particles were on exactly the same orbit would they keep the same relative position. If one particle is in a lower or higher orbit than the other, they will depart from each other over time. Even more interestingly, if a particle in orbit is displaced sideways, it will oscillate around the central position. (Can you confirm this?)

Challenge 190 ny

Gravitation leads to changes of relative distance. These changes evince another effect, shown in Figure 60: an extended body in free fall is slightly *squeezed*. This effect also tells us that it is an essential feature of gravity that free fall is *different* from point to point. That rings a bell. The squeezing of a body is the same effect as that which causes the tides. Indeed, the bulging oceans can be seen as the squeezed Earth in its fall towards the Moon. Using this result of universal gravity we can now affirm: the essence of gravity is the observation of tidal effects.

Page 160

Ref. 123

In other words, gravity is simple only *locally*. Only locally does it look like acceleration.

Only locally does a falling observer like Kittinger feel at rest. In fact, only a point-like observer does so! As soon as we take spatial extension into account, we find tidal effects.

▷ *Gravity is the presence of tidal effects.*

The absence of tidal effects implies the absence of gravity. Tidal effects are the everyday consequence of height-dependent time. Isn't this a beautiful conclusion from the invariance of the speed of light?

In principle, Kittinger could have *felt* gravitation during his free fall, even with his eyes closed, had he paid attention to himself. Had he measured the distance change between his two hands, he would have found a tiny decrease which could have told him that he was falling. This tiny decrease would have forced Kittinger to a strange conclusion. Two inertially moving hands should move along two parallel lines, always keeping the same distance. Since the distance changes, he must conclude that in the space around him lines starting out in parallel do not remain so. Kittinger would have concluded that the space around him was similar to the surface of the Earth, where two lines starting out north, parallel to each other, also change distance, until they meet at the North Pole. In other words, Kittinger would have concluded that he was in a *curved* space.

By studying the change in distance between his hands, Kittinger could even have concluded that the curvature of space changes with height. Physical space differs from a sphere, which has constant curvature. Physical space is more involved. The effect is extremely small, and cannot be felt by human senses. Kittinger had no chance to detect anything. However, the conclusion remains valid. Space-time is *not* described by Minkowski geometry when gravity is present. Tidal effects imply space-time curvature.

▷ *Gravity is the curvature of space-time.*

This is the main and final lesson that follows from the invariance of the speed of light.

## Bent space and mattresses

“Wenn ein Käfer über die Oberfläche einer Kugel krabbelt, merkt er wahrscheinlich nicht, daß der Weg, den er zurücklegt, gekrümmt ist. Ich dagegen hatte das Glück, es zu merken.* ”

Albert Einstein's answer to his son Eduard's question about the reason for his fame

On the 7th of November 1919, Albert Einstein became world-famous. On that day, an article in the *Times* newspaper in London announced the results of a double expedition to South America under the heading 'Revolution in science / new theory of the universe / Newtonian ideas overthrown'. The expedition had shown unequivocally – though not for the first time – that the theory of universal gravity, essentially given by $a = GM/r^2$, was wrong, and that instead space was *curved*. A worldwide mania started. Einstein was presented as the greatest of all geniuses. 'Space warped' was the most common headline.

* 'When an insect walks over the surface of a sphere it probably does not notice that the path it walks is curved. I, on the other hand, had the luck to notice it.'

Einstein's papers on general relativity were reprinted in full in popular magazines. People could read the field equations of general relativity, in tensor form and with Greek indices, in *Time* magazine. Nothing like this has happened to any other physicist before or since. What was the reason for this excitement?

The expedition to the southern hemisphere had performed an experiment proposed by Einstein himself. Apart from seeking to verify the change of time with height, Einstein had also thought about a number of experiments to detect the curvature of space. In the one that eventually made him famous, Einstein proposed to take a picture of the stars near the Sun, as is possible during a solar eclipse, and compare it with a picture of the same stars at night, when the Sun is far away. Einstein predicted a change in position of $1.75'$ (1.75 seconds of arc) for star images at the border of the Sun, a value *twice* as large as that predicted by universal gravity. The prediction was confirmed for the first time in 1919, and thus universal gravity was ruled out.

Ref. 124 Vol. I, page 164 Ref. 125

Does this result *imply* that space is curved? Not by itself. In fact, other explanations could be given for the result of the eclipse experiment, such as a potential differing from the inverse square form. However, the eclipse results are not the only data. We already know about the change of time with height. Experiments show that two observers at different heights measure the same value for the speed of light $c$ near themselves. But these experiments also show that if an observer measures the speed of light at the position of the *other* observer, he gets a value differing from $c$, since his clock runs differently. There is only one possible solution to this dilemma: metre bars, like clocks, also change with height, and in such a way as to yield the same speed of light everywhere.

If the speed of light is constant but clocks and metre bars change with height, the conclusion must be that *space is curved near masses*. Many physicists in the twentieth century checked whether metre bars really behave differently in places where gravity is present. And indeed, curvature has been detected around several planets, around all the hundreds of stars where it could be measured, and around dozens of galaxies. Many indirect effects of curvature around masses, to be described in detail below, have also been observed. All results confirm the curvature of space and space-time around masses, and in addition confirm the curvature values predicted by general relativity. In other words, metre bars near masses do indeed change their size from place to place, and even from orientation to orientation. Figure 61 gives an impression of the situation.

Challenge 191 e

But beware: the right-hand figure, although found in many textbooks, can be misleading. It can easily be mistaken for a reproduction of a *potential* around a body. Indeed, it is impossible to draw a graph showing curvature and potential separately. (Why?) We will see that for small curvatures, it is even possible to explain the change in metre bar length using a potential only. Thus the figure does not really cheat, at least in the case of weak gravity. But for large and changing values of gravity, a potential cannot be defined, and thus there is indeed no way to avoid using curved space to describe gravity. In summary, if we imagine space as a sort of generalized mattress in which masses produce deformations, we have a reasonable model of space-time. As masses move, the deformation follows them.

Ref. 126 Challenge 192 s

The acceleration of a test particle only depends on the curvature of the mattress. It does not depend on the mass of the test particle. So the mattress model explains why all bodies fall in the same way. (In the old days, this was also called the equality of the inertial and gravitational mass.)

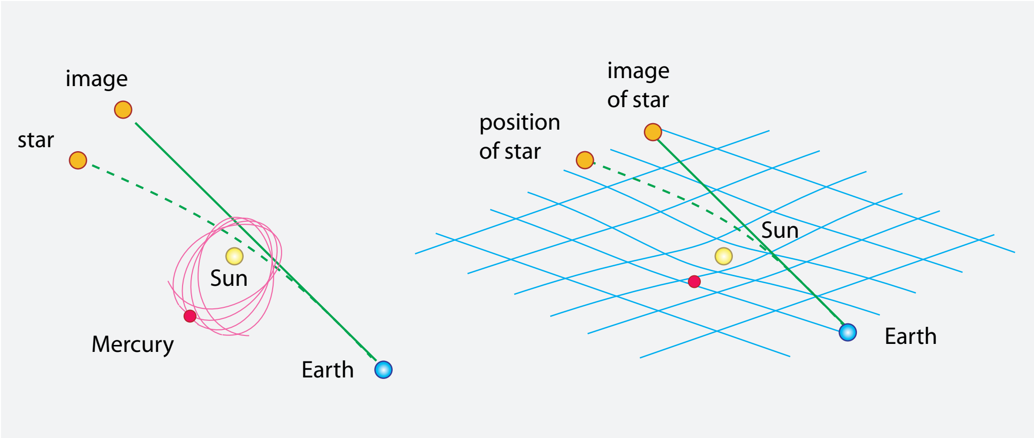



FIGURE 61 The mattress model of space: the path of a light beam and of a satellite near a spherical mass

Space thus behaves like a frictionless mattress that pervades everything. We live inside the mattress, but we do not feel it in everyday life. Massive objects pull the foam of the mattress towards them, thus deforming the shape of the mattress. More force, more energy or more mass imply a larger deformation. (Does the mattress remind you of the aether? Do not worry: physics eliminated the concept of aether because it is indistinguishable from vacuum.) Page 112

If gravity means curved space, then any accelerated observer, such as a man in a departing car, must also observe that space is curved. However, in everyday life we do not notice any such effect, because for accelerations and sizes of everyday life the curvature values are too small to be noticed. Could you devise a sensitive experiment to check the prediction? Challenge 193 ny

## Curved space-time

Figure 61 shows the curvature of space only, but in fact space-time is curved. We will shortly find out how to describe both the shape of space and the shape of space-time, and how to measure their curvature.

Let us have a first attempt to describe nature with the idea of curved space-time. In the case of Figure 61, the best description of events is with the use of the time $t$ shown by a clock located at spatial infinity; that avoids problems with the uneven running of clocks at different distances from the central mass. For the radial coordinate $r$, the most practical choice to avoid problems with the curvature of space is to use the circumference of a circle around the central body, divided by $2\pi$. The curved shape of space-time is best described by the behaviour of the space-time distance $\mathrm{d}s$, or by the wristwatch time $\mathrm{d}\tau = \mathrm{d}s/c$, between two neighbouring points with coordinates $(t, r)$ and $(t + \mathrm{d}t, r + \mathrm{d}r)$. As we saw above, gravity means that in spherical coordinates we have Page 41 Page 126

$$\mathrm{d}\tau^2 = \frac{\mathrm{d}s^2}{c^2} \neq \mathrm{d}t^2 - \mathrm{d}r^2/c^2 - r^2\mathrm{d}\varphi^2/c^2 \ . \tag{133}$$

The inequality expresses the fact that space-time is *curved*. Indeed, the experiments on time change with height confirm that the space-time interval around a spherical mass is given by

$$\mathrm{d}\tau^2 = \frac{\mathrm{d}s^2}{c^2} = \left(1 - \frac{2GM}{rc^2}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{c^2 - \frac{2GM}{r}} - \frac{r^2}{c^2}\mathrm{d}\varphi^2 \,. \tag{134}$$

This expression is called the *Schwarzschild metric* after one of its discoverers.* The metric (134) describes the curved shape of space-time around a spherical non-rotating mass. It is well approximated by the Earth or the Sun. (Why can their rotation be neglected?) Expression (134) also shows that gravity's strength around a body of mass $M$ and radius $R$ is measured by a dimensionless number $h$ defined as

Challenge 194 s

$$h = \frac{2G}{c^2}\frac{M}{R} \,. \tag{135}$$

This ratio expresses the gravitational strain with which lengths and the vacuum are deformed from the flat situation of special relativity, and thus also determines how much clocks slow down when gravity is present. (The ratio also reveals how far one is from any possible horizon.) On the surface of the Earth the ratio $h$ has the small value of $1.4 \cdot 10^{-9}$; on the surface of the Sun is has the somewhat larger value of $4.2 \cdot 10^{-6}$. The precision of modern clocks allows detecting such small effects quite easily. The various consequences and uses of the deformation of space-time will be discussed shortly.

We note that if a mass is highly concentrated, in particular when its radius becomes *equal* to its so-called *Schwarzschild radius*

$$R_\mathrm{S} = \frac{2GM}{c^2} \,, \tag{136}$$

the Schwarzschild metric behaves strangely: at that location, *time disappears* (note that $t$ is time at infinity). At the Schwarzschild radius, the wristwatch time (as shown by a clock at infinity) stops – and a horizon appears. What happens precisely will be explored below. This situation is not common: the Schwarzschild radius for a mass like the Earth is 8.8 mm, and for the Sun is 3.0 km; you might want to check that the object size for every system in everyday life is larger than its Schwarzschild radius. Bodies which reach this limit are called *black holes*; we will study them in detail shortly. In fact, general relativity states that *no* system in nature is smaller than its Schwarzschild size, in other words that the ratio $h$ defined by expression (135) is never above unity.

Page 238
Challenge 195 e
Ref. 128

In summary, the results mentioned so far make it clear that *mass generates curvature*. The mass–energy equivalence we know from special relativity then tells us that as a consequence, space should also be curved by the presence of any type of energy–momentum. Every type of energy curves space-time. For example, light should also curve space-time.

* Karl Schwarzschild (1873–1916), important German astronomer; he was one of the first people to understand general relativity. He published his formula in December 1915, only a few months after Einstein had published his field equations. He died prematurely, at the age of 42, much to Einstein's distress. We will deduce the form of the metric later on, directly from the field equations of general relativity. The other discoverer of the metric, unknown to Einstein, was the Dutch physicist J. Droste.

Ref. 127

However, even the highest-energy beams we can create correspond to extremely small masses, and thus to unmeasurably small curvatures. Even heat curves space-time; but in most systems, heat is only about a fraction of $10^{-12}$ of total mass; its curvature effect is thus unmeasurable and negligible. Nevertheless it is still possible to show experimentally that energy curves space. In almost all atoms a sizeable fraction of the mass is due to the electrostatic energy among the positively charged protons. In 1968 Kreuzer confirmed that energy curves space with a clever experiment using a floating mass. Ref. 129

It is straightforward to deduce that the temporal equivalent of spatial curvature is the uneven running of clock. Taking the two curvatures together, we conclude that when gravity is present, *space-time* is curved. Challenge 196 e

Let us sum up our chain of thoughts. Energy is equivalent to mass; mass produces gravity; gravity is equivalent to acceleration; acceleration is position-dependent time. Since light speed is constant, we deduce that *energy–momentum tells space-time to curve.* This statement is the first half of general relativity.

We will soon find out how to measure curvature, how to calculate it from energy–momentum and what is found when measurement and calculation are compared. We will also find out that different observers measure different curvature values. The set of transformations relating one viewpoint to another in general relativity, the *diffeomorphism symmetry*, will tell us how to relate the measurements of different observers.

Since matter moves, we can say even more. Not only is space-time curved near masses, it also bends back when a mass has passed by. In other words, general relativity states that space, as well as space-time, is *elastic*. However, it is rather stiff: quite a lot stiffer than steel. To curve a piece of space by 1 % requires an energy density enormously larger than to curve a simple train rail by 1 %. This and other interesting consequences of the elasticity of space-time will occupy us for a while. Ref. 130 Challenge 197 ny

## The speed of light and the gravitational constant

> “Si morior, moror.*”
> Antiquity

We continue on the way towards precision in our understanding of gravitation. All our theoretical and empirical knowledge about gravity can be summed up in just two general statements. The first principle states:

▷ *The speed $v$ of a physical system is bounded above:*

$$v \leqslant c \tag{137}$$

*for all observers, where $c$ is the speed of light.*

The theory following from this first principle, *special* relativity, is extended to *general* relativity by adding a second principle, characterizing gravitation. There are several equivalent ways to state this principle. Here is one.

* ‘If I rest, I die.’ This is the motto of the bird of paradise.

▷ *For all observers, the force $F$ on a system is limited by*

$$F \leqslant \frac{c^4}{4G} , \quad (138)$$

*where $G$ is the universal constant of gravitation.*

In short, there is a maximum force in nature. Gravitation leads to attraction of masses.
Challenge 198 e However, this force of attraction is limited. An equivalent statement is:

▷ *For all observers, the size $L$ of a system of mass $M$ is limited by*

$$\frac{L}{M} \geqslant \frac{4G}{c^2} . \quad (139)$$

In other words, a massive system cannot be more concentrated than a non-rotating black hole of the same mass. Another way to express the principle of gravitation is the following:

▷ *For all systems, the emitted power $P$ is limited by*

$$P \leqslant \frac{c^5}{4G} . \quad (140)$$

In short, there is a maximum power in nature.

The three limits given above are all equivalent to each other; and no exception is known or indeed possible. The limits include universal gravity in the non-relativistic case. They tell us *what* gravity is, namely curvature, and *how* exactly it behaves. The limits allow us to determine the curvature in all situations, at all space-time events. As we
Page 101 have seen above, the speed limit together with any one of the last three principles imply all of general relativity.*

For example, can you show that the formula describing gravitational red-shift com-
Challenge 199 ny plies with the general limit (139) on length-to-mass ratios?

We note that any formula that contains the speed of light $c$ is based on special relativity, and if it contains the constant of gravitation $G$, it relates to universal gravity. If a formula contains *both* $c$ and $G$, it is a statement of general relativity. The present chapter frequently underlines this connection.

Our mountain ascent so far has taught us that a precise description of motion requires the specification of all allowed viewpoints, their characteristics, their differences, and the transformations between them. From now on, *all* viewpoints are allowed, without exception: anybody must be able to talk to anybody else. It makes no difference whether an observer feels gravity, is in free fall, is accelerated or is in inertial motion. Furthermore, people who exchange left and right, people who exchange up and down or people who say that the Sun turns around the Earth must be able to talk to each other and to us. This

* This didactic approach is unconventional. It is possible that is has been pioneered by the present author.
Ref. 96 The British physicist Gary Gibbons also developed similar ideas independently.

gives a much larger set of viewpoint transformations than in the case of special relativity; it makes general relativity both difficult and fascinating. And since all viewpoints are allowed, the resulting description of motion is *complete*.*

## Why does a stone thrown into the air fall back to Earth? – Geodesics

> A genius is somebody who makes all possible mistakes in the shortest possible time.
>
> Anonymous

In our discussion of special relativity, we saw that inertial or free-floating motion is the motion which connecting two events that requires the *longest* proper time. In the absence of gravity, the motion fulfilling this requirement is *straight* (rectilinear) motion. On the other hand, we are also used to thinking of light rays as being straight. Indeed, we are all accustomed to check the straightness of an edge by looking along it. Whenever we draw the axes of a physical coordinate system, we imagine either drawing paths of light rays or drawing the motion of freely moving bodies.

Page 77

Page 54

In the absence of gravity, object paths and light paths coincide. However, in the presence of gravity, objects do not move along light paths, as every thrown stone shows. Light does not define spatial straightness any more. In the presence of gravity, both light and matter paths are bent, though by *different* amounts. But the original statement remains valid: even when gravity is present, bodies follow paths of longest possible proper time. For matter, such paths are called *timelike geodesics*. For light, such paths are called *lightlike* or *null geodesics*.

We note that in space-time, geodesics are the curves with *maximal* length. This is in contrast with the case of pure space, such as the surface of a sphere, where geodesics are the curves of *minimal* length.

In simple words, *stones fall because they follow geodesics*. Let us perform a few checks of this statement. Since stones move by maximizing proper time for inertial observers, they also must do so for freely falling observers, like Kittinger. In fact, they must do so for all observers. The equivalence of falling paths and geodesics is at least coherent.

If falling is seen as a consequence of the Earth's surface approaching – as we will argue below – we can deduce directly that falling implies a proper time that is as long as possible. Free fall indeed is motion along geodesics.

Page 143

Challenge 200 ny

We saw above that gravitation follows from the existence of a maximum force. The result can be visualized in another way. If the gravitational attraction between a central body and a satellite were *stronger* than it is, black holes would be smaller than they are; in that case the maximum force limit and the maximum speed could be exceeded by getting close to such a black hole. If, on the other hand, gravitation were *weaker* than it is, there would be observers for which the two bodies would not interact, thus for which they would not form a physical system. In summary, a maximum force of $c^4/4G$ implies universal gravity. There is no difference between stating that all bodies attract through gravitation and stating that there is a maximum force with the value $c^4/4G$. But at the same time, the maximum force principle implies that objects move on geodesics. Can you show this?

Challenge 201 ny

* Or it would be, were it not for a small deviation called quantum theory.

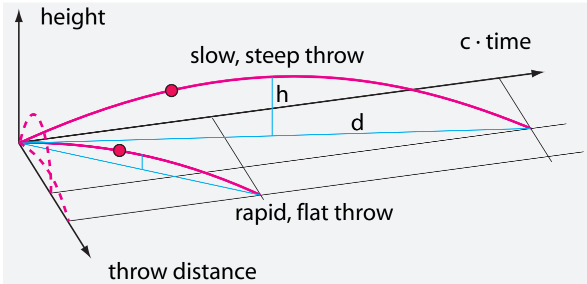



FIGURE 62 All paths of flying stones, independently of their speed and angle, have the same curvature in space-time (photograph © Marco Fulle)



Let us turn to an experimental check. If falling is a consequence of curvature, then the paths of *all* stones thrown or falling near the Earth must have the *same* curvature in space-time. Take a stone thrown horizontally, a stone thrown vertically, a stone thrown rapidly, or a stone thrown slowly: it takes only two lines of argument to show that *in space-time* all their paths are approximated to high precision by circle segments, as shown in Figure 62. All paths have the *same* curvature radius $r$, given by

Challenge 202 ny

$$r = \frac{c^2}{g} \approx 9.2 \cdot 10^{15}\ \mathrm{m}\ . \tag{141}$$

The large value of the radius, corresponding to a low curvature, explains why we do not notice it in everyday life. The parabolic shape typical of the path of a stone in everyday life is just the projection of the more fundamental path in 4-dimensional space-time into 3-dimensional space. The important point is that the value of the curvature does *not* depend on the details of the throw. In fact, this simple result could have suggested the ideas of general relativity to people a full century before Einstein; what was missing was the recognition of the importance of the speed of light as limit speed. In any case, this simple calculation confirms that falling and curvature are connected. As expected, and as mentioned already above, the curvature diminishes at larger heights, until it vanishes at infinite distance from the Earth. Now, given that the curvature of all paths for free fall is the same, and given that all such paths are paths of least action, it is straightforward that they are also geodesics.

If we describe fall as a consequence of the curvature of space-time, we must show that the description with geodesics reproduces all its features. In particular, we must be able to explain that stones thrown with small speed fall back, and stones thrown with high speed escape. Can you deduce this from space curvature?

Challenge 203 ny

In summary, the motion of any particle falling freely 'in a gravitational field' is described by the same variational principle as the motion of a free particle in special relativity: the path maximizes the proper time $\int \mathrm{d}\tau$. We rephrase this by saying that any particle in free fall from point $A$ to point $B$ minimizes the action $S$ given by

$$S = -mc^2 \int_A^B \mathrm{d}\tau\ . \tag{142}$$

That is all we need to know about the free fall of objects. As a consequence, any *deviation from free fall keeps you young*. The larger the deviation, the younger you stay.

Page 260 Ref. 131 As we will see below, the minimum action description of free fall has been tested extremely precisely, and no difference from experiment has ever been observed. We will also find out that for free fall, the predictions of general relativity and of universal gravity differ substantially both for particles near the speed of light and for central bodies of high density. So far, all experiments have shown that whenever the two predictions differ, general relativity is right, and universal gravity and other alternative descriptions are wrong.

All bodies fall along geodesics. This tells us something important. The fall of bodies does not depend on their mass. The geodesics are like 'rails' in space-time that tell bodies how to fall. In other words, space-time can indeed be imagined as a single, giant, deformed entity. Space-time is not 'nothing'; it is an entity of our thinking. The shape of this entity tells objects how to move. Space-time is thus indeed like an intangible mattress; this deformed mattress guides falling objects along its networks of geodesics.

Moreover, *bound* energy falls in the same way as mass, as is proven by comparing the fall of objects made of different materials. They have different percentages of bound energy. (Why?) Challenge 204 s For example, on the Moon, where there is no air, cosmonauts dropped steel balls and feathers and found that they fell together, alongside each other. The independence on material composition has been checked and confirmed over and over again. Ref. 132

## Can light fall?

How does radiation fall? Light, like any radiation, is energy without rest mass. It moves like a stream of extremely fast and light objects. Therefore deviations from universal gravity become most apparent for light. How does light fall? Light cannot change speed. Page 123 When light falls vertically, it only changes colour, as we have seen above. But light can also change direction. Long before the ideas of relativity became current, in 1801, the Ref. 133 Prussian astronomer Johann Soldner understood that universal gravity implies that light is *deflected* when passing near a mass. He also calculated how the deflection angle de- Page 164 pends on the mass of the body and the distance of passage. However, nobody in the nineteenth century was able to check the result experimentally.

Obviously, light has energy, and energy has weight; the deflection of light by itself is thus *not* a proof of the curvature of space. General relativity also predicts a deflection angle for light passing masses, but of *twice* the classical Soldner value, because the curvature of space around large masses adds to the effect of universal gravity. The deflection of light thus only confirms the curvature of space if the *value* agrees with the one predicted by general relativity. This is the case: observations do coincide with predictions. More Page 159 details will be given shortly.

Simply said, mass is not necessary to feel gravity; energy is sufficient. This result of the mass–energy equivalence must become second nature when studying general relativity. In particular, light is not light-weight, but heavy. Can you argue that the curvature of Challenge 205 ny light near the Earth must be the same as that of stones, given by expression (141)?

In summary, all experiments show that not only mass, but also energy falls along geodesics, whatever its type (bound or free), and whatever the interaction (be it electromagnetic or nuclear). Moreover, the motion of radiation confirms that space-time is curved.

Since experiments show that all particles fall in the same way, independently of their mass, charge or any other property, we can conclude that the system of all possible trajectories forms an independent structure. This structure is what we call *space-time*.

We thus find that *space-time tells matter, energy and radiation how to fall*. This statement is the second half of general relativity. It complements the first half, which states that energy tells space-time how to curve. To complete the description of macroscopic motion, we only need to add numbers to these statements, so that they become testable. As usual, we can proceed in two ways: we can deduce the equations of motion directly, or we can first deduce the Lagrangian and then deduce the equations of motion from it. But before we do that, let's have some fun.

## Curiosities and fun challenges about gravitation

> Wenn Sie die Antwort nicht gar zu ernst nehmen und sie nur als eine Art Spaß ansehen, so kann ich Ihnen das so erklären: Früher hat man geglaubt, wenn alle Dinge aus der Welt verschwinden, so bleiben noch Raum und Zeit übrig. Nach der Relativitätstheorie verschwinden aber auch Zeit und Raum mit den Dingen.*
>
> Albert Einstein in 1921 in New York

Take a plastic bottle and make some holes in it near the bottom. Fill the bottle with water, closing the holes with your fingers. If you let the bottle go, no water will leave the bottle during the fall. Can you explain how this experiment confirms the equivalence of rest and free fall? Challenge 206 s

* *

On his seventy-sixth birthday, Einstein received a birthday present specially made for him, shown in Figure 63. A rather deep cup is mounted on the top of a broom stick. The cup contains a weak piece of elastic rubber attached to its bottom, to which a ball is attached at the other end. In the starting position, the ball hangs outside the cup. The rubber is too weak to pull the ball into the cup against gravity. What is the most elegant way to get the ball into the cup? Challenge 207 s

* *

Gravity has the same properties in the whole universe – except in the US patent office. In 2005, it awarded a patent, Nr. 6 960 975, for an antigravity device that works by distorting space-time in such a way that gravity is 'compensated' (see patft.uspto.gov). Do you know a simpler device? Challenge 208 s

* *

* 'If you do not take the answer too seriously and regard it only for amusement, I can explain it to you in the following way: in the past it was thought that if all things were to disappear from the world, space and time would remain. But following relativity theory, space and time would disappear together with the things.'

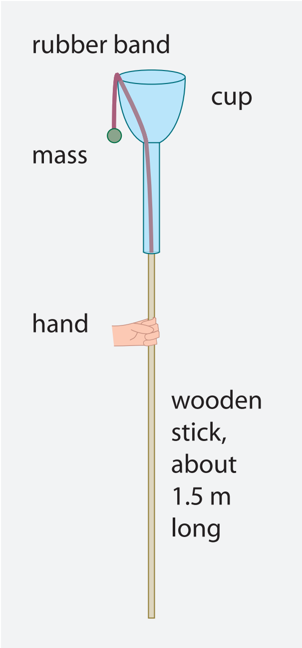



FIGURE 63 A puzzle: what is the simplest way to get the ball attached to the rubber band into the cup?

Challenge 209 e The radius of curvature of space-time at the Earth's surface is $9.2 \cdot 10^{15}$ m. Can you confirm this value?

* *

Challenge 210 s A piece of wood floats on water. Does it stick out more or less in a lift accelerating upwards?

* *

Page 52 We saw in special relativity that if two twins are identically accelerated in the same direction, with one twin some distance ahead of the other, then the twin ahead ages more than the twin behind. Does this happen in a gravitational field as well? And what happens when the field varies with height, as on Earth? Challenge 211 s

* *

A maximum force and a maximum power also imply a maximum flow of mass. Can you show that no mass flow can exceed $1.1 \cdot 10^{35}$ kg/s? Challenge 212 s

* *

The experiments of Figure 58 and 59 differ in one point: one happens in flat space, the other in curved space. One seems to be related energy conservation, the other not. Do these differences invalidate the equivalence of the observations? Challenge 213 s

* *

Challenge 214 s How can cosmonauts weigh themselves to check whether they are eating enough?

* *

Is a cosmonaut in orbit really floating freely? No. It turns out that space stations and satellites are accelerated by several small effects. The important ones are the pressure of the light from the Sun, the friction of the thin air, and the effects of solar wind. (Micrometeorites can usually be neglected.) These three effects all lead to accelerations of the order of $10^{-6}$ m/s$^2$ to $10^{-8}$ m/s$^2$, depending on the height of the orbit. Can you estimate how long it would take an apple floating in space to hit the wall of a space station, starting from the middle? By the way, what is the magnitude of the tidal accelerations in this situation?

Challenge 215 s

* *

Vol. I, page 92

There is no negative mass in nature, as discussed in the beginning of our walk (even antimatter has *positive* mass). This means that gravitation cannot be shielded, in contrast to electromagnetic interactions. Since gravitation cannot be shielded, there is no way to make a perfectly isolated system. But such systems form the basis of thermodynamics! We will study the fascinating implications of this later on: for example, we will discover an *upper limit* for the entropy of physical systems.

Page 99

* *

Can curved space be used to travel faster than light? Imagine a space-time in which two points could be connected either by a path leading through a flat portion, or by a second path leading through a partially curved portion. Could that curved portion be used to travel between the points faster than through the flat one? Mathematically, this is possible; however, such a curved space would need to have a *negative* energy density. Such a situation is incompatible with the definition of energy and with the non-existence of negative mass. The statement that this does not happen in nature is also called the *weak energy condition*. Is it implied by the limit on length-to-mass ratios?

Ref. 134
Challenge 216 ny

* *

The statement of a length-to-mass limit $L/M \geqslant 4G/c^2$ invites experiments to try to overcome it. Can you explain what happens when an observer moves so rapidly past a mass that the body's length contraction reaches the limit?

Challenge 217 ny

* *

There is an important mathematical property of three-dimensional space $\mathbb{R}^3$ that singles it from all other dimensions. A closed (one-dimensional) curve can form knots *only* in $\mathbb{R}^3$: in any higher dimension it can always be unknotted. (The existence of knots also explains why three is the smallest dimension that allows chaotic particle motion.) However, general relativity does not say *why* space-time has three plus one dimensions. It is simply based on the fact. This deep and difficult question will be settled only in the last part of our mountain ascent.

* *

Henri Poincaré, who died in 1912, shortly before the general theory of relativity was finished, thought for a while that curved space was not a necessity, but only a possibility. He imagined that one could continue using Euclidean space provided light was permitted

Challenge 218 s to follow curved paths. Can you explain why such a theory is impossible?

* *

Can two hydrogen atoms circle each other, in their mutual gravitational field? What would the size of this 'molecule' be? Challenge 219 s

* *

Challenge 220 s Can two light pulses circle each other, in their mutual gravitational field?

* *

The various motions of the Earth mentioned in the section on Galilean physics, such as its rotation around its axis or around the Sun, lead to various types of time in physics and astronomy. The time defined by the best atomic clocks is called *terrestrial dynamical time*. By inserting leap seconds every now and then to compensate for the bad definition of the second (an Earth rotation does not take 86 400, but 86 400.002 seconds) and, in minor ways, for the slowing of Earth's rotation, one gets the *universal time coordinate* or UTC. Then there is the time derived from this one by taking into account all leap seconds. One then has the – different – time which would be shown by a non-rotating clock in the centre of the Earth. Finally, there is *barycentric dynamical time*, which is the time that would be shown by a clock in the centre of mass of the solar system. Only using this latter time can satellites be reliably steered through the solar system. In summary, relativity says goodbye to Greenwich Mean Time, as does British law, in one of the rare cases were the law follows science. (Only the BBC continues to use it.)

Page 130
Page 350
Ref. 135

* *

Space agencies thus *have* to use general relativity if they want to get artificial satellites to Mars, Venus, or comets. Without its use, orbits would not be calculated correctly, and satellites would miss their targets and usually even the whole planet. In fact, space agencies play on the safe side: they use a generalization of general relativity, namely the so-called *parametrized post-Newtonian formalism*, which includes a continuous check on whether general relativity is correct. Within measurement errors, no deviation has been found so far.*

* To give an idea of what this means, the *unparametrized* post-Newtonian formalism, based on general relativity, writes the equation of motion of a body of mass $m$ near a large mass $M$ as a deviation from the inverse square expression for the acceleration $a$:

$$a = \frac{GM}{r^2} + f_2 \frac{GM}{r^2}\frac{v^2}{c^2} + f_4 \frac{GM}{r^2}\frac{v^4}{c^4} + f_5 \frac{Gm}{r^2}\frac{v^5}{c^5} + \cdots \qquad (143)$$

Here the numerical factors $f_n$ are calculated from general relativity and are of order one. The first two odd terms are missing because of the (approximate) reversibility of general relativistic motion: gravity wave emission, which is irreversible, accounts for the small term $f_5$; note that it contains the small mass $m$ instead of the large mass $M$. All factors $f_n$ up to $f_7$ have now been calculated. However, in the solar system, only the term $f_2$ has ever been detected. This situation might change with future high-precision satellite experiments. Higher-order effects, up to $f_5$, have been measured in the binary pulsars, as discussed below. Page 156

In a *parametrized* post-Newtonian formalism, all factors $f_n$, including the uneven ones, are fitted through the data coming in; so far all these fits agree with the values predicted by general relativity.

* *

Ref. 136 General relativity is also used by space agencies around the world to calculate the exact positions of satellites and to tune radios to the frequency of radio emitters on them. In addition, general relativity is essential for the so-called *global positioning system*, or GPS. This modern navigation tool* consists of 24 satellites equipped with clocks that fly around the world. Why does the system need general relativity to operate? Since all the satellites, as well as any person on the surface of the Earth, travel in circles, we have $\mathrm{d}r = 0$, and we can rewrite the Schwarzschild metric (134) as

$$\left(\frac{\mathrm{d}\tau}{\mathrm{d}t}\right)^2 = 1 - \frac{2GM}{rc^2} - \frac{r^2}{c^2}\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2 = 1 - \frac{2GM}{rc^2} - \frac{v^2}{c^2} . \tag{144}$$

Challenge 221 e For the relation between satellite time and Earth time we then get

$$\left(\frac{\mathrm{d}t_{\mathrm{sat}}}{\mathrm{d}t_{\mathrm{Earth}}}\right)^2 = \frac{1 - \frac{2GM}{r_{\mathrm{sat}}c^2} - \frac{v_{\mathrm{sat}}^2}{c^2}}{1 - \frac{2GM}{r_{\mathrm{Earth}}c^2} - \frac{v_{\mathrm{Earth}}^2}{c^2}} . \tag{145}$$

Challenge 222 s Can you deduce how many microseconds a satellite clock gains every day, given that the GPS satellites orbit the Earth once every twelve hours? Since only three microseconds would give a position error of one kilometre after a single day, the clocks in the satellites Ref. 137 must be adjusted to run slow by the calculated amount. The necessary adjustments are monitored, and so far have confirmed general relativity every single day, within experimental errors, since the system began operation.

* *

General relativity is the base of the sport of *geocaching*, the world-wide treasure hunt with the help of GPS receivers. See the www.terracaching.com and www.geocaching.com websites for more details.

* *

Ref. 138 The gravitational constant $G$ does not seem to change with time. The latest experiments limit its rate of change to less than 1 part in $10^{12}$ per year. Can you imagine how this can Challenge 223 d be checked?

* *

Challenge 224 s Could our experience that we live in only three spatial dimensions be due to a limitation of our senses? How?

* *

Challenge 225 ny Can you estimate the effect of the tides on the colour of the light emitted by an atom?

* *

The strongest possible gravitational field is that of a small black hole. The strongest grav-

* For more information, see the www.gpsworld.com website.

Ref. 139 itational field ever observed is somewhat less though. In 1998, Zhang and Lamb used the X-ray data from a double star system to determine that space-time near the 10 km sized neutron star is curved by up to 30 % of the maximum possible value. What is the corresponding gravitational acceleration, assuming that the neutron star has the same mass as the Sun? Challenge 226 ny

* *

Ref. 140 Challenge 227 e Light deflection changes the angular size $\delta$ of a mass $M$ with radius $r$ when observed at distance $d$. The effect leads to the pretty expression

$$\delta = \arcsin\left(\frac{r\sqrt{1-R_S/d}}{d\sqrt{1-R_S/r}}\right) \quad \text{where} \quad R_S = \frac{2GM}{c^2} . \tag{146}$$

Challenge 228 ny Page 249 What percentage of the surface of the Sun can an observer at infinity see? We will examine this issue in more detail shortly.

## What is weight?

There is no way for a *single* (and point-like) observer to distinguish the effects of gravity from those of acceleration. This property of nature allows making a strange statement: things *fall* because the surface of the Earth accelerates towards them. Therefore, the *weight* of an object results from the surface of the Earth accelerating upwards and pushing against the object. That is the principle of equivalence applied to everyday life. For the same reason, objects in free fall have no weight.

Let us check the numbers. Obviously, an accelerating surface of the Earth produces a weight for each body resting on it. This weight is proportional to the inertial mass. In other words, the inertial mass of a body is identical to the gravitational mass. This is indeed observed in experiments, and to the highest precision achievable. Ref. 141 Roland von Eötvös* performed many such high-precision experiments throughout his life, without finding any discrepancy. In these experiments, he used the fact that the inertial mass determines centrifugal effects and the gravitational mass determines free fall. (Can you imagine how he tested the equality?) Challenge 229 ny Recent experiments showed that the two masses agree to one part in $10^{-12}$. Ref. 141

However, the mass equality is not a surprise. Remembering the definition of mass ratio as negative inverse acceleration ratio, Page 89 independently of the origin of the acceleration, we are reminded that mass measurements cannot be used to distinguish between inertial and gravitational mass. As we have seen, the two masses are equal by definition in Galilean physics, Page 165 and the whole discussion is a red herring. Weight is an intrinsic effect of mass.

The equality of acceleration and gravity allows us to imagine the following. Imagine stepping into a lift in order to move down a few stories. You push the button. The lift is pushed upwards by the accelerating surface of the Earth somewhat less than is the build-

* Roland von Eötvös (b. 1848 Budapest, d. 1919 Budapest), Hungarian physicist. He performed many high-precision gravity experiments; among other discoveries, he discovered the effect named after him. The university of Budapest is named after him.

ing; the building overtakes the lift, which therefore remains behind. Moreover, because of the weaker push, at the beginning everybody inside the lift feels a bit lighter. When the contact with the building is restored, the lift is accelerated to catch up with the accelerating surface of the Earth. Therefore we all feel as if we were in a strongly accelerating car, pushed in the direction opposite to the acceleration: for a short while, we feel heavier, until the lift arrives at its destination.

## Why do apples fall?

> “Vires acquirit eundo.”
>
> Vergilius*

An accelerating car will soon catch up with an object thrown forward from it. For the same reason, the surface of the Earth soon catches up with a stone thrown upwards, because it is continually accelerating upwards. If you enjoy this way of seeing things, imagine an apple falling from a tree. At the moment when it detaches, it stops being accelerated upwards by the branch. The apple can now enjoy the calmness of real rest. Because of our limited human perception, we call this state of rest free fall. Unfortunately, the accelerating surface of the Earth approaches mercilessly and, depending on the time for which the apple stayed at rest, the Earth hits it with a greater or lesser velocity, leading to more or less severe shape deformation.

Falling apples also teach us not to be disturbed any more by the statement that gravity is the uneven running of clocks with height. In fact, this statement is *equivalent* to saying that the surface of the Earth is accelerating upwards, as the discussion above shows.

Can this reasoning be continued indefinitely? We can go on for quite a while. It is fun to show how the Earth can be of constant radius even though its surface is accelerating upwards everywhere. We can thus play with the equivalence of acceleration and gravity. However, this equivalence is only useful in situations involving only one accelerating body. The equivalence between acceleration and gravity ends as soon as *two* falling objects are studied. Any study of several bodies inevitably leads to the conclusion that gravity is not acceleration; *gravity is curved space-time*.

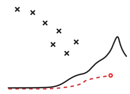
Challenge 230 ny

Many aspects of gravity and curvature can be understood with no or only a little mathematics. The next section will highlight some of the differences between universal gravity and general relativity, showing that only the latter description agrees with experiment. After that, a few concepts relating to the measurement of curvature are introduced and applied to the motion of objects and space-time. If the reasoning gets too involved for a first reading, skip ahead. In any case, the section on the stars, cosmology and black holes again uses little mathematics.

## A summary: the implications of the invariant speed of light on gravitation

The invariance of the speed of light implies that space and space-time are *curved* in all regions where gravity plays a role.

* ‘Going it acquires strength.’ Publius Vergilius Maro (b. 70 bce Andes, d. 19 bce Brundisium), from the *Aeneid* 4, 175.



Chapter 6

# OPEN ORBITS, BENT LIGHT AND WOBBLING VACUUM

Ref. 142

> “Einstein explained his theory to me every day, and on my arrival I was fully convinced that he understood it.”
>
> Chaim Weizmann, first president of Israel.

Before we tackle the details of general relativity, we explore the differences between the motion of objects in general relativity and in universal gravity, because the two descriptions lead to measurable differences. Since the invariance of the speed of light implies that space is curved near masses, we first of all have to check whether this curvature is indeed observed. After that, we explore how curvature is measured and how curvature measurements help to described motion with precision.

## Weak fields

Gravity is strong near horizons. This happens when the mass $M$ and the distance scale $R$ obey

$$\frac{2GM}{Rc^2} \approx 1 . \tag{147}$$

Therefore, gravity is strong mainly in three situations: near black holes, near the horizon of the universe, and at extremely high particle energies. The first two cases are explored below, while the last will be explored in the final part of our mountain ascent. In contrast, in most regions of the universe, including our own planet, there are *no* nearby horizons; in these cases, gravity is a *weak* effect.

Despite the violence of avalanches or of falling asteroids, in everyday life gravity is much weaker than the maximum force. On the Earth the ratio just mentioned is only about $10^{-9}$. In all cases of everyday life, gravitation can still be approximated by a field, i.e., with a potential in flat space-time, despite what was said above. These weak field situations are interesting because they are simple to understand; they mainly require for their explanation the different running of clocks at different heights. Weak field situations allow us to mention space-time curvature only in passing, and allow us to continue to think of gravity as a source of acceleration. Nevertheless, the change of time with height already induces many new and interesting effects that do not occur in universal gravity. To explore them, the only thing we need is a consistent relativistic treatment.

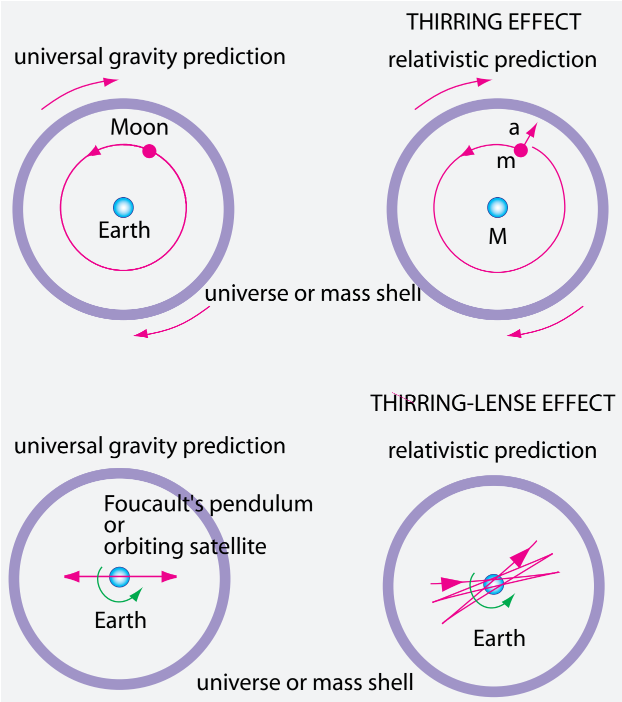



FIGURE 64 The Thirring and the Thirring–Lense effects

## The Thirring effects

In 1918, the Austrian physicist Hans Thirring published two 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 64 illustrates these predictions. Ref. 143

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 231 ny

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 along with it. Similarly, the Earth drags some vacuum with it, and thus turns the plane

* 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 65 The LAGEOS satellites: metal spheres with a diameter of 60 cm, a mass of 407 kg, and covered with 426 retroreflectors (NASA)

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 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. They followed the motion of two special artificial satellites – shown in Figure 65 – 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 low-budget experiment to beat by many years the efforts of much larger but much more sluggish groups.* The results confirm the predictions of general relativity with an error of about 25 %. Ref. 144

Frame dragging effects have also been measured in binary star systems. This is possible if one of the stars is a pulsar, because such stars 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 way these stars move and confirm that such subtle effects as frame dragging do take place. Ref. 145

## Gravitomagnetism**

Frame-dragging and the Thirring effect can be seen as special cases of gravitomagnetism. (We will show the connection below.) This approach to gravity, 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 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? Ref. 146

* One is the so-called Gravity Probe B satellite experiment, which should significantly increase the measurement precision; the satellite was put in orbit in 2005, after 30 years of planning. Despite several broken systems, in 2009 the experiment confirmed the existence of frame dragging.
** This section can be skipped at first reading.

Vol. III, page 46

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. 147, Ref. 148

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 mass *currents*.

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

Ref. 147

Vol. III, page 43

$$m\ddot{\boldsymbol{x}} = q\boldsymbol{E} + q\dot{\boldsymbol{x}} \times \boldsymbol{B} \ . \tag{148}$$

In other words, the change of *speed* 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

$$m\ddot{\boldsymbol{x}} = m\boldsymbol{G} + m\dot{\boldsymbol{x}} \times \boldsymbol{H} \ . \tag{149}$$

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

$$\boldsymbol{G} = \nabla\varphi = \nabla\frac{GM}{r} = -\frac{GM\boldsymbol{x}}{r^3} \ . \tag{150}$$

As usual, the quantity $\varphi$ 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 $\boldsymbol{\nabla G} = -4\pi G\rho$, where $\rho$ is the mass density. A *static* field $\boldsymbol{G}$ has no vortices; it obeys $\boldsymbol{\nabla} \times \boldsymbol{G} = 0$.

It is not hard to show that if gravito*electric* fields exist, 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 66. An observer at

Ref. 149

Vol. III, page 46

* The approximation requires low velocities, weak fields, and localized and stationary mass–energy distributions.

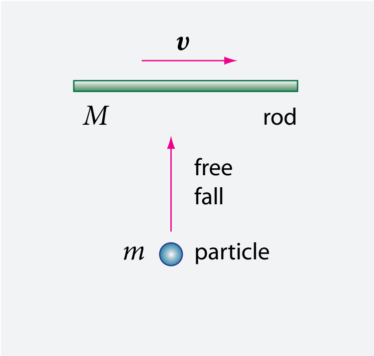



FIGURE 66 The reality of gravitomagnetism

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

$$m\boldsymbol{a} = m\boldsymbol{v} \times \boldsymbol{H} \tag{151}$$

Challenge 232 ny where, *almost* as in electrodynamics, the static *gravitomagnetic field* $\boldsymbol{H}$ obeys

$$\boldsymbol{H} = 16\pi N\rho\boldsymbol{v} \tag{152}$$

where $\rho$ 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 233 ny When the situation in Figure 66 is evaluated, we find that the proportionality constant $N$ is given by

$$N = \frac{G}{c^2} = 7.4 \cdot 10^{-28}\,\mathrm{m/kg}\;, \tag{153}$$

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 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. Challenge 234 s

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 66 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 can thus be seen 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 64. For an angular momentum $\boldsymbol{J}$ the gravitomagnetic field $\boldsymbol{H}$ is a dipole field; it is given by

$$\boldsymbol{H} = \nabla \times \left( -2 \frac{\boldsymbol{J} \times \boldsymbol{x}}{r^3} \right) \tag{154}$$

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

$$\boldsymbol{T} = \frac{\mathrm{d}\boldsymbol{S}}{\mathrm{d}t} = \tfrac{1}{2}\, \boldsymbol{S} \times \boldsymbol{H} \,. \tag{155}$$

The torque leads to the *precession of gyroscopes*. 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.

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 $\boldsymbol{J}$ yields a second effect. An orbiting mass will experience *precession of its orbital plane*. Seen from infinity one gets, for an orbit with semimajor axis $a$ and
Challenge 235 ny eccentricity $e$,

$$\dot{\boldsymbol{\Omega}} = -\frac{\boldsymbol{H}}{2} = -\frac{G}{c^2}\frac{\boldsymbol{J}}{|\boldsymbol{x}|^3} + \frac{G}{c^2}\frac{3(\boldsymbol{J}\boldsymbol{x})\boldsymbol{x}}{|\boldsymbol{x}|^5} = \frac{G}{c^2}\frac{2\boldsymbol{J}}{a^3(1-e^2)^{3/2}} \tag{156}$$

which is the prediction of Lense and Thirring.* The effect is extremely small, giving an angle change of only 8″ per orbit for a satellite near the surface of the Earth. Despite this smallness and a number of larger effects disturbing it, Ciufolini's team have managed to
Ref. 144 confirm the result.

As a third effect of gravitomagnetism, a rotating mass leads to the *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 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
Ref. 150 value observed so far.

The split into gravitoelectric and gravitomagnetic effects is thus a useful approximation to the description of gravity. It 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
Challenge 237 ny the Sun to us? To find the answer, thinking about the electromagnetic analogy can help.

Challenge 236 ny * A homogeneous spinning sphere has an angular momentum given by $J = \frac{2}{5} M \omega R^2$.

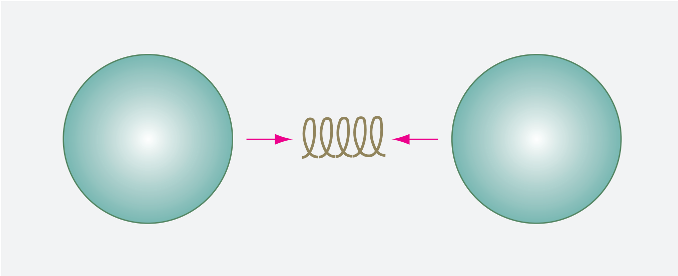


FIGURE 67 A Gedanken experiment showing the necessity of gravity waves

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 |

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.

Starting from the existence of a maximum energy speed, Jørgen Kalckar and Ole Ulf-
Ref. 151 beck have given a simple argument for the necessity of gravitational. 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 67. Such a spring

* 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'.

will make the masses bounce towards each other again and again. The central 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: it is taken away by space-time, in other words, it 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 general inertial observers, gravitodynamics can be deduced from universal gravity by boosting to other observers. One gets the four equations

Ref. 152
Challenge 238 ny

$$\nabla \boldsymbol{G} = -4\pi G\rho \quad , \quad \nabla \times \boldsymbol{G} = -\frac{1}{4}\frac{\partial \boldsymbol{H}}{\partial t}$$
$$\nabla \boldsymbol{H} = 0 \quad , \quad \nabla \times \boldsymbol{H} = -16\pi N\rho \boldsymbol{v} + 4\frac{N}{G}\frac{\partial \boldsymbol{G}}{\partial t} \; . \tag{157}$$

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 reflect the fact that the ratio between angular momentum and energy (the 'spin') of gravity waves is different from that of electromagnetic waves. Gravity waves have spin 2, whereas electromagnetic waves have spin 1. Note that 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. It is worth recalling that the spin of radiation is a *classical* property. The *spin of a wave* is the ratio $E/L\omega$, where $E$ is the energy, $L$ the angular momentum, and $\omega$ is the angular frequency. For electromagnetic waves, this ratio is equal to 1; for gravitational waves, it is 2.

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

$$m\ddot{\boldsymbol{x}} = m\boldsymbol{G} + m\dot{\boldsymbol{x}} \times \boldsymbol{H} \ . \tag{158}$$

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 same assumptions and arrive at the same conclusion.

Challenge 239 ny

Challenge 240 e

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

$$c = \sqrt{\frac{G}{N}} \ . \tag{159}$$

Vol. III, page 93

This result corresponds to the electromagnetic expression

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \ . \tag{160}$$

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 speed of gravitational waves has not yet been measured.

Ref. 153

Ref. 155

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 68,

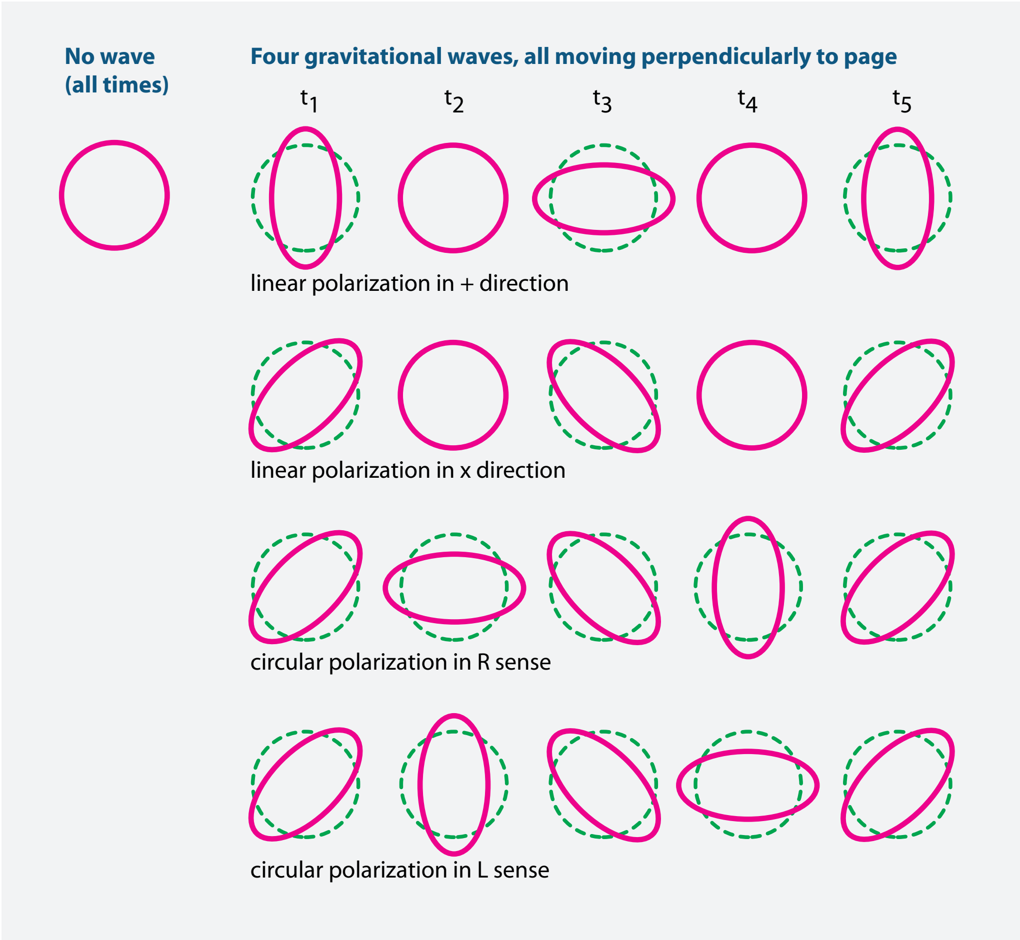



FIGURE 68 Effects on a circular or spherical body due to a plane gravitational wave moving in a direction perpendicular to the page

for both linear and circular polarization.* We note that the waves are invariant under

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

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 + h_{xx} & h_{xy} & 0 \\ 0 & h_{xy} & -1 + h_{xx} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \qquad (161)$$

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

$$h_{ab} = B_{ab} \sin(kz - \omega t + \varphi_{ab}) \qquad (162)$$

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

$$\frac{\omega}{k} = c \qquad (163)$$

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

a rotation by $\pi$ and that the two linear polarizations differ by an angle $\pi/4$; 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 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 $2\pi$ 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 68 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 Page 259 waves (instead of the spin 1 waves shown by ordinary latex mattresses).



## 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 Challenge 241 ny forbids mass dipoles from changing.

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 Challenge 242 ny to masses in mattresses. Can you point out the differences?

Einstein found that the amplitude $h$ of waves at a distance $r$ from a source is given, to Ref. 154 a good approximation, by the second derivative of the retarded quadrupole moment $Q$:

$$h_{ab} = \frac{2G}{c^4}\frac{1}{r}\mathrm{d}_{tt}Q^{\mathrm{ret}}_{ab} = \frac{2G}{c^4}\frac{1}{r}\mathrm{d}_{tt}Q_{ab}(t-r/c)\ . \qquad (165)$$

This expression shows that the amplitude of gravity waves *decreases only with* $1/r$, in contrast to naive expectations. However, this feature is the same as for electromagnetic waves. In addition, the small value of the prefactor, $1.6\cdot10^{-44}\,\mathrm{Wm/s}$, shows that truly

In another gauge, a plane wave can be written as

$$g = \begin{pmatrix} c^2(1+2\varphi) & A_1 & A_2 & A_3 \\ A_1 & -1+2\varphi & h_{xy} & 0 \\ A_2 & h_{xy} & -1+h_{xx} & 0 \\ A_3 & 0 & 0 & -1 \end{pmatrix} \qquad (164)$$

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

* 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 two spherical masses, 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.

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 $h = \delta l/l = 10^{-19}$. The production of detectable gravitational waves by humans is probably impossible.

Challenge 243 ny

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. 113

$$P = -\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{G}{45c^5}\,\dddot{Q}^{\mathrm{ret}}_{ab}\,\dddot{Q}^{\mathrm{ret}}_{ab} = \frac{32}{5}\frac{G}{c^5}\left(\frac{m_1 m_2}{m_1 + m_2}\right)^2 l^4 \omega^6 \qquad (166)$$

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

$$P = \frac{32}{5}\frac{G^4}{c^5}\frac{(m_1 m_2)^2 (m_1 + m_2)}{l^5} . \qquad (167)$$

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 that their effect cannot be detected at all.

Ref. 113

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 244 ny

As a result, the only observation of effects of gravitational waves to date is 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 69. 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 is the only case so far in which general relativity has been 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

Ref. 156 Ref. 157 Challenge 245 ny Page 140 Ref. 156

* Gravitomagnetism and gravitoelectricity allow one to define a gravitational Poynting vector. It is as easy to define and use as in the case of electrodynamics.

Vol. III, page 73 Ref. 149

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

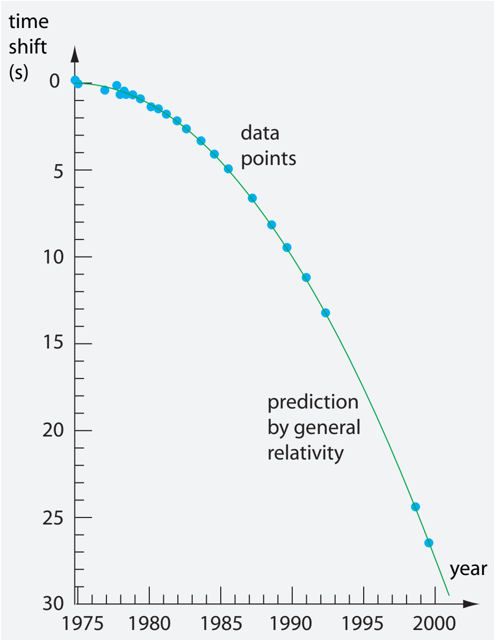



FIGURE 69 Comparison between measured time delay for the periastron of the binary pulsar PSR 1913+16 and the prediction due to energy loss by gravitational radiation

measurements were possible only because the two stars in this system are neutron stars with small size, 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. Ref. 113

The *direct* detection of gravitational waves is one of the aims of experimental general relativity. The race has been on since the 1990s. The basic idea is simple, as shown in Figure 70: 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. The merger of two black holes could be the source of such a strong gravitational wave, as also shown in Figure 70. Secondly, a system able to detect length changes of the order of $10^{-22}$ or better is needed – in other words, a lot of money. Any detection is guaranteed to make the news on television.* Essential for a successful detection are the techniques to eliminate noise in the detection signal. The worlds's best noise reduction experts are all working on gravitational wave detectors.

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 black hole collisions, are good candidates for detectable gravity wave sources.

Ref. 158
Challenge 246 ny
* The topic of gravity waves is full of interesting sidelines. For example, can gravity waves be used to power a rocket? Yes, maintain Bonnor and Piper. You might ponder the possibility yourself.

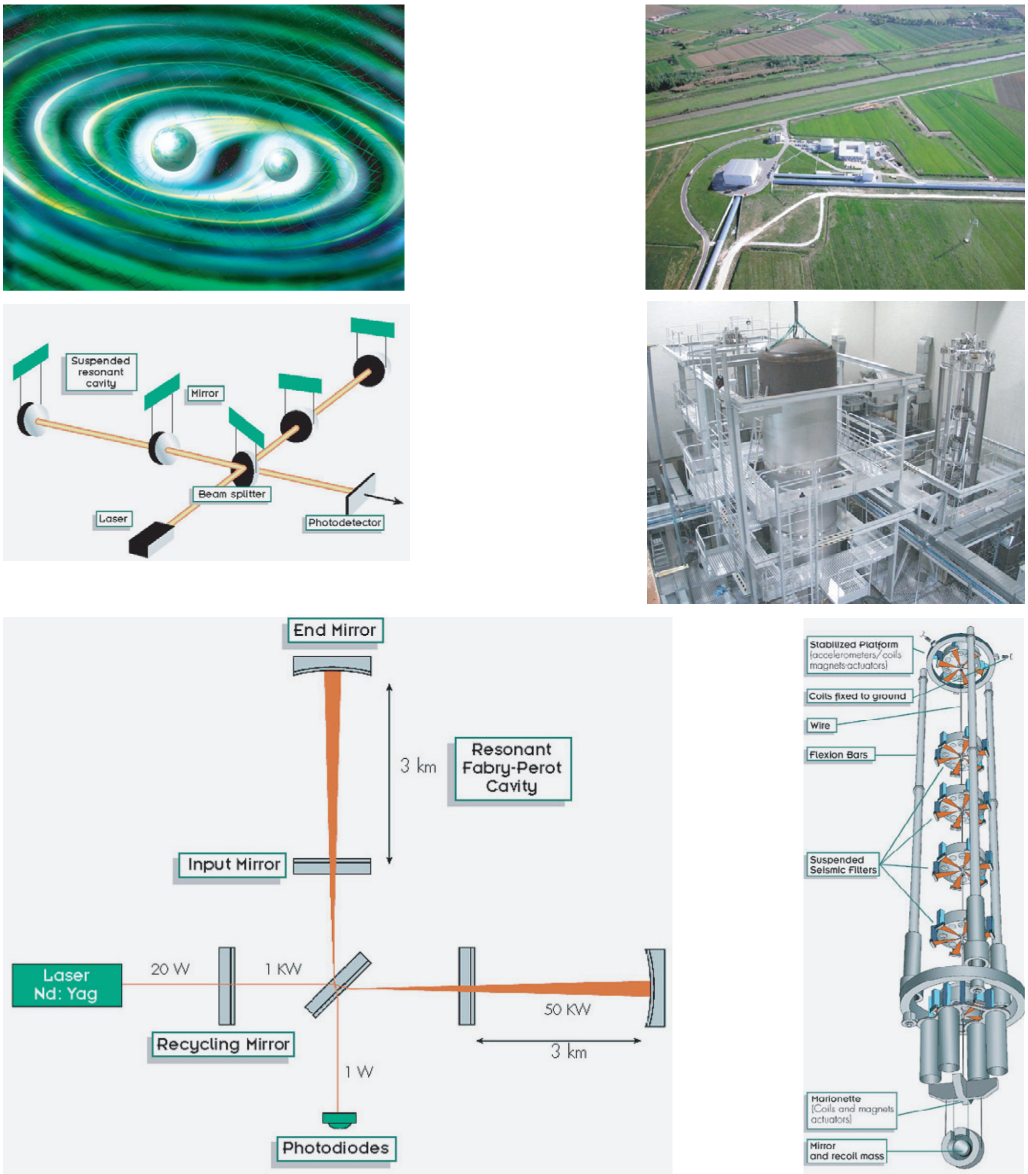



FIGURE 70 Detection of gravitational waves: an illustration of the merger of two black holes (top left) and the VIRGO detector in Cascina, Italy, with one of its huge mirror suspensions, the mirror suspension details, and two drawings of the laser interferometer (© INFN)

Gravitational waves are a fascinating area of study. They still provide many topics to explore. For example: can you find a method to measure their speed? No such measurement has been achieved, despite some serious attempts. Indeed, any measurement that does not simply use two spaced detectors of the type of Figure 70 would be a scientific sensation.

Challenge 247 r

Ref. 153

Another question on gravitational waves remains open at this point: If all change is due to motion of particles, as the Greeks maintained, how do gravity waves fit into the

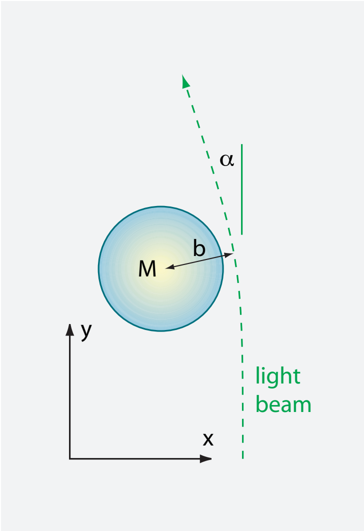



FIGURE 71 Calculating the bending of light by a mass

Challenge 248 ny picture? If gravitational waves were made of particles, space-time would also have to be. We have to wait until the beginning of the final part of our ascent to say more.

## Bending of light and radio waves

Gravity influences the motion of light. In particular, gravity bends light beams. The detection of the bending of light beams by the Sun made Einstein famous. Page 128

The bending of light by a mass is a pure gravitoelectric effect, and thus 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 infinity. The idea is to do all calculations to first order, as the value of the bending is very small. The angle of deflection $\alpha$, to first order, is simply Ref. 159

$$\alpha = \int_{-\infty}^{\infty} \frac{\partial v}{\partial x} \mathrm{d}y \ , \tag{168}$$

Challenge 249 e where $v$ is the speed of light measured by a distant observer. (Can you confirm this?) The next step is to use the Schwarzschild metric

$$\mathrm{d}\tau^2 = \left(1 - \frac{2GM}{rc^2}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{c^2 - \frac{2GM}{r}} - \frac{r^2}{c^2}\mathrm{d}\varphi^2 \tag{169}$$

Challenge 250 ny and transform it into $(x, y)$ coordinates to first order. This gives

$$\mathrm{d}\tau^2 = \left(1 - \frac{2GM}{rc^2}\right)\mathrm{d}t^2 - \left(1 + \frac{2GM}{rc^2}\right)\frac{1}{c^2}(\mathrm{d}x^2 + \mathrm{d}y^2) \qquad (170)$$

which again to first order leads to

$$\frac{\partial v}{\partial x} = \left(1 - \frac{2GM}{rc^2}\right)c\ . \qquad (171)$$

This 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 251 ny Inserting the last result into expression (168) and using a clever substitution, we get a deviation angle $\alpha$ given by

$$\alpha = \frac{4GM}{c^2}\frac{1}{b} \qquad (172)$$

where the distance $b$ is the so-called *impact parameter* of the approaching light beam. The resulting deviation angle $\alpha$ is *twice* the result we and Soldner found for universal gravity. Vol. I, page 164 For a beam just above the surface of the Sun, the result is the famous value of $1.75''$ which was confirmed by the measurement expedition of 1919. (How did they measure the deviation angle?) Challenge 252 ny This was the experiment that made Einstein famous, as it showed that universal gravity is wrong. 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 world war had just begun. Ref. 160 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. Page 164 Therefore Einstein became famous only because of the failure of the two expeditions that took place before he published his correct calculation.

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, over a dozen independent experiments have done so, using radio sources in the sky which lie on the path of the Sun. Ref. 136, Ref. 112 They have confirmed general relativity's prediction within a few per cent. Ref. 113

The bending of radiation has also been observed near Jupiter, near certain stars, near several galaxies and near galaxy clusters. Page 225 For the Earth, the angle 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 170 Of course, 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 253 ny

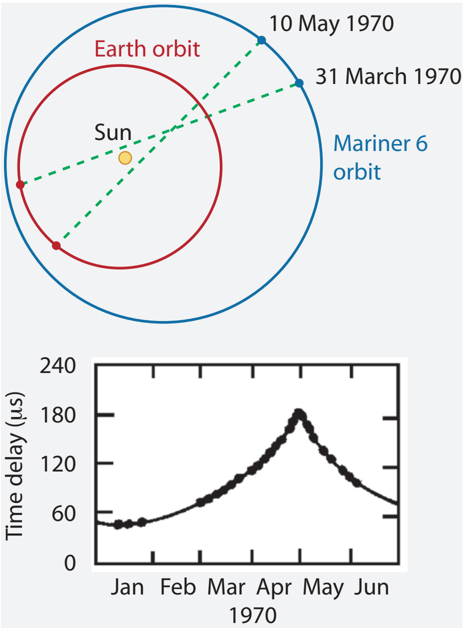



FIGURE 72 Time delay in radio signals – one of the experiments by Irwin Shapiro

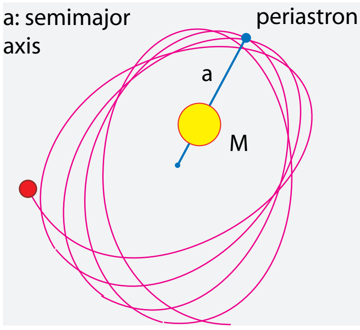



FIGURE 73 The orbit around a central body in general relativity

## 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!* 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 was to use an artificial satellite communicating with Earth. Ref. 161

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 72, have also confirmed the prediction within experimental errors, which nowadays are of the order of one part in a thousand. The delay has also 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. Ref. 162 Ref. 163

In short, relativistic gravitation is also confirmed by time delay measurements. 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 255 ny

* 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 254 e

## 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 73. 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 angular momentum as $e = E/mc^2$ and $j = J/m$. Next, the space curvature needs to be included. We use the Schwarzschild metric (169) mentioned above to deduce that the initial condition for the energy $e$, together with its conservation, leads to a relation between proper time $\tau$ and time $t$ at infinity:

Ref. 112, Ref. 113
Page 130
Challenge 256 e

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{e}{1 - 2GM/rc^2} , \tag{173}$$

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

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\tau} = \frac{j}{r^2} . \tag{174}$$

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

$$\left(\frac{\mathrm{d}r}{c\mathrm{d}\tau}\right)^2 + V^2(j,r) = e^2 \tag{175}$$

where the effective potential $V$ is given by

$$V^2(J,r) = \left(1 - \frac{2GM}{rc^2}\right)\left(1 + \frac{j^2}{r^2c^2}\right) . \tag{176}$$

Challenge 257 ny
Challenge 258 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

$$r_{\pm} = \frac{6GM/c^2}{1 \pm \sqrt{1 - 12(\frac{GM}{cj})^2}} \tag{177}$$

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 (175) and differentiating, the equation for $u(\varphi)$ becomes

$$u' + u = \frac{GM}{j^2} + \frac{3GM}{c^2}u^2 . \tag{178}$$

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

Challenge 259 e

$$u_0(\varphi) = \frac{GM}{j^2}(1 + \varepsilon \cos\varphi) , \tag{179}$$

i.e., ellipses, parabolas or hyperbolas. The type of conic section depends on the value of the parameter $\varepsilon$, 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 (178). Thus the solutions are not conic sections any more; however, as the correction is small, a good approximation is given by

Page 158

Challenge 260 e

$$u_1(\varphi) = \frac{GM}{j^2}\left(1 + \varepsilon \cos(\varphi - \frac{3G^2M^2}{j^2c^2}\varphi)\right) . \tag{180}$$

The hyperbolas and parabolas of universal gravity are thus slightly deformed. Instead of elliptical orbits we get the famous rosetta path shown in Figure 73. 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 261 e

$$\alpha \approx 6\pi \frac{GM}{a(1-\varepsilon^2)c^2} \tag{181}$$

for every orbit, where $a$ is the *semimajor axis*. For Mercury, the value is $43''$ 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 meantime, 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 (181) describes the motion within experimental errors.

Ref. 163

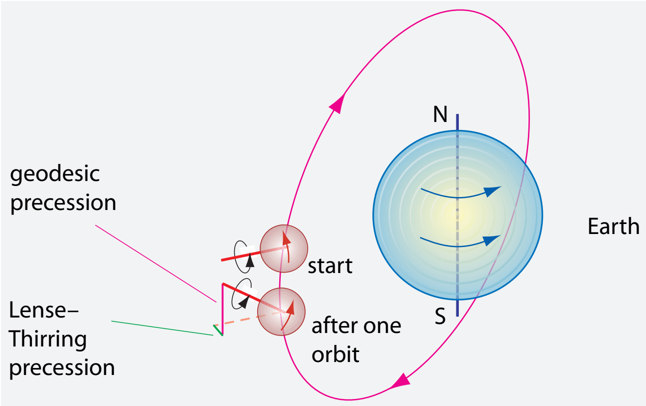



FIGURE 74 The geodesic effect



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 a good description of observations. Page 156

## The geodesic effect

Relativistic gravitation has a further effect on orbiting bodies. When a pointed body orbits a central mass $m$ at distance $r$, the *direction* of the tip will not be the same after a full orbit. This effect exists only in general relativity. The angle $\alpha$ describing the direction change after one orbit is given by

$$\alpha = 2\pi\left(1-\sqrt{1-\frac{3Gm}{rc^2}}\right) \approx \frac{3\pi Gm}{rc^2} \,. \qquad (182)$$

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 262 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 *precession* of the axis. Thus the effect is comparable to spin–orbit coupling in atomic theory. (The Thirring–Lense effect mentioned above is analogous to spin–spin coupling.)

The geodesic effect, or geodesic precession, was predicted by Willem de Sitter in 1916; in particular, 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. The effect was first detected 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 75, deposited by Lunokhod and Apollo on the Moon. Experiments to detect the geodesic effect in artificial satellites are also under way. Ref. 164 Ref. 165

At first sight, geodesic precession is similar to the Thomas precession found in special relativity. In both cases, a transport along a closed line results in the loss of the original Page 56

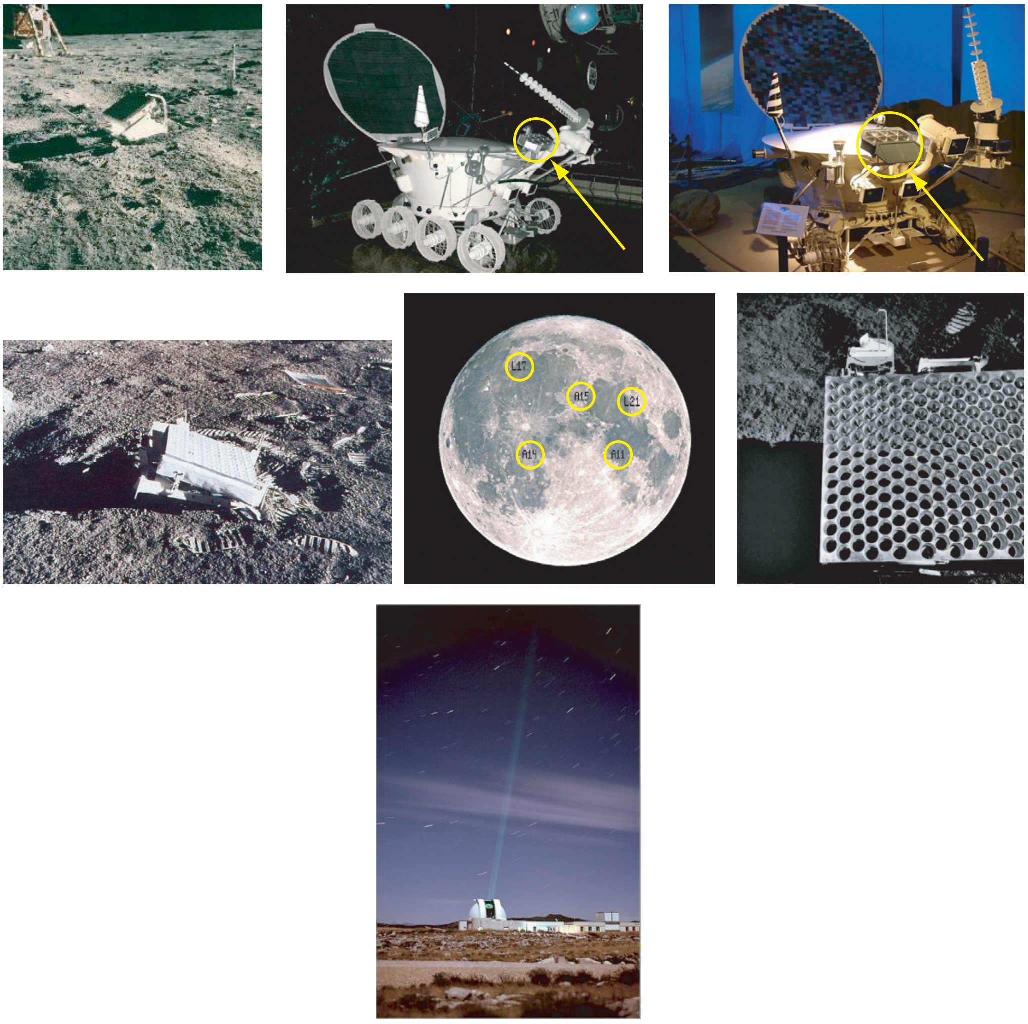



FIGURE 75 The lunar retroreflectors deposited by Apollo 11 (top left), Lunokhod (top centre and right), Apollo 14 (middle left) and Apollo 15 (middle right) together with their locations on the Moon and a telescope performing a distance measurement (© NASA, Observatoire de la Côte d'Azur)

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.

## Curiosities and fun challenges about weak fields

Challenge 263 ny

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

Page 25 be slowed down by emitting *Vavilov–Čerenkov radiation*, until they reach the lower speed. This is not observed.

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.)

* *

Challenge 264 ny Are narrow beams of gravitational waves, analogous to beams of light, possible?

* *

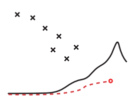
On effect that disturbs gravitational wave detectors are the tides. On the GEO600 detector in Hannover, tides change the distance of the mirrors, around 600 m, by 2 μm.

* *

Challenge 265 ny Would two parallel beams of gravitational waves attract each other?

## A summary on orbits and waves

In summary, the curvature of space and space-time implies that, in contrast to universal gravity, orbits are not closed, that orbiting objects change their orientation in space, that light is effectively slowed down near masses and therefore deflected by masses more than naively expected, and that empty vacuum can propagate gravitational waves. All experiments performed so far confirm these conclusions.



Chapter 7

# FROM CURVATURE TO MOTION

In the precise description of gravity, motion depends on space-time curvature. n order to quantify this idea, we first of all need to accurately describe curvature tself. To clarify the issue, we will start the discussion in two dimensions, and then move to three and four dimensions. Then 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 not 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 76. 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 266 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.

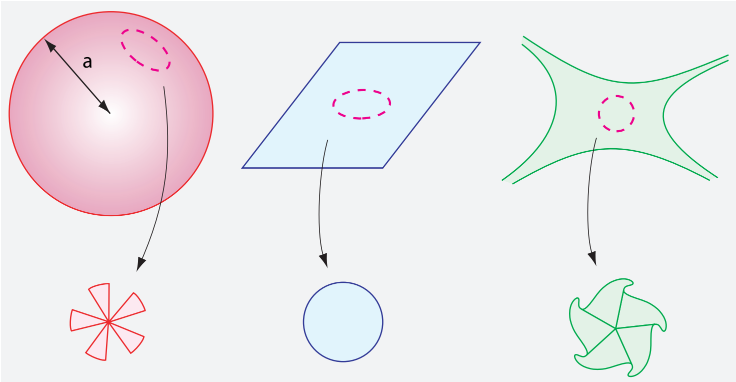



FIGURE 76 Positive, vanishing and negative curvature in two dimensions

Check your understanding: Can a one-dimensional space have intrinsic curvature? Is a torus intrinsically curved? Challenge 267 s

Alternatively, we can recognize intrinsic curvature also by checking whether two parallel lines 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 76 and Figure 78.

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 (two right angles) for positive curvature, and smaller than π for negative curvature.

Ref. 166 Let us see 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 Challenge 268 e

$$C = 2\pi r\left(1 - \frac{K}{6}r^2 + ...\right) \quad \text{and} \quad A = \pi r^2\left(1 - \frac{K}{12}r^2 + ...\right) \qquad (183)$$

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:

$$K = 6 \lim_{r\to 0}\left(1 - \frac{C}{2\pi r}\right)\frac{1}{r^2} \quad \text{or} \quad K = 12 \lim_{r\to 0}\left(1 - \frac{A}{\pi r^2}\right)\frac{1}{r^2} . \qquad (184)$$

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

* If the $n$-dimensional volume of a sphere is written as $V_n = C_n r^n$ and its $(n-1)$-dimensional 'surface' as

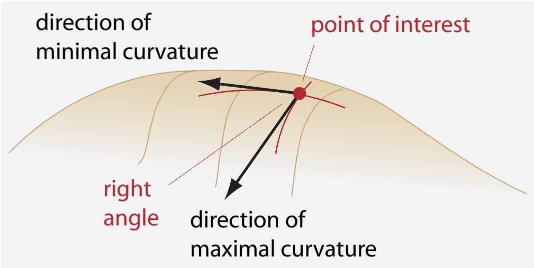



FIGURE 77 The maximum and minimum curvature of a surface are always at a right angle to each other.

From now on in this text, *curvature* will always mean *intrinsic* curvature. Like an ant on a surface, also observers in space can only detect intrinsic curvature. Therefore, only intrinsic curvature is of interest in the description of nature.

Note that the curvature 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 the latter case, 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; one needs at least four dimensions,
Challenge 270 e as you can find out if you try to imagine the situation.

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 77, 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
Challenge 271 e object from your surroundings, such as a Volkswagen Beetle. The Gaussian curvature $K$ defined in (184) 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.

Gaussian curvature is a measure of the intrinsic curvature of two-dimensional surfaces. Intrinsic measures of curvature are needed if we are forced to stay inside the surface or space that we are exploring. Physicists are thus particularly interested in Gaussian curvature and its higher-dimensional analogues.

## Three dimensions: curvature of space

For *three*-dimensional space, describing intrinsic curvature is a bit more involved. First of all, 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. In fact, we will describe three-dimensional curvature with help of 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

---

Ref. 167 $O_n = nC_n r^{n-1}$, we can generalize the expressions for curvature to

$$K = 3(n+2)\lim_{r\to 0}\left(1-\frac{V_n}{C_n r^n}\right)\frac{1}{r^2} \quad \text{or} \quad K = 3n\lim_{r\to 0}\left(1-\frac{O_n}{nC_n r^{n-1}}\right)\frac{1}{r^2}\,, \tag{185}$$

Challenge 269 ny as shown by Vermeil. A famous riddle is to determine the number $C_n$.

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

$$A = 4\pi r^2 \left(1 - \frac{K}{3} r^2 + ...\right) \quad \text{and} \quad V = \frac{4\pi}{3} r^3 \left(1 - \frac{K}{5} r^2 + ...\right) , \qquad (186)$$

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

$$K = 3 \lim_{r\to 0} \left(1 - \frac{A}{4\pi r^2}\right) \frac{1}{r^2} = 6 \lim_{r\to 0} \frac{r - \sqrt{A/4\pi}}{r^3} = 6 \lim_{r\to 0} \frac{r_{\text{excess}}}{r^3} , \qquad (187)$$

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.

If we apply the curvature definition with a small sphere to an arbitrary, non-isotropic point, we only get an *average* curvature at that point. For a non-isotropic point, the Gaussian curvature value 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
Challenge 273 ny
a tensor. (Why?) In other words, the 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 curvature values in three orthogonal directions, corresponding to the thee main axes of the ellipsoid.*

What are the curvature values for the 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 $\delta$ from the angle π on the surface of a body of mass $M$

* 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), German mathematician. Together with the Leonhard Euler, he was the most important mathematician of all times. A famous enfant prodige, 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. His notebooks are now available online at www.sub.uni-goettingen.de.

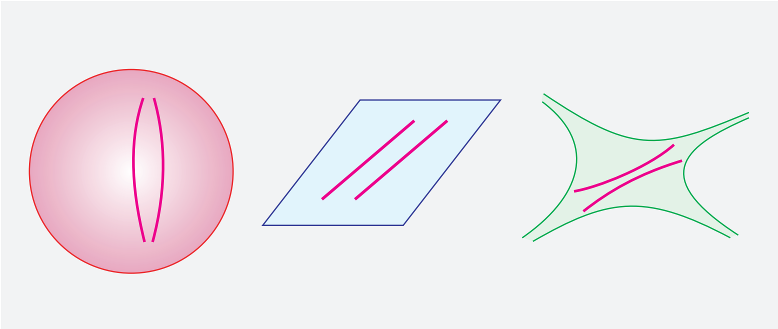


FIGURE 78 Positive, vanishing and negative curvature (in two dimensions) illustrated with the corresponding geodesic behaviour

and radius $r$ is given by

$$\delta = \pi - (\alpha + \beta + \gamma) \approx -A_{\text{triangle}} K = A_{\text{triangle}} \frac{GM}{r^3 c^2} \, . \qquad (188)$$

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 $\delta$ 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.

* 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) French dadaist and surrealist painter.

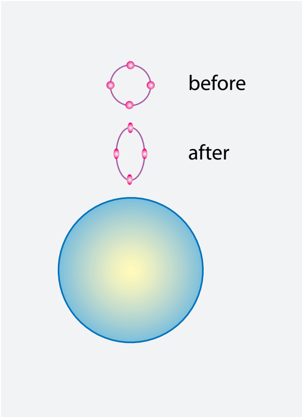



FIGURE 79 Tidal effects measure the curvature of space-time.



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 78. 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 274 e

$$\frac{\mathrm{d}^2 s}{\mathrm{d}l^2} = -K s + \text{higher orders} \tag{189}$$

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 $\tau$ (times the speed of light) for proper length. Thus separation and curvature are related by

$$\frac{\mathrm{d}^2 s}{\mathrm{d}\tau^2} = -K c^2 s + \text{higher orders} \,. \tag{190}$$

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: they describe tidal effects. In short, space-time curvature and tidal effects are precisely the same.

Page 160

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 79 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 275 ny to add indices to it. (How many?) The fun is that we can avoid indices for a while by
Ref. 168 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 276 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
Challenge 277 ny $K_{(12)}$, $K_{(23)}$ and $K_{(31)}$, is related to the excess radius defined above. Can you find out how?

If a surface has *constant* curvature, i.e., the same curvature at all locations, geometrical
Challenge 278 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 mo-
Ref. 169 tion 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.

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

$$K_{(12)} + K_{(23)} + K_{(31)} = \frac{8\pi G}{c^4} W^{(0)} \tag{191}$$

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 (191) specifies *how much* mass and energy curve space. We note that the factor on the right side is $2\pi$ divided by the maximum force.

Challenge 279 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

$$r_{\text{excess}} = r - \sqrt{A/4\pi} = \frac{G}{3c^2} M \,. \tag{192}$$

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 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.

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.

Challenge 280 ny

The quantities appearing in expression (191) are *independent* of the observer. But often people want to use observer-dependent quantities. The relation then gets more involved; the single equation (191) 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 $\varphi$. In other words, in everyday situations we can write

Challenge 281 e

$$K_{(0j)} = \frac{\partial^2 \varphi}{\partial (x^j)^2} \ . \tag{194}$$

In everyday situations, this approximation is excellent, and the function $\varphi$ 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

$$K_{(ij)} = 0 \quad \text{and} \quad K_{(01)} + K_{(02)} + K_{(03)} = \Delta\varphi = 4\pi G \rho \ . \tag{195}$$

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

Ref. 170

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

$$A = 4\pi r^2 \left(1 + \frac{1}{9} r^2 R\right) \tag{193}$$

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

Challenge 282 ny Can you show that relation (191) 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 123

## The Schwarzschild metric

Ref. 168 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 130 In spherical coordinates the line element is

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 \mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - \frac{2GM}{rc^2}} - r^2 \mathrm{d}\varphi^2 \,. \tag{196}$$

Challenge 283 ny The curvature of the Schwarzschild metric is then by

$$K_{r\varphi} = K_{r\theta} = -\frac{G}{c^2}\frac{M}{r^3} \quad \text{and} \quad K_{\theta\varphi} = 2\frac{G}{c^2}\frac{M}{r^3}$$
$$K_{t\varphi} = K_{t\theta} = \frac{G}{c^2}\frac{M}{r^3} \quad \text{and} \quad K_{tr} = -2\frac{G}{c^2}\frac{M}{r^3} \tag{197}$$

Ref. 168 Vol. I, page 160 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. Challenge 284 ny

## Curiosities and fun challenges about curvature

> “Il faut suivre sa pente, surtout si elle monte.*”
>
> André Gide

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 285 e

* *

Challenge 286 e Where are the points of highest and lowest Gaussian curvature on an egg?

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

## 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 tensors. 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
Ref. 171
of such a falling coffee cloud always stays constant, and this despite the deformation due
Page 127
to tidal forces. Figure 60 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 $\Box\varphi$.

## 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

$$R = -2K = \frac{-2}{r^2_{\text{curvature}}} \; . \tag{198}$$

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:

$$R = R^{\lambda}{}_{\lambda} = g^{\lambda\mu} R_{\lambda\mu} \; . \tag{199}$$

** '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.

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

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 287 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 *Einstein tensor* $G_{ab}$ is defined mathematically (for vanishing cosmological constant) as

$$G_{ab} = R_{ab} - \frac{1}{2} g_{ab} R \ . \tag{200}$$

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:

$$G_{00} = K_{(12)} + K_{(23)} + K_{(31)} \ . \tag{201}$$

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

$$G_{11} = K_{(02)} + K_{(03)} - K_{(23)} \ . \tag{202}$$

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 288 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

$$\mathrm{d}\boldsymbol{p} = T\ \mathrm{d}\boldsymbol{A}\ . \tag{203}$$

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:

$$T = \left(\begin{array}{c|ccc} w & S_1 & S_2 & S_3 \\ \hline S_1 & t_{11} & t_{12} & t_{13} \\ S_2 & t_{21} & t_{22} & t_{23} \\ S_3 & t_{31} & t_{32} & t_{33} \end{array}\right) = \left(\begin{array}{c|c} \text{energy} & \text{energy flow or} \\ \text{density} & \text{momentum density} \\ \hline \text{energy flow or} & \text{momentum} \\ \text{momentum density} & \text{flow density} \end{array}\right) \tag{204}$$

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 vanishing divergence of the tensor $T$, often written as

$$\partial_a T^{ab} = 0 \quad \text{or abbreviated} \quad T^{ab}{}_{,a} = 0\ , \tag{205}$$

expresses the fact 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 289 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 $\rho$, a pressure $p$ and a 4-velocity $\boldsymbol{u}$, we have

$$T^{ab} = (\rho_0 + p)u^a u^b - p g^{ab} \tag{206}$$

where $\rho_0$ is the density measured in the comoving frame, the so-called *proper* density.* Obviously, $\rho$, $\rho_0$ and $p$ depend on space and time.

Of course, for a particular material fluid, we need to know how pressure $p$ and density

* In the *comoving* frame we thus have

$$T^{ab} = \begin{pmatrix} \rho_0 c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}\ . \tag{207}$$

$\rho$ are related. A full material characterization thus requires the knowledge of the relation

$$p = p(\rho) . \tag{208}$$

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 at all. Its energy–momentum tensor is given by

$$T^{ab} = \rho_0 u^a u^b . \tag{209}$$

Challenge 290 ny Can you explain the difference from the liquid case?

Challenge 291 ny The divergence of the energy–momentum tensor vanishes for all times and positions, as you may want to check. 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 101 Another way is to deduce the equation from the Hilbert action, as done below. A third way is we are doing at present, namely to generalize the relation between curvature and energy to general observers.

* 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.

** This approximation leads to the famous speculation that the total energy of the universe is zero. Do you agree? Challenge 292 s

Einstein's field equations are given by

$$G_{ab} = -\kappa\, T_{ab}$$
or
$$R_{ab} - \frac{1}{2} g_{ab} R - \Lambda g_{ab} = -\kappa\, T^{ab} \,. \tag{210}$$

The constant $\kappa$, called the *gravitational coupling constant*, has been measured to be

$$\kappa = \frac{8\pi G}{c^4} = 2.1 \cdot 10^{-43}\,/\mathrm{N} \tag{211}$$

and its small value – the value $2\pi$ 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 $\Lambda$, the so-called *cosmological constant*, corresponds to a vacuum energy volume density, or pressure $\Lambda/\kappa$. Its low value is quite hard to measure. The currently favoured value is Page 217

$$\Lambda \approx 10^{-52}\,/\mathrm{m}^2 \quad \text{or} \quad \Lambda/\kappa \approx 0.5\,\mathrm{nJ/m^3} = 0.5\,\mathrm{nPa}\,. \tag{212}$$

Ref. 174 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.**



* 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 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. This unphysical idea is in contrast with the idea of *inter*action, as explained above. Ref. 175 Page 231

- *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 of general relativity. Page 231

- *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.

## Universal gravitation – again

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

Challenge 293 ny

$$\nabla^2\varphi = 4\pi\rho \quad \text{and} \quad \frac{d^2x}{dt^2} = -\nabla\varphi \tag{213}$$

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

$$a = G\frac{M}{r^2} , \tag{214}$$

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*. 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.

Challenge 294 e

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

$$R = -\kappa T . \tag{219}$$

This result also implies the relation between the excess radius and the mass inside a

* 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

$$\nabla_e b_a = R_{ceda} v^c v^d . \tag{215}$$

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

$$\nabla_e b^a = \nabla_e \nabla^a \varphi = R^a_{ced} v^c v^d \tag{216}$$

which implies that

$$\Delta\varphi = \nabla_a\nabla^a\varphi = R^a_{cad} v^c v^d = R_{cd} v^c v^d = \kappa(T_{cd} v^c v^d - T/2) \tag{217}$$

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

$$\Delta\varphi = 4\pi G\rho \tag{218}$$

as we wanted to show.

Challenge 295 ny sphere.

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. 176

Albert Einstein used to say that general relativity only provides the understanding of one side of the field equations (210), but not of the other. Can you see which side he meant? Challenge 296 ny

What can we do of interest with these 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* are predicted to appear. 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 produces 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 do things fall?

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, like all other examples of motion. 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. Obviously, the measure must be observer-invariant; in particular, it must be invariant under all possible changes of viewpoints.

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

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

constant $\Lambda$. 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

$$S = \frac{c^4}{16\pi G} \int (R + 2\Lambda)\, \mathrm{d}V \; . \tag{220}$$

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

A lengthy calculation confirms that the Hilbert action allows to deduce Einstein's field equations and vice versa. Both formulations are completely 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 in such a way that this integral changes as little as possible.

In addition to the Hilbert action, for a full description of motion we need initial conditions. The various ways to do this define a specific research field. This topic however, leads too far from our path. Ref. 112

In summary, 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*.

Challenge 297 ny

Can you show that the Hilbert action follows from the maximum force?

## The symmetries of general relativity

The main symmetry of the Lagrangian of general relativity is called *diffeomorphism invariance* or *general covariance*. 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. Diffeomorphism invariance is the essential symmetry of the Hilbert action.

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. 172

Apart from diffeomorphism symmetry, full general relativity, including mass–energy, has an additional symmetry which is not yet fully elucidated. This symmetry connects the various possible initial conditions of the field equations; the symmetry is extremely complex and is still a topic of research. These fascinating investigations should give new insights into the classical description of the big bang. Ref. 173

## Mass in general relativity

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 Page 257

is measured by the time a probe takes to orbit the unknown body.*

The intuitive mass definition *requires* flat space-time at infinity; it cannot be extended
Challenge 298 ny 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
Ref. 177 limit makes sense only if the constant $\Lambda$ is positive; this is the case for the currently
Ref. 112, Ref. 113 measured value, which is $\Lambda \approx 10^{-52}/\mathrm{m}^2$. Indeed, the radius–mass relation of black holes

$$2GM = Rc^2\left(1 - \frac{\Lambda}{3}R^2\right) \tag{223}$$

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
Page 121 the additional principle of a minimum force in nature, was proposed above.

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 derivation 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.

* 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

$$m = \frac{c^2}{32\pi G}\int_{S_R}(g_{ij,i}\nu_j - g_{ii,j}\nu_j)\mathrm{d}A \tag{221}$$

where $S_R$ is the coordinate sphere of radius $R$, $\nu$ 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
Ref. 180 distribution is sufficiently concentrated. Mathematical physicists have also shown that for any manifold whose metric changes at infinity as

$$g_{ij} = (1 + f/r + O(1/r^2))\delta_{ij} \tag{222}$$

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

## 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 231 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.

However, an interaction also transports energy between systems. Now, we have seen Page 179 that gravity can be said to transport energy only approximately. The properties of gravi- Challenge 299 s tational energy confirm this argument. Even in its energy aspect, gravitation is an interaction only approximately.

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 Challenge 300 ny along its path $s$, we get

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}s} = m\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}s} = m\left(\boldsymbol{e}_a\frac{\mathrm{d}\boldsymbol{u}^a}{\mathrm{d}s} + \frac{\mathrm{d}\boldsymbol{e}_a}{\mathrm{d}s}\boldsymbol{u}^a\right) = m\boldsymbol{e}_a\left(\frac{\mathrm{d}\boldsymbol{u}^a}{\mathrm{d}s} + \Gamma^a{}_{bd}\boldsymbol{u}^b\boldsymbol{u}^c\right) = 0 \qquad (224)$$

where $\boldsymbol{e}$ describes the unit vector along a coordinate axis. The energy–momentum Challenge 301 ny 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 (224) the second term alone is the real gravita- Ref. 181 tional force. But this term can be made to vanish along the entirety of any given world Challenge 302 ny 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

* 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 Vol. III, page 123 same applies to diffraction and to reflection. A full list of ways to bend light can be found elsewhere.

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 deflect 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

$$\mathrm{d}s^2 = \mathrm{d}x_a\,\mathrm{d}x^a = g_{ab}(x)\,\mathrm{d}x^a\,\mathrm{d}x^b \,. \tag{225}$$

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 timelike or spacelike (metric) *geodesic*, i.e.,
Challenge 303 ny the longest possible path between the two events,** only if

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(g_{ad}\frac{\mathrm{d}x^d}{\mathrm{d}s}\right) = \frac{1}{2}\frac{\partial g_{bc}}{\partial x^a}\frac{\mathrm{d}x^b}{\mathrm{d}s}\frac{\mathrm{d}x^c}{\mathrm{d}s} \,, \tag{226}$$

as long as $\mathrm{d}s$ is different from zero along the path.*** All bodies in free fall follow such
Page 134 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 (226) 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.

The path does not depend on the mass or on the material of the body. Therefore *an-*
Ref. 179 *timatter* also falls along geodesics. In other words, antimatter and matter do not repel;

---

* 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.

*** This is often written as

$$\frac{\mathrm{d}^2x^a}{\mathrm{d}s^2} + \Gamma^a_{bc}\frac{\mathrm{d}x^b}{\mathrm{d}s}\frac{\mathrm{d}x^c}{\mathrm{d}s} = 0 \tag{227}$$

where the condition

$$g_{ab}\frac{\mathrm{d}x^a}{\mathrm{d}s}\frac{\mathrm{d}x^b}{\mathrm{d}s} = 1 \tag{228}$$

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 $\Gamma$ appearing above are given by

$$\Gamma^a{}_{bc} = \left\{ {a \atop bc} \right\} = \frac{1}{2}g^{ad}(\partial_b g_{dc} + \partial_c g_{db} - \partial_d g_{bc}) \,, \tag{229}$$

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

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?

Challenge 304 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

$$\frac{d^2x^a}{du^2} + \Gamma^a{}_{bc}\frac{dx^b}{du}\frac{dx^c}{du} = 0 \tag{230}$$

with the different condition

$$g_{ab}\frac{dx^a}{du}\frac{dx^b}{du} = 0 \,. \tag{231}$$

Given all these definitions of various types of geodesics, what are the lines drawn in Figure 61 on page 130?

Challenge 305 ny

## 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 306 e

The *Riemann-Christoffel curvature tensor* $R$ is then defined as a quantity which allows us to calculate the curvature $K(\boldsymbol{p},\boldsymbol{q})$ for the surface spanned by $\boldsymbol{p}$ and $\boldsymbol{q}$, with area $A$, through

$$K(\boldsymbol{p},\boldsymbol{q}) = \frac{R\,\boldsymbol{pqpq}}{A^2(\boldsymbol{p},\boldsymbol{q})} = \frac{R_{abcd}\,p^a q^b p^c q^d}{(g_{\alpha\delta}g_{\beta\gamma} - g_{\alpha\gamma}g_{\beta\delta})p^\alpha q^\beta p^\gamma q^\delta} \tag{232}$$

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.***

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

** Bernhard Riemann (b. 1826 Breselenz, d. 1866 Selasca), important German 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

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
Challenge 307 e
velocity and the same direction, the distance between them changes. In other words, the local effect of gravity is *relative acceleration* of nearby particles.

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
Challenge 308 ny
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:

$$\boldsymbol{b} = R\,\boldsymbol{u}\,\boldsymbol{u}\,\boldsymbol{d} \quad \text{or, more clearly,} \quad b^a = R^a{}_{bcd}\, u^b\, u^c\, d^d \; . \tag{235}$$

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 $R$
Challenge 309 ny
vanishes in a region, space-time in that region is flat. This connection is easily deduced from this second definition.*

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 $\eta_{ab}$ from special relativity. In its neighbourhood, we have

$$\begin{aligned} g_{ab} &= \eta_{ab} + \frac{1}{3} R_{acbd} x^c x^d + O(x^3) \\ &= \frac{1}{2}(\partial_c \partial_d g_{ab}) x^c x^d + O(x^3) \; . \end{aligned} \tag{237}$$

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

have

$$g_{ab} = 1/g^{ab} \quad \text{and} \quad g_a{}^b = g^a{}_b = \delta^a_b \; . \tag{233}$$

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

$$R^a{}_{bcd} = \frac{\partial \Gamma^a{}_{bd}}{\partial x^c} - \frac{\partial \Gamma^a{}_{bc}}{\partial x^d} + \Gamma^a{}_{ec}\Gamma^e{}_{bd} - \Gamma^a{}_{fd}\Gamma^f{}_{bc} \; . \tag{234}$$

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
Ref. 183
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 $\varepsilon$, the vector $\boldsymbol{w}$ is changed to $\boldsymbol{w} + \delta\boldsymbol{w}$. One then has

$$\delta\boldsymbol{w} = -\varepsilon^2 R\,\boldsymbol{u}\,\boldsymbol{v}\,\boldsymbol{w} + \quad \text{higher-order terms} \quad . \tag{236}$$

More can be learned about the geodesic deviation by studying the behaviour of the famous south-pointing
Page 196
carriage which we have encountered before. This device, common 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
Challenge 310 ny
show a different direction than at the start of the trip. Can you explain why?

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

$$R_{bc} = R^{a}{}_{bac} \; . \tag{240}$$

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
Challenge 312 e Minkowski metric, in contrast to $R_{abcd}$.
Challenge 313 ny Can you confirm the relation $R_{abcd}R^{abcd} = 48m^2/r^6$ for the Schwarzschild solution?

## Curiosities and fun challenges about general relativity

For a long time, 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 $8 \cdot 10^{-10}$ m/s$^2$ (towards the Sun) since they passed the orbit of Saturn. This effect is called the *Pioneer anomaly*. The origin is not clear and still a subject of research. But several investigations have shown that the reason is *not* a deviation from the inverse
Ref. 184 square dependence of gravitation, as is sometimes proposed. In other words, the effect must be electromagnetic.

There are many hints that point to an asymmetry in heat radiation emission of the satellites. The on-board generators produce 2.5 kW of heat. A front-to-back asymmetry
Ref. 185 of only 80 W is sufficient to explain the anomaly. But the precise source of the asymmetry is still being discussed. Finding the asymmetry – or another explanation for the
Challenge 314 r deceleration – is one of the challenges of modern space physics.
Vol. III, page 102 Here is a test: Why can't the light mill effect, namely that impinging light pulls certain
Challenge 315 ny objects towards the source, be the reason for the anomaly?

* *

Maximum power or force appearing on horizons is the basis for general relativity. Are physical systems other than space-time that can also be described in this way?
Page 32 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.

* The free-fall definition shows that the Riemann tensor is symmetric in certain indices and antisymmetric
Challenge 311 ny in others:

$$R_{abcd} = R_{cdab} \quad , \quad R_{abcd} = -R_{bacd} = -R_{abdc} \; . \tag{238}$$

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

$$R_{abcd} + R_{adbc} + R_{acdb} = 0 \; . \tag{239}$$

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.

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. 102

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. 182

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 246

* *

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.

If a particular scalar-tensor theory obeys the general horizon equation (109) 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. Page 103

## A summary of the field equations

The field equations of general relativity state that (1) the local curvature of space is given by the local energy density divided by the maximum force, and (2) that objects move along the geodesics defined by this local curvature.

This description is confirmed 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. 188 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 is based on the full 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. 189 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 84. 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 Ref. 187 thought considers them: the starred sky above me and the moral law inside me.' 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 relgion-based governments.


FIGURE 80 A modern photograph of the night sky, showing a few thousand stars and the milky way. The Milky Way is positioned horizontally. (© Axel Mellinger, from Ref. 186)

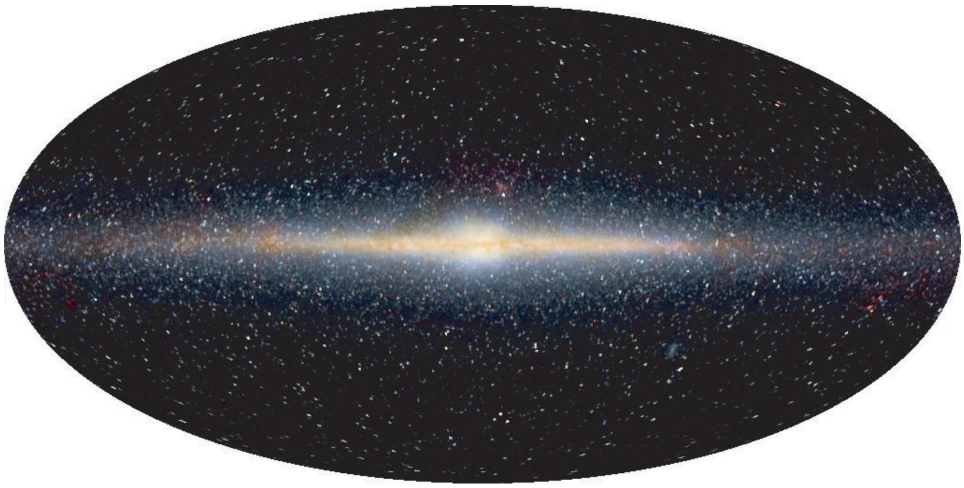


FIGURE 81 How the night sky, and our galaxy in particular, looks in the near infrared (NASA false colour image)

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 316 ny

Vol. III, page 126

Many more extragalactic objects are visible with telescopes. Nowadays, this is one of

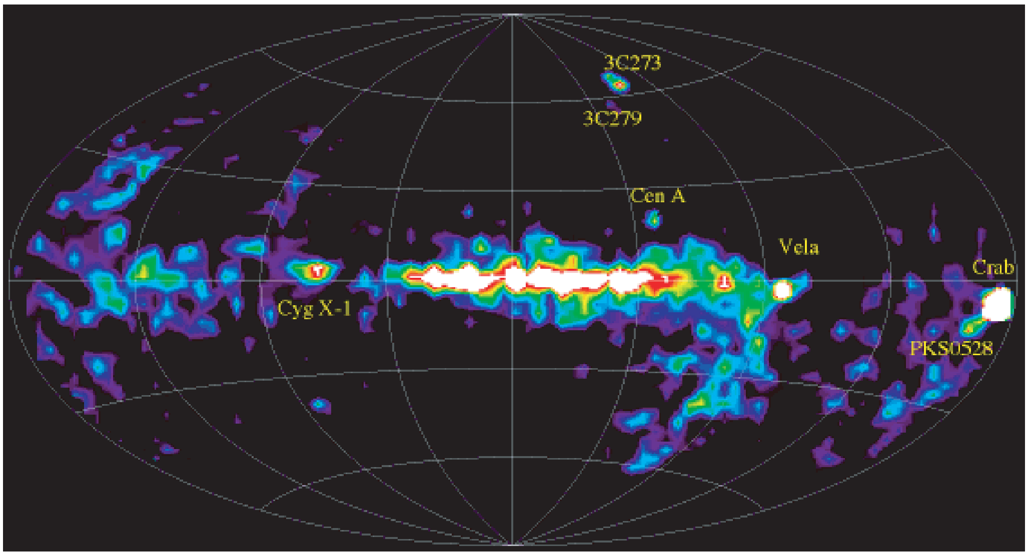



FIGURE 82 The X-rays observed in the night sky, for energies between 1 and 30 MeV (NASA)

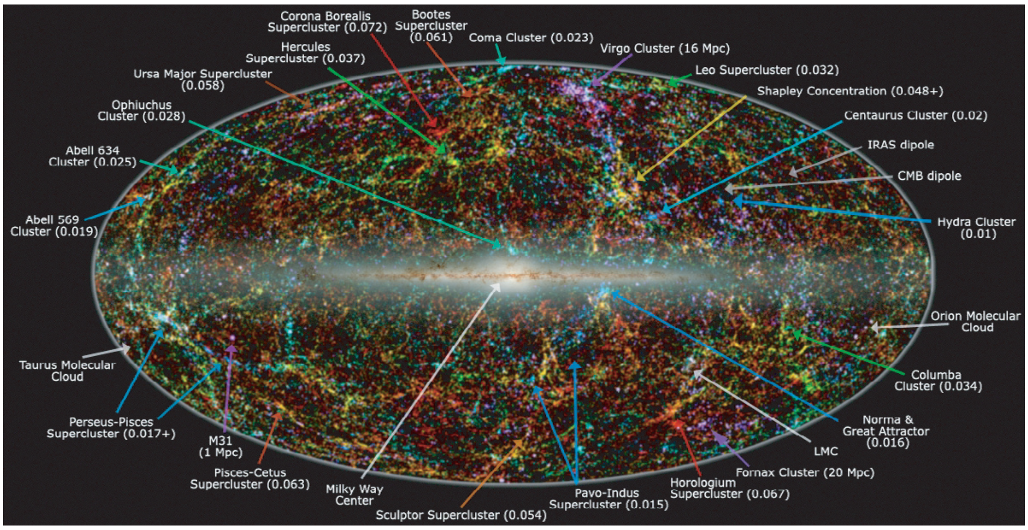



FIGURE 83 A false colour image, composed from infrared data, showing the large-scale structure of the universe around us; the colour of each galaxy represents its distance and the numbers in parentheses specify the redshift; an infrared image of the Milky Way is superposed (courtesy Thomas JarretIPAC/Caltech)

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 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 84,

FIGURE 84 The Andromeda nebula M31, on of our neighbour galaxies (and the 31st member of the Messier object listing) (NASA)

Figure 85 and Figure 86. 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 photograph in Figure 81, 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 317 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. 190

## 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*. *Clouds* are systems in which the matter density diminishes with the distance from the centre, with no definite border and with no definite size. Most astrophysical objects are best described as 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

* 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 85 The elliptical galaxy NGC 205 (the 205th member of the New Galactic Catalogue) (NASA)

FIGURE 86 The colliding galaxies M51 and M51B, 65 000 al across, 31 Mal away, show how a galaxy dies (NASA).

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,

FIGURE 87 The universe is full of galaxies – this photograph shows the Perseus cluster (NASA)

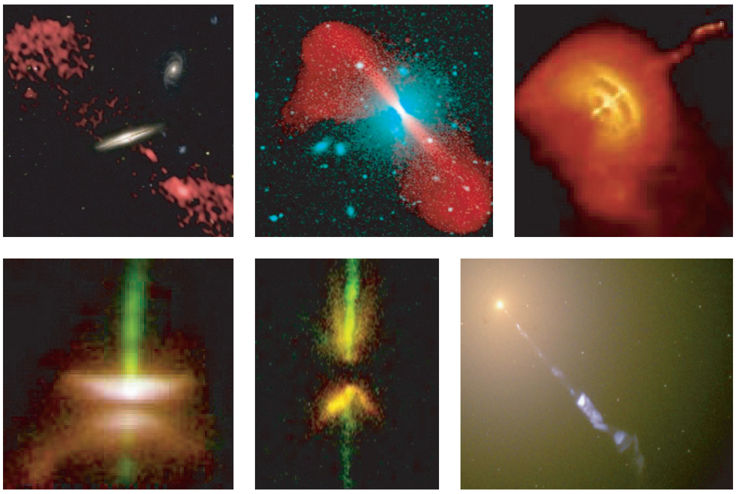


FIGURE 88 Rotating clouds emitting jets along their axis; top row: a composite image (visible and infrared) of the galaxy 0313-192, the galaxy 3C296, and the Vela pulsar; bottom row: the star in formation HH30, the star in formation DG Tauri B, and a black hole jet from the galaxy M87 (all NASA)

little dust particles from skin, vapour, etc.

Ref. 191 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 88 gives some examples. (Does the Sun have a jet? Does the Milky Way have a jet? So far, none has been detected – there is still room for discovery.) Challenge 318 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. 192

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 | form stars between 0.04 and 200 solar masses |
| | frequency | between 0 and 1000 solar masses per year per galaxy; around 1 solar mass per year in the Milky Way |
| 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 |
| 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** | | |

TABLE 5 (Continued) Some observations about the universe

| Aspect | Main properties | Value |
|---|---|---|
| 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. 190 |
| | speed | 600 km/s towards Hydra-Centaurus |
| | containing | about 30 000 pulsars Ref. 193 |
| | 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 |
| | 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 (except when stars fuse) Ref. 194 |
| 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. 195 |
| L dwarfs | low temperature | 1200 to 2800 K |
| T dwarfs | low temperature | 900 to 1100 K |

TABLE 5 (Continued) Some observations about the universe

| Aspect | Main properties | Value |
| --- | --- | --- |
| white dwarfs | small radius | $r \approx 5000$ km |
| | high temperature | cools from 100 000 to 5000 K |
| neutron stars | nuclear mass density | $\rho \approx 10^{17}$ kg/m$^3$ |
| | small size | $r \approx 10$ 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. 196 |
| some are gamma repeaters, others are anomalous X-ray pulsars | | |
| | mass | above 25 solar masses Ref. 197 |
| 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.* $10^{26}$ m = 100 Ym |
| Expansion | Hubble's constant | 71(4) km s$^{-1}$ Mpc$^{-1}$ or $2.3(2) \cdot 10^{-18}$ s$^{-1}$ |
| 'Age' of the universe | | 13.7(2) Ga |
| Vacuum | energy density | 0.5 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 211 |
| | topology | simple at all measured scales |
| Dimensions | number | 3 for space, 1 for time, at all measured energies and scales |
| Matter | density | 2 to $11 \cdot 10^{-27}$ 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 $\Omega_M$) |
| Dark matter | density | $\Omega_{DM} = 0.21$ (included in $\Omega_M$), unknown |
| Dark energy | density | $\Omega_{DM} = 0.75$, unknown |
| Photons | number density | 4 to $5 \cdot 10^8$ /m$^3$ |
| | | $= 1.7$ to $2.1 \cdot 10^{-31}$ kg/m$^3$ |
| | energy density | $\Omega_R = 4.6 \cdot 10^{-5}$ |
| Neutrinos | energy density | $\Omega_\nu$ unknown |
| Average temperature | photons | 2.725(2) K |
| | neutrinos | not measured, predicted value is 2 K |
| Perturbations | photon anisotropy | $\Delta T/T = 1 \cdot 10^{-5}$ |

FIGURE 89 The beauty of astronomy: the Cygnus Bubble, discovered in 2008, a nebula expelled from a central star (false colour image courtesy T.A. Rector, H. Schweiker)



TABLE 5 (Continued) Some observations about the universe

| Aspect | Main properties | Value |
|---|---|---|
| | 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.

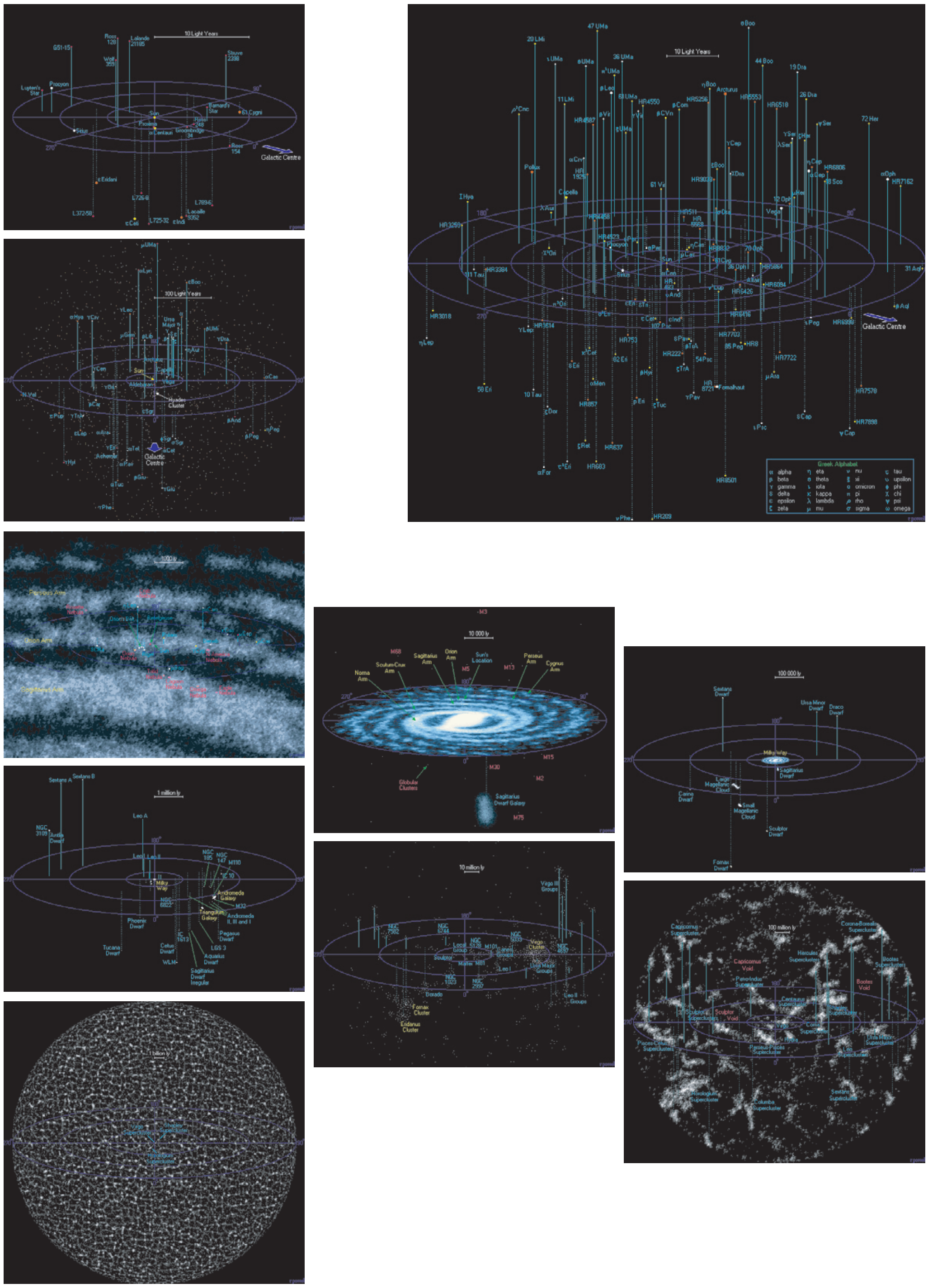


FIGURE 90 An atlas of our cosmic environment: illustrations at scales up to 12.5, 50, 250, 5 000, 50 000, 500 000, 5 million, 100 million, 1 000 million and 14 000 million light years (© Richard Powell, www.atlasoftheuniverse.com)

## 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*) *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
Vol. VI, page 265
this increase? We will come back to this issue in the final leg of our adventure.)

— 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
Challenge 319 e
this belief below. (Do you agree with it?)

— 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 the complete ascent of Motion Mountain becomes impossible. (For example: Is the amount of matter and en-
Challenge 320 s
ergy in the full universe the same as in the observable universe?)

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 one took all the iron from the Earth's core and made it into
Challenge 321 s
a wire reaching to the edge of the observable universe, how thick would it be? The answer might surprise you. 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
Challenge 322 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.

* '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.

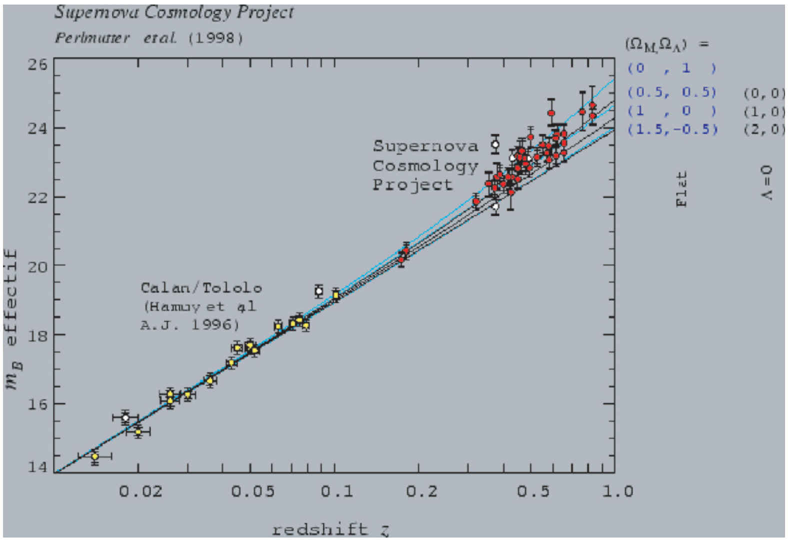



FIGURE 91 The relation between star distance and star velocity

Challenge 323 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 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*. Ref. 198

The second point about the universe is even more important. In the 1920s, independently, 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 Ref. 199

$$v = H\ d\ , \tag{241}$$

where the proportionality constant $H$ is today called the *Hubble constant*. A modern graph of the relation is given in Figure 91. The Hubble constant is known today to have a value around 71 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.

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

In fact, the discovery by Wirtz, Lundmark and Stromberg implies that *every* galaxy 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 (241)! Some deviations are observed for nearby galaxies, as mentioned above, and for faraway galaxies, as we will see.

Challenge 324 ny

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 13 700 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 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, 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 for 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 92. In fact, the spectrum follows the black-body dependence to a precision of less than 1 part in $10^4$.

Ref. 200

Challenge 325 ny

Ref. 201

Ref. 202

In summary, 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.

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

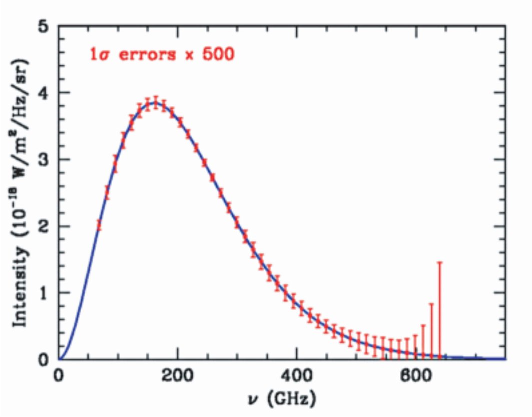



FIGURE 92 The measured spectrum of the cosmic background radiation, with the error bars multiplied by 500, compared to the calculated Planck spectrum for 2.728 K (NASA).

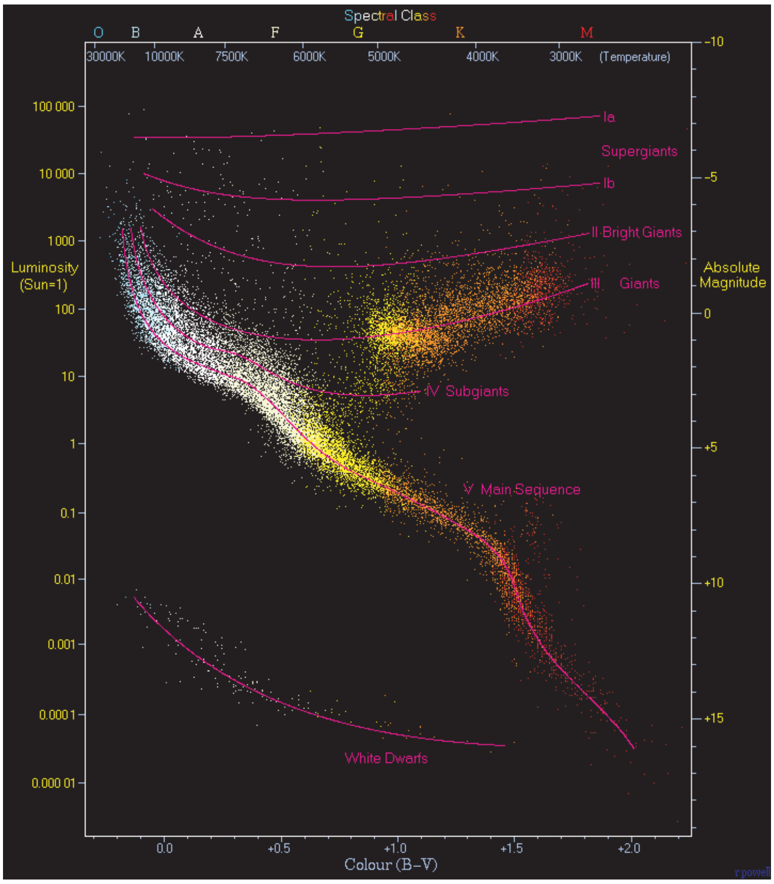



FIGURE 93 The Hertzsprung–Russell diagram (© Richard Powell)

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 93. 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 the diagram, to estimate the age of stellar clusters, 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. 203 a central part of astrophysics, but we will not explore them here.

Yet we do not have the full answer to our question. Why do stars shine at all? Clearly, they shine because they are hot. They are hot because of nuclear reactions in their interior.
Vol. V, page 150 We will discuss these processes in more detail in the volume on the nucleus.

## A short history of the universe

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

Ref. 204

Not only stars are born, shine and die. Also galaxies do so. What about the universe? The most important adventures that the matter and radiation around us have experienced
Ref. 205 are summarized in Table 6. The steps not yet discussed will be studied in the rest of our ascent of Motion Mountain. The history table is awe-inspiring. This history table even has applications no theoretical physicist would have imagined. 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.* $14 \cdot 10^9$ a | $\approx t_{Pl}$ [b] | Time, space, matter and initial conditions are indeterminate | $10^{32}$ K $\approx T_{Pl}$ |
| $13 \cdot 10^9$ a | *c.* $1000\ t_{Pl}$ $\approx 10^{-42}$ 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 |

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

TABLE 6 (Continued) A short history of the universe

| Time before now[a] | Time from big bang[b] | Event | Temperature |
|---|---|---|---|
| | | 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 |
| $z = 1100$ | 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 | $T_\gamma = T_o(1+z)$ |
| $z = 10$ 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 | |
| $4.7 \cdot 10^9$ 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; water condenses | |
| $3.5 \cdot 10^9$ 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. 206 | |
| $1.3 \cdot 10^9$ a | | Macroscopic, multicellular life appears, fungi conquer land | |

TABLE 6 (Continued) A short history of the universe

| Time before now[a] | Time from big bang[b] | Event | Temperature |
|---|---|---|---|
| $800 \cdot 10^6$ a | | Earth is completely covered with ice for the first time (reason still unknown) Ref. 207 | |
| 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: 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) | |
| $450 \cdot 10^6$ a | | Land plants appear | |
| $370 \cdot 10^6$ a | | Wooden trees 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 | |
| $65.5 \cdot 10^6$ a | | **Cenozoic** era (Caenozoicum, 'age of new life') starts: dinosaurs become extinct after an asteroid hits the Earth in the Yucatan, 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. 208 | |
| 500 000 a | | Formation of youngest stars in galaxy | |
| 500 000 a | | Homo sapiens appears | |

TABLE 6 (Continued) A short history of the universe

| Time before now[a] | Time from big bang[b] | Event | Temperature |
|---|---|---|---|
| 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 | $T_\gamma = 2.73$ K, $T_\nu \approx 1.6$ K, $T_b \approx 0$ 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 213. A year is abbreviated 'a' (Latin 'annus'). Errors in the last digits are given between parentheses.

*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 85

*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 through radioactive dating.
Vol. V, page 135

Despite its length and its interest, the history table has its limitations. For example, 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.

Research in astrophysics is directed at discovering and understanding all phenomena observed in the skies. 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 in the universe.

## 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 326 s

Page 201

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. 209 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 327 ny

$$ds^2 = c^2 dt^2 - a^2(t)(dx^2 + dy^2 + dz^2) \,. \tag{242}$$

The quantity $a(t)$ is called the *scale factor*. Matter is described by a density $\rho_M$ and a pressure $p_M$. Inserting all this into the field equations, we get two equations that any school student can grasp; they are

$$\left(\frac{\dot a}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho_M + \frac{\Lambda}{3} \tag{243}$$

and

$$\ddot a = -\frac{4\pi G}{3}(\rho_M + 3p_M)\, a + \frac{\Lambda}{3}\, a \,. \tag{244}$$

Together, they imply

$$\dot\rho_M = -3\frac{\dot a}{a}(\rho_M + p_M) \,. \tag{245}$$

At the present time $t_0$, the pressure of matter is negligible. (In the following, the index 0 refers to the present time.) In this case, the expression $\rho_M a^3$ is constant in time. Challenge 328 e

Equations (243) and (244) depend on only two constants of nature: the gravitational constant $G$, related to the maximum force or power in nature, and the cosmological constant $\Lambda$, describing the energy density of the vacuum, or, if one prefers, the smallest force in nature.

Before we discuss the equations, first a few points of vocabulary. It is customary to relate all mass densities to the so-called *critical mass density* $\rho_c$ given by Challenge 329 ny

$$\rho_c = \frac{3H_0^2}{8\pi G} \approx (8 \pm 2) \cdot 10^{-27}\,\mathrm{kg/m^3} \tag{246}$$

* Aleksander Aleksandrowitsch Friedmann (1888–1925), Russian 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), Belgian priest and cosmologist, took it up and expanded it in 1927, focusing, as his job required, 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 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 220, page 221

corresponding to about 8, give or take 2, hydrogen atoms per cubic metre. On Earth, one 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 $\Omega_M$ defined as

$$\Omega_M = \rho_0/\rho_c \, . \tag{247}$$

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

$$\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c} = \frac{\Lambda c^2}{8\pi G \rho_c} = \frac{\Lambda c^2}{3H_0^2} \, . \tag{248}$$

A third dimensionless parameter $\Omega_K$ 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

$$\Omega_K = \frac{-k}{R_0^2 \, H_0^2} \tag{249}$$

and its sign is opposite to the one of the curvature $k$; $\Omega_K$ 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 be simply or multiply connected. In these cases the topology is not completely fixed by the curvature.

The present-time Hubble parameter is defined by $H_0 = \dot{a}_0/a_0$. From equation (243) we then get the central relation

Challenge 330 ny

$$\Omega_M + \Omega_\Lambda + \Omega_K = 1 \, . \tag{250}$$

In the past, when data were lacking, physicists 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 94 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

$$H_0 = 71 \pm 4 \,\mathrm{km/sMpc} = 2.3 \pm 2 \cdot 10^{-18} \,/\mathrm{s} \tag{251}$$

which corresponds to an age of the universe of $13.7 \pm 2$ thousand million years. In other words, the age deduced from the history of space-time agrees with the age, given above,

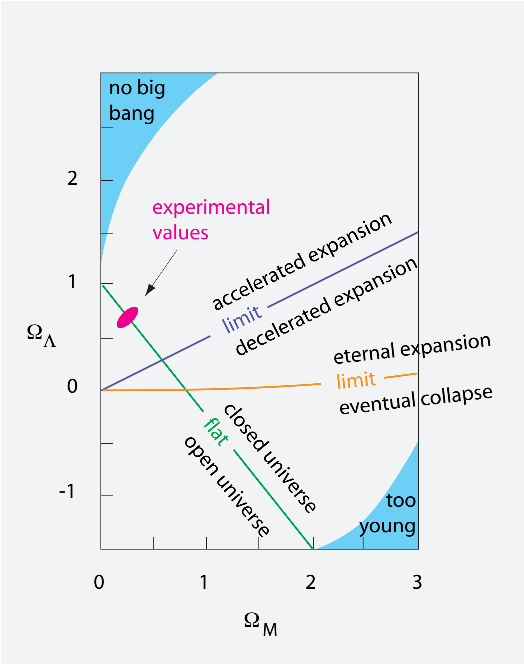



FIGURE 94 The ranges for the Ω parameters and their consequences

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

$$q_0 = -\frac{\ddot{a}_0}{a_0 H_0^2} = \frac{1}{2}\Omega_M - \Omega_\Lambda . \tag{252}$$

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 94.

An even clearer way to picture the expansion of the universe for vanishing pressure is to rewrite equation (243) using $\tau = t\, H_0$ and $x(\tau) = a(t)/a(t_0)$, yielding

$$\left(\frac{dx}{d\tau}\right)^2 + U(x) = \Omega_K$$
$$\text{where} \quad U(x) = -\Omega_\Lambda x - \Omega_\Lambda x^2 . \tag{253}$$

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

For vanishing $\Omega_\Lambda$, the universe either expands for ever, or recollapses, depending on the value of the mass–energy density.

For non-vanishing (positive) $\Omega_\Lambda$, the potential has exactly one maximum; if the particle has enough energy to get over the maximum, it will accelerate continuously. That is the situation the universe seems to be in today.

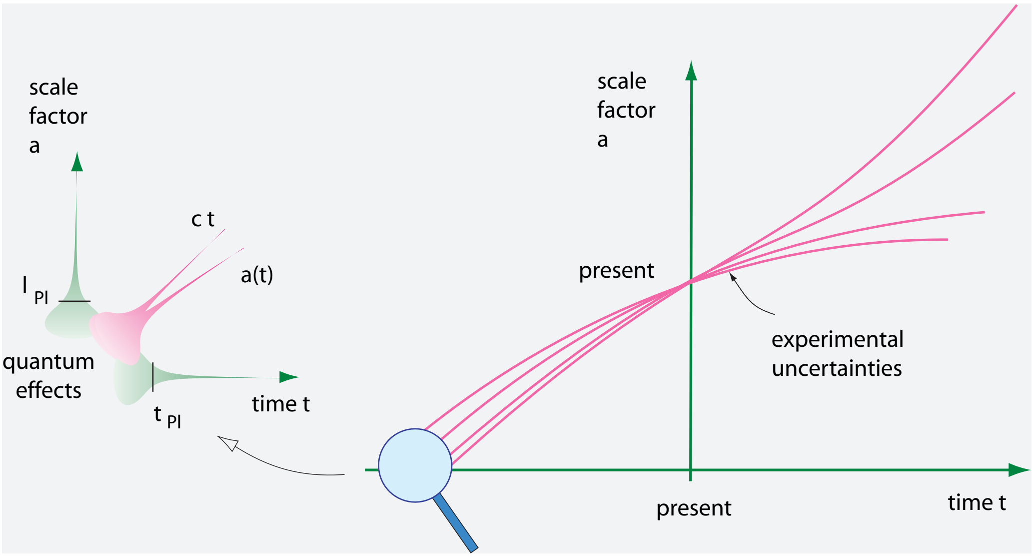



FIGURE 95 The evolution of the universe's scale $a$ for different values of its mass density

For a certain time range, the result is shown in Figure 95. There are two points to be noted: first the set of possible curves is described by *two* parameters, not one. In addition, lines cannot be drawn down to zero size. 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.

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

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 is the one we assume when we talk about the *age* of the universe.

An overview of the possibilities for the long time evolution is given in Figure 96. 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 $U(x)$.

In 1917, the Dutch physicist Willem de Sitter had found, 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 331 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 became the standard model in textbooks. However, in a sort of collective blindness that lasted from

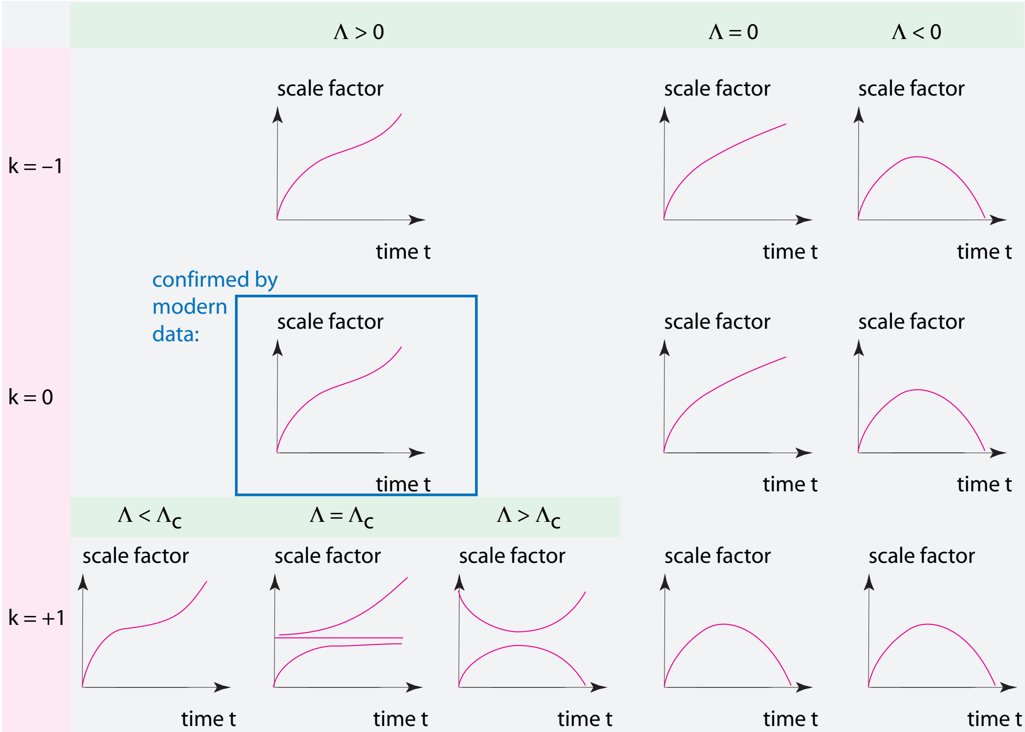



FIGURE 96 The long-term evolution of the universe's scale factor $a$ for various parameters

around 1950 to 1990, almost everybody believed that $\Lambda = 0$.* Only towards the end of the twentieth century did experimental progress allow one 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 97, shows. But that 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 eye, the sky is pitch black. But measurements show that even the empty sky is not completely

Challenge 332 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 one has $t_0 = 1/H_o$.
** '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), German writer.


FIGURE 97 The sky is blue at night as well, as this long-time exposure shows. On the top left, the bright object is Mars; the bottom shown a rare coloured fog bow created by moonlight. (© Wally Pacholka)

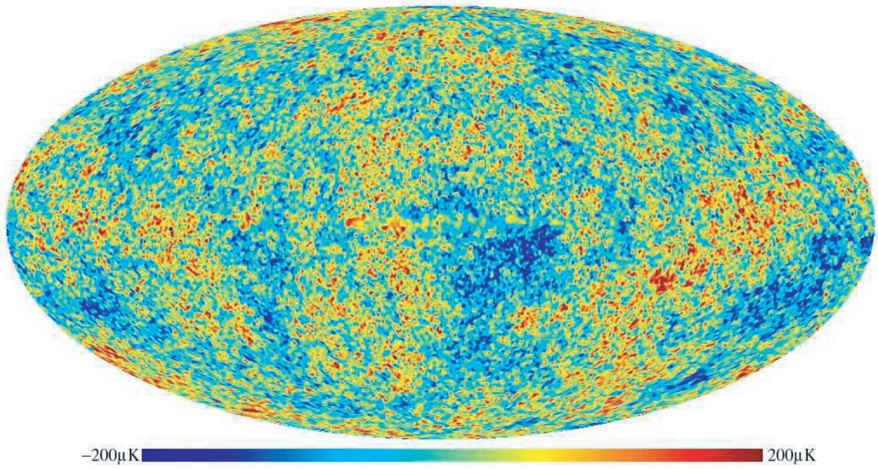



FIGURE 98 A false colour image of the fluctuations of the cosmic background radiation, after the Doppler shift from our local motion and the signals from the Milky Way have been subtracted (WMAP/NASA)

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. 210 Thus the universe is indeed colder than the stars. But why is this so? If the universe were homogeneous on large scales and 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 333 s

In a deep forest, one sees a tree in every direction. 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! It would be impossible 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' 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 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. 211

Calculations are necessary to decide which effect is the greater one. This issue has been studied in great detail by Paul Wesson; he explains that the first effect is larger than the second by a factor of about three. We may thus state correctly that the sky is dark at night *mostly* because the universe has a finite age. We can add that the sky would be somewhat brighter if the universe were not expanding. Ref. 212

We note that the darkness of the sky arises only because the speed of light is finite. Can you confirm this? Ref. 210 Challenge 335 ny

The darkness of the sky also tells us that the universe has a *large* (but finite) age. 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. 213

$$T \sim \frac{1}{a(t)} . \tag{254}$$

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

* Heinrich Wilhelm Matthäus Olbers (b. 1758 Arbergen, d. 1840 Bremen), 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 others had made similar points before, such as the Swiss astronomer Jean Philippe Loÿs de Cheseaux in 1744 and Johannes Kepler in 1610. Page 125

** 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 334 ny 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. 211

Challenge 336 ny

In summary, the sky is black 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 337 ny

## 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. 214

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 225

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. 215

$$\Omega_M \approx 0.3 \quad , \quad \Omega_\Lambda \approx 0.7 \quad , \quad \Omega_K \approx 0.0 \qquad (255)$$

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 227

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

* 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

term. For the cosmological constant $\Lambda$ the present measurements yield

$$\Lambda = \Omega_\Lambda \frac{3H_0^2}{c^2} \approx 10^{-52}\,/\mathrm{m}^2 \,. \tag{256}$$

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

$$\rho_\Lambda c^2 = \frac{\Lambda c^4}{8\pi G} \approx 0.5\,\mathrm{nJ/m^3} \approx \frac{10^{-46}\,(\mathrm{GeV})^4}{(\hbar c)^3} \,. \tag{257}$$

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 Page 181 relativity

$$\Delta\varphi = 4\pi G(\rho + 3p/c^2) \tag{258}$$

Ref. 216 we get

$$\Delta\varphi = 4\pi G(\rho_\mathrm{M} - 2\rho_\Lambda) \,. \tag{259}$$

Challenge 338 ny Thus the gravitational acceleration around a mass $M$ is

$$a = \frac{GM}{r^2} - \frac{\Lambda}{3}c^2 r = \frac{GM}{r^2} - \Omega_\Lambda H_0^2\, r \,, \tag{260}$$

which shows that a *positive* vacuum energy indeed leads to a *repulsive* gravitational effect. Inserting the mentioned value (256) for the cosmological constant $\Lambda$ 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 339 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 this cosmological constant is *not* explained by general relativity. This mystery will be solved only with the help of quantum theory. In any case, the cosmological constant is the first 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. 217 popular books in order to explain rain? No. Even if it were filled with air, the total mass would never have allowed the universe to reach the present size; it would have recollapsed Challenge 340 ny much earlier and we would not exist.

matter of a type unknown so far. This issue is called the *dark matter problem*; it is one of the important unsolved questions of cosmology.

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
Page 189
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 dust particle.

If we remember that the average density of the universe is $10^{-26}$ 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
Page 82
contain so-called *virtual* charges. However, 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
Page 204
Page 244
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
Ref. 218
to nature. Nevertheless, the term caught on. Since the past smallness of the universe be checked directly, we need to look for other, verifiable consequences. The central ones are the following:

— all matter moves away from all other matter;
— the mass of the universe is made up of about 75% hydrogen and 23% helium;
— there is thermal background radiation of about 2.7 K;
— the maximal age for any system in the universe is around fourteen thousand million years;
— there are background neutrinos with a temperature of about 2 K;*
— for non-vanishing cosmological constant, Newtonian gravity is slightly reduced.

All predictions except the last two have been confirmed by observations. Technology will probably not allow us to check the last two in the foreseeable future; however, there is no evidence against them.

* The theory states that $T_\nu/T_\gamma \approx (4/11)^{1/3}$. These neutrinos appeared about 0.3 s after the big bang.

Competing descriptions of the universe have not been successful in matching Ref. 218 observations. In addition, theoretical arguments state that with matter distributions such as the observed one, and some rather weak general assumptions, there is no way Ref. 219 to avoid a period in the *finite* past in which the universe was extremely small and hot. Therefore it is worth having a close look at the situation.

### 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. The mechanism and the origin of the expansion is *unknown* at this point of our mountain ascent. Because of the importance of spatial expansion, the whole phenomenon cannot be called an explosion at all. 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 it big? The visible universe was rather small about fourteen thousand million years ago, 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?

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 nonlinearities in general relativity at situations of high energy densities is not even completely clarified yet.

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 Vol. VI, page 95 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.

### Was the big bang a beginning?

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

* Many physicists are still wary of making such strong statements on this point. The first sections of the Vol. VI, page 52 final part of our mountain ascent give the precise arguments leading to them.

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 Vol. VI, page 58 in the first chapter of the final part of our mountain ascent. 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.

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 Ref. 220 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 Page 45 already in the beginning of our walk.

A similar mistake lies behind the idea that the universe had certain 'initial condi- Page 190 tions.' 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, Ref. 221 Stephen Hawking sold millions of copies of a book explaining that a description without initial conditions is the most appealing, without mentioning the fact that there is no other possibility anyway.*

In summary, the big bang is not a beginning, nor does it imply one. We will uncover Page 264 the correct way to think about it in the final part of our mountain ascent.

## 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 Page 238 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 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.

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

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

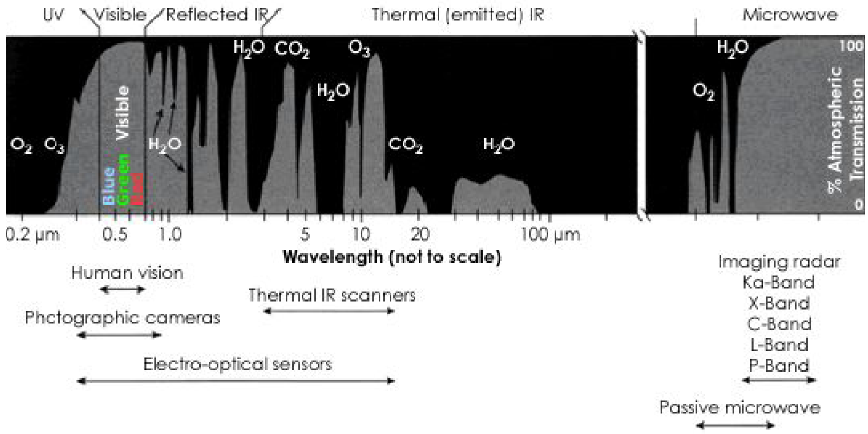



FIGURE 99 The transmittance of the atmosphere (NASA)

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.

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 the ascent of Motion Mountain 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.

Challenge 341 ny

## 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 99 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.)

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.

Vol. III, page 164

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

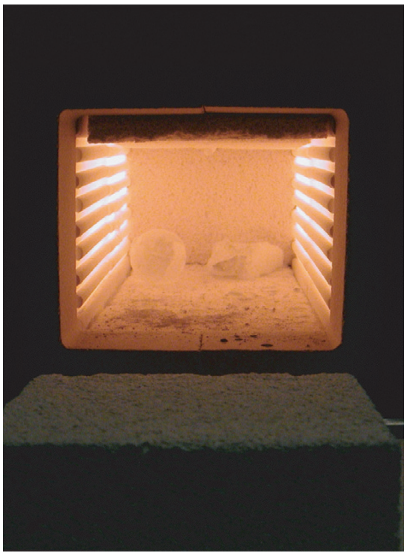


FIGURE 100 A hot red oven shows that at high temperature, objects and their environment cannot be distinguished from each other



body, nothing can be seen at all. Any oven, such as the shown in Figure 100 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 are the colours of the stars different?

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
Page 118 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. 222 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
Page 79 photographs beautifully show this.

The basic colour determined by temperature is changed by two effects. The first, the
Challenge 342 ny *Doppler red-shift* $z$, depends on the speed $v$ between source and observer as

$$z = \frac{\Delta\lambda}{\lambda} = \frac{f_S}{f_O} - 1 = \sqrt{\frac{c+v}{c-v}} - 1 \,. \tag{261}$$

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

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 all other 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.

Challenge 343 ny

to a recessional speed of more than 94 % of the speed of light. Note that in the universe, the red-shift is also related to the scale factor $R(t)$ by

$$z = \frac{R(t_0)}{R(t_{\text{emission}})} - 1 \,. \tag{262}$$

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 is given by

$$z_g = \frac{\Delta\lambda}{\lambda} = \frac{f_S}{f_0} - 1 = \frac{1}{\sqrt{1 - \frac{2GM}{c^2 R}}} - 1 \,. \tag{263}$$

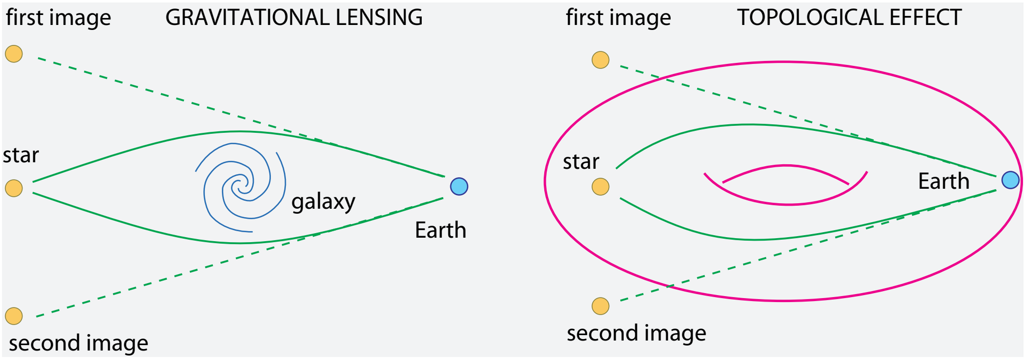



FIGURE 101 Two ways in which one star can lead to several images

Challenge 344 ny 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 all the known properties of nature. But the colour issue leads to the next question. Page 233

## 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. They are called *brown dwarfs*. 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 235 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 quasars in the system Q0957+561. Ref. 223

The two images are the result of *gravitational lensing*, an effect illustrated in Figure 101. Indeed, a large galaxy can be seen between the two images observed by Pelt, and much nearer to the Earth that 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 is Fritz Zwicky, who predicted in 1937 that the effect would be quite common and easy to Ref. 224

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

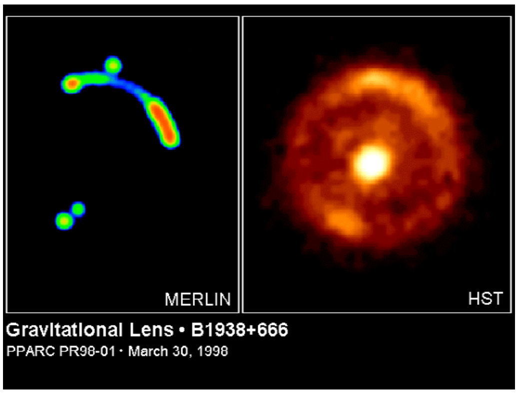



FIGURE 102 The Zwicky–Einstein ring B1938+666, seen in the radio spectrum (left) and in the optical domain (right) (NASA)

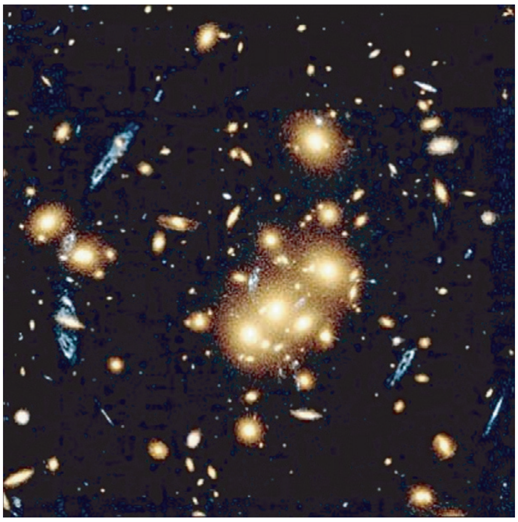


FIGURE 103 Multiple blue images of a galaxy formed by the yellow cluster CL0024+1654 (NASA)

observe, if lined-up galaxies instead of lined-up stars were considered, as indeed turned out to be the case.

Challenge 345 ny

Interestingly, when the time delay is known, astronomers are able to determine the 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 102, is one of the most beautiful examples. In 2005, several cases of gravitational lensing 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 103, 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.

## What is the shape of the universe?

A popular analogy for the expansion of the universe is the comparison to a rubber balloon that increass 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. 225 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 standard 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 212 In addition, Figure 94 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, 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 101. 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.

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

** 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

Ref. 226 $k = 0$, and to *any* hyperbolic manifold in the case $k = -1$.

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 Vol. VI, page 94 more complex options appear; they will be discussed in the last part of our walk.

## What is behind the horizon?

> "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 Ref. 227 moves rapidly away from us. A detailed investigation shows that for a matter-dominated Challenge 346 ny universe the horizon moves away from us with a velocity

$$v_{\text{horizon}} = 3c\ . \tag{264}$$

A pretty result, isn't it? Obviously, since the horizon does not transport any signal, this is not a contradiction of relativity. But what is behind the horizon?

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, an open or marginal universe implies that there is an infinite amount of matter Challenge 347 ny behind the horizon. Is such a statement verifiable?

In a *closed* universe, matter is still predicted to exist behind the horizon; however, in this case it is only a finite amount.

In short, the standard 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.

## 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 person 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. 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. 228 Guth argued that the precise flatness, the homogeneity and the isotropy 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:

$$E = E_{\mathrm{b}} + E_{\gamma} + E_{\nu} \approx \frac{c^2 M_0}{T_0} + \ldots + \ldots \approx \frac{c^2}{G} + \ldots \,. \qquad (265)$$

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

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. 229

$$S = \frac{kc^3}{G\hbar}\frac{A}{4} = \frac{kG}{\hbar c} 4\pi M^2 \,, \qquad (266)$$

Challenge 348 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.

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

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

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. 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 349 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 350 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 Ref. 230 coming years, might provide some information on the way to settle the issue.

## 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 Page 254 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.

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

Can we distinguish between space expanding and galaxies moving apart? Yes, we can. Challenge 351 ny Can you find an argument or devise an experiment to do so?

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. 231

## 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 answer the question positively.

Challenge 352 ny

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 quantum theory does not change this conclusion, despite recurring reports to the contrary.

Page 140

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 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?

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

Ref. 232

— *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. 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. 233

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

** 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.

Ref. 234 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 it. Challenge 353 e

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 incorrectness of Mach's principle will persist, like that of the derision of Columbus. Ref. 234

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 about 20 million years. As the physicist Dennis Sciama pointed out, from the shape of our galaxy we can draw a powerful conclusion: there must be a lot of other matter, i.e., a lot of other stars and galaxies in the universe. Can you confirm his reasoning? Challenge 354 s

## 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 fact that the night sky is dark in all directions is a consequence 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

* 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.

move with these 'low' speeds across the universe have already been studied in our walk.
Challenge 355 ny Can you give a summary?

By the way, is the term 'universe' correct? Does the universe rotate, as its name implies? If by universe one means 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, Ref. 235 if by universe one only means '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, it cannot have rotated by more than a hundredth of a millionth of a turn! In short, 'universe' is a misnomer.

## 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 grav- Ref. 236 itation does *not* alter the path of two parallel light beams, even though it *does* alter the path of antiparallel light beams.* The reason is that for parallel beams moving at light Challenge 356 ny speed, the gravitomagnetic component exactly *cancels* the gravitoelectric component.

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 357 ny 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 Bronstein demonstrated in the 1930s that any light decay Ref. 237 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. 238 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

* Antiparallel beams are parallel beams travelling in opposite directions.

Challenge 358 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 and gravitational red-shifts.
Page 225

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

## Summary on cosmology

In summary, asking what precisely we see at night leads to many awe-inspiring insights. And if you ever have the chance to look through a telescope, do so!

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 he limit of length-to-mass ratios in nature. In other words, they produce he highest force value possible in nature at their surface, the black hole event horizon. Black holes also produce the highest space-time curvature values. 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. 128 *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 Ref. 239 there is a black hole at the centre of almost all galaxies, including our own. Black holes are also suspected at the heart of quasars, of *active galatic 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, the most impressive, the most powerful and Ref. 240 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. 241 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, it would appear to be completely black. The object would not allow any light to leave it; in addition, it would block all light

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

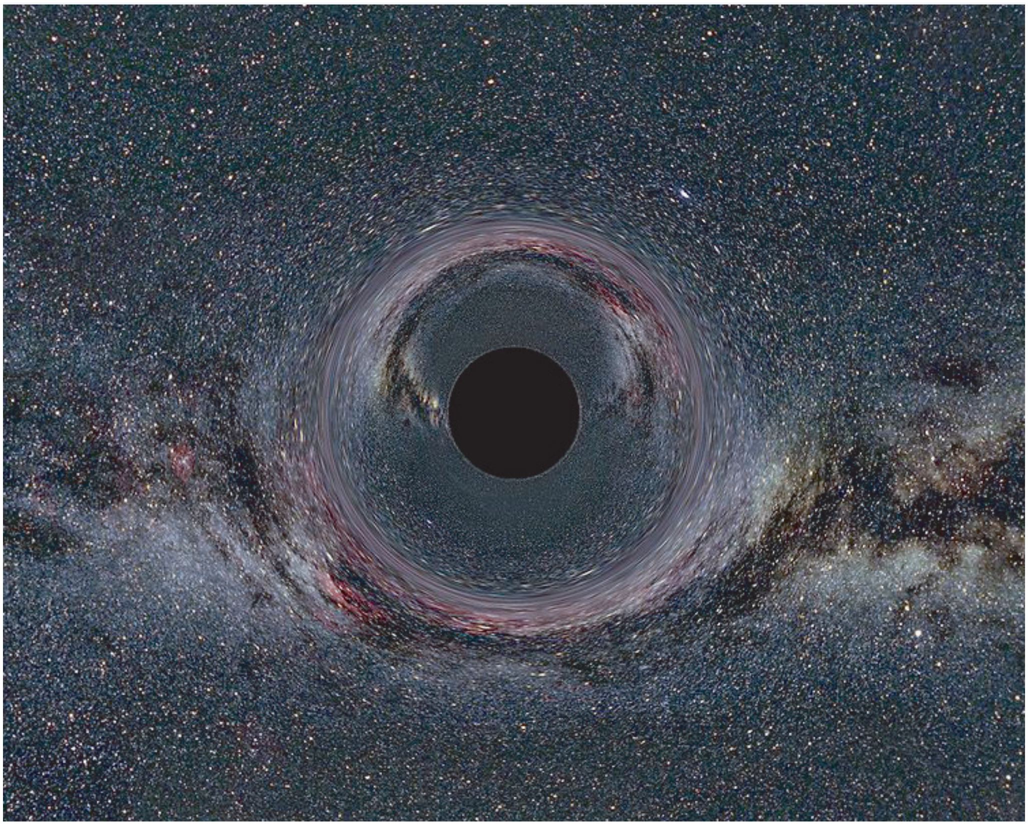


FIGURE 104 A simplified simulated image of how a black hole of ten solar masses, with Schwarzschild radius of 30 km, seen from a constant distance of 600 km, will distort an image of the Milky Way in the background. Note the Zwicky–Einstein ring formed at around twice the black hole radius and the thin bright rim. (Image © Ute Kraus at www.tempolimit-lichtgeschwindigkeit.de.)

Ref. 128 coming from behind it. In 1967, John Wheeler* made the now standard term *black hole*, due to Anne Ewing, popular in physics.

Challenge 359 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

$$R_S = \frac{2GM}{c^2} \tag{267}$$

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 $2\pi$. 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 addition, the

* John Archibald Wheeler (1911–), 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.

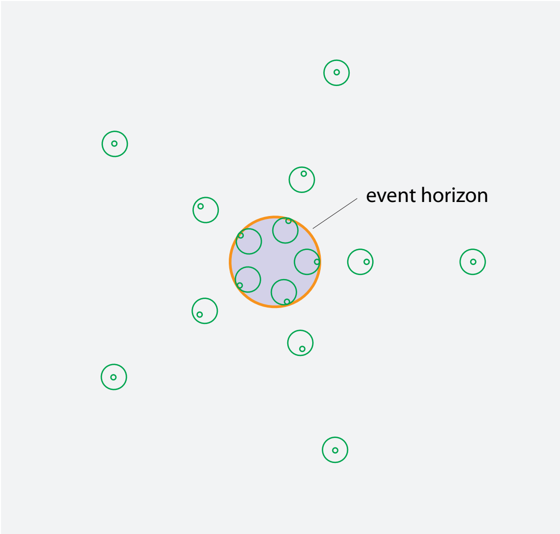



FIGURE 105 The light cones in the equatorial plane around a non-rotating black hole, seen from above

term 'size' has to be taken with some grain of salt.) 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. Nevertheless, 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 shring and to approache the critical radius $R_S$, two things happen. First, the local proper acceleration for (imaginary) point masses increases 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 360 ny

$$g_{\text{surf}} = \frac{GM}{R_S^2} = \frac{c^4}{4GM} = \frac{c^2}{2R_S} . \qquad (268)$$

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 example: What happens when a light beam is sent upwards from the horizon? And from slightly above the horizon?

Challenge 361 ny

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 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, Ref. 242 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.

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.

## Black hole horizons as limit surfaces

The characterizing property of a black hole is thus its *horizon*. The first time we encoun- Page 86 tered 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.

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 de- Page 131 scribed by

$$\mathrm{d}i^2 = \left(1 - \frac{2GM}{rc^2}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - \frac{2GM}{rc^2}} - r^2\mathrm{d}\varphi^2/c^2 . \tag{269}$$

This is the so-called *Schwarzschild metric*. As mentioned above, $r$ is the circumference divided by $2\pi$; $t$ is the time measured at infinity. No *outside* observer will ever receive any signal emitted from a radius value $r = 2GM/c^2$ or smaller. Indeed, as the proper time $i$ of an observer at radius $r$ is related to the time $t$ of an observer at infinity through

$$\mathrm{d}i = \sqrt{1 - \frac{2GM}{rc^2}}\,\mathrm{d}t , \tag{270}$$

we find that an observer at the horizon would have vanishing proper time. In other words,

* 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.

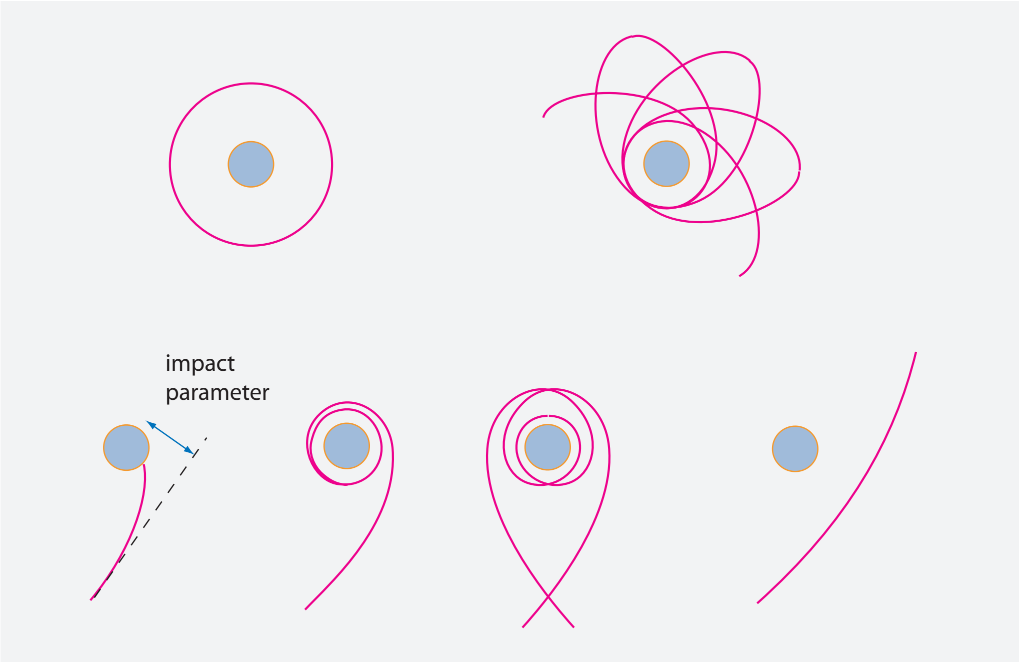



FIGURE 106 Motions of *massive* objects around a non-rotating black hole – for different impact parameters and initial velocities

at the horizon the red-shift is infinite. (In fact, 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.

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 245

## Orbits around black holes

Ref. 237 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.*

Challenge 363 ny 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 106. Such a path shows the famous periastron shift in all its glory.

Challenge 364 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 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 106. 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 365 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 106; 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 107. 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

* For such paths, Kepler's rule connecting the average distance and the time of orbit

$$\frac{GMt^3}{(2\pi)^2} = r^3 \qquad (271)$$

Challenge 362 ny still holds, provided the proper time and the radius measured by a distant observer are used.

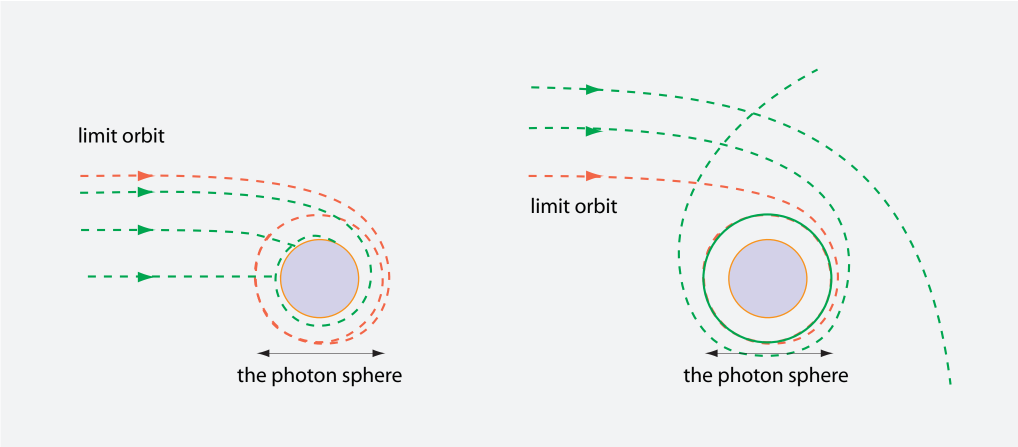



FIGURE 107 Motions of *light* passing near a non-rotating black hole

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! 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. Are there any rosetta paths for light around a black hole?

Challenge 366 ny

Challenge 367 ny

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 368 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

* 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 mountain ascent.

Vol. VI, page 146

** 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. 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. 243

Ref. 244

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.* 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. (However, the uniqueness theorem is not valid any more if the black hole carries nuclear quantum numbers, such as weak or strong charges.) Ref. 245 Ref. 246

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 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 369 ny

$$\left(\frac{J}{cM}\right)^2 + \frac{GQ^2}{4\pi\varepsilon_0 c^4} \leqslant \left(\frac{GM}{c^2}\right)^2 . \tag{272}$$

Challenge 370 ny This follows from the limit on length-to-mass ratios at the basis of general relativity. Rotating black holes realizing the limit (272) are called *extremal* black holes. The limit (272) implies that the horizon radius of a general black hole is given by

$$r_\mathrm{h} = \frac{GM}{c^2}\left(1 + \sqrt{1 - \frac{J^2 c^2}{M^4 G^2} - \frac{Q^2}{4\pi\varepsilon_0 G M^2}}\right) \tag{273}$$

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. Because of their rotation, their surface of infinite gravity or infinite red-shift, called the *static limit*, is different from their (outer) horizon. 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. Ref. 247

Ref. 128 * Wheeler claims that he was inspired by the difficulty of distinguishing between bald men; however, 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.'

Page 47 ** More about the still hypothetical magnetic charge later on. In black holes, it enters like an additional type of charge into all expressions in which electric charge appears.

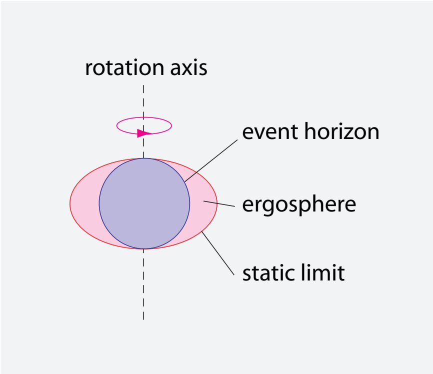



FIGURE 108 The ergosphere of a rotating black hole

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 371 e

$$A = \frac{16\pi G^2}{c^4} M^2 . \tag{274}$$

The relation between surface area and mass for a rotating and charged black hole is more complex: it is given by

$$A = \frac{8\pi G^2}{c^4} M^2 \left( 1 + \sqrt{1 - \frac{J^2 c^2}{M^4 G^2} - \frac{Q^2}{4\pi\varepsilon_0 G M^2}} \right) \tag{275}$$

where $J$ is the angular momentum and $Q$ the charge. In fact, the relation

$$A = \frac{8\pi G}{c^2} M r_{\mathrm{h}} \tag{276}$$

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. 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

Ref. 248

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
Challenge 373 ny process. For black holes rotating more slowly, the percentage is obviously smaller.

For *charged* black holes, such irreversible energy extraction processes are also possible.
Challenge 374 ny Can you think of a way? Using expression (272), we find that up to 50 % of the mass of
Challenge 375 ny a non-rotating black hole can be due to its charge. In fact, in the quantum part of our mountain ascent we will encounter an energy extraction process which nature seems to
Page 109 use quite frequently.

The Penrose process allows one to determine how angular momentum and charge
Ref. 249 increase the mass of a black hole. The result is the famous mass–energy relation

$$M^2 = \frac{E^2}{c^4} = \left( m_{\text{irr}} + \frac{Q^2}{16\pi\varepsilon_0 G m_{\text{irr}}} \right)^2 + \frac{J^2}{4m_{\text{irr}}^2}\frac{c^2}{G^2} = \left( m_{\text{irr}} + \frac{Q^2}{8\pi\varepsilon_0 \rho_{\text{irr}}} \right)^2 + \frac{J^2}{\rho_{\text{irr}}^2}\frac{1}{c^2} \qquad (277)$$

which shows how the electrostatic and the rotational energy enter the mass of a black hole. In the expression, $m_{\text{irr}}$ is the *irreducible mass* defined as

$$m_{\text{irr}}^2 = \frac{A(M, Q=0, J=0)}{16\pi}\frac{c^4}{G^2} = \left( \rho_{\text{irr}} \frac{c^2}{2G} \right)^2 \qquad (278)$$

and $\rho_{\text{irr}}$ 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
Ref. 250 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.

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

* And it would be much more dangerous, since any small object would hit such an against-the-stream
Challenge 372 ny satellite at about 15.8 km/s, thus transforming the object into a dangerous projectile. 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.

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 |

Challenge 376 ny entropy $S$ is proportional to the horizon area. Later it was found, using quantum theory, that

$$S = \frac{A}{4}\frac{kc^3}{\hbar G} = \frac{A\,k}{4\,l_{\mathrm{Pl}}^2} \; . \qquad (279)$$

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 mountain ascent. Page 110

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. Page 104 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-Page 104 rotating, uncharged black holes. The interesting connections between black holes, thermodynamics, and quantum theory will be presented in the upcoming parts of our mountain ascent. Can you imagine other mechanisms that make black holes shine? Challenge 377 ny

## 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 Ref. 252 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. Page 107

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

Ref. 239 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. 253 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, 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. Ref. 239

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, 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 signal is being looked for at the gravitational wave detectors that are in operation around the globe. Page 158

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 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 mewthod 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 in fact have been observed recently. Ref. 254 Ref. 255

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 around 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 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.

Ref. 256

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 378 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. 257

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

Vol. VI, page 95 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 Vol. VI, page 58 reason is that arbitary small time and space values do not exist in nature.

## 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. Page 113

* *

Following universal gravity, light could climb upwards from the surface of a black hole 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 379 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? The falling person, however, reaches the horizon in a *finite* amount of her own time. Can you calculate it? Challenge 380 ny Challenge 381 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

$$\frac{L}{M} \geqslant \frac{4G}{c^2} \; . \qquad (280)$$

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 382 ny

* No translation possible.

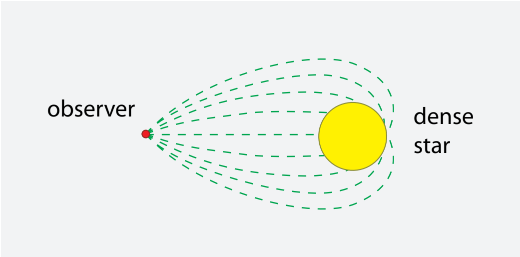



FIGURE 109 Motion of some light rays from a dense body to an observer

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 109. Can you imagine what the image looks like? Note that in addition to the paths shown in Figure 109, 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.

Challenge 383 ny

Page 142

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. 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. Only a rocket with engines switched on and pushing towards the sky can orbit a black hole at $3R/2$.

Ref. 251

Challenge 384 ny

∗ ∗

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 385 s

∗ ∗

Do *white holes* exist, i.e., time-inverted black holes, in which everything flows out of, instead of into, some bounded region?

Challenge 386 ny

∗ ∗

Challenge 387 ny

Show that a cosmological constant Λ leads to the following metric for a black hole:

$$d\tau^2 = \frac{ds^2}{c^2} = \left(1 - \frac{2GM}{rc^2} - \frac{\Lambda}{3}r^2\right)dt^2 - \frac{dr^2}{c^2 - \frac{2GM}{r} - \frac{\Lambda c^2}{3}r^2} - \frac{r^2}{c^2}d\varphi^2 \,. \qquad (281)$$

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:

$$F = -\frac{Gm}{r^2} + \frac{c^2\Lambda}{6}r \,. \qquad (282)$$

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 charged system. What is the gyromagnetic ratio for rotating black holes?

Challenge 388 ny

* *

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 104. 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 389 s

* *

Challenge 390 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.

* *

Challenge 391 ny

Are black holes made of space or of matter? Both answers are correct. Can you show this?

* *

Challenge 392 ny

Can you confirm that black holes imply a limit to power? Power is energy change over time. General relativity limits power 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.

## Summary on black holes

Black holes realize the maximum force, and realize maximum density, maximum blackness and maximum entropy for a given mass. Black holes also deflect and capture 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

$$r_0 \approx 3ct_0 \approx 4 \cdot 10^{26}\ \mathrm{m} \tag{283}$$

and its matter content is about

$$m_0 \approx \frac{4\pi}{3}\rho_o r_0^3 \quad \text{whence} \quad \frac{2Gm_0}{c^2} = 72\pi G\rho_0 c t_0^3 = 6 \cdot 10^{26}\ \mathrm{m} \tag{284}$$

for a density of $3 \cdot 10^{-27}\ \mathrm{kg/m^3}$. Thus we have

$$r_0 \approx \frac{2Gm_0}{c^2}\ , \tag{285}$$

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 393 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

$$\begin{aligned} x_0' &= x_0 + x_1 \\ x_1' &= -x_0 + x_1 \\ x_2' &= x_2 \\ x_3' &= x_3 \end{aligned} \tag{286}$$

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 394 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 invariant quantities are necessary for communication! In fact, we can *talk* to each other only be-

* 'Care about time.' Lucius Annaeus Seneca (*c.* 4 BCE–65), *Epistolae* 88, 39.

cause 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 Ref. 213 issue.

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 pace 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, Page 206, page 253 namely the one 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.

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 mountain ascent.

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 Vol. VI, page 104 space). Is this just coincidence? We will come back to this issue in the last part of our mountain ascent.

## Can space and time be measured?

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 Page 335 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.

In fact, we need *electrodynamics* to solve it. It is only be using the elementary charge Ref. 258 $e$ that we can form length scales, of which the simplest one is

$$l_{\text{emscale}} = \frac{e}{\sqrt{4\pi\varepsilon_0}} \frac{\sqrt{G}}{c^2} \approx 1.4 \cdot 10^{-36}\ \text{m} . \tag{287}$$

Page 24 Here, $\varepsilon_0$ 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

$$l_{\text{qtscale}} = \sqrt{\frac{\hbar G}{c^3}} \approx 1.6 \cdot 10^{-35}\ \text{m} , \tag{288}$$

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:

$$e = \sqrt{4\pi\varepsilon_0 c\hbar\alpha} . \tag{289}$$

Here, $\alpha \approx 1/137.06$ is the fine-structure constant that characterizes the strength of electromagnetism. In terms of $\alpha$, expression (287) becomes

$$l_{\text{scale}} = \sqrt{\frac{\alpha\hbar G}{c^3}} = \sqrt{\alpha}\ l_{\text{Pl}} . \tag{290}$$

Summing up, every length measurement is based on the electromagnetic coupling constant $\alpha$ and on the Planck length. Of course, the same is true for every time and every Challenge 396 e mass measurement. There is thus no way to define or measure lengths, times and masses using gravitation or general relativity only.*

Given this sobering result, we can ask whether in general relativity space and time are really required at all.

Ref. 259 * 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; that is not correct, Challenge 397 s however. Can you confirm this?

## Are space and time necessary?

Ref. 260 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.

He 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

$$(a_3 - v) = (a_1 - v) + (a_2 - v) \; . \tag{291}$$

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 timelike 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. The standard reference is the text by Hawking and Ellis; Ref. 219 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 observation of closed timelike curves is unlikely. Indeed, no candidate has ever been suggested. Later on, we will find that searches for such curves at the microscopic scale have also failed to find any example in nature. Page 117

The impossibility of closed timelike curves seems to point to a difference between space and time. But in fact, this difference is only apparent. All these investigations are based on the behaviour of matter. Thus these 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. Let us look at the issue in another way.

## 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*

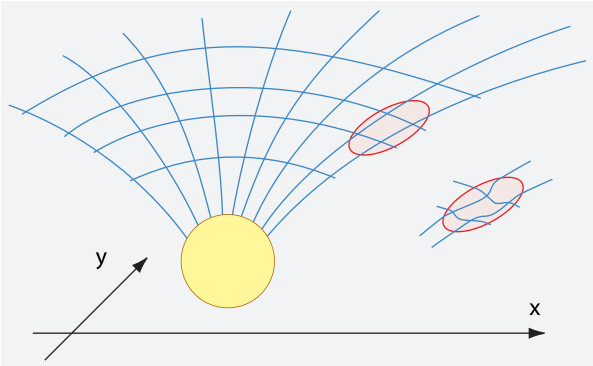



FIGURE 110 A 'hole' in space

*paradox*. Take the situation shown in Figure 110, 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. 261

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 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 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 will now discuss a startling consequence of diffeomorphism invariance.

FIGURE 111 A model of the hollow Earth theory (© Helmut Diehl)

## 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 conjecture that we live on the *inside* of a sphere, was popular in paranormal 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 111, 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. Page 55

Interestingly, the reasoning is partially correct. There is *no way* to disprove this sort of description of the universe. In fact, as the great Austrian 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 hypothesis 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?

Ref. 262
Challenge 398 e
Challenge 399 s

Such investigations show that 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. Page 104

## 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 use matter, radiation or space-time at infinity.

Vol. I, page 334 In the beginning of our mountain ascent 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 ascent we found that space and time are required to define matter; now we have found that we even need *flat* space-time to define it. In this fundamental issue, general relativity has brought no improvement over the results of Galilean physics. Vol. I, page 183

In the rest of our adventure, quantum physics will confirm that matter is needed to distinguish between space and time. No distinction is possible in principle. Still later, we will discover that mass and space-time are on an equal footing in nature. The fact that either is defined with the other implies particles and vacuum are made of the same substance. Distinctions between space and time turn out to be possible only at low, everyday energies. Vol. IV, page 141 Vol. VI, page 72

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 matter and radiation. Continuing our mountain ascent is the only way out.

Chapter 11

# GENERAL RELATIVITY IN A NUTSHELL – A SUMMARY FOR THE LAYMAN

> Sapientia felicitas.*
>
> Antiquity

General relativity is the final description of *macroscopic motion*. General relativity describes, for all observers, all macroscopic motion due to gravity, and in particular, describes how the observations of motion of *any* two observers are related to each other. General relativity is based on two observations:

— All observers agree that there is a 'perfect' speed in nature, namely a common maximum energy speed relative to (nearby) matter. This speed value 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. This force value is realized on event horizons.

These two observations contain the full theory of relativity. From these observation 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 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

* 'Wisdom is happiness.' This old saying once was the motto of Oxford University.

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 first approximation.
— Space and time are not distinguished globally, but only locally. *Matter* is 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 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 summary, the principles of maximum force and of maximum speed – and the theory of general relativity that follows from them – describe all motion due to gravity and all macroscopic motion that is observed in the universe.

## 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. 263 Ref. 264

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 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. 263, Ref. 264 Challenge 400 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. 263 Page 156

TABLE 9 Types of tests of general relativity

| Measured effect | Confirmation | Type | Reference |
| --- | --- | --- | --- |
| Equivalence principle | $10^{-12}$ | motion of matter | Ref. 141, Ref. 263, Ref. 265 |
| $1/r^2$ dependence (dimensionality of space-time) | $10^{-10}$ | motion of matter | Ref. 266 |
| Time independence of $G$ | $10^{-19}$ /s | motion of matter | Ref. 263 |
| Red-shift (light and microwaves on Sun, Earth, Sirius) | $10^{-4}$ | space-time curvature | Ref. 119, Ref. 117, Ref. 263 |
| Perihelion shift (four planets, Icarus, pulsars) | $10^{-3}$ | space-time curvature | Ref. 263 |
| Light deflection (light, radio waves around Sun, stars, galaxies) | $10^{-3}$ | space-time curvature | Ref. 263 |
| Time delay (radio signals near Sun, near pulsars) | $10^{-3}$ | space-time curvature | Ref. 263, Ref. 150 |
| Gravitomagnetism (Earth, pulsar) | $10^{-1}$ | space-time curvature | Ref. 144 |
| Geodesic effect (Moon, pulsars) | $10^{-1}$ | space-time curvature | Ref. 165, Ref. 263 |
| Gravity wave emission delay (pulsars) | $10^{-3}$ | space-time curvature | Ref. 263 |

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 gold or of oranges with a precision of a single kilogram!

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. 264, Ref. 267

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. Albert Einstein's achievement has no flaws. Ref. 263

We note that general relativity has not been tested for microscopic motion. In this context, *microscopic motion* is any motion for which the action is around the quantum of action, namely $10^{-34}$ Js. This issue is central to the last part of our adventure.

## 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 Ref. 268 theory. Research in cosmology and astrophysics, however, is at a high point at present. Here is a short overview.

* *

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.

* *

The investigation of cosmic collisions and many-body problems, especially those involving neutron stars and black holes, helps astrophysicists to improve their understanding Ref. 247 of the rich behaviour they observe in their telescopes.

* *

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 Ref. 269 open problems in physics.

* *

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 back- Ref. 270 ground anisotropies due to protogalaxies.

* *

The determination of the cosmological parameters, such as the matter density, the curva- Ref. 215 ture and the vacuum density, is a central effort of modern astrophysics. 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 Ref. 271 understood. 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 Vol. V, page 109 later on.

* *

A computer database of all solutions of the field equations is being built. Among other Ref. 272 things, researchers are checking whether they really are all different from each other.

* *

Ref. 274 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 Vol. V, page 117 are made impossible by quantum effects.

* *

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 Ref. 275 modern description.

* *

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 Ref. 273, Ref. 276 seconds after the big bang, is still an important topic of investigation.

* *

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 Vol. VI, page 17 complete description of motion. This is the topic of the final part of this adventure.

* *

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 Ref. 277 formalism.

In short, general relativity, astrophysics and astronomy are extremely interesting fields of research and important advances are expected in the near future.

## Could general relativity be different?

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 Page 93 wrong, as it was wrong for the speed of light $c$.

Since the constant of gravitation enters into our definition of gravity and acceleration, and thus, even if we do not notice it, into the construction of all rulers, all measurement standards and all measuring set-ups, there is *no way* to detect whether its value actu- Challenge 401 ny ally varies. No imaginable experiment could detect a variation. Every measurement of force is, whether we like it or not, a comparison with the limit force. There is no way, in principle, to check the invariance of a standard. This is even more astonishing because Page 261 measurements of this type are regularly reported, as in Table 9. But the result of any such experiment is easy to predict: no change will ever be found.

Could the number of space dimension 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 give deviations from the inverse square 'law' of gravitation. So far, there are no data pointing in this direction.

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. However, two candidates might exist.

First, the inclusion of torsion in the field equations, a possible extension of the theory, is one of the more promising attempts to include particle 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 Raychaudhuri equation. The use of the extended Raychaudhuri equation, which includes torsion, should allow one to deduce the full Einstein–Cartan theory from the maximum force principle. This issue is a topic of present research.

Ref. 279

Ref. 278

Secondly, one experimental result still is unexplained. The rotation speed of visible matter far from the centre of galaxies might imply either the existence of dark matter or some deviation from the inverse square dependence of universal gravity. The latter option would imply a modification in the field equations for 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. Both options might be compatible with maximum force. The issue is still open.

Ref. 280

## The limits of general relativity

Despite its successes, the description of motion presented so far is unsatisfactory; maybe you already have some gut feeling about certain unresolved issues.

Challenge 402 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. 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. In particular, 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.

Finally, we know too little about the vacuum. We need to understand the magnitude of the cosmological constant 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. Worse, the 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 Vol. VI, page 53 to scale before we reach the top of Motion Mountain. We also 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? Also this issue remains 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 Vol. V, page 51 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. At small scales, the curvature of space is negligible. We therefore take a step backwards, to situations *without* gravity, and explore the microscopic details of light, matter and vacuum. And despite the simplification to flat space-time, a lot of fun awaits us there.

> “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

Appendix A

# UNITS, MEASUREMENTS AND CONSTANTS

Measurements are comparisons with standards. Standards are based on a *unit*. Many different systems of units have been used throughout the world. Most 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 units, this happened in the eighteenth century: 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, 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
Ref. 281 Mesures' (BIPM), which all originated in the times just before the French revolution.

## SI units

All SI units are built from seven *base units*, whose official definitions, translated from French into English, are given below, together with the dates of their formulation:

- '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 *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 *kilogram* is the unit of mass; it is equal to the mass of the international prototype of the kilogram.' (1901)*
- 'The *ampere* is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 metre apart in vacuum, would produce between these conductors a force equal to $2 \cdot 10^{-7}$ newton per metre of length.' (1948)
- 'The *kelvin*, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.' (1967)*
- 'The *mole* is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12.' (1971)*
- '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)*

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 each observation and measurements.* 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 misuse by the standard-setting organization. (The kilogram, still defined with the help of an artefact, is the last exception to this requirement; extensive research is under way to eliminate this artefact from the definition – an international race that will take a few more years. There are two approaches: counting particles, or fixing $\hbar$. The former can be achieved in crystals, the latter using any formula where $\hbar$ appears, such as the formula for the de Broglie wavelength or that of the Josephson effect.)
- 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, 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:

* The respective symbols are s, m, kg, A, K, mol and cd. The international prototype of the kilogram is a platinum–iridium cylinder kept at the BIPM in Sèvres, in France. For more details on the levels of the caesium atom, consult a book on atomic physics. The Celsius scale of temperature $\theta$ is defined as: $\theta/°\mathrm{C} = T/\mathrm{K} - 273.15$; note the small difference with the number appearing in the definition of the kelvin. SI also states: 'When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles.' In the definition of the mole, it is understood that the carbon 12 atoms are unbound, at rest and in their ground state. 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.

Vol. I, page 88
Ref. 282

* Jacques Babinet (1794–1874), French physicist who published important work in optics.

| Name | Abbreviation | Name | Abbreviation |
|---|---|---|---|
| hertz | $Hz = 1/s$ | newton | $N = kg\,m/s^2$ |
| pascal | $Pa = N/m^2 = kg/m\,s^2$ | joule | $J = Nm = kg\,m^2/s^2$ |
| watt | $W = kg\,m^2/s^3$ | coulomb | $C = As$ |
| volt | $V = kg\,m^2/As^3$ | farad | $F = As/V = A^2s^4/kg\,m^2$ |
| ohm | $\Omega = V/A = kg\,m^2/A^2s^3$ | siemens | $S = 1/\Omega$ |
| weber | $Wb = Vs = kg\,m^2/As^2$ | tesla | $T = Wb/m^2 = kg/As^2 = kg/Cs$ |
| henry | $H = Vs/A = kg\,m^2/A^2s^2$ | degree Celsius | °C (see definition of kelvin) |
| lumen | $lm = cd\,sr$ | lux | $lx = lm/m^2 = cd\,sr/m^2$ |
| becquerel | $Bq = 1/s$ | gray | $Gy = J/kg = m^2/s^2$ |
| sievert | $Sv = J/kg = m^2/s^2$ | katal | $kat = mol/s$ |

We note that in all definitions of units, the kilogram only appears to the powers of 1, 0 and −1. The final explanation for this fact appeared only recently. Can you try to formulate the reason?

Challenge 403 ny

The *admitted* non-SI units are *minute*, *hour*, *day* (for time), *degree* $1° = \pi/180$ rad, *minute* $1' = \pi/10\,800$ rad, *second* $1'' = \pi/648\,000$ 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 | | Power | Name | | Power | Name | | Power | Name | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $10^1$ | deca | da | $10^{-1}$ | deci | d | $10^{18}$ | Exa | E | $10^{-18}$ | atto | a |
| $10^2$ | hecto | h | $10^{-2}$ | centi | c | $10^{21}$ | Zetta | Z | $10^{-21}$ | zepto | z |
| $10^3$ | kilo | k | $10^{-3}$ | milli | m | $10^{24}$ | Yotta | Y | $10^{-24}$ | yocto | y |
| $10^6$ | Mega | M | $10^{-6}$ | micro | µ | unofficial: | | | Ref. 283 | | |
| $10^9$ | Giga | G | $10^{-9}$ | nano | n | $10^{27}$ | Xenta | X | $10^{-27}$ | xenno | x |
| $10^{12}$ | Tera | T | $10^{-12}$ | pico | p | $10^{30}$ | Wekta | W | $10^{-30}$ | weko | w |
| $10^{15}$ | Peta | P | $10^{-15}$ | femto | f | $10^{33}$ | Vendekta | V | $10^{-33}$ | vendeko | v |
| | | | | | | $10^{36}$ | Udekta | U | $10^{-36}$ | udeko | u |

▪ SI units form a *complete* system: they cover in a systematic way the complete set of observables of physics. Moreover, they fix the units of measurement for all other sciences

* 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 404 e

as well.

▪ 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 *coherent* 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.*

Since every measurement is a comparison with a standard, 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.

Challenge 405 e

## 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.

* *

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

* 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 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.

*** 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.

constant $G$ and the strong coupling constant $\alpha_s$. Even less precisely known are the age of the universe and its density (see Table 14).

Page 275

* *

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. 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. 284
Ref. 285

* *

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 406 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 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 $\sigma$; it is defined through

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \, , \tag{292}$$

where $\bar{x}$ is the average of the measurements $x_i$. (Can you imagine why $n-1$ is used in the formula instead of $n$?)

Challenge 407 s

* 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.

Ref. 286

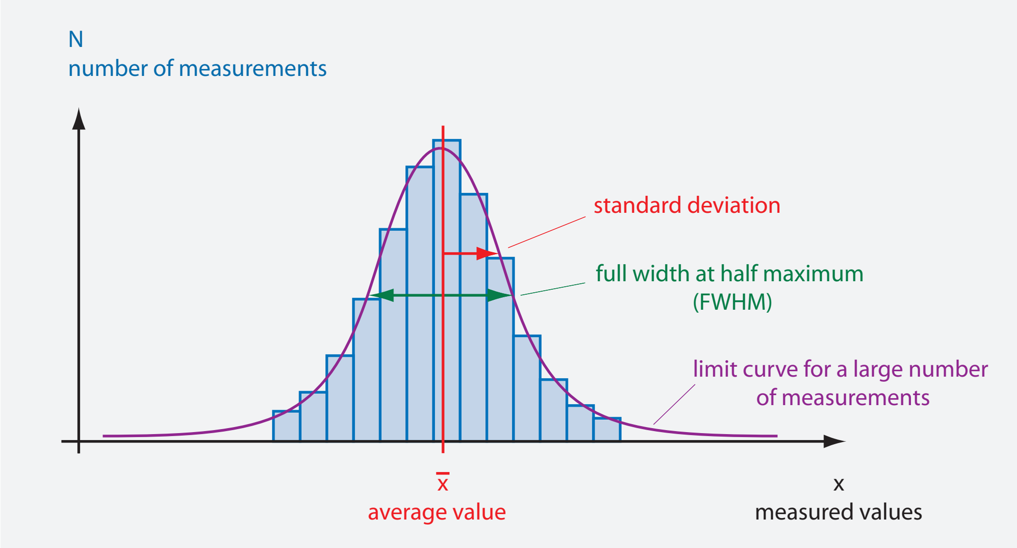



FIGURE 112 A precision experiment and its measurement distribution

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 226, is described by the expression

$$N(x) \approx \mathrm{e}^{-\frac{(x-\bar{x})^2}{2\sigma^2}} . \tag{293}$$

The square $\sigma^2$ of the standard deviation is also called the *variance*. For a Gaussian distribution of measurement values, $2.35\sigma$ is the full width at half maximum. Challenge 408 e

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*. Ref. 287

The tables below give the 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; e.g. 0.31(6) means – roughly speaking – 0.31 ± 0.06. 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. 288 Ref. 289

## 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 the ratio between the largest and the smallest measurable length is $\Delta x/x > l_{\mathrm{Pl}}/d_{\mathrm{horizon}} = 10^{-61}$. (Is this ratio valid also for force or for volume?) In the final volume of our text, studies of clocks Challenge 409 e

Vol. VI, page 87 and metre bars strengthen this theoretical limit.

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

Ref. 288 In principle, all quantitative properties of matter can be calculated with quantum theory. For example, colour, density and elastic properties can be predicted using the values of the following constants using the equations of the standard model of high-energy physics.

Vol. V, page 191

TABLE 11 Basic physical constants

| Quantity | Symbol | Value in SI units | Uncert.[a] |
|---|---|---|---|
| number of space-time dimensions | | 3 + 1 | 0[b] |
| vacuum speed of light[c] | $c$ | 299 792 458 m/s | 0 |
| vacuum permeability[c] | $\mu_0$ | $4\pi \cdot 10^{-7}$ H/m | 0 |
| | | = 1.256 637 061 435 ... μH/m | 0 |
| vacuum permittivity[c] | $\varepsilon_0 = 1/\mu_0 c^2$ | 8.854 187 817 620 ... pF/m | 0 |
| original Planck constant | $h$ | $6.626\,068\,76(52) \cdot 10^{-34}$ Js | $7.8 \cdot 10^{-8}$ |
| reduced Planck constant | $\hbar$ | $1.054\,571\,596(82) \cdot 10^{-34}$ Js | $7.8 \cdot 10^{-8}$ |
| positron charge | $e$ | 0.160 217 646 2(63) aC | $3.9 \cdot 10^{-8}$ |
| Boltzmann constant | $k$ | $1.380\,650\,3(24) \cdot 10^{-23}$ J/K | $1.7 \cdot 10^{-6}$ |
| gravitational constant | $G$ | $6.673(10) \cdot 10^{-11}$ Nm$^2$/kg$^2$ | $1.5 \cdot 10^{-3}$ |
| gravitational coupling constant | $\kappa = 8\pi G/c^4$ | $2.076(3) \cdot 10^{-43}$ s$^2$/kg m | $1.5 \cdot 10^{-3}$ |
| fine structure constant,[d] | $\alpha = \frac{e^2}{4\pi\varepsilon_0\hbar c}$ | 1/137.035 999 76(50) | $3.7 \cdot 10^{-9}$ |
| e.m. coupling constant | $= g_{em}(m_e^2 c^2)$ | = 0.007 297 352 533(27) | $3.7 \cdot 10^{-9}$ |
| Fermi coupling constant,[d] | $G_F/(\hbar c)^3$ | $1.166\,39(1) \cdot 10^{-5}$ GeV$^{-2}$ | $8.6 \cdot 10^{-6}$ |
| weak coupling constant | $\alpha_w(M_Z) = g_w^2/4\pi$ | 1/30.1(3) | $1 \cdot 10^{-2}$ |
| weak mixing angle | $\sin^2\theta_W(\overline{MS})$ | 0.231 24(24) | $1.0 \cdot 10^{-3}$ |
| weak mixing angle | $\sin^2\theta_W$ (on shell) | 0.2224(19) | $8.7 \cdot 10^{-3}$ |
| | $= 1 - (m_W/m_Z)^2$ | | |
| strong coupling constant[d] | $\alpha_s(M_Z) = g_s^2/4\pi$ | 0.118(3) | $25 \cdot 10^{-3}$ |

*a*. Uncertainty: standard deviation of measurement errors.
*b*. Only down to $10^{-19}$ m and up to $10^{26}$ m.
*c*. Defining constant.

*d*. All coupling constants depend on the 4-momentum transfer, as explained in the section on
Page 89 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. $g_{em}(Q^2 = M_W^2 c^2) \approx 1/128$. In contrast, the strong coupling constant has lover values at higher momentum transfers; e.g., $\alpha_s(34\,\text{GeV}) = 0.14(2)$.

Why do all these constants have the values they have? For any 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 $c$, $G$, $\hbar$ and $\alpha$. 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 ratios, and thus of all dimensionless constants. The quest of understanding all ratios, all dimensionless constants, including the fine structure constant $\alpha$ 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 wave resistance | $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ | 376.730 313 461 77... Ω | 0 |
| Avogadro's number | $N_A$ | $6.022\,141\,99(47) \cdot 10^{23}$ | $7.9 \cdot 10^{-8}$ |
| Rydberg constant [a] | $R_\infty = m_e c\alpha^2/2h$ | $10\,973\,731.568\,549(83)\,\text{m}^{-1}$ | $7.6 \cdot 10^{-12}$ |
| conductance quantum | $G_0 = 2e^2/h$ | 77.480 916 96(28) μS | $3.7 \cdot 10^{-9}$ |
| magnetic flux quantum | $\varphi_0 = h/2e$ | 2.067 833 636(81) pWb | $3.9 \cdot 10^{-8}$ |
| Josephson frequency ratio | $2e/h$ | 483.597 898(19) THz/V | $3.9 \cdot 10^{-8}$ |
| von Klitzing constant | $h/e^2 = \mu_0 c/2\alpha$ | 25 812.807 572(95) Ω | $3.7 \cdot 10^{-9}$ |
| Bohr magneton | $\mu_B = e\hbar/2m_e$ | 9.274 008 99(37) yJ/T | $4.0 \cdot 10^{-8}$ |
| cyclotron frequency of the electron | $f_c/B = e/2\pi m_e$ | 27.992 4925(11) GHz/T | $4.0 \cdot 10^{-8}$ |
| classical electron radius | $r_e = e^2/4\pi\varepsilon_0 m_e c^2$ | 2.817 940 285(31) fm | $1.1 \cdot 10^{-8}$ |
| Compton wavelength | $\lambda_c = h/m_e c$ | 2.426 310 215(18) pm | $7.3 \cdot 10^{-9}$ |
| of the electron | $\lambdabar_c = \hbar/m_e c = r_e/\alpha$ | 0.386 159 264 2(28) pm | $7.3 \cdot 10^{-9}$ |
| Bohr radius [a] | $a_\infty = r_e/\alpha^2$ | 52.917 720 83(19) pm | $3.7 \cdot 10^{-9}$ |
| nuclear magneton | $\mu_N = e\hbar/2m_p$ | $5.050\,783\,17(20) \cdot 10^{-27}$ J/T | $4.0 \cdot 10^{-8}$ |
| proton–electron mass ratio | $m_p/m_e$ | 1 836.152 667 5(39) | $2.1 \cdot 10^{-9}$ |
| Stefan–Boltzmann constant | $\sigma = \pi^2 k^4/60\hbar^3 c^2$ | $56.704\,00(40)\,\text{nW/m}^2\text{K}^4$ | $7.0 \cdot 10^{-6}$ |
| Wien's displacement constant | $b = \lambda_{max} T$ | 2.897 768 6(51) mmK | $1.7 \cdot 10^{-6}$ |
| bits to entropy conversion const. | | $10^{23}$ bit = 0.956 994 5(17) J/K | $1.7 \cdot 10^{-6}$ |
| 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 | $d$ | $23^h56'4.090\,53''$ |
| astronomical unit [b] | AU | 149 597 870.691(30) km |
| light year | al | 9.460 528 173 ... 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 |
| 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 - shortens line 0.55° |
| Moon's average density | $\rho_{☾}$ | $3.3\,\mathrm{Mg/m^3}$ |
| 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 |
| Sun's mass | $M_{\odot}$ | $1.988\,43(3) \cdot 10^{30}$ kg |
| Sun's gravitational length | $l_{\odot} = 2GM_{\odot}/c^2$ | 2.953 250 08 km |
| Sun's luminosity | $L_{\odot}$ | 384.6 YW |
| Solar equatorial radius | $R_{\odot}$ | 695.98(7) Mm |
| Sun's angular size | | 0.53° average; minimum on fourth of July (aphelion) 1888″, maximum on fourth of January (perihelion) 1952″ |

TABLE 13 (Continued) Astronomical constants

| Quantity | Symbol | Value |
|---|---|---|
| Sun's average density | $\rho_{\odot}$ | 1.4 $\mathrm{Mg/m^3}$ |
| 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 |
| 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 410 s
Ref. 290

*b*. Average distance Earth–Sun. The truly amazing precision of 30 m results from time averages of signals sent from Viking orbiters and Mars landers taken over a period of over twenty years.

*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 = $1^{\circ} = \pi/180$ rad, 1 (first) minute = $1' = 1^{\circ}/60$, 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 challenge, can you determine whether any property of the universe itself is listed?)
Challenge 411 s

TABLE 14 Cosmological constants

| Quantity | Symbol | Value |
|---|---|---|
| gravitational constant | $G$ | $6.672\,59(85) \cdot 10^{-11}\ \mathrm{m^3/kg\,s^2}$ |
| cosmological constant | $\Lambda$ | *c.* $1 \cdot 10^{-52}\ \mathrm{m^{-2}}$ |
| age of the universe [a] (determined from space-time, via expansion, using general relativity) | $t_0$ | $4.333(53) \cdot 10^{17}$ s = $13.73(0.17) \cdot 10^{9}$ a |
| age of the universe [a] | $t_0$ | over $3.5(4) \cdot 10^{17}$ s = $11.5(1.5) \cdot 10^{9}$ a |

TABLE 14 (Continued) Cosmological constants

| Quantity | Symbol | Value |
|---|---|---|
| (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 | $q_0 = -(\ddot{a}/a)_0/H_0^2$ | −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 $10^{26}$ m |
| number of space dimensions | | 3, for distances up to $10^{26}$ m |
| critical density | $\rho_c = 3H_0^2/8\pi G$ | $h_0^2 \cdot 1.878\,82(24) \cdot 10^{-26}\,\mathrm{kg/m^3}$ |
| of the universe | | $= 0.95(12) \cdot 10^{-26}\,\mathrm{kg/m^3}$ |
| (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 | $m_b$ | $1.67 \cdot 10^{-27}$ 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) K |
| photons in the universe | $n_\gamma$ | $10^{89}$ |
| 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 | $\sigma_8$ | 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 | $l_{Pl} = \sqrt{\hbar G/c^3}$ | $1.62 \cdot 10^{-35}$ m |
| Planck time | $t_{Pl} = \sqrt{\hbar G/c^5}$ | $5.39 \cdot 10^{-44}$ s |
| Planck mass | $m_{Pl} = \sqrt{\hbar c/G}$ | 21.8 µg |
| instants in history [a] | $t_0/t_{Pl}$ | $8.7(2.8) \cdot 10^{60}$ |
| space-time points | $N_0 = (R_0/l_{Pl})^3 \cdot$ | $10^{244\pm1}$ |
| inside the horizon [a] | $(t_0/t_{Pl})$ | |
| 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; Page 206 the fluctuations $\Delta T_0$ which led to galaxy formation are today about $16 \pm 4\,\mu\text{K} = 6(2) \cdot 10^{-6}\,T_0$.

## Useful numbers

Ref. 291

| | |
|---|---|
| $\pi$ | 3.14159 26535 89793 23846 26433 83279 50288 41971 69399 37510$_5$ |
| e | 2.71828 18284 59045 23536 02874 71352 66249 77572 47093 69995$_9$ |
| $\gamma$ | 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 9: 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 hyperboloid also looks straight from all directions, provided the positioning is correct. One thus needs 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 coherent beam of light. If the fringes are straight, the surface is planar. (How do you ensure the wave front of the light beam is planar?)

**Challenge 3**, page 16: A fraction of infinity is still infinite.

**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: To compensate for the aberration, the telescope has to be inclined *along* the direction of motion of the Earth; to compensate for parallaxis, *against* the motion.

**Challenge 6**, page 18: Otherwise the velocity sum would be larger than $c$.

**Challenge 7**, page 18: 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 ration can be determined, though not easily, as the angle at the observer is very close to a right angle as well.

**Challenge 8**, page 18: 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.

**Challenge 9**, page 18: 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 19: The 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 shutter and its neat trigger technique, see the paper by the authors. For even faster shutters, see also the discussion on page 121.

**Challenge 11**, page 20: 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 it 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.

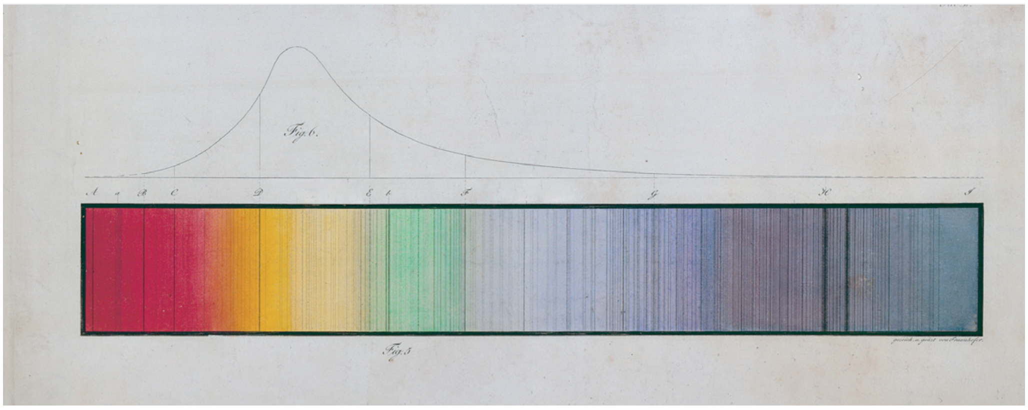


FIGURE 113 The original lines published by Fraunhofer (© Fraunhofer Gesellschaft)



**Challenge 12**, page 21: 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 21: 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
Ref. 36 a trip of 170 000 light years from a supernova explosion.

**Challenge 15**, page 26: This is best discussed by showing that other possibilities make no sense.

**Challenge 16**, page 26: 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 18**, page 28: 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 20**, page 30: $\lambda_R/\lambda_S = \gamma$.

**Challenge 21**, page 30: To change from bright red (650 nm) to green (550 nm), $v = 0.166c$ is necessary.

**Challenge 22**, page 31: People measure the shift of spectral lines, such as the shift of the so-called Lyman-$\alpha$ line of hydrogen, that is emitted (or absorbed) when a free electron is captured
Page 154 (or ejected) by a proton. It is one of the famous Fraunhofer lines.

**Challenge 23**, page 31: The speeds are given by

$$v/c = \frac{(z+1)^2 - 1}{(z+1)^2 + 1} \tag{294}$$

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 24**, page 32: 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 25**, page 32: 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 28**, page 33: Inside colour television tubes (they use higher voltages than black and white ones), electrons are described by $v/c \approx \sqrt{2 \cdot 30/511}$ or $v \approx 0.3c$.

**Challenge 29**, page 34: If you can imagine this, publish it. Readers will be delighted to hear the story.

**Challenge 31**, page 34: 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!

Page 24

**Challenge 34**, page 36: 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 to make the statement.

**Challenge 38**, page 39: The human value is achieved in particle accelerators; the value in nature is found in cosmic rays of the highest energies.

**Challenge 40**, page 40: Redrawing Figure 10 on page 26 for the other observer makes the point.

**Challenge 41**, page 40: 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 42**, page 41: 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 43**, page 42: The light cone remains unchanged; thus causal connection as well.

**Challenge 46**, page 42: 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 51**, page 45: The ratio predicted by naive reasoning is $(1/2)^{(6.4/2.2)} = 0.13$.

**Challenge 52**, page 46: 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 54**, page 46: 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 10.

**Challenge 56**, page 47: Not with present experimental methods.

**Challenge 57**, page 47: Hint: think about different directions of sight.

**Challenge 59**, page 48: Hint: be careful with the definition of 'rigidity'.

**Challenge 61**, page 48: 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 $l_{glider}$ and the gap length $l_{gap}$ obey $l_{glider}/l_{gap} < \gamma(u + v)/u$. 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 S.R. 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 62**, page 48: Yes, the rope breaks; in accelerated cars, distance changes, as shown later on in the text.

**Challenge 63**, page 49: 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 64**, page 49: A relativistic submarine would instantly melt due to friction with the water. If not, it would fly of the planet because it moves faster than the escape velocity. And produce several other disasters.

**Challenge 65**, page 51: The question confuses observation of Lorentz contraction and its measurement. A relativistic pearl necklace does get shorter, but the shortening can only be measured, not photographed. The measured sizes of the pearls are flattened ellipsoids relativistic speeds. The observed necklace consists of overlapping spheres.

**Challenge 66**, page 51: No: think about it!

**Challenge 69**, page 52: 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 70**, page 53: 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 71**, page 54: Just use plain geometry to show this.

**Challenge 72**, page 54: Most interestingly, the horizon can easily move faster than light, if you move your head appropriately, as can the end of the rainbow.

**Challenge 74**, page 58: The expression does not work for a photon hitting a mirror, for example.

**Challenge 75**, page 58: Relativity makes the arguments of challenge 167 watertight.

**Challenge 80**, page 62: The lower collision in Figure 38 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 90**, page 66: Just turn the left side of Figure 42 a bit in anti-clockwise direction.

**Challenge 91**, page 67: In collisions between relativistic charges, part of the energy is radiated away as light, so that the particles effectively lose energy.

**Challenge 92**, page 68: 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 94**, page 71: 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 95**, page 71: *No* example of motion of a massive body! The motion of light waves has null phase 4-velocity and null group 4-velocity, as explained on page 77.

**Challenge 99**, page 73: For ultrarelativistic particles, like for massless particles, one has $E = pc$.

**Challenge 100**, page 74: Hint: evaluate $P_1$ and $P_2$ in the rest frame of one particle.

**Challenge 102**, page 74: Use the definition $\boldsymbol{F} = \mathrm{d}\boldsymbol{p}/\mathrm{d}t$ and the relation $\boldsymbol{KU} = 0 = \boldsymbol{Fv} - \mathrm{d}E/\mathrm{d}t$ valid for rest-mass preserving forces.

**Challenge 104**, page 75: The story is told on page 96.

**Challenge 109**, page 76: 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 116**, page 77: The relation for the frequency follows from the definition of the phase.

**Challenge 135**, page 85: 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 137**, page 86: Constant acceleration and gravity are similar in their effects, as discussed in the section on general relativity.

**Challenge 140**, page 87: Yes, it is true.

**Challenge 141**, page 87: It is flat, like a plane.

**Challenge 142**, page 87: Despite the acceleration towards the centre of the carousel, no horizon appears.

**Challenge 144**, page 88: 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 146**, page 89: Locally, light always moves with speed $c$.

**Challenge 147**, page 89: Away from Earth, $g$ decreases; it is effectively zero over most of the distance.

**Challenge 150**, page 90: 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. 12 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. Moreover, almost all so-called 'one-way' experiments are in fact still hidden 'two-way' experiments (see his page 150). Ref. 90

Ref. 91 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 151**, page 91: As shown in the cited reference, the limit follows from the condition $l\gamma^3 a \leqslant c^2$.

**Challenge 153**, page 92: Yes.

**Challenge 154**, page 92: Yes. Take $\Delta f\ \Delta t \geqslant 1$ and substitute $\Delta l = c/\Delta f$ and $\Delta a = c/\Delta t$.

**Challenge 156**, page 93: 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 157**, page 95: 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 158**, page 100: If you hear about a claim to surpass the force or power limit, let me know.

**Challenge 159**, page 100: Take a surface moving with the speed of light, or a surface defined with a precision smaller than the Planck length.

**Challenge 160**, page 105: Also shadows do not remain parallel in 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 161**, page 107: If you find one, publish it and then send it to me.

**Challenge 163**, page 112: 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 164**, page 112: If so, publish it; then send it to me.

**Challenge 165**, page 114: 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 167**, page 115: Many do not believe the limits yet; so any proposed counter-example or any additional paradox is worth a publication.

**Challenge 169**, page 119: If so, publish it; then send it to me.

**Challenge 172**, page 121: If so, publish it; then send it to me.

**Challenge 174**, page 123: They are accelerated upwards.

**Challenge 175**, page 123: 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 179**, page 124: 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 180**, page 124: Yes, light absorption and emission are always lossless conversions of energy into mass.

**Challenge 183**, page 125: 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 187**, page 126: Length is time times the speed of light. If time changes with height, so do lengths.

**Challenge 189**, page 127: 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 192**, page 129: Such a graph would need four or even 5 dimensions.

**Challenge 194**, page 131: The energy due to the rotation can be neglected compared with all other energies in the problem.

**Challenge 204**, page 136: Different nucleons, different nuclei, different atoms and different molecules have different percentages of binding energies relative to the total mass.

**Challenge 206**, page 137: In free fall, the bottle and the water remain at rest with respect to each other.

**Challenge 207**, page 137: 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 208**, page 137: Apart the chairs and tables already mentioned, important anti-gravity devices are suspenders, belts and plastic bags.

**Challenge 210**, page 138: The same amount.

**Challenge 211**, page 138: Yes, in gravity the higher twin ages more. The age difference changes with height, and reaches zero for infinite height.

**Challenge 212**, page 138: The mass flow limit is $c^3/4G$.

**Challenge 213**, page 138: No, the conveyer belt can be built into the train.

**Challenge 214**, page 138: 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.)

**Challenge 215**, page 139: The apple hits the wall after about half an hour.

**Challenge 219**, page 140: With $\hbar$ as smallest angular momentum one get about 100 Tm.

**Challenge 218**, page 140: Approaches with curved light paths, or with varying speed of light do not describe horizons properly.

**Challenge 220**, page 140: 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 222**, page 141: The orbital radius is 4.2 Earth radii; that makes *c.* 38 μs every day.

**Challenge 223**, page 141: 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 263.

**Challenge 224**, page 141: 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 234**, page 149: 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 247**, page 158: One needs to measure the timing of pulses which cross the Earth at different gravitational wave detectors on Earth.

**Challenge 267**, page 168: No; a line cannot have intrinsic curvature. A torus is indeed intrinsically curved; it cannot be cut open to a flat sheet of paper.

**Challenge 288**, page 177: 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 292**, page 179: 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 229.

**Challenge 299**, page 185: Indeed, in general relativity gravitational energy cannot be *localized* in space, in contrast to what one expects and requires from an interaction.

**Challenge 318**, page 197: There is a good chance that some weak form of a jet exists; but a detection will not be easy.

**Challenge 320**, page 202: 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 322**, page 202: If we assme a diameter of 150 μm and a density of $1000\,\mathrm{kg/m^3}$ 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 321**, page 202: A few millimetres.

**Challenge 323**, page 203: 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 326**, page 209: The rabbit observes that all other rabbits seem to move away from him.

**Challenge 333**, page 216: 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 349**, page 230: The universe does not allow observation from outside. It thus has no state properties.

**Challenge 354**, page 232: Flattening due to rotation requires other masses to provide the background against which the rotation takes place.

**Challenge 385**, page 249: 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 389**, page 250: The description says it all. A visual impression can be found in the room on black holes in the 'Deutsches Museum' in München.

**Challenge 393**, page 251: 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.

**Challenge 397**, page 254: Any device that uses mirrors requires electrodynamics; without electrodynamics, mirrors are impossible.

**Challenge 399**, page 257: The hollow Earth theory is correct if usual distances are consistently changed according to $r_{he} = R^2_{Earth}/r$. 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 406**, page 270: Probably the quantity with the biggest variation is mass, where a prefix for 1 eV/c$^2$ would be useful, as would be one for the total mass in the universe, which is about $10^{90}$ times larger.

**Challenge 407**, page 270: The formula with $n - 1$ is a better fit. Why?

**Challenge 411**, page 275: No, only properties of parts of the universe are listed. The universe itself has no properties, as shown in the last volume..

Vol. VI, page 103

**Challenge 413**, page 313: 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 414**, page 313: There are still many discoveries to be made in modern mathematics, especially in topology, number theory and algebraic geometry. Mathematics has a good future.

# 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 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 17.

3 F. Tuinstra, Rømer and the finite speed of light, *Physics Today* 57, pp. 16–17, December 2004. Cited on page 17.

4 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.

5 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 18.

6 Aristarchos, 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 18.

7 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 website on reconstrutions of historical science experiments at www.uni-oldenburg.de/histodid/forschung/nachbauten. Cited on page 18.

8 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 19 is taken from it. Cited on page 19.


* Samuel Johnson (1709–1784), famous English poet and intellectual.

**9** 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 20.

**10** 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 source?, *Physics Letters* **39**, pp. 1051–1054, Errata 1236, 1977. Cited on page 21.

**11** 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 K. Brecher, *Bulletin of the American Physical Society* **45**, 2000. He assumed that both sides of the burster emit light. The large speed difference and the pulse sharpness then yield this result. Measuring the light speed from rapidly moving stars is another way; see the previous reference. Some of these experiments are not completely watertight, however. There is a competing theory of electrodynamics, due to Ritz, which 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 21.

**12** An overview of experimental results is given in Yuan Zhong Zhang, *Special Relativity and its Experimental Foundations*, World Scientific, 1998. Cited on pages 21, 28, 35, 46, 60, 282, and 291.

**13** See e.g. C. Will, *Theory and Experiment in Gravitational Physics*, Revised edition, Cambridge University Press, 1993. Cited on pages 21 and 25.

**14** B. E. Schaefer, Severe limits on variations of the speed of light with frequency, *Physical Review Letters* **82**, pp. 4964–4966, 21 June 1999. Cited on page 21.

**15** 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. Cited on pages 21, 24, and 68.

**16** JEAN VAN BLADEL, *Relativity and Engineering*, Springer, 1984. Cited on page 22.

**17** ALBERT EINSTEIN, *Mein Weltbild*, edited by CARL SELIG, Ullstein Verlag, 1998. Cited on page 23.

**18** ALBRECHT FÖLSING, *Albert Einstein – eine Biographie*, Suhrkamp p. 237, 1993. Cited on pages 24 and 36.

**19** 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 24 and 77.

**20** WOLFGANG RINDLER, *Relativity – Special, General and Cosmological*, Oxford University Press, 2001. A beautiful book by one of the masters of the field. Cited on pages 24 and 76.

**21** 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, or arxiv.org/abs/gr-qc/0504109. 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. Cited on page 25.

**22** The slowness of the speed of light inside stars is due to the frequent scattering of photons by the star matter. The most common estimate for the Sun is an escape time of 40 000 to 1 million years, but estimates between 17 000 years and 50 million years can be found in the literature. Cited on page 25.

**23** 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 25.

**24** 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 26.

**25** 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 27.

**26** ROD S. LAKES, Experimental limits on the photon mass and cosmic vector potential, *Physical Review Letters* **80**, pp. 1826–1829, 1998. The speed of light is independent of frequency within a factor of $6 \cdot 10^{-21}$, as was shown from gamma ray studies by B. E. SCHAEFER, Severe limits on variations of the speed of light with frequency, *Physical Review Letters* **82**, pp. 4964–4966, 1999. Cited on page 28.

**27** F. Tuinstra, De lotgevallen van het dopplereffect, *Nederlands tijdschrift voor natuurkunde* 75, p. 296, August 2009. Cited on page 30.

**28** 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 30.

**29** R. Lambourne, The Doppler effect in astronomy, *Physics Education* **32**, pp. 34–40, 1997, Cited on page 31.

**30** 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 32.

**31** 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 32.

**32** 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 33.

**33** 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 33.

**34** 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 33.

**35** 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 33.

**36** 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 34 and 279.

**37** 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 35.

**38** 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 35.

**39** Max Born, *Die Relativitätstheorie Einsteins*, Springer, 2003, a commented reedition of the original text of 1920. Cited on page 36.

**40** 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 36.

**41** 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 36.

**42** 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 39.

**43** 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 40.

**44** 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 40.

**45** 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 43.

**46** 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 44.

**47** 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. 12. Cited on page 44.

**48** 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 44.

**49** 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 $\mu$-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 45.

**50** Paul J. Nahin, *Time Machines – Time Travel in Physics, Metaphysics and Science Fiction*, Springer Verlag and AIP Press, second edition, 1999. Cited on page 45.

**51** 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 45.

**52** 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 46.

**53** Search for 'fuel' and 'relativistic rocket' on the internet. Cited on page 46.

**54** A. Harvey & E. Schucking, A small puzzle from 1905, *Physics Today*, pp. 34–36, March 2005. Cited on page 46.

**55** 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 48.

**56** van Lintel & C. Gruber, The rod and hole paradox re-examined, *European Journal of Physics* **26**, pp. 19–23, 2005. Cited on page 48.

**57** 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 48.

**58** S. P. Boughn, The case of the identically accelerated twins, *American Journal of Physics* **57**, pp. 791–793, 1989. Cited on pages 48 and 52.

**59** 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 48.

**60** 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 49.

**61** G. R. Rybicki, Speed limit on walking, *American Journal of Physics* **59**, pp. 368–369, 1991. Cited on page 53.

**62** 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 Edwin F. Taylor & John A. Wheeler, *Spacetime Physics – Introduction to Special Relativity*, second edition, Freeman, 1992, pages 89-92. This excellent book was mentioned already in the text. Cited on page 55.

**63** 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 E. Recami, editor, *Tachyons, Monopoles and Related Topics*, North-Holland, Amsterdam, 1978. Cited on page 56.

**64** 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 56.

**65** Planck wrote this in a letter in 1908. Cited on page 57.

**66** 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 57.

**67** 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 58.

**68** 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 63.

**69** 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 64.

**70** 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. 87. Cited on page 65.

**71** 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 66.

**72** Umberto Bartocci, *Albert Einstein e Olinto De Pretto: la vera storia della formula più famosa del mondo*, Ultreja, 1998. Cited on page 68.

**73** The references preceding Einstein's $E = mc^2$ are: 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 68.

**74** A jewel among the textbooks on special relativity is the booklet by Ulrich E. Schröder, *Spezielle Relativitätstheorie*, Verlag Harri Deutsch, 1981. Cited on pages 71 and 74.

**75** 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 74.

**76** Christian Møller, *The Theory of Relativity*, Clarendon Press, 1952, 1972. This standard textbook has been translated in several languages. Cited on page 74.

**77** 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 75.

**78** 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 76.

**79** R. J. Low, When moving clocks run fast, *European Journal of Physics* **16**, pp. 228–229, 1995. Cited on pages 81 and 82.

**80** 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. Cited on page 72.

**81** 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 page 83.

**82** E. A. Desloge & R. J. Philpott, Uniformly accelerated reference frames in special relativity, *American Journal of Physics* **55**, pp. 252–261, 1987. Cited on pages 83 and 84.

**83** R. H. Good, Uniformly accelerated reference frame and twin paradox, *American Journal of Physics* **50**, pp. 232–238, 1982. Cited on pages 84, 85, and 88.

**84** J. D. Hamilton, The uniformly accelerated reference frame, *American Journal of Physics* **46**, pp. 83–89, 1978. Cited on page 85.

**85** The best and cheapest mathematical formula collection remains the one by K. Rottmann, *Mathematische Formelsammlung*, BI Hochschultaschenbücher, 1960. Cited on page 85.

**86** 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 86.

**87** 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 86 and 293.

**88** Edward A. Desloge, The gravitational red-shift in a uniform field, *American Journal of Physics* 58, pp. 856–858, 1990. Cited on page 88.

**89** L. Mishra, The relativistic acceleration addition theorem, *Classical and Quantum Gravity* **11**, pp. L97–L102, 1994. Cited on page 89.

**90** 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 90 and 282.

**91** H. C. Ohanian, The role of dynamics in the synchronization problem, *American Journal of Physics* 72, pp. 141–148, 2004. Cited on pages 90 and 282.

**92** 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 91.

**93** Clear statements against such a change are made by Michael Duff in several of his publications. See, for example, M. J. Duff, Comment on time-variation of fundamental constants, arxiv.org/abs/hep-th/0208093. Cited on page 93.

**94** 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 96.

**95** It seems that the first published statement of the principle was in the year 2000 edition of this text, in the chapter on gravitation and relativity. The present author discovered the maximum force principle in 1998, when searching for a way to derive the results of chapter 3 that would be so simple that it would convince even a secondary-school student. The reference is Christoph Schiller, *Motion Mountain – The Adventure of Physics*, found at www.motionmountain.net. The idea of a maximum force was also proposed by Gary Gibbons in 2002 (see reference below). Nowadays Gary Gibbons is more cautious than me about whether the maximum force can be seen as an actual physical principle (despite the title of his paper). The approach of a maximum force was discussed in various usenet discussion groups in the early twenty-first century. These discussion showed that the idea of a maximum force (and a maximum power) were known to some people, but that before Gibbons and me only few had put it in writing. Also this physics discovery was thus made much too late. In short, only the idea to raise maximum force or power to a *principle* seems to be original; it was published first in the reference following this one and then in 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. Cited on page 96.

**96** G. W. Gibbons, The maximum tension principle in general relativity, *Foundations of Physics* 32, pp. 1891–1901, 2002, or arxiv.org/abs/hep-th/0210109. Gary Gibbons explains that the maximum force follows from general relativity; he does not make a statement about the converse. See also L. Kostro & B. Lange, Is $c^4/G$ the greatest possible force in nature?, *Physics Essays* 12, pp. 182–189, 1999. See also C. Massa, Does the gravitational constant increase?, *Astrophysics and Space Science* 232, pp. 143–148, 1995. Cited on pages 96 and 133.

**97** C. Schiller, Maximum force and minimum distance: physics in limit statements, part of this text and downloadable at www.motionmountain.net/MotionMountain-Part6.pdf, preprint at arxiv.org/abs/physics/0309118. Cited on pages 98, 101, 110, and 119.

**98** H. C. Ohanian & R. 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 100, 109, 111, 113, 117, 118, and 303.

**99** 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 102.

**100** 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 102.

**101** 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 103.

**102** 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 106 and 190.

**103** Edwin F. Taylor & John A. Wheeler, *Spacetime Physics – Introduction to Special Relativity*, second edition, Freeman, 1992. Cited on page 107.

**104** This counter-example was suggested by Steve Carlip. Cited on page 109.

**105** E. R. Caianiello, *Lettere al Nuovo Cimento* **41**, p. 370, 1984. Cited on page 111.

**106** 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 115.

**107** R. Penrose, Naked singularities, *Annals of the New York Academy of Sciences* **224**, pp. 125–134, 1973. Cited on page 117.

**108** G. Huisken & T. Ilmanen, The Riemannian Penrose inequality, *International Mathematics Research Notices* **59**, pp. 1045–1058, 1997. Cited on page 117.

**109** 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 117.

**110** C. 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 page 118.

**111** 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 119.

**112** 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 HANS 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 122, 160, 162, 183, and 184.

**113** 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 Hans 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 122, 156, 157, 160, 162, 184, and 301.

**114** 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 123.

**115** This story is told e.g. by W. G. Unruh, Time, gravity, and quantum mechanics, preprint available at arxiv.org/abs/gr-qc/9312027. Cited on page 123.

**116** H. Bondi, Gravitation, *European Journal of Physics* **14**, pp. 1–6, 1993. Cited on page 124.

**117** 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 125 and 261.

**118** 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 125.

**119** 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 126 and 261.

**120** J. C. Hafele & Richard 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. Cited on page 126.

**121** 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 126.

**122** 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 126.

**123** More information about tides can be found in E. P. Clancy, *The Tides*, Doubleday, New York, 1969. Cited on page 127.

**124** 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 seen as a gross violation of scientific ethics. 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 129.

**125** 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 shiftswere 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 129.

**126** 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 129.

**127** 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 131.

**128** 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. An interesting comment by a physicist who used to write his papers in topless bars. Cited on pages 131, 235, 236, and 242.

**129** 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 132.

**130** A good and accessible book on the topic is David Blair & Geoff McNamara, *Ripples on a cosmic sea*, Allen & Unwin, 1997. Cited on page 132.

**131** 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 136.

**132** 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. 136. Cited on page 136.

**133** J. Soldner, *Berliner Astronomisches Jahrbuch auf das Jahr 1804*, 1801, p. 161. Cited on page 136.

**134** 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 139.

**135** 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 140.

**136** An overview is given in C. Will, *Theory and Experiment in Gravitational Physics*, chapter 14.3, Cambridge University Press, revised edition, 1993. (Despite being a standard reference, his view the role of tides and the role of gravitational energy within the principle of equivalence has been criticised by other researchers.) See also C. 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 141, 160, and 299.

**137** 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 141.

**138** A. G. Lindh, Did Popper solve Hume's problem?, *Nature* **366**, pp. 105–106, 11 November 1993, Cited on page 141.

**139** 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. Some beautiful graphics at the research.physics.uiuc.edu/CTA/movies/spm website show the models of this star system. Cited on page 142.

**140** 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 142.

**141** 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 142 and 261.

**142** Nigel Calder, *Einstein's Universe*, Viking, 1979. Weizmann and Einstein once crossed the Atlantic on the same ship. Cited on page 145.

**143** 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. See also Ref. 164. Cited on page 146.

**144** The feat used the LAGEOS and LAGEOS II satellites and is told in Ignazio 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. Cited on pages 147, 150, and 261.

**145** 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 147.

**146** 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 147.

**147** 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 148.

**148** A. Tartaglia & M. L. Ruggiero, Gravito-electromagnetism versus electromagnetism, *European Journal of Physics* **25**, pp. 203–210, 2004. Cited on page 148.

**149** 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 148 and 156.

**150** M. Kramer & al., Tests of general relativity from timing the double pulsar, preprint at arxiv.org/abs/astro-ph/0609417. Cited on pages 150 and 261.

**151** This is told in John A. Wheeler, *A Journey into Gravity and Spacetime*, W.H. Freeman, 1990. Cited on page 151.

**152** 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 152.

**153** 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 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 pages 153 and 158.

**154** The quadrupole formula is explained clearly in the text by Goenner. See Ref. 113. Cited on page 155.

**155** 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 153.

**156** The beautiful summary by Daniel Kleppner, The gem of general relativity, *Physics To-*

*day* **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 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 156.

**157** 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 156.

**158** 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 157.

**159** 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 159.

**160** A. Einstein, Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes, *Annalen der Physik* **35**, p. 898, 1911. Cited on page 160.

**161** I.I. Shapiro & al., Fourth test of general relativity, *Physical Review Letters* **13**, pp. 789–792, 1964. Cited on page 161.

**162** I.I. Shapiro & al., Fourth test of general relativity: preliminary results, *Physical Review Letters* **20**, pp. 1265–1269, 1968. Cited on page 161.

**163** J. H. Taylor, Pulsar timing and relativistic gravity, *Proceedings of the Royal Society, London A* **341**, pp. 117–134, 1992. Cited on pages 161 and 163.

**164** 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 pages 164 and 300.

**165** 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 164 and 261.

**166** Wolfgang Rindler, *Essential Relativity*, Springer, revised second edition, 1977. Cited on page 168.

**167** 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 Mannigfalktigkeit, *Göttinger Nachrichten, mathematische–physikalische Klasse* p. 334, 1917. Cited on page 169.

**168** 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 173 and 175.

**169** MICHAEL H. SOFFEL, *Relativity in Astronomy, Celestial Mechanics and Geodesy*, Springer Verlag, 1989. Cited on page 173.

**170** RICHARD P. FEYNMAN, FERNANDO B. MORINIGO, WILLIAM G. WAGNER & BRIAN HATFIELD, *Feynman Lectures on Gravitation*, Westview Press, 1995. Cited on page 174.

**171** J. C. BAEZ & E. F. BUNN, The meaning of Einstein's equation, *American Journal of Physics* 73, pp. 644–652, 2005. Cited on page 176.

**172** 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 183.

**173** H. NICOLAI, Gravitational billiards, dualities and hidden symmetries, arxiv.org//abs/gr-qc/0506031. Cited on page 183.

**174** 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 180.

**175** 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 180.

**176** 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 182.

**177** See for example R.A. KNOP & al., New constraints on $\Omega_M$, $\Omega_\Lambda$, and $w$ from an independent set of eleven high-redshift supernovae observed with HST, *Astrophysical Journal* **598**, pp. 102–137, 2003. Cited on page 184.

**178** RICHARD P. FEYNMAN, ROBERT B. LEIGHTON & MATTHEW SANDS, *The Feynman Lectures on Physics*, Addison Wesley, 1977, volume II, p. 42–14. No citations.

**179** 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 186.

**180** 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 184.

**181** 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 185.

**182** 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 190.

**183** The equivalence of the various definitions of the Riemann tensor is explained in most texts on general relativity; see Ref. 98. Cited on page 188.

**184** K. Tangen, Can the Pioneer anomaly have a gravitational origin?, arxiv.org/abs/gr-qc/0602089. Cited on page 189.

**185** H. Dittus & C. Lämmerzahl, Die Pioneer-Anomalie, *Physik Journal* 5, pp. 25–31, January 2006. Cited on page 189.

**186** A. Mellinger, A color all-sky panorama of the Milky Way, preprint at arxiv.org/abs/0908.4360. Cited on page 192.

**187** This famous quote is the first sentence of the final chapter, the 'Beschluß', of Immanuel Kant, *Kritik der praktischen Vernunft*, 1797. Cited on page 191.

**188** 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 191.

**189** Aetius, Opinions, III, I, 6. See Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, 1991, p. 445. Cited on page 191.

**190** 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 194 and 198.

**191** A beautiful introduction to modern astronomy was Paolo Maffei, *I mostri del cielo*, Mondadori Editore, 1976. Cited on page 196.

**192** 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 197.

**193** 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 198.

**194** D. Figer, An upper limit to the masses of stars, *Nature* **434**, pp. 192–194, 2005. Cited on page 198.

**195** G. Basri, The discovery of brown dwarfs, *Scientific American* **282**, pp. 77–83, April 2001. Cited on page 198.

**196** P. M. Woods & C. Thompson, Soft gamma repeaters and anomalous X-ray pulsars: magnetar candidates, arxiv.org/abs/astro-ph/0406133. Cited on page 199.

**197** 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 199.

**198** 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 203.

**199** 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 203.

**200** G. Gamow, The origin of the elements and the separation of galaxies, *Physical Review* **74**, p. 505, 1948. Cited on page 204.

**201** 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 204.

**202** 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 204.

**203** See for example, D. Prialnik, *An Introduction to the Theory of Stellar Structure and Evolution*, Cambridge University Press, 2000. Cited on page 206.

**204** Macrobius, *Somnium Scipionis*, XIV, 19. See Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, 1991, p. 61. Cited on page 206.

**205** 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 206.

**206** 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 207.

**207** Gabriele Walker, *Snowball Earth – The Story of the Great Global Catastrophe That Spawned Life as We Know It*, Crown Publishing, 2003. Cited on page 208.

**208** 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 208.

**209** 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 210.

**210** H. Knutsen, Darkness at night, *European Journal of Physics* **18**, pp. 295–302, 1997. Cited on pages 215 and 216.

**211** See for example P.D. Peşić, Brightness at night, *American Journal of Physics* **66**, pp. 1013–1015, 1998. Cited on pages 216 and 217.

**212** Paul Wesson, Olbers' paradox and the spectral intensity of extra-galactic background light, *Astrophysical Journal* **367**, p. 399, 1991. Cited on page 216.

**213** 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 mountain ascent. Cited on pages 216 and 253.

**214** Supernova searches are being performed by many research groups at the largest optical and X-ray telescopes. Cited on page 217.

**215** 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 217 and 262.

**216** A. Harvey & E. Schucking, Einstein's mistake and the cosmological contant, *American Journal of Physics* **68**, pp. 723–727, 2000. Cited on page 218.

**217** 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 218.

**218** Up to his death, Fred Hoyle defended his belief that the universe is static, e.g. in G. Burbidge, F. Hoyle & J. V. Narlikar, A different approach to cosmology, *Physics Today* **52**, pp. 38–44, 1999. This team has also written a book with the same title, published in 2000 by Cambridge University Press. Cited on pages 219 and 220.

**219** 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 220, 255, and 309.

**220** Augustine, *Confessions*, 398, writes: '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. [...] 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 221.

**221** 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 221.

**222** 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 223.

**223** 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 225.

**224** 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 225.

**225** 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 227.

**226** Thanks to Steve Carlip for clarifying this point. Cited on page 227.

**227** G. F. R. Ellis & T. Rothman, Lost horizons, *American Journal of Physics* **61**, pp. 883–893, 1993. Cited on page 228.

**228** A. Guth, *Die Geburt des Kosmos aus dem Nichts – Die Theorie des inflationären Universums*, Droemer Knaur, 1999. Cited on page 228.

**229** 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 $10^{120}$ $k$, whereas the actual value is 'estimated' to be $10^{100}$ $k$. However, other authors give $10^{84}$ $k$. In 1974, Roger Penrose also made statements about the entropy of the universe. Cited on page 229.

Challenge 412 ny

**230** 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 230.

**231** 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 231.

**232** 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 231.

**233** 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 231.

**234** 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 232.

**235** 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 233.

**236** 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 233.

**237** 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. Bronstein, Die Ausdehnung des Weltalls, *Physikalische Zeitschrift der Sowjetunion* **3**, pp. 73–82, 1933. Cited on pages 233 and 239.

**238** 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 233.

**239** 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 235 and 246.

**240** 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 235.

**241** 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 235.

**242** The beautiful paper is R. Oppenheimer & H. Snyder, On continued gravitational contraction, *Physical Review* **56**, pp. 455–459, 1939. Cited on page 238.

**243** 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 241.

**244** 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 241.

**245** 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 242.

**246** 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 242.

**247** 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 242 and 262.

**248** R. Penrose & R. M. Floyd, Extraction of rotational energy from a black hole, *Nature* **229**, pp. 177–179, 1971. Cited on page 243.

**249** 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 244.

**250** J. D. Bekenstein, Black holes and entropy, *Physical Review* **D7**, pp. 2333–2346, 1973. Cited on page 244.

**251** 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 Don 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 249.

**252** 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 245.

**253** For information about black holes formation via star collapse, see the Wikipedia article at en.wikipedia.org/wikie/Stellar_black_hole. Cited on page 246.

**254** 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 246.

**255** The first direct evidence for matter falling into a black hole was published in early 2001. Cited on page 246.

**256** 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. 219. Cited on page 247.

**257** 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 247.

**258** G. J. Stoney, On the physical units of nature, *Philosophical Magazine* **11**, pp. 381–391, 1881. Cited on page 254.

**259** 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 254.

**260** R. Geroch, Einstein algebras, *Commun. Math. Phys.* **26**, pp. 271–275, 1972. Cited on page 255.

**261** A. Macdonald, Einstein's hole argument, *American Journal of Physics* **69**, pp. 223–225, 2001. Cited on page 256.

**262** 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 257.

**263** 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 260 and 261.

**264** H. Dittus, F. Everitt, C. Lämmerzahl & G. Schäfer, Die Gravitation im Test, *Physikalische Blätter* 55, pp. 39–46, 1999. Cited on pages 260 and 261.

**265** 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 261.

**266** 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 261.

**267** For theories competing with general relativity, see for example C. M. Will, The confrontation between general relativity and experiment, *Living Reviews of Relativity* 2001, www.iews.org/lrr-2001-4. 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, 'killed' 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, *Astrophysical Journal Letters* 382, p. L101, 1991. Cited on page 261.

**268** 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 262.

**269** 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 262.

**270** The ESA satellite called 'Planck' will measure the polarization of the cosmic microwave background. Cited on page 262.

**271** 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 262.

**272** 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 262.

**273** 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 263.

**274** 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 classical general relativity, *Classical and Quantum Gravity* **17**, pp. 4187–4194, 2000, or arxiv.org/abs/gr-qc/0004029. Cited on page 263.

**275** 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 263.

**276** 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 263.

**277** 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 263.

**278** 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 264.

**279** Torsion is presented in R. T. Hammond, New fields in general relativity, *Contemporary Physics* **36**, pp. 103–114, 1995. Cited on page 264.

**280** 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 264.

**281** 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 266.

**282** 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 64. On high-precision *temperature* measurements, see page 425. Cited on page 267.

**283** The unofficial prefixes were first proposed in the 1990s by Jeff K. Aronson of University of Oxford, and might come into general usage in the future. Cited on page 268.

**284** 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 270.

**285** 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 270.

**286** 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 270.

**287** 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 271.

**288** 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 pages 271 and 272.

**289** 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 271.

**290** The details are given in the well-known astronomical reference, Kenneth Seidelmann, *Explanatory Supplement to the Astronomical Almanac*, 1992. Cited on page 275.

**291** For information about the number π, and about some other mathematical constants, the website oldweb.cecm.sfu.ca/pi/pi.html provides the most extensive information and references. It also has a link to the many other sites on the topic, including the overview at mathworld.wolfram.com/Pi.html. Simple formulae for π are

$$\pi + 3 = \sum_{n=1}^{\infty} \frac{n\,2^n}{\binom{2n}{n}} \qquad (295)$$

or the beautiful formula discovered in 1996 by Bailey, Borwein and Plouffe

$$\pi = \sum_{n=0}^{\infty} \frac{1}{16^n} \left( \frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6} \right) . \qquad (296)$$

The mentioned site also explains the newly discovered methods for calculating specific binary digits of π without having to calculate all the preceding ones. The known digits of π pass all tests of randomness, as the mathworld.wolfram.com/PiDigits.html website explains. However, this property, called *normality*, has never been proven; it is the biggest open question about π. It is possible that the theory of chaotic dynamics will lead to a solution of this puzzle in the coming years.

Another method to calculate π and other constants was discovered and published by D. V. Chudnovsky & G. V. Chudnovsky, The computation of classical constants, *Proceedings of the National Academy of Sciences (USA)* **86**, pp. 8178–8182, 1989. The Chudnowsky brothers have built a supercomputer in Gregory's apartment for about 70 000 euros, and for many years held the record for calculating the largest number of digits of π. They have battled for decades with Kanada Yasumasa, who held the record in 2000, calculated on an industrial supercomputer. However, the record number of (consecutive) digits in 2010 was calculated in 123 days on a simple desktop PC by Fabrice Bellard, using a Chudnovsky formula. Bellard calculated over 2.7 million million digits, as told on bellard.org. New formulae to calculate π are still occasionally discovered.

For the calculation of Euler's constant $\gamma$ see also D. W. DeTemple, A quicker convergence to Euler's constant, *The Mathematical Intelligencer*, pp. 468–470, May 1993.

Note that little is known about the basic properties of some numbers; for example, it is
Challenge 413 r still not known whether $\pi + e$ is a rational number or not! (It is believed that it is not.) Do
Challenge 414 s you want to become a mathematician? Cited on page 277.

# CREDITS

Many people who have kept their gift of curiosity alive have helped to make this project come true. Most of all, Saverio Pascazio has been – present or not – a constant reference for this project. Fernand Mayné, Anna Koolen, 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 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, Heinrich Neumaier and Peter Brown 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, 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.

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.

The typesetting and book design is due to the professional consulting of Ulrich Dirr. The typography was improved with the help of Johannes Küster. The design of the book and its website owe also much to the suggestions and support of my wife Britta.

Since May 2007, the electronic edition and distribution of the Motion Mountain text is generously supported by the Klaus Tschira Foundation.

## Film credits

The beautiful animation of a dice flying at relativistic speed, on page 52, 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 81, is copyright Anthony Searle and Australian National University, and courtesy of Craig Savage. It is from the splendid website at www.anu.edu.au/Physics/Savage/TEE. The equally beautiful animation of an observer accelerating between houses, on page 83, is also copyright Anthony Searle and Australian National University, and courtesy of Craig Savage. It is from the wonderful website at www.anu.edu.au/Physics/Searle.

## Image credits

The mountain photograph on the front cover is courtesy and copyright by Dave Thompson and found on his website www.daveontrek.co.uk. The photograph of the night sky on page 14 is copyright and courtesy of Anthony Ayiomamitis; it is found on his splendid website www.perseus.gr. The photograph of the reconstruction of Fizeau's experiment on page 19 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 19 is courtesy and copyright of Tom Mattick. The image of the historical Michelson experiment is courtesy and copyright of the Astrophysikalisches Institut Potsdam, the images of the modern high-precision experiment on page 37 are copyright and courtesy of Stephan Schiller. The relativistic images of the travel through the simplified Stonehenge on page 50 are copyright of Nicolai Mokros and courtesy of Norbert Dragon. The relativistic views on page 51 and 51 are courtesy and copyright of Daniel Weiskopf. The stalactite photograph on page 96 is courtesy and copyright of Richard Cindric and found on the website www.kcgrotto.org. The figures of galaxies on pages 194, 192, 195, 195, 193, 196, 196, 215, 226 and 226 are courtesy of NASA. The photo of the night sky on page 192 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 maps of the universe on page 201 and the Hertzsprung–Russell diagram on page 205 are courtesy and copyright of Richard Powell, and taken from his websites www.anzwers.org/free/universe and www.atlasoftheuniverse.com. The picture of the universe on page 193 is courtesy of Thomas Jarret, IPAC and Caltech, and is found on the spider.ipac.caltech.edu/staff/jarret/lss/index.html website. The simulated view of a black hole on page 236 is copyright and courtesy of Ute Kraus and can be found on her splendid website www.tempolimit-lichtgeschwindigkeit.de. The photograph on page 215 is couresy and copyright of Wally Pacholka and found on the wonderful website www.twanlight.org that collects pictures of the world at night. 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.terra.vzw.org. All drawings are copyright by Christoph Schiller. If you suspect that your copyright is not correctly given or obtained, this has not done on purpose; please contact me in this case.

# NAME INDEX



Page numbers in *italic* typeface refer to pages where the person is presented in more detail.

B

Bronstein

F

FASCHING

S
SHORT

# SUBJECT INDEX

Page numbers in *italic* typeface refer to pages where the keyword is defined or presented in detail. The subject index thus acts as a glossary.

F

FREE

M

MECHANICS

P

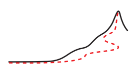
PLANCK

S

SATELLITE



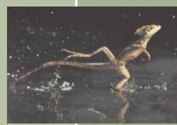
**T**



# MOTION MOUNTAIN

## The Adventure of Physics – Vol. II

### Relativity



Why do change and motion exist?
How does a rainbow form?
What is the most fantastic voyage possible?
Is ‘empty space’ really empty?
How can one levitate things?
At what distance between two points does it become
impossible to find room for a third one in between?
What does ‘quantum’ mean?
Which problems in physics are unsolved?

Answering these and other questions on motion, the book 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 the recent results in mechanics, thermodynamics, electrodynamics, relativity, quantum theory, quantum gravity and unification. It is written for undergraduate students and for anybody interested in physics.

Christoph Schiller, PhD Université Libre de Bruxelles, is a physicist with more than 25 years of experience in the presentation of physical topics.

ISBN 978-3-00-021946-7

9 783000 219467

pdf file available free of charge at www.motionmountain.net