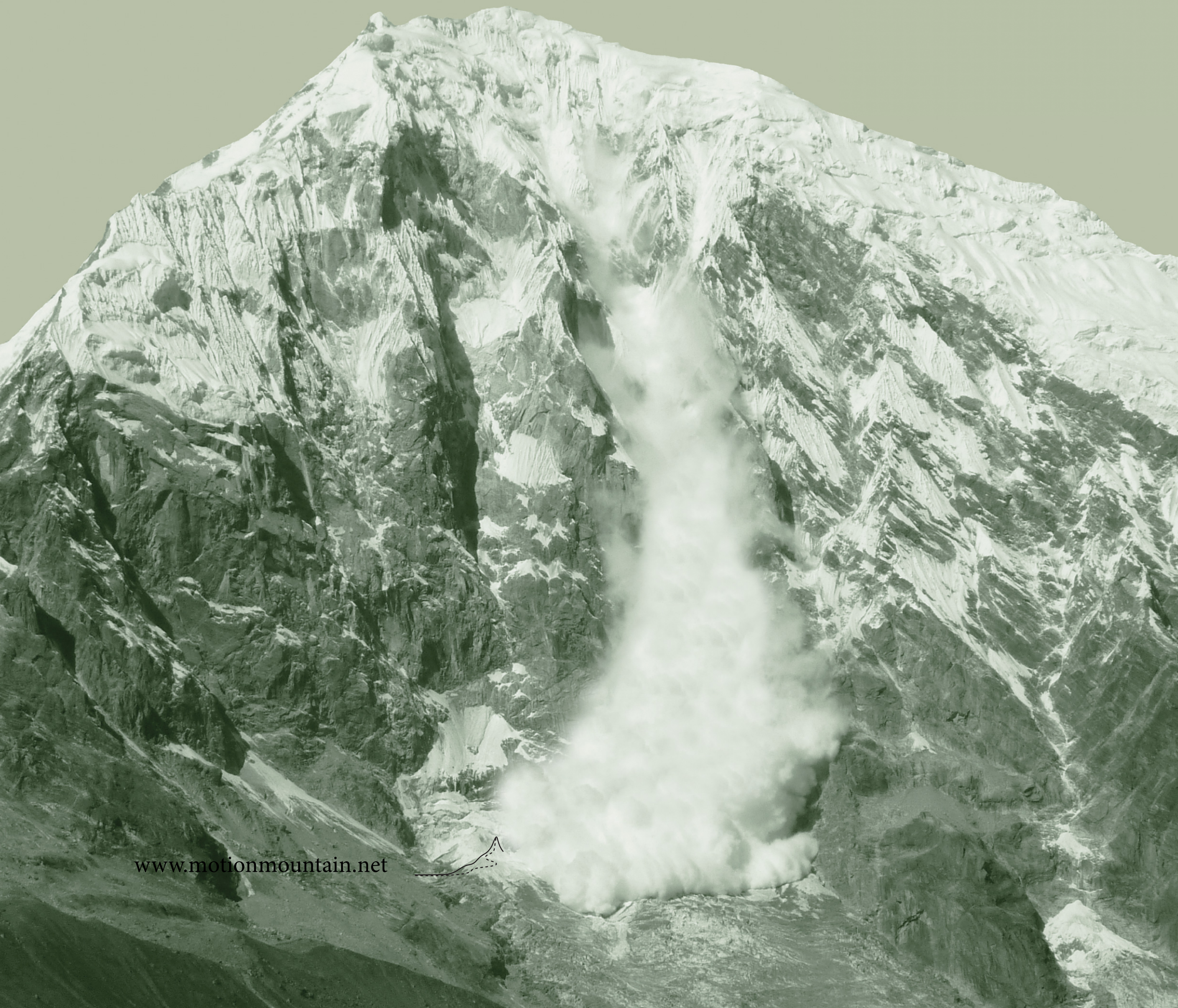
Christoph Schiller

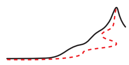
MOTION MOUNTAIN

THE ADVENTURE OF PHYSICS – VOL.II

RELATIVITY AND COSMOLOGY



www.motionmountain.net



Christoph Schiller

Motion Mountain

The Adventure of Physics
Volume II

Relativity and Cosmology

Edition 31, available as free pdf
with films at www.motionmountain.net

Editio trigesima prima.

Thirty-first edition.

To Britta, Esther and Justus Aaron

τῷ ἐμοὶ δαὶμονι

Die Menschen stärken, die Sachen klären.

PREFACE

> Primum movere, deinde docere.*
>
> Antiquity

This book series is for anybody who is curious about motion in nature. How do things, people, animals, images and empty space move? The answer leads to many adventures. This volume presents the best ones about extremely fast, powerful and distant motion. In the exploration of motion – physics – special and general relativity make up two important stages, as shown in Figure 1.

Special relativity is the exploration of nature's speed limit c. *General* relativity is the exploration of the force limit $c^4/4G$. The text shows that in both domains, all results follow from these two limit values. In particular, *cosmology* is the exploration of motion near nature's distance limit $1/\sqrt{\Lambda}$. This simple, intuitive and unusual way of learning relativity should reward the curiosity of every reader – whether student or professional.

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.

The motto of the text, *die Menschen stärken, die Sachen klären*, a famous statement on pedagogy, translates as: 'To fortify people, to clarify things.' Clarifying things – and adhering only to the truth – requires courage, as changing the 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. Enjoy it.

Christoph Schiller

* '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.

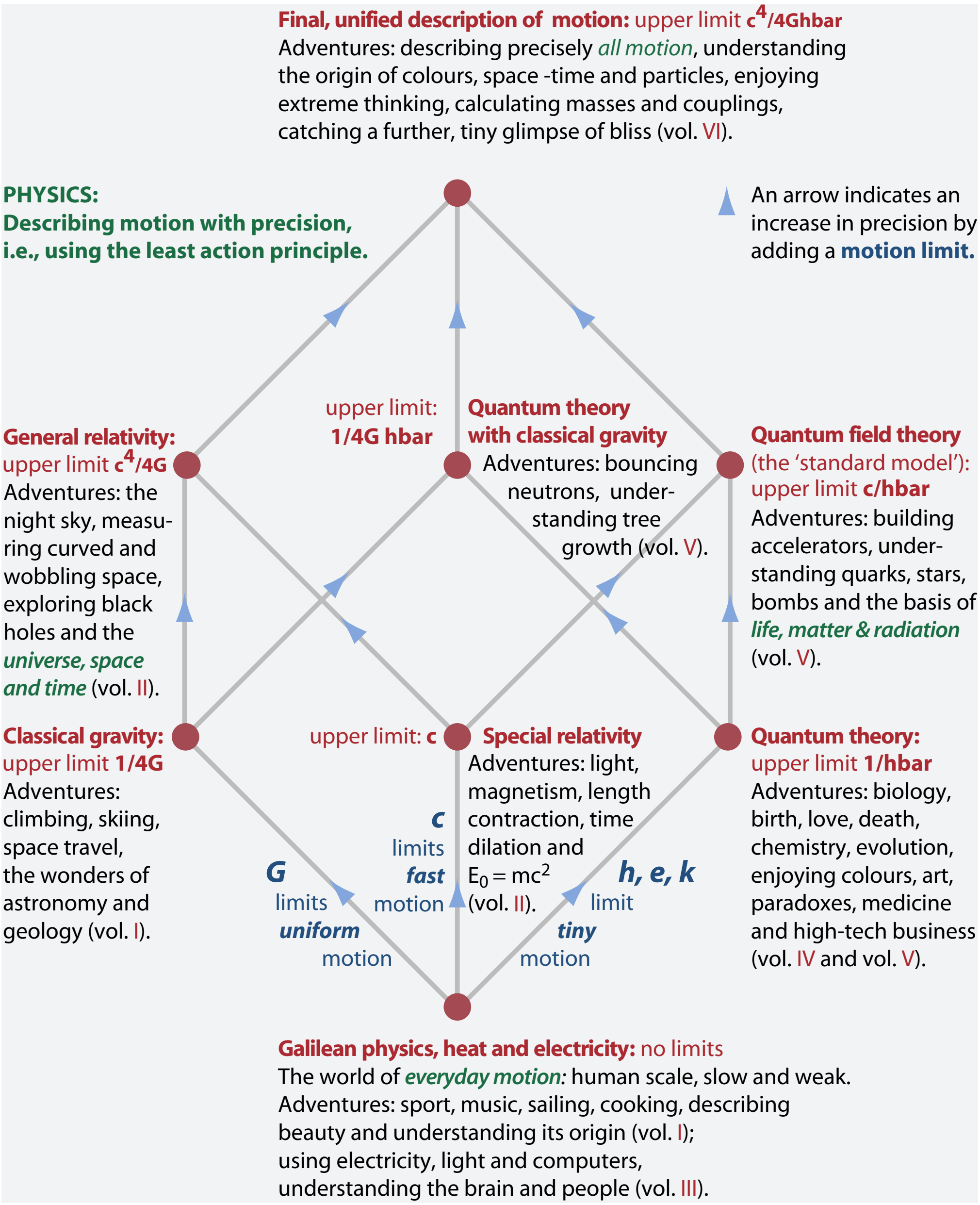


FIGURE 1 A complete map of physics, the science of motion, as first proposed by Matvei Bronshtein (b. 1907 Vinnytsia, d. 1938 Leningrad). The Bronshtein cube starts at the bottom with everyday motion, and shows the connections to the fields of modern physics. Each connection increases the precision of the description and is due to a limit to motion that is taken into account. The limits are given for uniform motion by the gravitational constant G, for fast motion by the speed of light c, and for tiny motion by the Planck constant h, the elementary charge e and the Boltzmann constant k.

Using this book

Marginal notes refer to bibliographic references, to other pages or to challenge solutions. In the colour edition, marginal notes, pointers to footnotes and links to websites are typeset in green. Over time, links on the internet tend to disappear. Most links can be recovered via www.archive.org, which keeps a copy of old internet pages. In the free pdf edition of this book, available at www.motionmountain.net, all green pointers and links are clickable. The pdf edition also contains all films; they can be watched directly in Adobe Reader.

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

Advice for learners

Learning allows us to discover what kind of person we can be. Learning widens knowledge, improves intelligence and provides a sense of achievement. Therefore, learning from a book, especially one about nature, should be efficient and enjoyable. Avoid bad learning methods like the plague! Do not use a marker, a pen or a pencil to highlight or underline text on paper. It is a waste of time, provides false comfort and makes the text unreadable. Add notes and comments instead! And do not learn from a screen. In particular, do not learn from videos, from games or from a smartphone. All games and almost all videos are drugs for the brain. Smartphones are drug dispensers that make people addicted and prevent learning. Learn from paper – at your speed, and allow your mind to wander! Nobody marking paper or looking at a screen is learning efficiently.

In my experience as a pupil and teacher, one learning method never failed to transform unsuccessful pupils into successful ones: if you read a text for study, summarize every section you read, *in your own words and images, 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 and images, aloud. And *enjoy* the telling aloud! You can do this alone or with friends, in a room or while walking. If you do this with everything you read, you will reduce your learning and reading time significantly; 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.

Advice for teachers

A teacher likes pupils and likes to lead them into exploring the field he or she chose. His or her enthusiasm is the key to job satisfaction. If you are a teacher, before the start of a lesson, picture, feel and tell yourself how you enjoy the topic of the lesson; then picture, feel and tell yourself how you will lead each of your pupils into enjoying that topic as much as you do. Do this exercise consciously, every day. You will minimize trouble in your class and maximize your teaching success.

This book is not written with exams in mind; it is written to make teachers and students *understand* and *enjoy* physics, the science of motion.

Feedback

The latest pdf edition of this text is and will remain free to download from the internet. 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 film did you miss?

Also help on the specific points listed on the www.motionmountain.net/help.html web page is welcome. All feedback will be used to improve the next edition. You are welcome to send feedback by mail or by sending in a pdf with added yellow notes, to provide illustrations or photographs, or to contribute to the errata wiki on the website. If you would like to translate a chapter of the book in your language, please let me know.

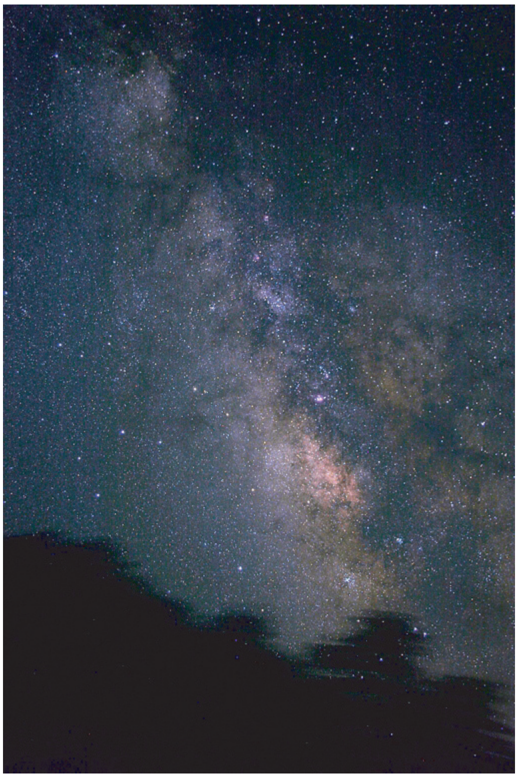
On behalf of all readers, thank you in advance for your input. For a particularly useful contribution you will be mentioned – if you want – in the acknowledgements, receive a reward, or both.

Support

Your donation to the charitable, tax-exempt non-profit organisation that produces, translates and publishes this book series is welcome. For details, see the web page www.motionmountain.net/donation.html. The German tax office checks the proper use of your donation. If you want, your name will be included in the sponsor list. Thank you in advance for your help, on behalf of all readers across the world.

The paper edition of this book is available, either in colour or in black and white, from www.amazon.com, in English and in certain other languages. And now, enjoy the reading.

CONTENTS



Relativity

In our quest to learn how things move,
the experience of hiking and seeing leads us to discover
that there is a maximum energy speed in nature,
that two events that occur at the same time for one observer
may not for another, and
that acceleration limits observation distance by a horizon.
We discover that empty space can bend, wobble and move,
we experience the fascination of black holes,
we find that there is a maximum force in nature,
we perceive why we can see the stars
and we understand why the sky is dark at night.

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 Page 295 atoms.

In short, light is important because

▷ Light is the standard for *ideal, 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

** ‘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 Challenge 2 s needs other methods to check flatness with light. Can you specify one?

**** 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 *α-rays* (helium nuclei), *β-rays* (again electrons), and *γ-rays* (high-energy X-rays) also produce shadows. All these discoveries were made between 1890 and 1910: those were the ‘ray days’ of physics.

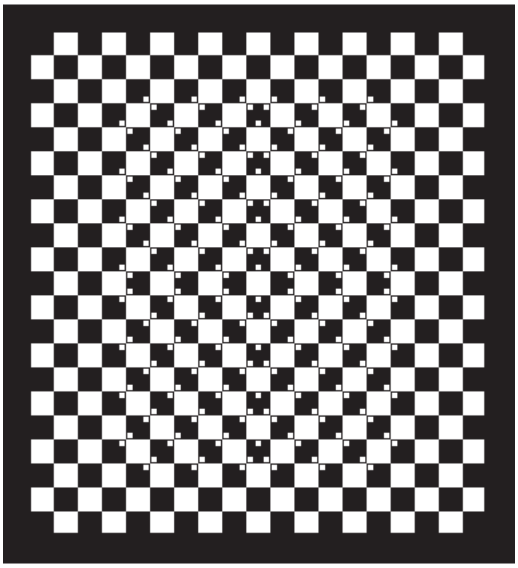


FIGURE 2 How do you check whether the lines are curved or straight?

▷ 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* or ideal speed, which is used as the implicit measurement standard, must have a *finite* value. An infinite velocity standard would not allow measurements at all. (Why?) In nature, lighter bodies 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.

Challenge 3 s

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 – but teaches us to learn to enjoy the past.

The speed of light is *high*; therefore it was not measured until the years 1668 to 1676, even though many, including Isaac Beeckman in 1629 and Galileo in 1638, had tried to do so earlier. ** The first measurement method was realized and published by the Danish astronomer Ole Rømer*** when he was studying the orbits of Io and the other Galilean

Ref. 3

* 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.

** During his whole life, and still in 1638, René Descartes argued publicly that the speed of light was infinite for reasons of principle. But in 1637, he had assumed a finite value in his explanation of Snell's 'law'. This shows how confused philosophers can be. In fact, Descartes wrote to Beeckman in 1634 that if one could prove that the speed of light is finite, he would be ready to admit directly that he 'knew nothing of philosophy.' We should take him by his word.

Ref. 2

*** Ole (Olaf) Rømer (b. 1644 Aarhus, d. 1710 Copenhagen), important 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 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

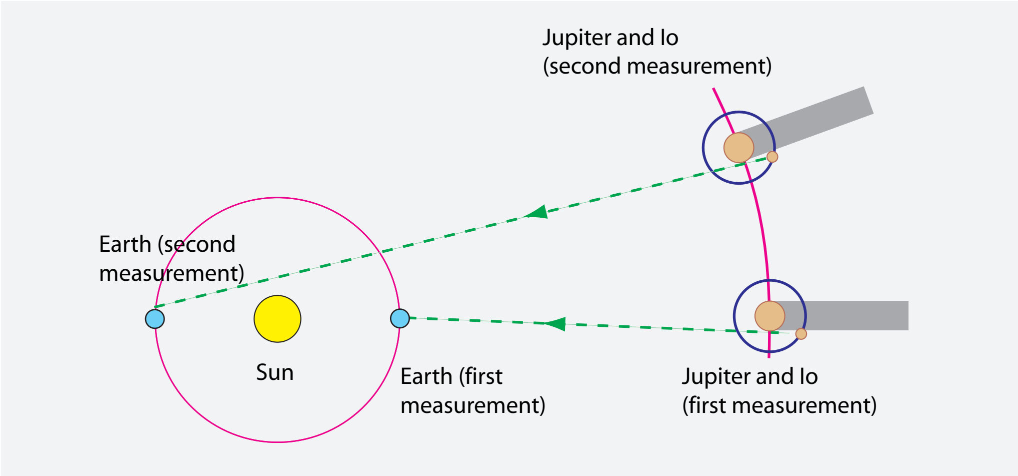


FIGURE 3 Rømer's method of measuring the speed of light.

Vol. I, page 210
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 the 1720s, independently, by the astronomers Eustachio Manfredi (b. 1674 Bologna , d. 1739 Bologna) and James Bradley (b. 1693 Sherborne , d. 1762 Chalford). Their measurements allowed the 'rain method' to measure the speed of light.

Ref. 4
Challenge 4 s
Vol. I, page 152
Ref. 5

Aberration and the speed of rain

How can we measure the speed of falling rain? We walk rapidly with an umbrella, measure the angle α 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_r of the *rain* is then given by

$$c_r = v/\tan\alpha \,. \qquad (1)$$

In the same way we can measure the speed of wind when on a surfboard or on a ship. The same method can be applied to the speed of 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,

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.

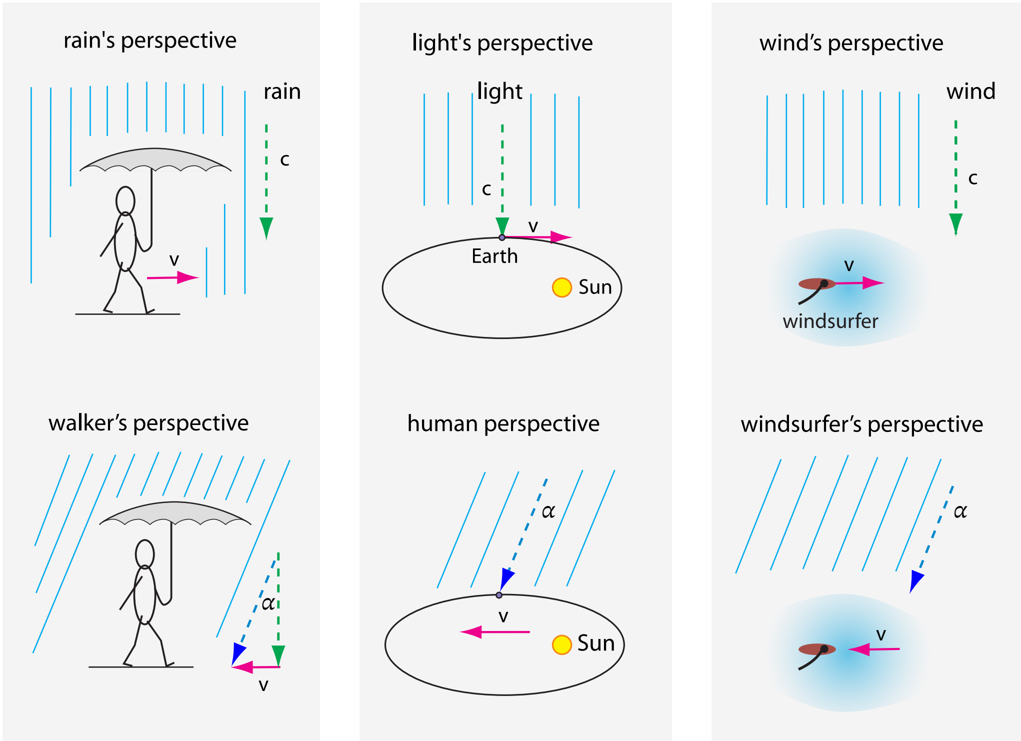


FIGURE 4 The rainwalker's or windsurfer's method of measuring the speed of light.

the angle is not 90°. For the speed of *light* c, we get

$$c = v/\sin\alpha \,. \qquad (2)$$

Challenge 5 s
Ref. 7

(Why is the expression for light different?) The deviation from the geometrically expected angle was called the *aberration* of light by Eustachio Manfredi. The aberration is determined by comparing measurements over the course of a year, in particular, six months apart. The explanation of aberration was also found by James Bradley, who independently, made similar measurements.* The measured value of the aberration angle

* Umbrellas were not common in Europe in 1719 or 1726; they became fashionable later. The umbrella part of the story is made up. It is said that Bradley understood aberration while sailing on the Thames, when he noted that on a moving ship the apparent wind, showed by an on-board flag, has a direction that depends on the sailing direction and thus differs from that on land. For many years, independently, Manfredi and Bradley had observed numerous stars, notably Gamma Draconis, and during that time they had been puzzled by the *sign* of the aberration, which was *opposite* to the effect they were looking for, namely that of the star *parallax*. Both the parallax and the aberration for a star above the ecliptic make them describe a small ellipse in the course of an Earth year, though the ellipses differ by their orientation and their rotation sign. Can you see why? Today we know that the largest known parallax for a star is 0.77″, whereas the major axis of the aberration ellipse is 20.5″ for all stars. The discovery by Bradley and Manfredi convinced even church officials that the Earth moves around the Sun, and Galileo's books were eventually taken from the index of forbidden books. Since the church delayed the publication of Manfredi's discovery, Bradley is often named as the sole discoverer of aberration. But the name of the effect recalls Manfredi's priority. Because of the discovery, Manfredi became member of the Académie des Sciences and the Royal Society.

Challenge 6 s

for a star exactly above the ecliptic is $20.49552(1)'' \approx 0.1$ mrad – a really small angle. It is called the *aberration constant*. Its existence clearly shows that the Earth orbits the Sun, when observed by a distant observer. Yes, the Earth moves.

Using the aberration angle, we can deduce the speed of light if we know the speed of the Earth when travelling around the Sun. For this, 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
Vol. I, page 179
between the distance to the Moon (determined as explained earlier on) and the distance
Challenge 7 s
to the Sun. The explanation is left as a puzzle for the reader.

The angle of Aristarchus * is almost a right angle (which would yield an infinite dis-
Ref. 6
tance), and good instruments 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 Sun–Moon distance ratio of about 400. Today, thanks to radar distance
Page 308
measurements of planets, the average distance to the Sun is known with the incredible precision of 30 metres;** its value is 149 597 870.691(30) km, or roughly 150 million kilometres.

The speed of light

Using the distance between the Earth and the Sun, the Earth's orbital speed is $v = 2\pi R/T = 29.7$ km/s. Therefore, the aberration angle gives us the following result

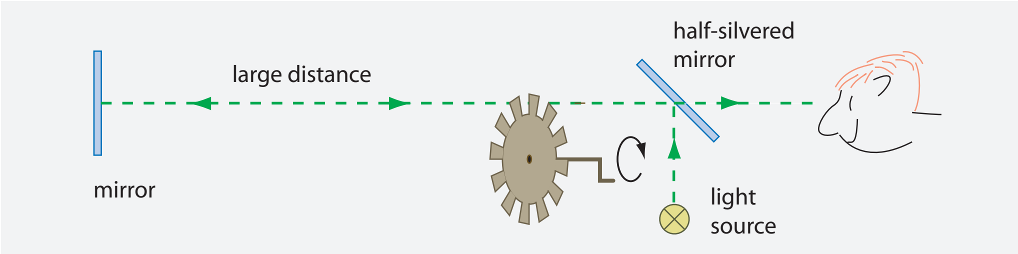
▷ The speed of light (in vacuum) is $c = 0.300$ Gm/s, or 0.3 m/ns, or 0.3 mm/ps, or 1080 million km/h.

This is an astonishing speed value, especially when compared with the highest speed ever achieved by a man-made object, namely the Helios II satellite, which travelled around the Sun at 253 Mm/h = 70.2 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 Hippolyte Fizeau (b. 1819 Paris, d. 1896 Venteuil). 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,
Vol. I, page 61
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 his
Ref. 9
experiment by Jan Frercks has even achieved a precision of 2 %. Today, the measurement

Ref. 8
* 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 monk and astronomer Nicolaus Copernicus (b. 1473 Thorn, d. 1543 Frauenburg) reproposed the heliocentric system two thousand years later, he did not mention Aristarchus, even though he got the idea from him.

** Moon distance variations can even be measured to the nearest centimetre; can you guess how this is
Challenge 8 s
achieved?



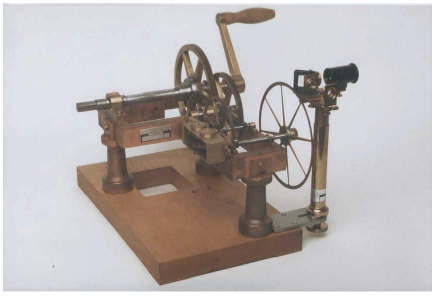


FIGURE 5 Fizeau's set-up to measure the speed of light (photo © AG Didaktik und Geschichte der Physik, Universität Oldenburg).

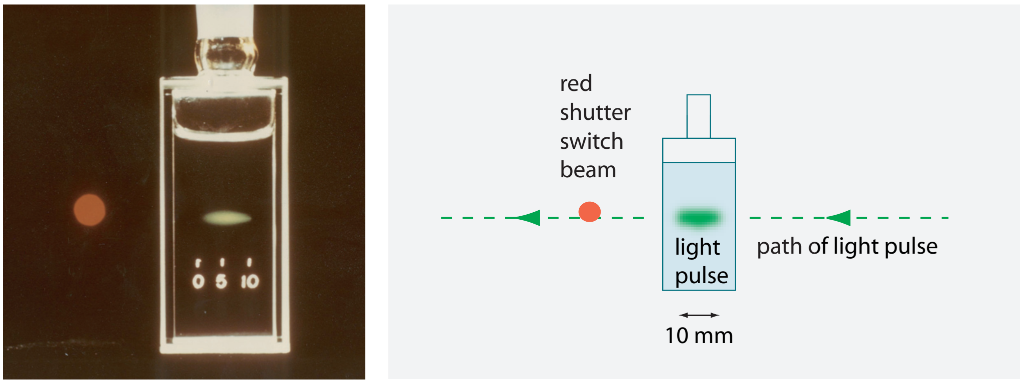


FIGURE 6 The first photograph of a green light pulse moving from right to left through a bottle with milky water, marked in millimetres (photograph © Tom Mattick).

is 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 Vol. III, page 32 the 'ping' command.

The speed of light is so high that in everyday life 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 Ref. 10 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 Challenge 10 s 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?)

A finite speed of light also implies that a rapidly rotating light beam bends, as shown

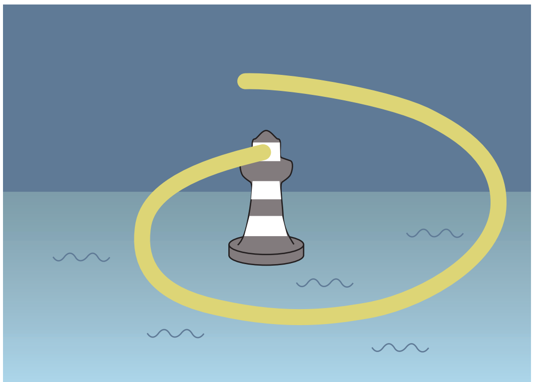


FIGURE 7 A consequence of the finiteness of the speed of light. Watch out for the tricky details – light *does* travel straight from the source, it does *not* move along the drawn curved line; the same occurs for water emitted by a rotating water sprinkler.

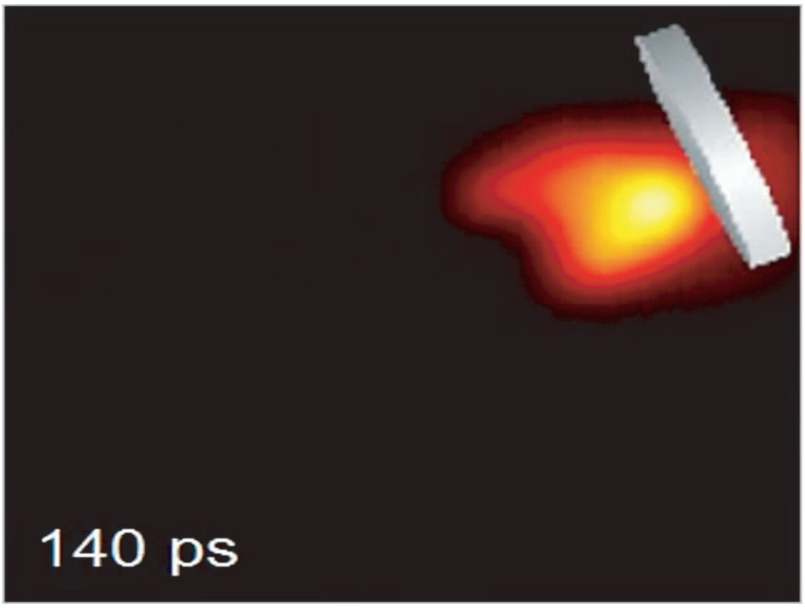


FIGURE 8 A film taken with a special ultrafast camera showing a short light pulse that bounces off a mirror (QuickTime film © Wang Lihong and Washington University at St. Louis).

as in Figure 7. In everyday life, the high speed of light and the slow rotation of lighthouses make the effect unnoticeable. But maybe, one day, ...

Finally, in the twenty-first century, *films* of moving light pulses started to appear. A beautiful example is shown in Figure 8. Such films again confirm that light has a finite speed.

In summary, light moves extremely rapidly, but with a *finite speed*. For example, light is much faster than lightning, as you might like to check yourself. A century of increasingly precise measurements of the speed of light in all its forms have culminated in the modern value

Challenge 11 s

$$c = 299\,792\,458 \text{ m/s}. \tag{3}$$

In fact, this value has now been fixed *exactly*, by definition, and the metre has been

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.
In vacuum, the speed of light is $c = 299\,792\,458\,\mathrm{m/s}$, or roughly 30 cm/ns – always and everywhere.
Light in vacuum always moves faster than any material body does.
The proper speed of light is infinite. Page 48
The speed of light pulses, their true signal speed, is the forerunner speed, not the group velocity. In vacuum, the forerunner speed is always and everywhere c. Vol. III, page 135
Light beams are approximations when the wavelength is neglected.
Light beams move in a straight line when far from matter.
Shadows can move without any speed limit.
Normal and high-intensity light is a wave. Light of extremely low intensity is a stream of particles.
In matter, both the forerunner speed and the energy speed of light are at most c.
In matter, the group velocity of light pulses can be negative, zero, positive or infinite.

defined in terms of the speed of light c since 1983. The good approximate values 0.3 Gm/s or 0.3 μm/fs are obviously easier to remember. A summary of what is known today about the motion of light is given in Table 1. Two of the most surprising properties of light motion were discovered in the late nineteenth century. They form the basis of what is called
Ref. 11 the *theory of special relativity*.

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

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

All experiments ever performed show: the speed of electromagnetic radiation in vacuum does *not* depend on the frequency of the radiation, nor on its polarization, nor on its intensity.

Ref. 12 For example, electromagnetic pulses from the Crab nebula pulsar have been shown to have the same speed over 13 decades of frequencies, from radio waves to γ-rays. The speed value is the same to a precision of 14 digits. Observations using γ-ray bursts have improved this precision to 20 digits. After starting together and travelling together for
Ref. 13 thousands of millions of years across the universe, light pulses with different frequencies and polarizations still arrive side by side.

Comparisons between the speed of γ-rays and the speed of visible light have also been

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

performed in accelerators. Also the speed of radio waves of different frequencies when
Ref. 14 travelling around the Earth can be compared. All such experiments found no detectable change of the speed of light with frequency. Additional experiments show that the speed
Ref. 15 of light is the same in all directions of space, to at least 21 digits of precision.

Light from the most powerful lasers, light from the weakest pocket lamps and light from the most distant stars has the same speed. In the same way, linearly polarized, circularly polarized and elliptically polarized light, but also thermal, i.e., unpolarized light has the same speed.

In summary,

▷ Nature provides *no way* to accelerate or decelerate the motion of light in vacuum.

Watching pulsating stars in the sky proves it. The speed of light in vacuum is always the same: it is *invariant*. But this invariance is puzzling.

We all know that in order to throw a stone as fast and 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. We also know that hitting a tennis ball more rapidly makes it faster.

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 observe the elliptical shape of the orbits. However, such bizarre effects are not seen, and perfect
Ref. 16 ellipses are observed. Willem de Sitter gave this beautiful argument already in 1913; he confirmed its validity with a large number of double stars.

In other words, light in vacuum is never faster than light:

▷ All light beams in vacuum have the same speed.

Ref. 13, Ref. 17 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 the speed of the emitted light does not change. (Can
Challenge 12 s you guess what lamps were used?)

In everyday life, we also know that a stone or a tennis ball 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 in a vacuum, 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
Ref. 18 highest precision possible. Even for the highest observer speeds, the speed of the arriving light remains the same.

Both sets of experiments, those with moving lamps and those with moving observ-

FIGURE 9 All devices based on electric motors prove that the speed of light is invariant (© Miele, EasyGlide).

ers, thus show that the velocity of light has exactly the *same magnitude* for everybody, everywhere and always – even if observers are moving with respect to each other or with respect to the light source.

▷ The speed of light in vacuum is *invariant*.

The speed of light in vacuum is indeed the ideal, perfect measurement standard for speed. By the way, an equivalent alternative term for 'speed of light' is 'radar speed' or 'radio speed'; we will see in the part on electrodynamics why this is the case. Vol. III, page 108

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. (The difference is probably due to a tiny speed difference and 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

Ref. 19 There is also a further set of experimental evidence for the invariance of the speed of light. Every electromagnetic device, such as an electric vacuum cleaner, shows that the speed of light is invariant. Vol. III, page 53 We will discover that magnetic fields would not result from electric currents, as they do every day in every electric motor 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 Albert Einstein show that the invariance of the speed of light is also in agreement with the observed motion of bodies. Ref. 20 We will check this agreement in this chapter. The connection between relativity and electric vacuum cleaners, as well as other machines, will be explored in the chapters on electrodynamics. Vol. III, page 53

The motion of light and the motion of bodies are deeply connected. 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 *almost* as fast as such a light wave would see a light wave moving *slowly*. And an observer moving at the same speed as the wave would

FIGURE 10 Albert Einstein (1879–1955).

see a *frozen* wave. However, experiment and the properties of electromagnetism *prevent* both observations; observers and bodies *cannot* reach the speed of light. Vol. III, page 53

▷ The speed of light in vacuum is a *limit speed*.

Observers and bodies thus always move *slower* than light.

In summary, the speed of light in vacuum is an invariant limit speed. Therefore, there is no way to accelerate a light pulse. And, in contrast to a tennis ball, there is no way to see a light pulse before it actually arrives. Thus, playing tennis with light is neither possible nor is it fun – at least in vacuum. But what about other situations? Challenge 14 d

Albert Einstein

Albert Einstein (b. 1879 Ulm, d. 1955 Princeton) was one of the greatest physicists ever. (By the way, the 's' in his family name is pronounced 'sh' and the two instances of 'ei' are pronounced like 'eye', so that the full pronunciation is [ˈalbɛrt ˈaɪnʃtaɪn].) 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. In 1906, he published the proof of the famous formula $E = c^2 m$, after a few others also had proposed it. Page 76 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 quantum theory in its most counter-intuitive aspects. Later, he explained the Einstein–de Haas effect which proves that magnetism is due to motion inside materials. After many other discoveries, in 1915 and 1916 Einstein published his highest achievement: the general theory of relativity, one of the most beautiful and remarkable works of science. Page 136 In the remaining forty years of his life, he searched for the unified theory of motion, without success.

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. Another of his enemies was the philosopher Henri Bergson. An influential figure of the time, he somehow achieved, with his confused thinking, to prevent that Einstein received the Nobel Prize in Physics. Until his death, Einstein kept

TABLE 2 How to convince yourself and others that there is a maximum energy speed c in nature. Compare this table with the table about maximum force, on page 109 below, and with the table about a smallest action, on page 19 in volume IV.

Statement	Test
The maximum energy speed value c is observer-invariant.	Check all observations.
Local energy speed values $> c$ are not observed.	Check all observations.
Local energy speed values $> c$ cannot be produced.	Check all attempts.
Local energy speed values $> c$ cannot even be imagined.	Solve all paradoxes.
The maximum local energy speed value c is a principle of nature.	Deduce the theory of special relativity from it.
	Check that all consequences, however weird, are confirmed by observation.

his Swiss passport in his bedroom. He was not only a great physicist, but also a great Ref. 21 thinker; his collection of thoughts about topics outside physics are well worth reading. However, his family life was disastrous, and he made each of his family members deeply unhappy.

Anyone interested in emulating Einstein should know first of all that he published *many* papers.* He was both ambitious and hard-working. Moreover, many of his papers Ref. 22 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 the speed of 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., of any localized mass or energy – is bound by

$$v \leqslant c \,. \tag{4}$$

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

* All his papers and letters are now freely available online, at einsteinpapers.press.princeton.edu.

The existence of an invariant limit speed c is not as surprising at we might think: we
Page 104 need such an invariant value in order to 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, to the existence of event horizons and to the existence of antimatter, 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 was called this way; he would have preferred the name 'Invarianztheorie', i.e., 'theory of Ref. 23 invariance', but was not able to change the name any more. Thus Einstein called the Ref. 19 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, as we will find out.**

Can an invariant limit speed really 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 this and the next 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 mirror on an aeroplane and reflect a light beam with it, the light still moves away with the same speed, both for the pilot and for an observer on Earth. All experiments confirm this weird behaviour of light.

If we accelerate a bus that 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. Even with the current measurement precision of $2 \cdot 10^{-13}$, Ref. 15 we cannot discern any changes of the speed of light for different speeds of the observer. 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 Vol. I, page 441 '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 measured to be the same for

* Henri Poincaré (1854 Nancy–1912 Paris), important mathematician and physicist. Poincaré was one of the most productive scientists of his time, advancing relativity, quantum theory and many parts of mathematics.
Ref. 24 ** Among the most beautiful introductions to relativity are still those given by Albert Einstein himself. It has taken almost a century for books almost as beautiful to appear, such as the texts by Schwinger or by Ref. 25, Ref. 26 Taylor and Wheeler.

both, we have

$$\frac{x}{t} = c = \frac{\xi}{\tau} . \tag{5}$$

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 relative to each other.

Challenge 15 e Time is thus not unique. This surprising result, which has been confirmed by many Ref. 27 experiments, was first stated clearly in 1905 by Albert Einstein. Every observer has its own time. Two observers' times agree only if they do not move against each other. 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 c 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 Ref. 28 the centre of the Sun, the speed of light is estimated to be around 30 km/year = 1 mm/s, and even in the laboratory, for some materials, the speed of light has been measured to Ref. 29 be as low as 0.3 m/s.

Vol. I, page 327 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 than 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 air, the speed of light 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.

Special relativity with a few lines

The speed of light is invariant and constant for all observers. We can thus deduce all Ref. 30 relations between what two different observers measure with the help of Figure 11. 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

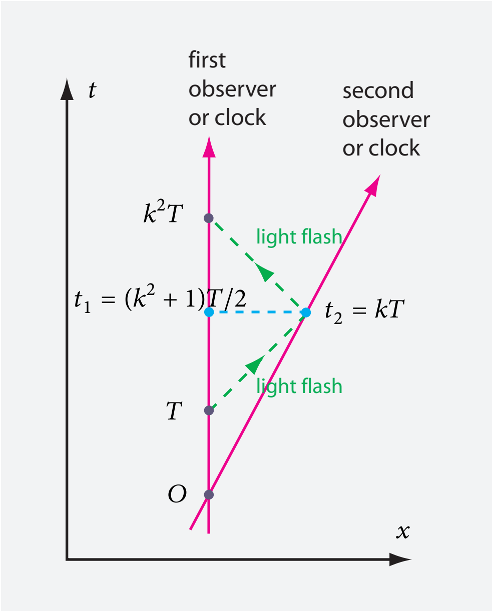


FIGURE 11 A drawing containing most of special relativity, including the expressions for time dilation and for the Lorentz transformation.

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.

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 16 s
Challenge 17 s

$$k = \sqrt{\frac{c+v}{c-v}} \quad \text{or} \quad \frac{v}{c} = \frac{k^2-1}{k^2+1} \ . \tag{6}$$

Page 31
This factor will appear again in the Doppler effect.*

Figure 11 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. This effect is called *time dilation*. In other terms, *time is relative*. Figure 12 shows a way to illustrate the result.

The *time dilation factor* between the two observers is found from Figure 11 by comparing the values t_1 and t_2; it is given by

$$\frac{t_1}{t_2} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} = \gamma(v) \ . \tag{7}$$

Time intervals for a moving observer are *shorter* by this factor γ; the time dilation factor is always larger than 1. In other words,

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

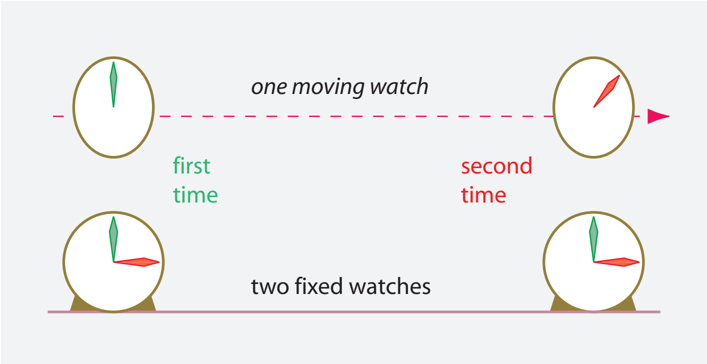


FIGURE 12 Moving clocks go slow: moving clocks mark time more slowly than do stationary clocks.

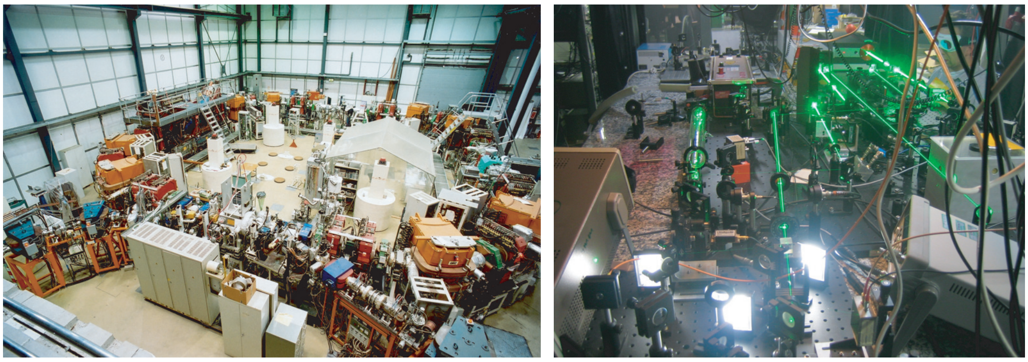


FIGURE 13 Moving clocks go slow: moving lithium atoms in a storage ring (left) read out with lasers (right) confirm the prediction to highest precision (© TSR relativity team at the Max Planck Gesellschaft).

▷ Moving clocks go slower.

Challenge 18 e

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 (7) has been tested to a precision better than one part in 10 million, with an experiment shown in Figure 13. The same factor γ also appears in the formula $E = c^2\gamma m$ for the equivalence of mass and energy, which we will deduce below. Expressions (6) or (7) are the only pieces of mathematics needed in special relativity: all other results derive from it.

Ref. 31

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, as shown in Figure 14. A man on either ladder will always observe that the steps of the *other* ladder are shorter. There is nothing deeper than this observation at the basis of time dilation and length contraction.

Page 52

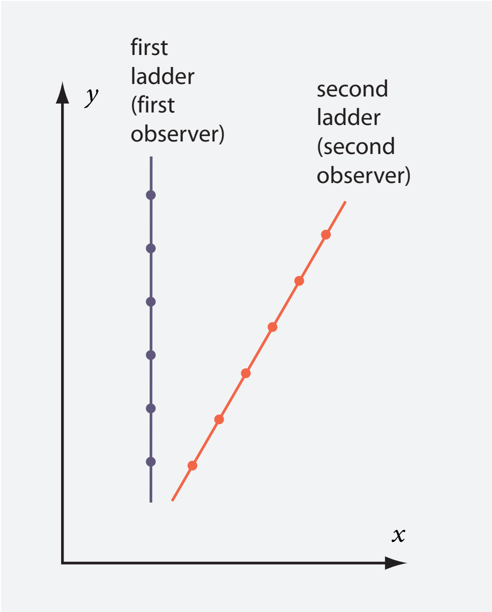


FIGURE 14 The observers on both ladders claim that the other ladder is shorter.

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 relative. Let us have a look at some of these experiments.

Acceleration of light and the Doppler effect

Can light in vacuum be accelerated? It depends on what you mean. Most physicists 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, Vol. III, page 157 and 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 19 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. Such 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.*

* 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.

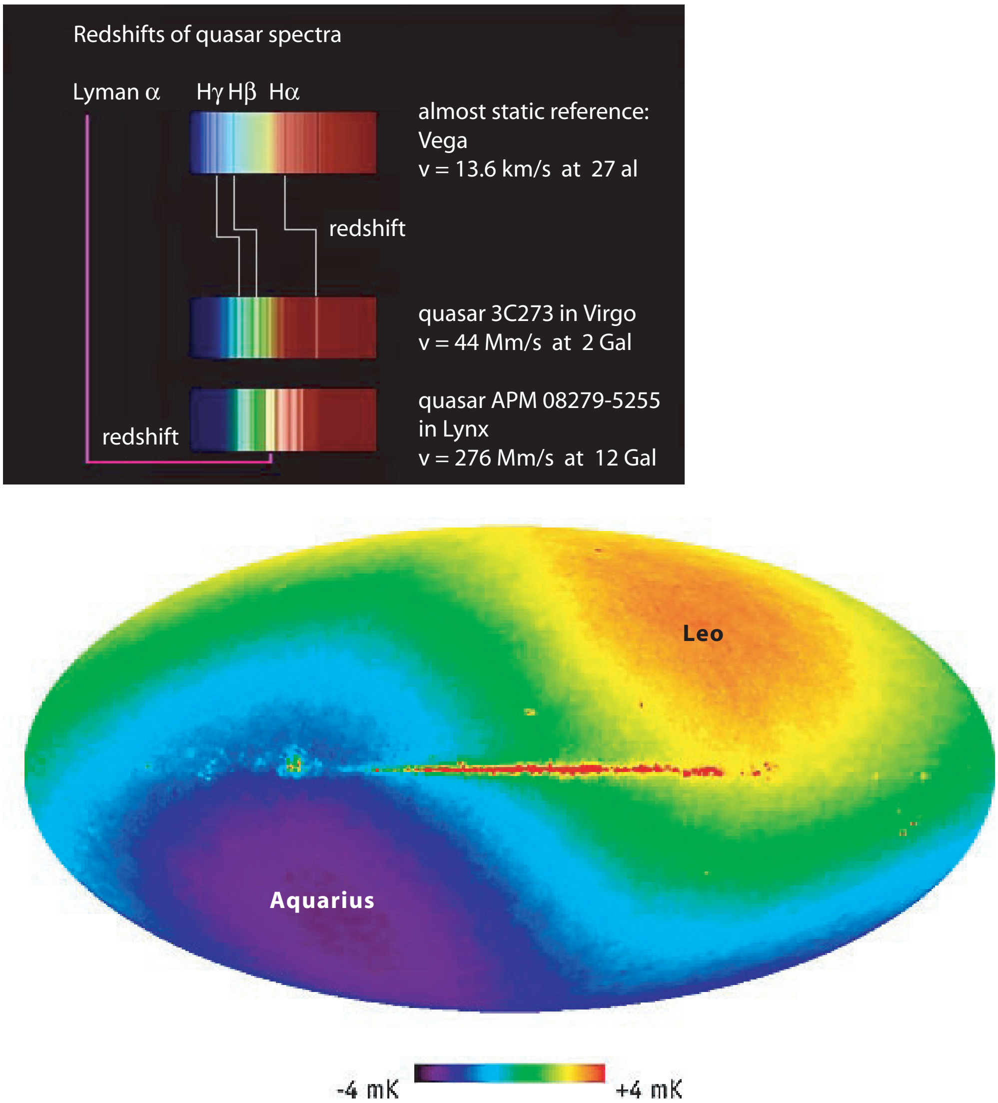


FIGURE 15 Top: the Doppler effect for light from two quasars. Below: the – magnified, false colour – Doppler effect for the almost black colour of the night sky – the cosmic background radiation – due to the Earth travelling through space. In the latter case, the Doppler shift implies a tiny change of the effective temperature of the night sky (© Maurice Gavin, NASA).

Physicists have tested the possibility of massive light in quite some detail. Observations now put any possible mass of light particles, or photons, at less than $1.3 \cdot 10^{-52}$ kg from terrestrial 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*.

Ref. 32, Ref. 18

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*.

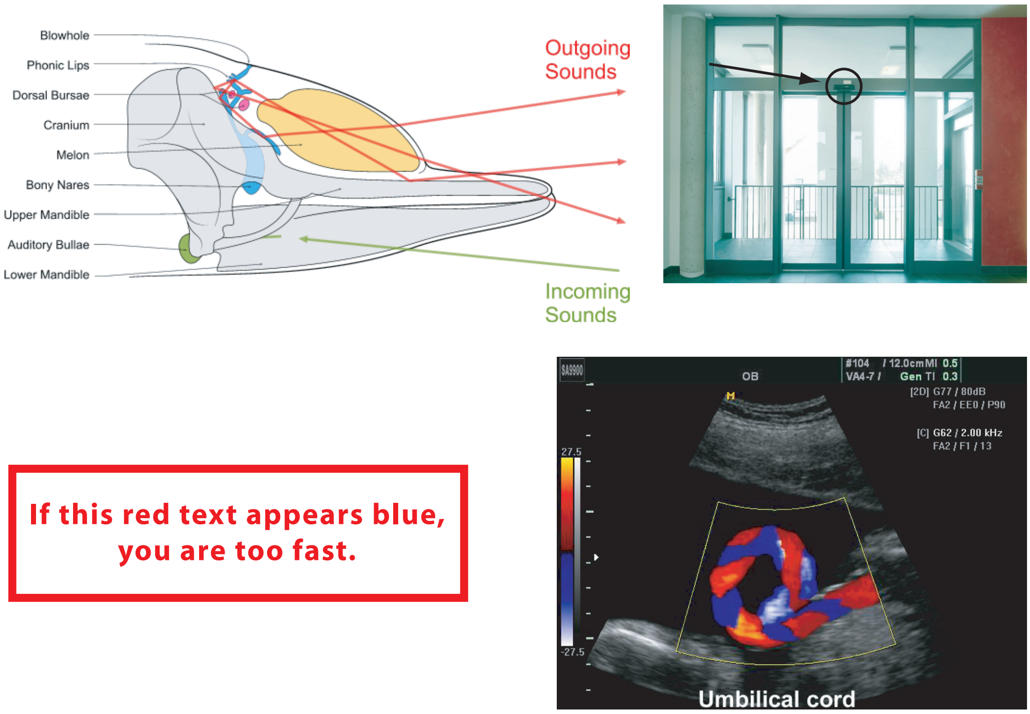


FIGURE 16 The Doppler sonar system of dolphins, the Doppler effect system in a sliding door opener, the Doppler effect as a speed warning and Doppler sonography to detect blood flow (coloured) in the umbilical cord of a foetus (© Wikimedia, Hörmann AG, Medison).

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 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 the effect is used to measure blood flow and heart beat of unborn babies in ultrasound systems (despite being extremely loud for the babies), as shown in Figure 16.

Vol. I, page 313

Doppler was also the first person to extend the concept of frequency shift 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 λ. Like the tone change for moving trains, Doppler realized that a moving light source produces a colour at the receiver that *differs* 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 106

Challenge 20 e

$$\frac{\lambda_r}{\lambda_s} = \frac{1}{\sqrt{1 - v^2/c^2}}\left(1 - \frac{v}{c}\cos\theta_r\right) = \gamma\left(1 - \frac{v}{c}\cos\theta_r\right) . \qquad (8)$$

* Christian Andreas Doppler (b. 1803 Salzburg, d. 1853 Venezia), important physicist. Doppler studied the effect named after him for sound and light. Already in 1842 he predicted (correctly) that one day we would 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. 33

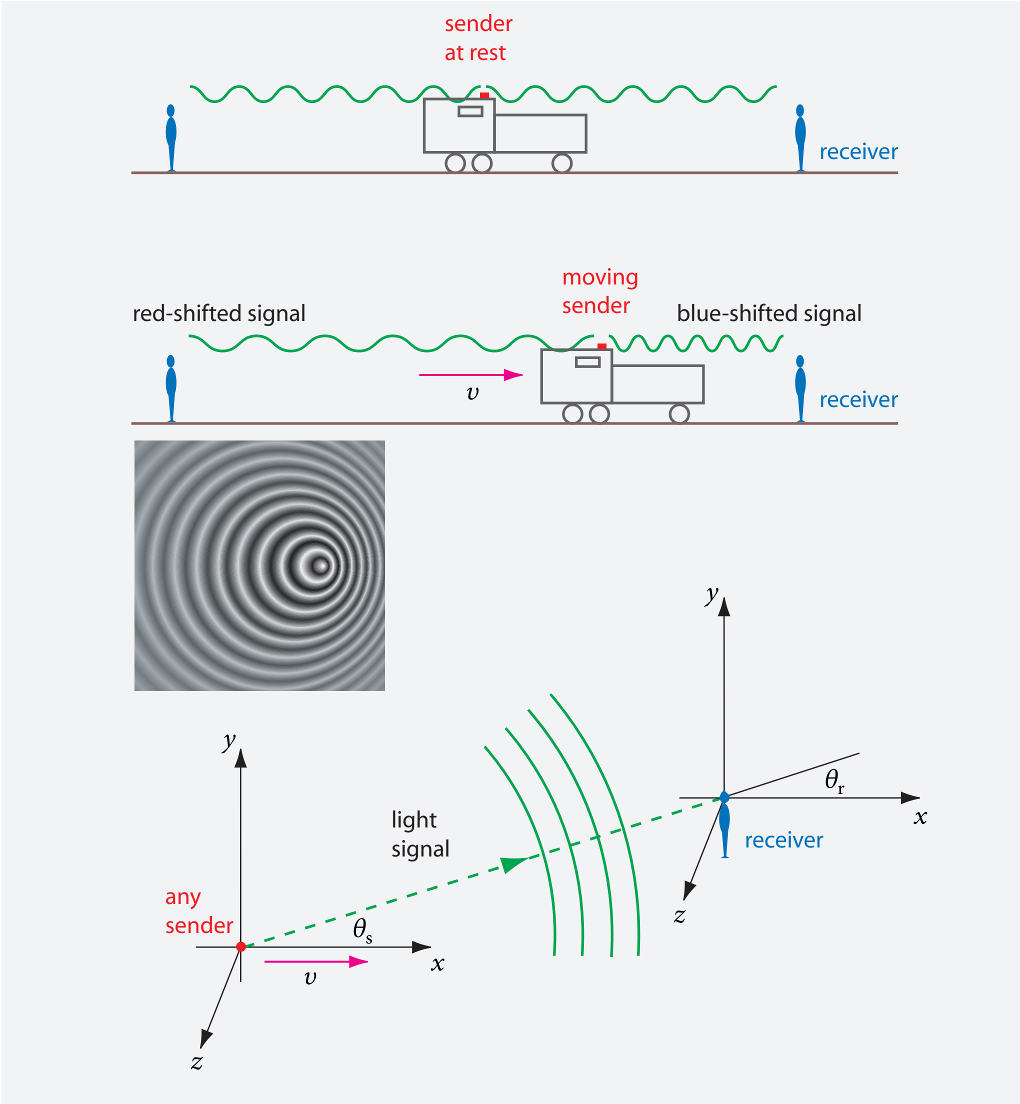


FIGURE 17 The set-up for the observation of the Doppler effect in one and three dimensions: waves emitted by an approaching source arrive with higher frequency and shorter wavelength, in contrast to waves emitted by a departing source (wave graph © Pbroks13).

The variables v and θ_r in this expression are defined in Figure 17. 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, also called the *colour shift*, was made by Johannes Stark* in 1905, who studied the light emitted by moving atoms. All

* Johannes Stark (b. 1874 Schickenhof, d. 1957 Eppenstatt), 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 in Physics. He left his professorship in 1922 and later turned into a full-blown National Socialist. A member of the National Socialist party from 1930 onwards, he became known for aggressively criticizing other people's statements about nature purely

subsequent experiments confirmed the calculated colour shift within measurement er-
Ref. 34 rors; the latest checks have found agreement to within two parts per million.

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 $\theta_{\text{unshifted}}$ the colour will stay the same. (How does the wavelength change in the purely transverse case? What is the expression for $\theta_{\text{unshifted}}$ in terms of the speed v?) Challenge 21 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.* Vol. III, page 108

As predicted by Doppler himself, the Doppler effect is regularly used to measure the speed of distant stars, as shown in Figure 15. In these cases, the Doppler shift is often characterized by the *red-shift number* z, defined with the help of wavelength λ or frequency f by

$$z = \frac{\Delta\lambda}{\lambda} = \frac{f_{\text{S}}}{f_{\text{R}}} - 1 = \sqrt{\frac{c+v}{c-v}} - 1 \ . \qquad (9)$$

Challenge 23 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 24 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. 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. Ref. 35

Even the rotation of our galaxy was discovered using the Doppler effect of its stars. Astronomers thus discovered that the Sun takes about 220 million years for a rotation around the centre of the Milky Way.

What happens if one really tries to play tennis with light, using a racket that moves at really high, thus relativistic speed? Such passionate tennis players actually exist; the fastest rackets built so far had a speed over 80 % per cent of the speed of light. They were produced in 2013 by shooting extremely powerful and short laser pulses, with a power of 0.6 ZW and a duration of 50 fs, onto a 10 nm thin diamond-like carbon foil. Such pulses eject a flat and rapid electron cloud into the vacuum; for a short time, this cloud acted as a *relativistic mirror*. When a second laser beam was reflected from this Ref. 36

for ideological reasons; he became rightly despised by the academic community all over the world, already during his lifetime.

Challenge 22 s * At what speed does a red traffic light appear green?

relativistic racket, the light speed remained unchanged, but its frequency was increased by a factor of about 14, changing the beam colour from the near infrared to the extreme ultraviolet. This relativistic electron mirror had a reflectivity far less than 1 %, though, its lifetime was only a few picoseconds, and its size only about 2 μm; therefore calling it a racket is a slight exaggeration.

In summary, whenever we try to change the vacuum *speed* of light, we only manage to change its *colour*. That is the Doppler effect. In other terms, attempts to accelerate or decelerate light only lead to colour change. And a colour change does not change the speed of light at all, as shown above. Page 22

Modern Doppler measurements are extremely precise. Our Sun moves with up to 9 cm/s with respect to the Earth, due to the planets that orbit it. Nowadays, the Doppler shift due to this speed value is measured routinely, using a special laser type called a *frequency comb*. This device allows to measure light frequencies within fractions of 1 Hz. Frequency combs allow the detection of even smaller speed values through the induced Doppler shifts. This method is used on a regular basis to detect exoplanets orbiting distant stars. Ref. 37

The connection between colour change and light acceleration attempts leads to a puzzle: we know from classical physics that when light passes a large mass, such as a star, it is deflected. Does this deflection lead to a Doppler shift? Vol. I, page 201 Challenge 25 s

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 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 (6) 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-dependent time breaks down in the case of the Doppler effect for *sound*?) Ref. 38 Page 29 Challenge 26 s

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 *dislocations* in crystalline solids. (We discuss this motion 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 transformations, show length contraction, and obey the famous energy formula $E = c^2 \gamma m$. In Vol. V, page 298 Ref. 39

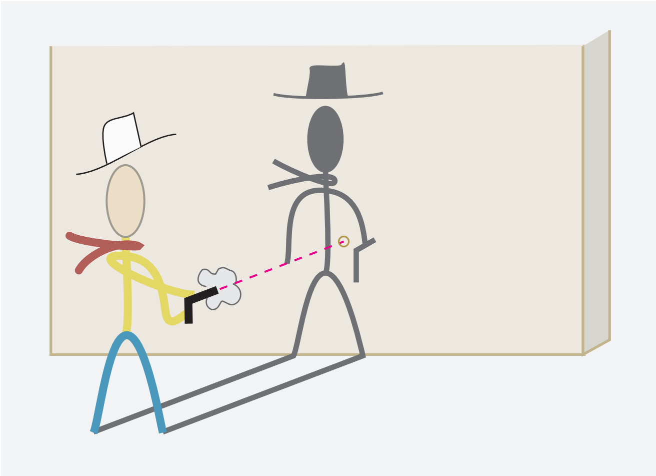


FIGURE 18 Lucky Luke.

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?

For Lucky Luke to achieve the feat shown in Figure 18, his bullet has to move faster than the speed of light. (What about his hand?) Challenge 27 e 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; an example was the Large Electron Positron ring, the LEP, of 27 km circumference, located partly in France and partly in 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. Ref. 40 The result is shown in Figure 19: 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?) Challenge 28 e

The speed–energy relation of Figure 19 is a consequence of the maximum speed, and its precise details are deduced below. Page 69 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

* 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.) Ref. 41

Crackpots can be found, for example, via the internet, in the sci.physics.relativity newsgroup. Ref. 42 See also the www.crank.net website. Crackpots are sometimes interesting, mainly because they demonstrate the importance of *precision* in language and in reasoning, which they all, without exception, neglect.

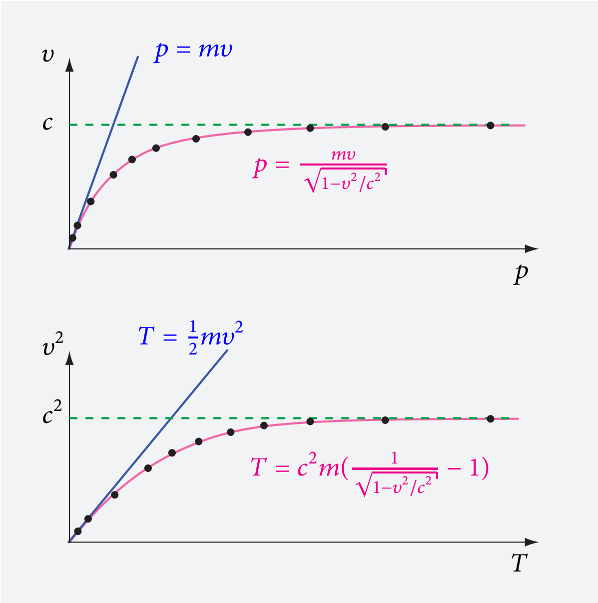


FIGURE 19 Experimental values (black dots) for the electron velocity v as function of their momentum p and as function of their kinetic energy T. The predictions of Galilean physics (blue) and the predictions of special relativity (red) are also shown.

taken for granted for more than two centuries, so that nobody ever thought of checking Ref. 43 it; but when this was finally done, as in Figure 19, it was found to be wrong.

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 reaching 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 cathode ray tube inside an old colour television, given that the transformer Challenge 29 s inside produces 30 kV.

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 Challenge 30 d 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.

We conclude that

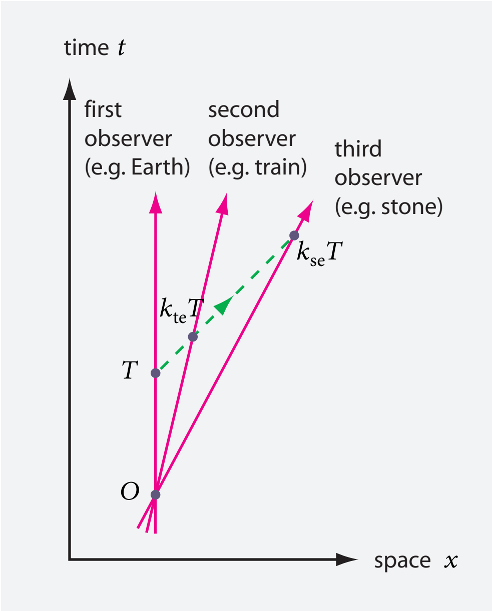


FIGURE 20 How to deduce the composition of velocities.

▷ The maximum speed is the speed of *massless* entities.

Electromagnetic waves, including light, and gravitational waves are the only known entities that travel at the maximum speed. Though the speed of neutrinos cannot be distinguished experimentally from the maximum speed, recent experiments showed that they do have a tiny mass. Ref. 44

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 31 e Challenge 32 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 26 The existence of a maximum speed, together with Figure 20, implies that the k-factors must satisfy $k_{se} = k_{st}k_{te}$.* Then we only need to insert the relation (6) between each k-

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

Challenge 33 e factor and the respective speed to get

$$v_{se} = \frac{v_{st} + v_{te}}{1 + v_{st} v_{te}/c^2} . \qquad (10)$$

Challenge 34 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 (10) has been confirmed by each of the millions of cases for which it has been checked. You may check that it simplifies with high precision to the naive sum for everyday life speed values.

Page 68 Ref. 18

Observers and the principle of special relativity

Special relativity is built on a simple principle:

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

Ref. 46 Or, as Hendrik Lorentz** liked to say, the equivalent:

▷ The speed v of a physical system is bound by

$$v \leqslant c \qquad (11)$$

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 $c = 1/\sqrt{\varepsilon_0 \mu_0}$, 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 of c, including its 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 show without exception that the speed of light in vacuum is invariant, whether they were performed *before* or

Vol. III, page 106

Ref. 45 * 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.

*** All these experiments, which Einstein did not bother to cite in his 1905 paper, were performed by the complete who's who of 19th century physics, such as Wilhelm Röntgen, Alexander Eichenwald, François Arago, Augustin Fresnel, Hippolyte Fizeau, Martin Hoek, Harold Wilson, Albert Michelson, (the first

Ref. 47 Ref. 48

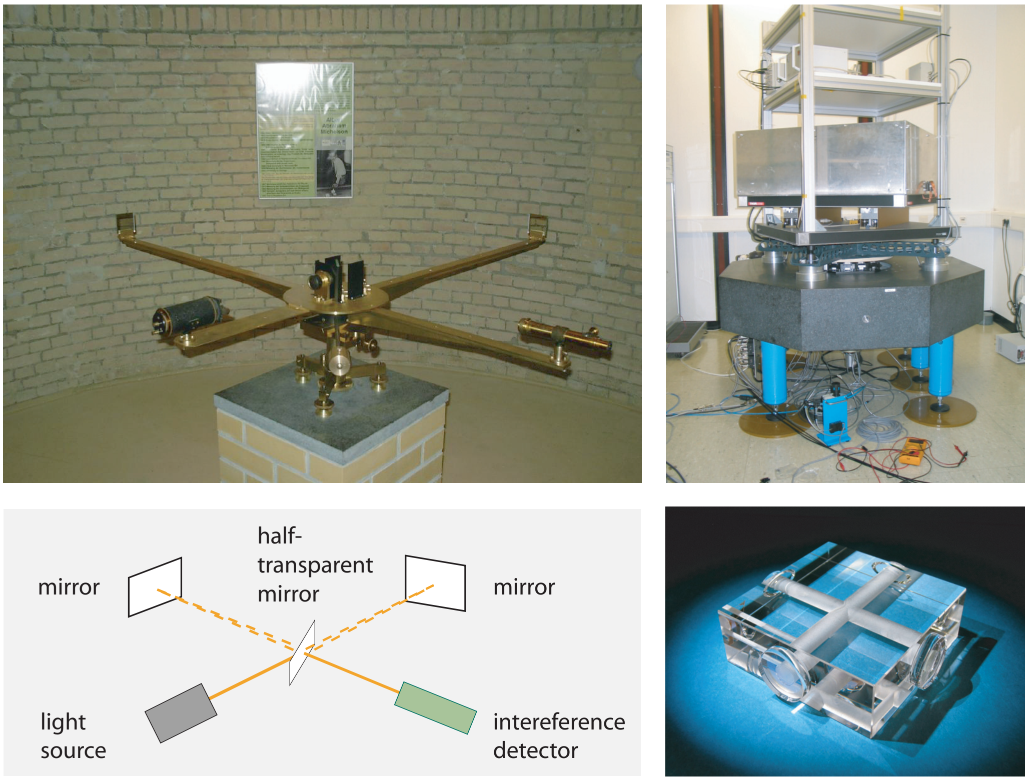


FIGURE 21 Testing the invariance of the speed of light on the motion of the observer: the reconstructed set-up of the first experiment by Albert Michelson in Potsdam, performed in 1881, and a modern high-precision, laser-based set-up that keeps the mirror distances constant to less than a proton radius and constantly rotates the whole experiment around a vertical axis (© Astrophysikalisches Institut Potsdam, Stephan Schiller).

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 21. All such experiments found no change of the speed of light with the motion Ref. 49 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 53 this is explained in the section on electrodynamics.

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

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.

Vol. I, page 156 * This point is essential. For example, Galilean physics states that only *relative* motion is observable. 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. 23 Poincaré – after the old principle, instead of calling it 'invariance theory', as Einstein would have preferred.

▷ 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.

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 start 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'. Inertial observers are thus much rarer than non-inertial observers. Can you confirm this? Nevertheless, inertial observers are the most simple ones, and they form a special set:

Challenge 36 e

▷ 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 space and time coordinates x, y, z and t, and a Greek one, using coordinates ξ, υ, ζ and τ,** that move with constant velocity $\boldsymbol{v}$ relative to each other, as shown in Figure 22. The invariance of the speed of light in any direction for any two observers means that the coordinate differences found by two observers are related by

Challenge 37 e

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

Challenge 35 s

* Can you give the precise argument leading to this deduction?
** They are read as 'xi', 'upsilon', 'zeta' and 'tau'. The names, correspondences and pronunciations of all Greek letters are explained in Appendix A in the first volume.

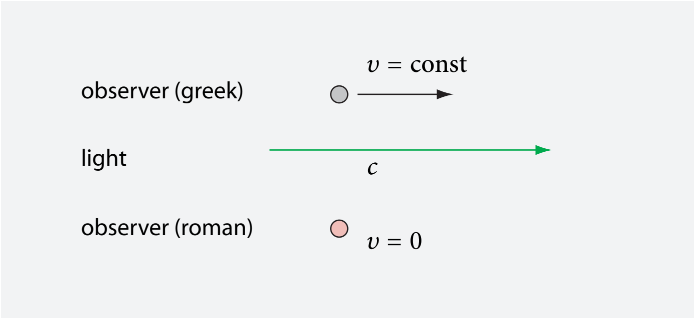


FIGURE 22 Two inertial observers and a beam of light. Both measure the same speed of light c.

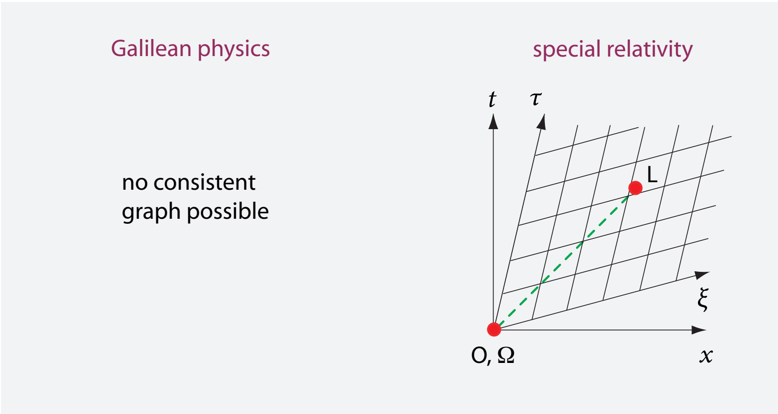


FIGURE 23 The space-time diagram for light seen from two inertial observers, using coordinates (t, x) and (τ, ξ).

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

$$(c\mathrm{d}t)^2 - (\mathrm{d}x)^2 = (c\mathrm{d}\tau)^2 - (\mathrm{d}\xi)^2 . \qquad (13)$$

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

Challenge 38 e

expression, we deduce

$$\mathrm{d}t = \frac{\mathrm{d}\tau}{\sqrt{1 - v^2/c^2}} = \gamma \mathrm{d}\tau . \qquad (14)$$

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*, *Lorentz factor* or *relativistic contraction* γ 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 γ increases. The largest value humans have ever achieved is about $2 \cdot 10^5$; the largest observed value in nature is about

Challenge 39 s

10^{12}. Can you imagine where they occur?

For a relativistic correction γ larger than 1 – thus in principle for any relative speed different from zero – the time measurements of the two observers give different values. Because time differs from one observer to another, moving observers observe *time dilation*.

But that is not all. Once we know how clocks behave, we can easily deduce how *coordinates* change. Figures 22 and 23 show that the x coordinate of an event L is the sum of two intervals: the ξ coordinate plus any distance between the two origins. In other words, we have

$$\xi = \gamma(x - vt) \ . \tag{15}$$

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

$$\tau = \gamma(t - xv/c^2) \ . \tag{16}$$

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. 50 beautiful discoveries of physics, in 1892 and 1904, Lorentz deduced these relations from Vol. III, page 76 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 τ, as well as x and ξ, 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 (15) and (16) 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 (15) and (16) contain many curious results. Again they Challenge 40 e show that time depends on the observer. They also show that length depends on the Page 52 observer: in fact, moving observers observe *length contraction*. Space and time are thus indeed relative.

The Lorentz transformations (15) and (16) are also strange in another respect. When two observers look at each other, each of them claims to measure shorter intervals than Challenge 41 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 Page 52 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. 51 velocity. However, all alternatives had 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.

* For information about Hendrik Antoon Lorentz, see page 40.

** The same discovery had been published first in 1887 by Woldemar Voigt (b. 1850 Leipzig, d. 1919 Göttingen); Voigt – pronounced 'Fohgt' – was also the discoverer of the Voigt effect and the Voigt tensor. Later, in 1889, George Fitzgerald (b. 1851 Dublin, d. 1901 Dublin) also found the result.

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 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.

Challenge 42 s

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* $\mathrm{d}i$ between two events, defined as

Ref. 52

$$\mathrm{d}i^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = c^2\mathrm{d}t^2\left(1 - \frac{v^2}{c^2}\right) , \qquad (17)$$

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 later that when gravitation is present, the whole of 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 (17) 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.)

* '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 365

** The term 'manifold' is defined in all mathematical details later in our walk.

*** Hermann Minkowski (b. 1864 Aleksotas, d. 1909 Göttingen) was mainly a mathematician. He had developed, independently, similar ideas to Einstein, but the latter was faster. Minkowski then developed the concept of space-time. Unfortunately, Minkowski died suddenly at the age of 44.

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 43 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 23. 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 tabloid 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 24.) 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 24.

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 24 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*. 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 authors sloppily 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 44 s

In short, time orders events only *partially*. In particular, for two events that are not

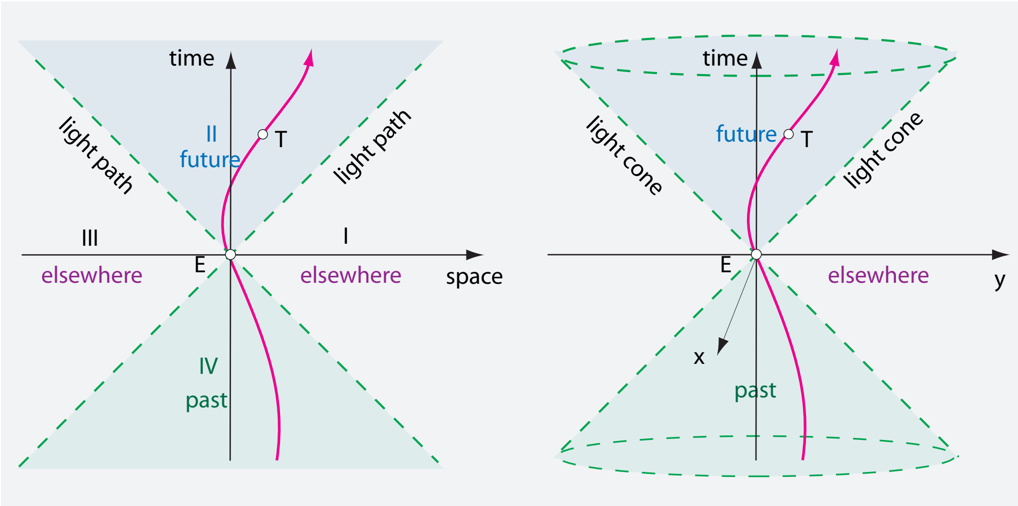


FIGURE 24 A space-time diagram for a moving object T seen from an inertial observer O in the case of one and two spatial dimensions; the slope of the world-line at a point is the speed at that point, and thus is never steeper than that of light.

Challenge 45 e causally connected, their temporal order (or their simultaneity) depends on the observer!

A vector inside the light cone is called *time-like*; one on the light cone is called *light-like* or *null*; and one outside the cone is called *space-like*. 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 time-like 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 Challenge 46 e why? In other words, if the speed of light could be surpassed in some way, the future Challenge 47 s 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 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,

* Another related result is slowly becoming common knowledge. Even if space-time had a non-trivial shape, such as a cylindrical topology with closed time-like curves, one still would not be able to travel into the past, in contrast to what many science fiction novels suggest. The impossibility of this type of time travel is made Ref. 53 clear by Steven Blau in a recent pedagogical paper.

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 kilometres away, the open sea. A second observer is a pilot in a relativistic aeroplane and happens to be just above the island when the lightning hits the island. Which lightning hits first for the pilot?

Challenge 48 e

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 distant lightning hit first.

Challenge 49 e

When you wave your hand in front of a mirror, your image waves with the same frequency. What happens if the mirror moves away with relativistic speed?

Challenge 50 e

We will discover in the section on quantum theory that the yellow colour of gold is a relativistic effect; also the liquid state of mercury at room temperature is a consequence of relativity. Both effects are due to the high speed of the outer electrons of these atoms.

Vol. IV, page 195

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 away is given by

Challenge 51 e

$$d = \frac{vt}{\sqrt{1 - v^2/c^2}} . \tag{18}$$

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 entire universe in less than a second. (The fuel issue is discussed below.)

Page 51

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

$$w = \frac{d}{t} = \frac{v}{\sqrt{1 - v^2/c^2}} = \gamma\, v . \tag{19}$$

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.*

* Using proper velocity, the relation given in equation (10) for the composition of two velocities $\mathbf{w}_a = \gamma_a \mathbf{v}_a$

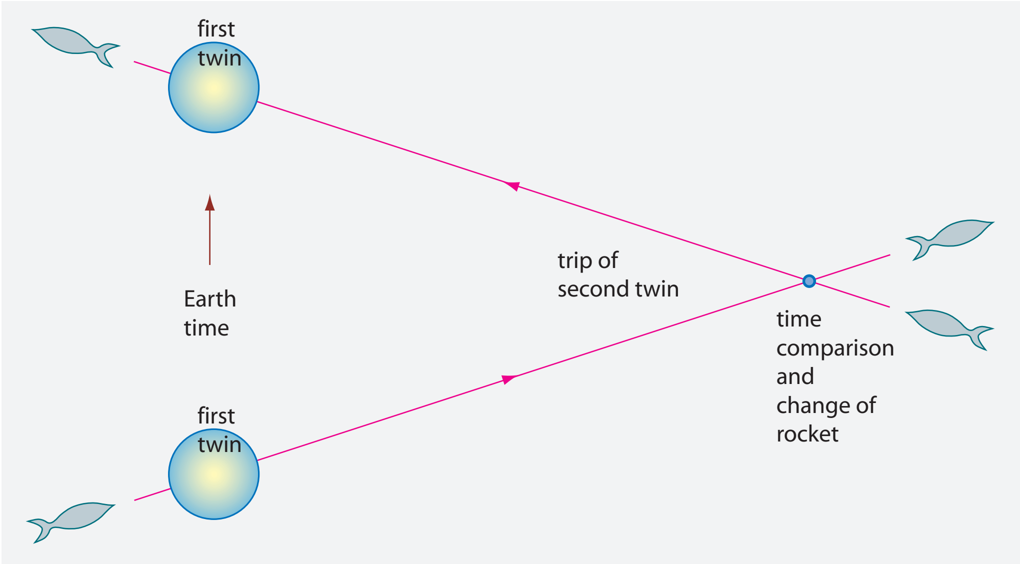


FIGURE 25 The twin paradox.

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. 55, Ref. 56

$$\frac{t}{t'} = \gamma \,. \tag{21}$$

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. In short, the invariance of c implies: *Travellers remain younger*.

The most extreme illustration of this is the famous *twin paradox*. An adventurous

Challenge 52 e

and $\mathbf{w}_b = \gamma_b \mathbf{v}_b$ simplifies to

$$w_{s\parallel} = \gamma_a \gamma_b (v_a + v_{b\parallel}) \quad \text{and} \quad w_{s\perp} = w_{b\perp} \,, \tag{20}$$

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. 54

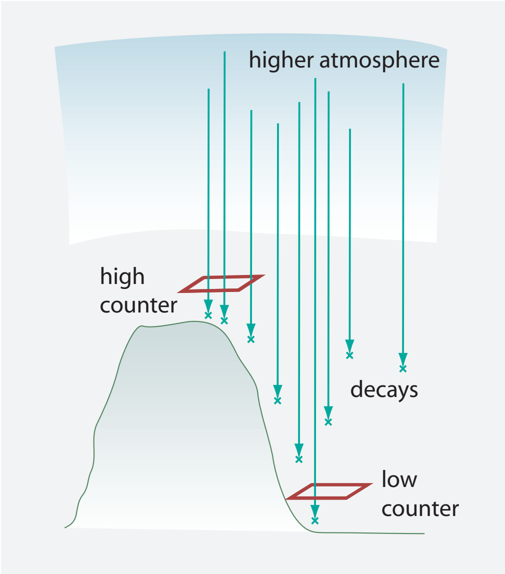


FIGURE 26 More muons than expected arrive at the ground because fast travel keeps them young.

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 25. At his arrival, he notes that his twin brother on Earth is much older than himself. This result has also been confirmed in many experiments – though not with real twins yet. Can you explain the result, especially the asymmetry between the two twins?

Ref. 57
Challenge 53 s

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 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

Vol. V, page 162

Ref. 58

* 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.

in the valley below, as shown in Figure 26. The first time this experiment was performed, 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 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 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 γ 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.

Ref. 59 Challenge 54 s Challenge 55 s Ref. 60

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. The effect is so common that for fast particles one speaks of the *apparent lifetime* τ_{app} through the relation $\tau_{app} = \gamma\tau$. 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. 18

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 rotation velocity. This story serves as a reminder to be careful when applying special relativity in situations involving gravity: pure special relativity is only applicable when space-time is flat, i.e., when gravity is *not* present.

Ref. 61

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 56 e Ref. 62

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 57 s

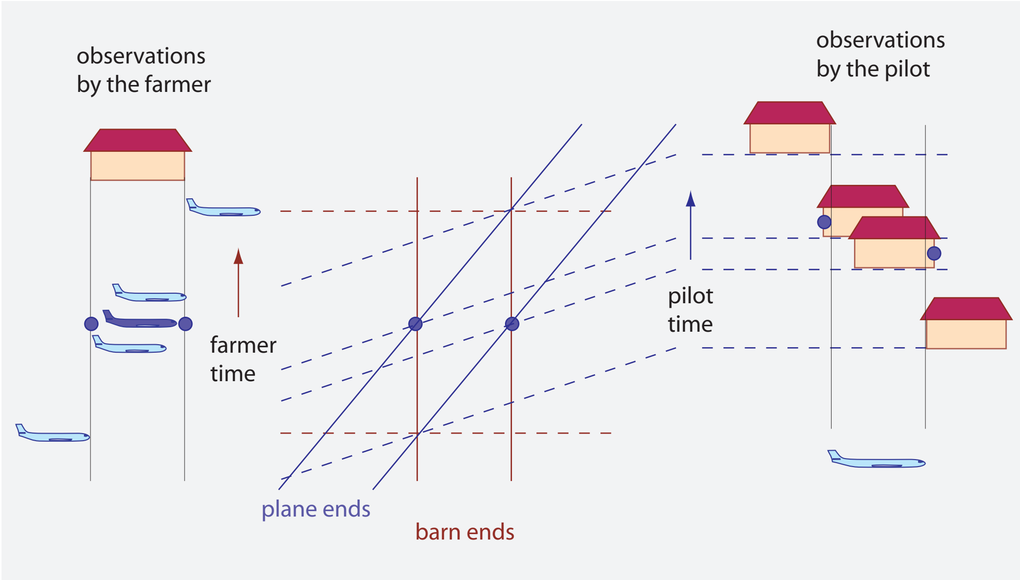


FIGURE 27 The observations of the pilot and the barn owner.

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

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 27. 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 completely true: can you work out the details?) For obvious reasons, this experiment has never been realized. Challenge 60 s

Let us explore some different length contraction experiments. 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 γ 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 59 s

* Is the Earth contraction value measurable at all?

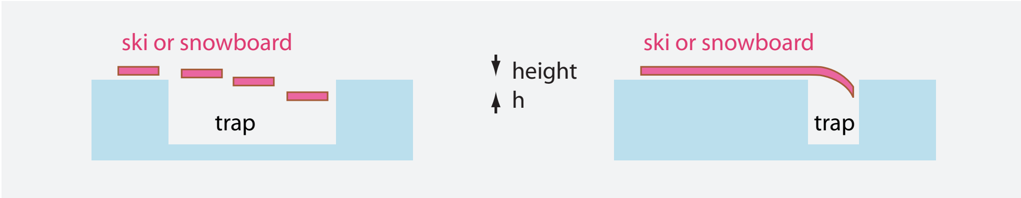


FIGURE 28 The observations of the trap digger (left) and of the snowboarder (right), as often (misleadingly) published in the literature.

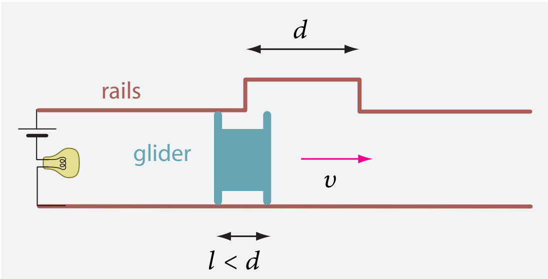


FIGURE 29 Does the conducting glider keep the lamp lit at large speeds?

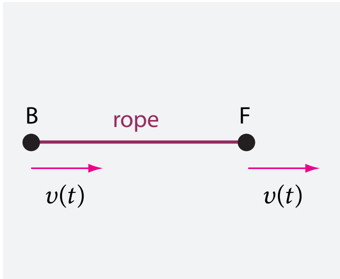


FIGURE 30 What happens to the rope?

Ref. 63 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 28. Can you confirm this? (Challenge 61 e) In other words, shape is not an observer-invariant concept. (However, rigidity *is* observer-invariant, if defined properly; can you confirm this?) (Challenge 62 s)

The snowboard explanation and figure however, though published, are *not* correct, as Harald van Lintel and Christian Gruber have pointed out. (Ref. 64) We 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. (Challenge 63 e) Figure 28 is so exaggerated that it is incorrect. The snowboarder would simply speed over the hole.

In fact, we can simplify the discussion of such examples of length contraction by exploring what happens when a rod moves on an inclined path towards a slot, without any gravity. (Ref. 65) A careful exploration shows that if the slot and the rod are parallel for the rod observer, they are *not* parallel for the slot observer, and vice versa. The concept of *parallel* is relative!

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 29. (Ref. 66) 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? (Challenge 64 s) (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? Be warned: this problem gives rise to *heated* debates! What is unrealistic in this experiment?

Ref. 67 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 30. 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. Who is right? You can check by yourself that this prediction is confirmed by all observers, in the cars and on Earth. Challenge 65 s

A funny – but again 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 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 correct buoyancy force? Alternatively, answer the following question: why is it impossible for a submarine to move at relativistic speed? (Ref. 68, Challenge 66 s, Challenge 67 s)

In summary, for macroscopic bodies, length contraction is interesting but will probably never be observed. However, length contraction 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 31.

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 32 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

* 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 31 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).

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. Figure 33 shows this clearly. But 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
Ref. 69 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 34. The total deformation is the result of the angle-dependent aberration. We discussed the aberration of star pos-
Page 18 itions at the beginning of this chapter. In complete images, aberration transforms circles into circles: such transformations are called *conformal*. As a result, a sphere is seen to have a circular outline even at relativistic speeds – though its thickness changes.

Aberration leads to the *pearl necklace paradox*. If the relativistic motion keeps intact the circular shape of spheres, but transforms rods into shorter rods, what happens to a
Challenge 68 s pearl necklace moving along its own long axis? Does it get shorter or not?

A further puzzle: imagine that a sphere moves and rotates at high speed. Can all the
Challenge 69 r mentioned effects lead to an apparent, observer-dependent sense of rotation?

Which is the best seat in a bus?

Ref. 67 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
Challenge 70 e 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 con-

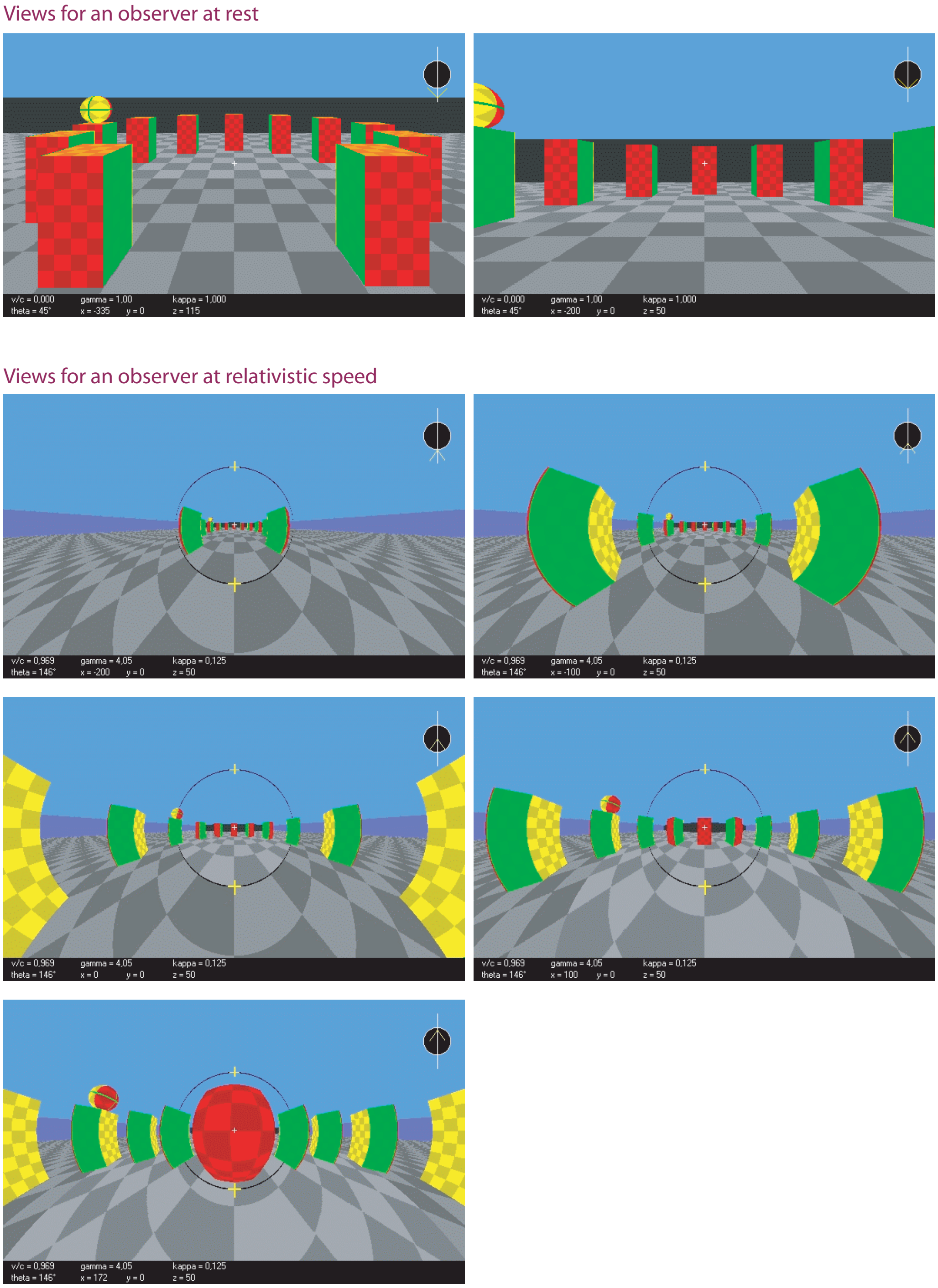


FIGURE 32 Flying through twelve vertical columns (shown in the two uppermost images) with 0.9 times the speed of light as visualized by Nicolai Mokros and Norbert Dragon, showing the effect of speed and position on distortions (© Nicolai Mokros).

FIGURE 33 What a researcher standing and one running rapidly through a corridor observe (ignoring colour and brightness effects) (© Daniel Weiskopf).

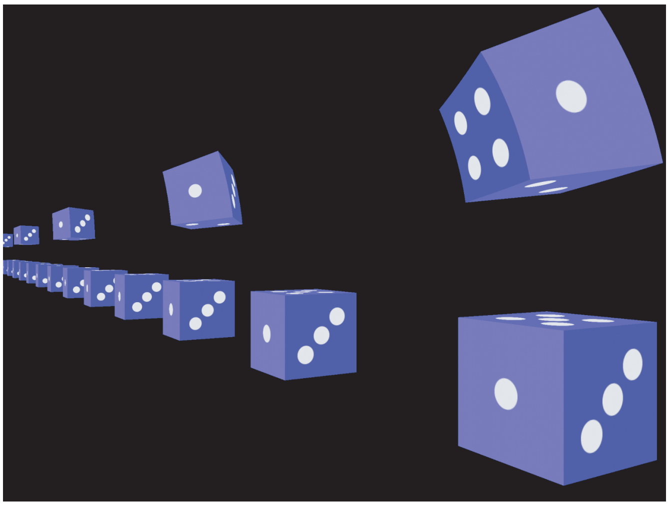


FIGURE 34 A stationary row of dice (below), and the same row, flying above it at relativistic speed towards the observer, though with Doppler and brightness effects switched off. (Mpg film © Ute Kraus at www.tempolimit-lichtgeschwindigkeit.de).

stant 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

Challenge 71 e for the frame of the freely rolling twins the relation

$$t'_{\mathrm{b}} = \gamma \Delta x\, v/c^2 + t'_{\mathrm{f}}\ , \qquad (22)$$

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 72 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 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 35; 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. Ref. 70 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.

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

> “Quid celerius umbra?*”
>
> Antiquity

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. Page 48 (Why?) Challenge 73 s The following examples show speeds that are genuinely higher than the speed of light in vacuum.

As first example, consider the point at which scissors cut paper, marked X in Figure 36. 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 –

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

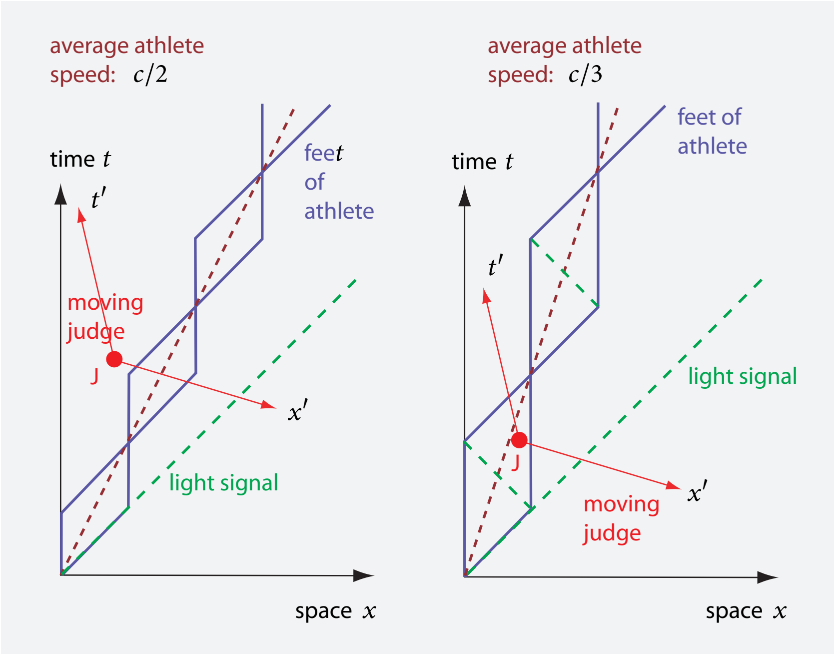


FIGURE 35 For the athlete on the left, the competition judge moving in the opposite direction sees both feet off the ground at certain times, but not for the athlete on the right.

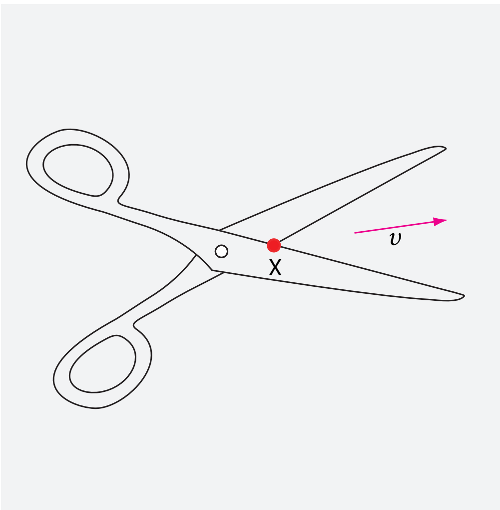


FIGURE 36 A simple example of motion that can be faster than light.

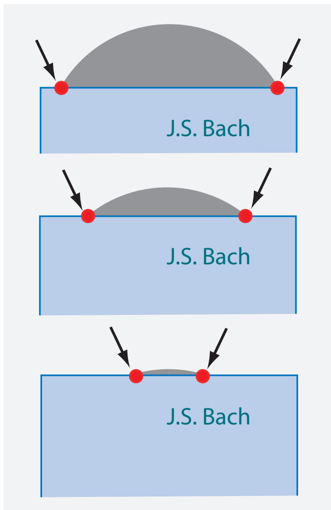


FIGURE 37 Another example of faster-than-light motion.

disappearing into its sleeve, as shown in Figure 37. 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 the horizon moves away from you can be larger than that of light? Challenge 74 s

A further standard example is the motion of a spot of light produced by shining a laser

Ref. 71 beam onto the Moon. If the laser beam is deflected, 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. In fact, when a beam is swept across an inclined surface, the spot can move backwards, split and recombine. Researchers are still looking for such events both in the universe and in the laboratory.

Finally, here is the simplest example of all. Imagine to switch on a light bulb in front of a wall. Ref. 71 During the switch-on process, the boundary between the illuminated surface and the surface that is still dark moves with a speed higher than the speed of light. *Light bulbs produce superluminal speeds.*

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 75 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'. Vol. III, page 133 We will discuss this surprising phenomenon in more detail later on. 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

$$t = l/v - l/c \,. \tag{23}$$

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

$$v_{appr} = \frac{l}{t} = \frac{vc}{c-v} \,, \tag{24}$$

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. 72

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

$$v_{leav} = \frac{vc}{c+v} \,, \tag{25}$$

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

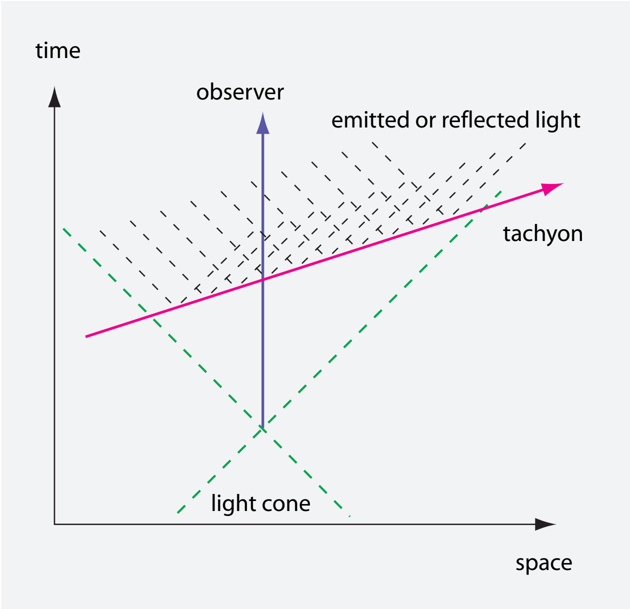


FIGURE 38 Hypothetical space-time diagram for tachyon observation.

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 38. One image would be red-shifted, the other blue-shifted. 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? Nobody has ever seen such phenomena.

Ref. 73

Challenge 76 e

Tachyons, if they existed, would be strange objects: they would accelerate when they 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 adventure. 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.

Ref. 74
Page 73

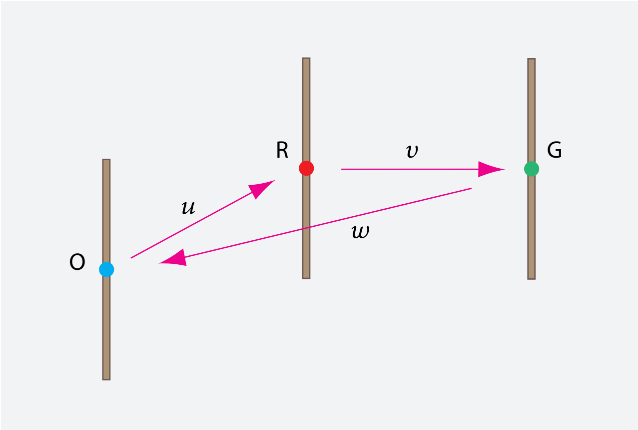


FIGURE 39 If O's stick is parallel to R's and R's is parallel to G's, then O's stick and G's stick are not.

Parallel to parallel is not parallel – Thomas precession

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 39. In special relativity, a general concatenation 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. Ref. 75

An example of this effect appears in rotating motion. Imagine that we walk in a circle with relativistic speed and hold 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. The effect 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

$$T = T_0\sqrt{1 - v^2/c^2}\ . \qquad (26)$$

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 that the Boltzmann constant is a relativistic invariant. Ref. 76

Not all researchers agree on the expression for the transformation of energy, however. (They do agree on the invariance 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 proposed. Ref. 77 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 for a moving observer 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. And similar issues appear for the relativistic transformation of entropy.) 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.

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 40. 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

$$t_3 - t_1 = 2\,(t_2 - t_1)\;. \qquad (27)$$

Here it is assumed that the clocks have been synchronised according to the prescription on page 51. 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. 78, Ref. 79 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? Challenge 77 s

Summary

For all physical systems, the locally measured energy speed, the forerunner speed and the measured signal speed are limited by $c = 299\,792\,458$ m/s, the speed of light in vacuum. As a result, time, age, distance, length, colour, spatial orientation, angles and temperature

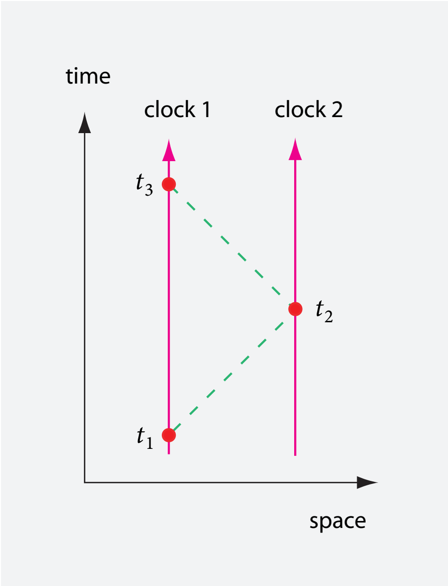
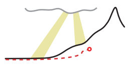


FIGURE 40 Clocks and the measurement of the speed of light as two-way velocity.

– as long as it can be defined – depend on the observer. In contrast, the speed of light in vacuum c is invariant.



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*.

The exploration of relativistic mechanics will first lead us to the *equivalence of mass and energy*, a deep relation that is the basis of the understanding of motion. Relativistic mechanics will also lead us to the concept of *horizon*, a concept that we will need later to grasp the details of black holes, the night sky and the universe as whole.

MASS IN RELATIVITY

Vol. I, page 100

In Galilean physics, the mass ratio between two bodies was defined using *collisions*. More precisely, *mass* was given by the negative inverse of the velocity change ratio

$$\frac{m_2}{m_1} = -\frac{\Delta v_1}{\Delta v_2} \,. \qquad (28)$$

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

Challenge 78 s

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

Ref. 80

$$\sum_i \gamma_i m_i \boldsymbol{v}_i = \text{const} \qquad (29)$$

and

$$\sum_i \gamma_i m_i = \text{const} \,. \qquad (30)$$

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

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

Challenge 79 s

Obviously, in order to recover Galilean physics, the relativistic correction factors γ_i

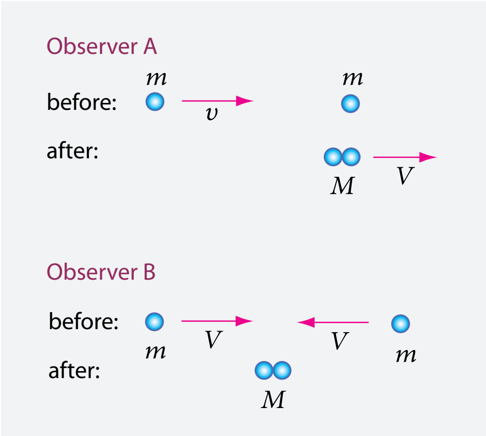


FIGURE 41 An *inelastic* collision of two identical particles seen from two different inertial frames of reference.

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 41.

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 V gives v, in other words, that

Challenge 80 e

$$v = \frac{2V}{1 + V^2/c^2} \; . \tag{31}$$

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

$$\gamma_v = \frac{1}{\sqrt{1 - v^2/c^2}} \; . \tag{32}$$

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

$$\frac{m_1}{m_2} = -\frac{\Delta(\gamma_2 v_2)}{\Delta(\gamma_1 v_1)} \; . \tag{33}$$

This is the generalization of the definition of mass ratio from Galilean physics. The correction factors γ_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. (In the chapter on Galilean mechanics we also used a second, generalized mass definition based on acceleration ratios. We do not explore its relativistic generalization because it contains some subtleties which we will encounter shortly.)

Vol. I, page 103

Challenge 81 e

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

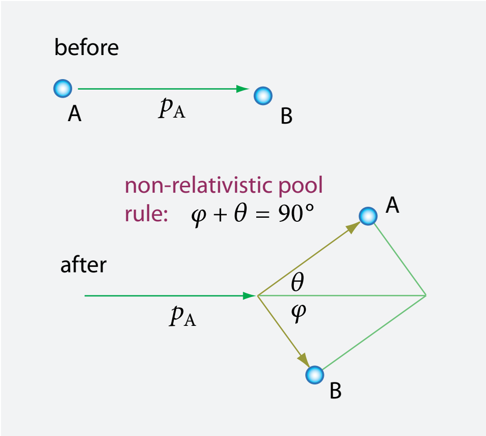


FIGURE 42 A useful rule for playing non-relativistic snooker – and to predict non-relativistic *elastic* collisions.

Following the example of Galilean physics, we call the quantity

$$\boldsymbol{p} = \gamma m \boldsymbol{v} \tag{34}$$

the (*linear*) *relativistic* (*three-*) *momentum* of a particle. Total momentum is a *conserved* 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 by experimental data, as was shown in Figure 19. Page 38

Now that we have the correct definitions of mass and momentum, we can explore collisions in more detail.

Why relativistic snooker is more difficult

A well-known property of collisions between a moving sphere or particle and a resting one of the *same mass* 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 42, the two angles φ and θ add up to a right angle. (The only exception to this rule is the case that the collision is exactly head on; in that case the first sphere simply stops.) Challenge 82 e

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 calculate that Challenge 83 e

$$\tan\theta \tan\varphi = \frac{2}{\gamma + 1} , \tag{35}$$

where the angles are defined in Figure 43. 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,

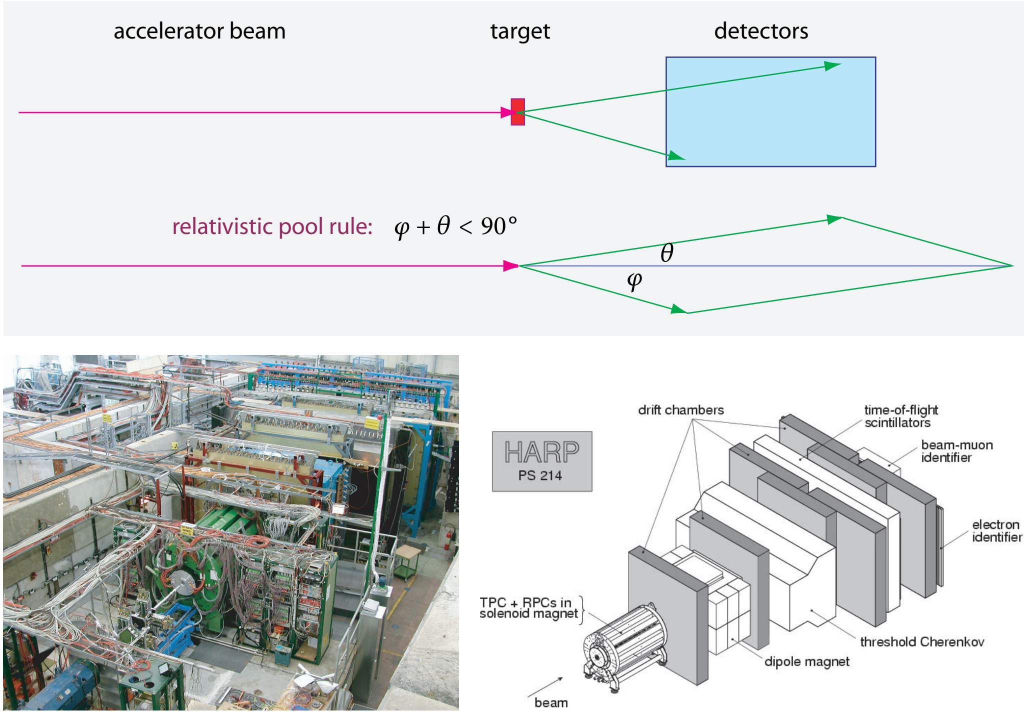


FIGURE 43 The dimensions of detectors for particle accelerators with single beams are based on the relativistic snooker angle rule – as an example, the HARP experiment at CERN (© CERN).

which show the tracks left by particles when they move through them, as shown in Figure 44. All such photographs confirm the relativistic expression. In fact, the *shapes* of detectors are chosen according to expression (35), as sketched in Figure 43. 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 particle experiments also prove the formula for the composition of velocities. Can you show this?

Ref. 18

Challenge 84 e

Mass and energy are equivalent

Page 66

Challenge 85 s

Let us go back to the collinear and inelastic collision of Figure 41. What is the mass M of the final system? Calculation shows that

$$M/m = \sqrt{2(1+\gamma_v)} > 2 \,. \tag{36}$$

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

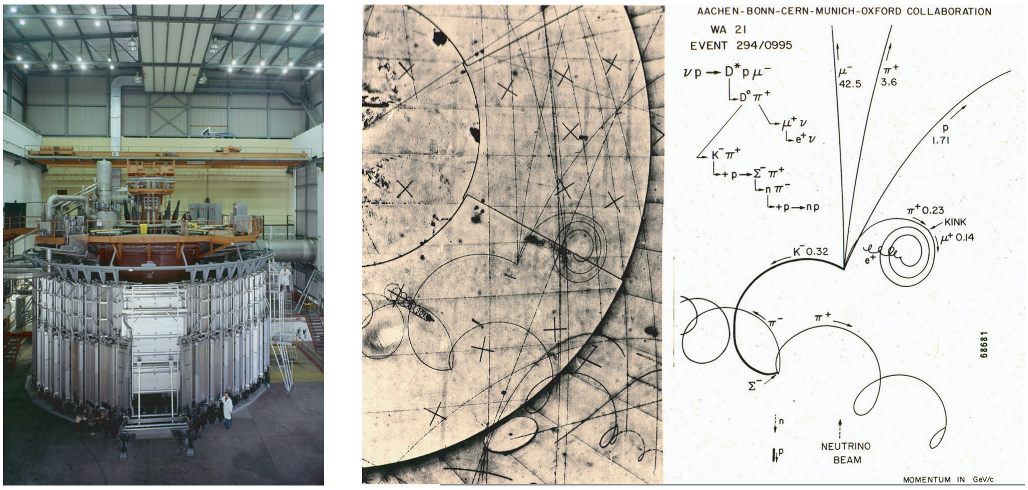


FIGURE 44 The 'Big European Bubble Chamber' and an example of tracks of relativistic particles it produced, with the momentum values deduced from the photograph (© CERN).

energy E of an object of mass m and velocity v, we use the expression

$$E = c^2 \gamma m = \frac{c^2 m}{\sqrt{1 - v^2/c^2}} \,, \qquad (37)$$

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

$$E_0 = c^2 m \,. \qquad (38)$$

Why do we write $c^2 m$ instead of mc^2? Because in formulae, constant factors come always *first*. The factor c^2 is not central; the essence of the expression is the relation between energy E and mass m. c^2 is simply the conversion factor between the two quantities.

The mass-energy relation $E = c^2 \gamma m$ is one of the most beautiful and famous discoveries of modern physics. In simple 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:

▷ Mass and energy 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. They can fall.

Conversely, *mass has all properties of energy.* For example, we can use mass to make engines run. But this is no news, as the process is realized in every engine we know of! Muscles, car engines, and 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.

The conversion factor c^2 is large: 1 kg of rock, if converted to electric energy, would be worth around 8 000 million Euro. In this unit, even the largest financial sums correspond to modest volumes of rock. Since c^2 is so large, we can also say:

▷ 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

$$T = c^2\gamma m - c^2 m = \frac{1}{2}mv^2 + \frac{1\cdot 3}{2\cdot 4}m\frac{v^4}{c^2} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}m\frac{v^6}{c^4} + ... \qquad (39)$$

Challenge 86 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 $E = c^2\gamma m$ implies that extracting *any* energy from a material system results in a mass decrease. When a person plays the piano, thinks or runs, her 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. Indeed, the energy content of even a speck of dust is so substantial that the annihilation with the same amount of antimatter would already be a dangerous event. Challenge 87 e

The equivalence of mass and energy suggests that it is possible to 'create' massive particles by manipulating light or by extracting kinetic energy in collisions. This is indeed correct; the transformation of other energy forms into matter particles is occurring, as we speak, in the centre of galaxies, in particle accelerators, and whenever a cosmic ray hits the Earth's atmosphere. The details of these processes will become clear when we explore quantum physics.

The mass–energy equivalence $E = c^2\gamma m$ 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. All had a negative result. *There is no freely available energy in nature.*

In summary, the mass-energy equivalence is a fact of nature. But many scientists cannot live long without inventing mysteries. Two different, extremely diluted forms of energy, called *dark matter* and (confusingly) *dark energy*, were found to be distributed throughout the universe in the 1990s, with a density of about 1 nJ/m^3. Their existence Page 220 is deduced from quite delicate measurements in the sky that detected their mass. Both dark energy and dark matter must have mass and particle properties. But so far, their nature and origin has not yet been resolved.

Weighing light

The mass–energy equivalence $E = c^2\gamma m$ also implies that one needs about 90 thousand Challenge 88 e 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.

The mass–energy equivalence $E = c^2\gamma m$ has been confirmed by all experiments performed so far. The measurement is simplest for the nuclear mass defect. The most precise Ref. 81 experiment, from 2005, compared the masses difference of nuclei before and after neutron capture on the 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.

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 research 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 ex- Ref. 82 periments is sufficient to confirm $\Delta E_0 = c^2\Delta m$ 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.

In fact, 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 non-relativistic 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 89 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

$$E_0 = c^2 m \,. \qquad (40)$$

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

How are momentum and energy related? The definitions of momentum (34) and energy (37) lead to two basic relations. First of all, their *magnitudes* are related by

Challenge 90 e

$$c^4 m^2 = E^2 - c^2 p^2 \qquad (41)$$

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

$$\boldsymbol{p} = \frac{E}{c^2}\boldsymbol{v} \,, \qquad (42)$$

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 adventure, including the following discussion.

Challenge 91 e

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 therefore described by a space-time diagram such as the left-hand one in Figure 45; it is reminiscent of the Orion constellation. It is easy to check that the process described by such a diagram is, according to the above definition, a collision.

Ref. 83

Challenge 92 e

The right-hand side of Figure 45 shows the same process seen from another, Greek, frame of reference. The Greek observer says that the first object has changed its mo-

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

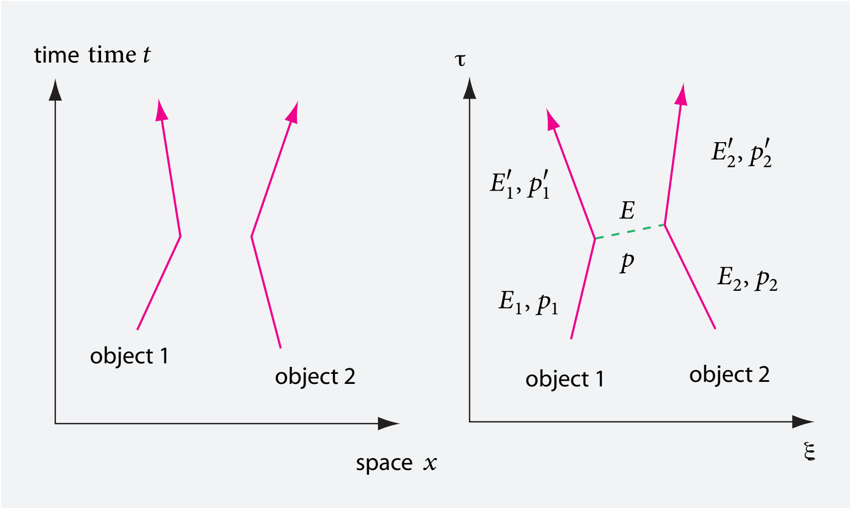


FIGURE 45 Space-time diagrams of the same collision for two different observers.

mentum *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, and give them a prime after the collision. Then the unknown mass m obeys

Challenge 93 e

$$m^2c^4 = (E_1 - E_1')^2 - (p_1 - p_1')^2c^2 = 2m_1^2c^4 - 2E_1E_1'\left(\frac{1 - v_1v_1'}{c^2}\right) < 0\ . \qquad (43)$$

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 particles later on, when we come to discuss quantum theory.

Vol. IV, page 64

* It is usual to change the mass–energy and mass–momentum relation of tachyons to $E = \pm c^2 m/\sqrt{v^2/c^2 - 1}$ 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.

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 94 e

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 64

Ref. 84

Challenge 95 e

As a curiosity, we mention that the notion of *relative velocity* exists also in relativity. Given two particles 1 and 2, the magnitude of the relative velocity is given by

$$v_{\text{rel}} = \frac{\sqrt{(\mathbf{v}_1 - \mathbf{v}_2)^2 - (\mathbf{v}_1 \times \mathbf{v}_2)^2/c^2}}{1 - \mathbf{v}_1 \cdot \mathbf{v}_2/c^2} . \tag{44}$$

The value is never larger than c, even if both particles depart into opposite directions with ultrarelativistic speed. The expression is also useful for calculating the relativistic cross sections for particle collisions. If we determine the relative 4-velocity, we get the interesting result that in general, $\mathbf{v}_{12} \neq -\mathbf{v}_{21}$, i.e., the two relative velocities do not point in opposite directions – except when the particle velocities are collinear. Nevertheless, the relation $v_{12} = v_{21} = v_{\text{rel}}$ is satisfied in all cases.

Challenge 96 e

There is an additional secret hidden in collisions. In the right-hand side of Figure 45, 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. 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.

Challenge 97 s

Vol. IV, page 192

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 46 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. 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.

Ref. 85

Will the issues with the centre of mass hinder us in our adventure? No. We are more interested in the motion of single particles than that of composite objects or systems.

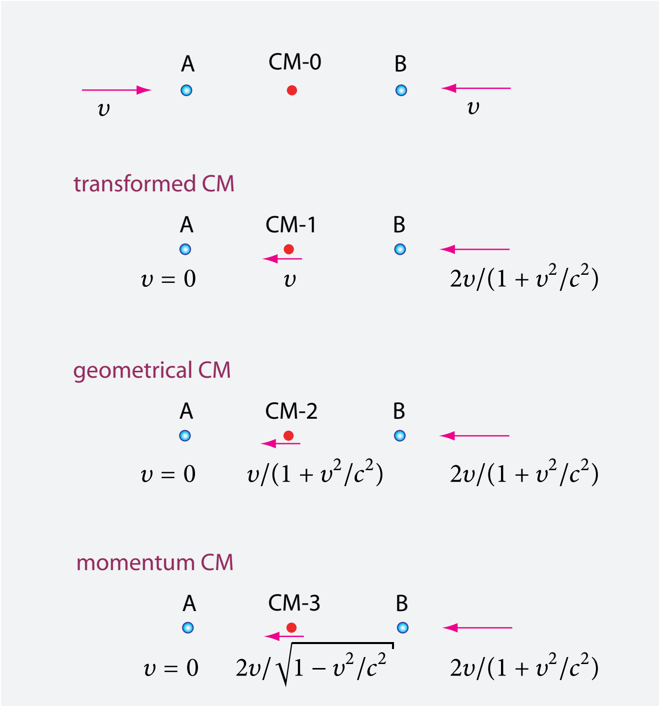


FIGURE 46 There is no consistent way to define a relativistic centre of mass.

Why is most motion so slow?

For most everyday systems, dilation factors γ 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 particle accelerator or inside a cathode ray tube found in the first colour televisions. The particles making up cosmic radiation are another example; cosmic rays are 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 $E = c^2\gamma m$. 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 98 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 230 with slow motion because it is *old*. In fact, 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

$$E = c^2 \gamma m \qquad (45)$$

which is often called the most famous formula of physics. We write it in this slightly unusual, but clear way to stress that c^2 is a unit-dependent and thus *unimportant* factor. Such factors are always put first in physical formulae.* Einstein published this formula
Ref. 19 in a 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, *be-*
Ref. 86 *fore* Einstein's first relativity paper, Olinto De Pretto, a little-known Italian engineer, calculated, discussed and published the formula $E = c^2 m$. 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. 86 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 $E = c^2 m$ 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. Also Tolver Preston discussed the equivalence of mass and energy, already in 1875, in his book *Physics of the Ether*. The real hero of the story might be the Swiss chemist Jean Charles Gallisard de Marignac; already in 1861 he published the now accepted idea about the formation of the elements: whenever protons form elements, the condensation leads to a lower total mass, and the energy difference is emitted as energy. The mass–energy equivalence was thus indeed floating in the air, waiting to be understood and put into the correct context.

Vol. V, page 146 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 99 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, even for those moving at speeds near that of light? We have to introduce a simple idea: 4-vectors. We already know

* Examples are $A = 4\pi r^2$, $a = Gm/r^2$, $U = RI$, $F = (1/4\pi\varepsilon_0)Q^2/r^2$, $pV = RT$ or $S = k \ln W$.

** Umberto Bartocci, mathematics professor of the University of Perugia in Italy, published the details of
Ref. 87 this surprising story in several papers and in a book.

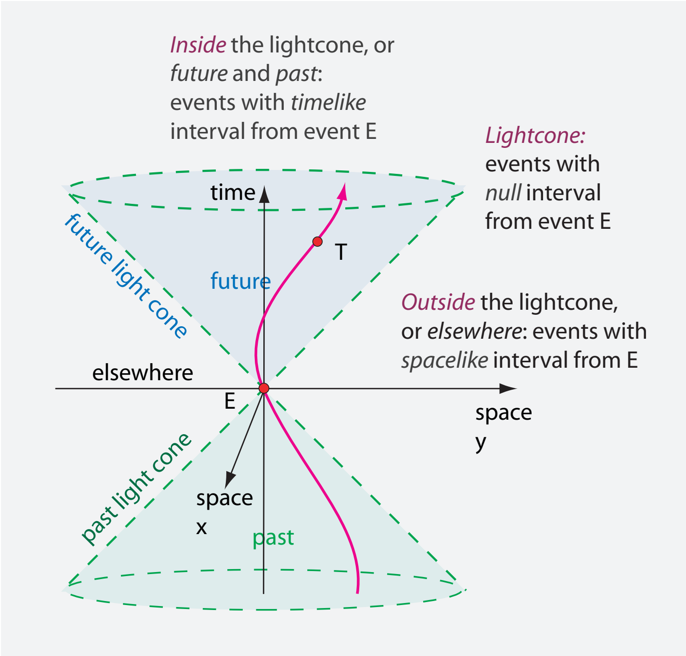


FIGURE 47 The space-time diagram of a moving object T, with one spatial dimension missing.

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

$$\boldsymbol{X} = (ct, \boldsymbol{x}) = (ct, x, y, z) = X^i \ . \tag{46}$$

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 example of a *4-vector*. The usual 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

$$\boldsymbol{X}^2 = {X_0}^2 - {X_1}^2 - {X_2}^2 - {X_3}^2 = c^2t^2 - x^2 - y^2 - z^2 = X_a X^a = \eta_{ab} X^a X^b = \eta^{ab} X_a X_b \ . \tag{47}$$

The squared space-time interval is thus the squared time interval *minus* the squared Page 42 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 *light-like* intervals. We call the set of all null vector end points the light cone; it is shown in Figure 47. 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 time-like. For negative space-time intervals the interval is called space-like. 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 47

We note that the definition of the light cone, its interior and its exterior, are *observer-invariant*. We therefore use these concepts regularly.

Challenge 100 e

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 *super*scripts and coordinates with *sub*scripts. (For 3-vectors, we only use subscripts.) They are related by the following general relation

$$\begin{aligned} X^b &= (ct, x, y, z) \\ X_a &= (ct, -x, -y, -z) = \eta_{ab} X^b \; , \end{aligned} \tag{48}$$

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

$$\eta^{ab} = \eta_{ab} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{49}$$

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* τ, 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

Ref. 88

* This is the so-called *time-like convention*, used in about 70 % of all physics texts worldwide. Note that 30 % of all physics textbooks use the negative of η as the metric, the so-called *space-like convention*, and thus have opposite signs in this definition.

change of its 4-coordinates $\boldsymbol{X} = (ct, \boldsymbol{x})$ with respect to proper time, i.e., as

$$\boldsymbol{U} = \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}\tau} . \tag{50}$$

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

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma\frac{\mathrm{d}x}{\mathrm{d}t} , \quad \text{where as usual} \quad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}} , \tag{51}$$

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

$$U^0 = \gamma c\, , U^i = \gamma v_i \quad \text{or} \quad \boldsymbol{U} = (\gamma c, \gamma \boldsymbol{v}) . \tag{52}$$

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 time-like 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

$$\begin{aligned} H^0_V &= \gamma_V (H^0 - H^1 V/c) \\ H^1_V &= \gamma_V (H^1 - H^0 V/c) \\ H^2_V &= H^2 \\ H^3_V &= H^3 \end{aligned} \tag{53}$$

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 (10) from this definition? Challenge 101 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 102 s

4-acceleration and proper acceleration

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

$$\boldsymbol{B} = \frac{\mathrm{d}\boldsymbol{U}}{\mathrm{d}\tau} = \frac{\mathrm{d}^2\boldsymbol{X}}{\mathrm{d}\tau^2} . \tag{54}$$

Using $\mathrm{d}\gamma/\mathrm{d}\tau = \gamma \mathrm{d}\gamma/\mathrm{d}t = \gamma^4 \boldsymbol{v}\boldsymbol{a}/c^2$, we get the following relations between the four com-
Ref. 89 ponents of $\boldsymbol{B}$ and the 3-acceleration $\boldsymbol{a} = \mathrm{d}\boldsymbol{v}/\mathrm{d}t$:

$$B^0 = \gamma^4 \frac{\boldsymbol{v}\boldsymbol{a}}{c} \quad , \quad B^i = \gamma^2 a_i + \gamma^4 \frac{(\boldsymbol{v}\boldsymbol{a})v_i}{c^2} \ . \tag{55}$$

Challenge 103 e The magnitude B of the 4-acceleration is easily found via $\boldsymbol{B}\boldsymbol{B} = \eta_{cd} B^c B^d = -\gamma^4(a^2 + \gamma^2(\boldsymbol{v}\boldsymbol{a})^2/c^2) = -\gamma^6(a^2 - (\boldsymbol{v} \times \boldsymbol{a})^2/c^2)$. 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 space-like vector. We also note that $\boldsymbol{B}\boldsymbol{U} = \eta_{cd} B^c U^d = 0$, 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* ob-
Ref. 91 server is related to the proper acceleration $\boldsymbol{a}_c$ measured by the comoving observer using expressions (55) and (53). In this case $\boldsymbol{v}$ is both the relative speed of the two observers

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

$$\boldsymbol{J} = \mathrm{d}\boldsymbol{B}/\mathrm{d}\tau = \mathrm{d}^2\boldsymbol{U}/\mathrm{d}\tau^2 \ . \tag{56}$$

Challenge 104 e For the relation with the 3-jerk $\boldsymbol{j} = \mathrm{d}\boldsymbol{a}/\mathrm{d}t$ we then get

$$\boldsymbol{J} = (J^0, J^i) = \left(\frac{\gamma^5}{c}(\boldsymbol{j}\boldsymbol{v} + a^2 + 4\gamma^2 \frac{(\boldsymbol{v}\boldsymbol{a})^2}{c^2}) \ , \ \gamma^3 j_i + \frac{\gamma^5}{c^2}((\boldsymbol{j}\boldsymbol{v})v_i + a^2 v_i + 4\gamma^2 \frac{(\boldsymbol{v}\boldsymbol{a})^2 v_i}{c^2} + 3(\boldsymbol{v}\boldsymbol{a})a_i) \right) \tag{57}$$

Challenge 105 e Page 94 which we will use later on. Surprisingly, $\boldsymbol{J}$ does not vanish when the 3-jerk $\boldsymbol{j}$ vanishes. Why not? For this
Ref. 90 reason, slightly amended definitions of 4-jerk have been proposed.

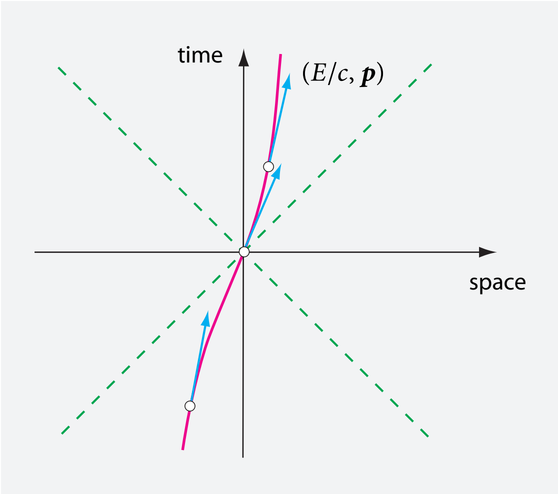


FIGURE 48 Energy–momentum is *tangent* to the world line.

and the speed of the accelerated particle. We get

$$a^2 = \frac{1}{\gamma_v^4}\left(a_c^2 - \frac{(\boldsymbol{a}_c \boldsymbol{v})^2}{c^2}\right) , \qquad (58)$$

Page 80 which we know already in a slightly different form. It shows (again):

▷ The comoving or proper 3-acceleration is always *larger* than the 3-acceleration measured by any other inertial observer.

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 $1/\gamma_v^2$, whereas a speed parallel to the acceleration yields a factor $1/\gamma_v^3$. Challenge 106 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.

4-momentum or energy–momentum or momenergy

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

$$\boldsymbol{P} = m\boldsymbol{U} \qquad (59)$$

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

$$\boldsymbol{P} = (\gamma mc, \gamma m\boldsymbol{v}) = (E/c, \boldsymbol{p}) . \qquad (60)$$

For this reason 4-momentum is also called the *energy–momentum* 4-vector. In short,

▷ The 4-momentum or energy–momentum of a body is given by its mass times the 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 48, follows directly from the definition, since

$$(E/c, \boldsymbol{p}) = (\gamma mc, \gamma m\boldsymbol{v}) = m(\gamma c, \gamma \boldsymbol{v}) = m(c\mathrm{d}t/\mathrm{d}\tau, \mathrm{d}\boldsymbol{x}/\mathrm{d}\tau) \ . \qquad (61)$$

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

$$E^2/c^2 - p^2 = c^2 m^2 \ , \qquad (62)$$

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 $E_0 = c^2 m$. A particle whose kinetic energy is *much* higher than its rest mass is called *ultrarelativistic*. Particles in accelerators or in cosmic rays fall into
Challenge 107 s this category. What is their energy–momentum relation?

The conservation of energy, momentum and mass of Galilean mechanics thus merge, in special relativity, into the conservation of momenergy:

▷ In nature, energy–momentum, or 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 $c^2 m$ from a system of mass m at rest. In particular, an absolute 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 γ_{12} is the Lorentz factor due to their relative velocity
Challenge 108 s v_{12}?

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 γm the *relativistic mass*. Workers in the field usually (but not
Ref. 92 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.

4-force – and the nature of mechanics

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

$$\boldsymbol{K} = \mathrm{d}\boldsymbol{P}/\mathrm{d}\tau = m\boldsymbol{B}\ , \tag{63}$$

Ref. 89, Ref. 93
where $\boldsymbol{B}$ is 4-acceleration. 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*

$$\boldsymbol{K} = (K^0, K^i) = \left(\gamma^4 \frac{m\boldsymbol{va}}{c}, \gamma^2 m a_i + \gamma^4 v_i \frac{m\boldsymbol{va}}{c^2}\right) = \left(\frac{\gamma}{c}\frac{\mathrm{d}E}{\mathrm{d}t}, \gamma\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}\right) = \left(\gamma\frac{\boldsymbol{Fv}}{c}, \gamma\boldsymbol{F}\right)\ . \tag{64}$$

Challenge 109 e
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 110 s
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, we saw above that 3-momentum is parallel, but not proportional to 3-velocity.

Challenge 111 e
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, and reduces the components in the perpendicular directions. In particular, boost cannot be used to increase 3-force values beyond all bounds. (Though they appear to allow to increase the value of 4-force beyond all bounds.) The situation somewhat resembles the situation for 3-acceleration, though the transformation behaviour differs.
Page 80

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 for all observers – except if the potentials are related to fields that can carry energy and momentum. In other terms,
Ref. 94

▷ Relativity only allows potentials related to radiation.

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

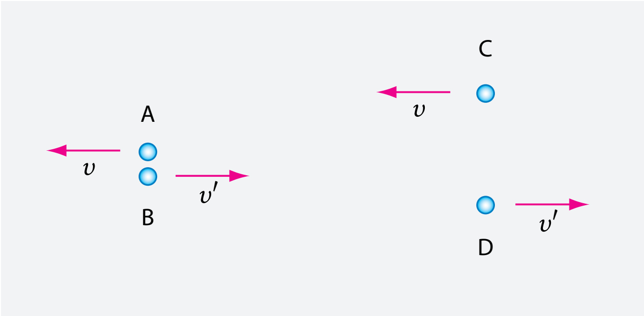


FIGURE 49 On the definition of relative velocity (see text).

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 any 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 $\boldsymbol{F}\boldsymbol{v}$ are limited in magnitude. Can you guess how?

Challenge 112 d

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 49. 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 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 more examples shortly.

Page 60
Page 100

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

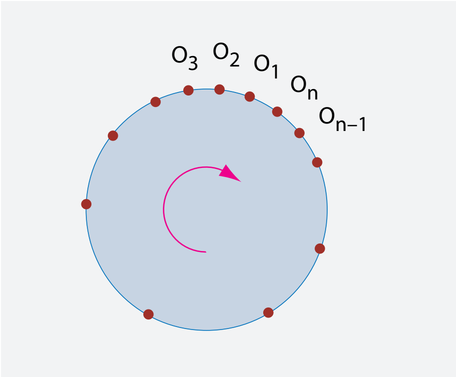


FIGURE 50 Observers on a rotating object.

Challenge 113 e

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 114 e

Ref. 95

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 50. 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. 26

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 115 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 it 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 116 s

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 117 s

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

$$l^{ab} = x^a p^b - x^b p^a \,. \qquad (65)$$

The two indices imply that the 4-angular momentum is a *tensor*, not a vector. Angular

Challenge 118 e 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 seen, is rotating? Challenge 119 s

For a rotating particle, the rotational energy is part of the rest mass. You may want to calculate the fraction for the Earth and the Sun. It is not large. Challenge 120 e

Here are some puzzles about relativistic rotation. We know that velocity is relative: its measured value depends on the observer. Is this the case also for angular velocity? Challenge 121 s What is the expression for relativistic rotational energy, and for its relation to 4-angular momentum? Challenge 122 s

Rotation also yields the *rotational Doppler effect*. To observe it is tricky but nowadays a regular feat in precision laser laboratories. To see it, one needs a circularly polarized light beam; such beams are available in many laboratories. When such a light beam is reflected from a polarizable rotating surface, the frequency of the reflected beam is shifted in a certain percentage of the light. This rotational Doppler shift is given by the rotation frequency of the surface. The effect is important in the theory of the Faraday effect; it has already been used to measure the rotation of various optical elements and even the rotation of molecules. One day, the effect might be useful in engineering or astronomy, to measure the rotation velocity of distant or delicate spinning objects.

Wave motion

Vol. I, page 293 Waves also move. 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

$$L^a = \left(\frac{\omega}{c}, \boldsymbol{k}\right) . \tag{66}$$

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 4-vector is time-like. Challenge 123 e

The *phase* φ of a wave can now be defined as

$$\varphi = L_a x^a = L^a x_a . \tag{67}$$

Being a scalar, as expected, the phase of any wave, be it light, sound or any other type, is the same for all observers: *the phase is a relativistic invariant.** Challenge 124 e

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

$$\nu = \boldsymbol{L}\boldsymbol{U} \tag{68}$$

must be obeyed. Challenge 125 s

* 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 $(\mathrm{d}\omega/c, \mathrm{d}\mathbf{k})(c, \mathbf{v}_{\text{group}}) = 0$.

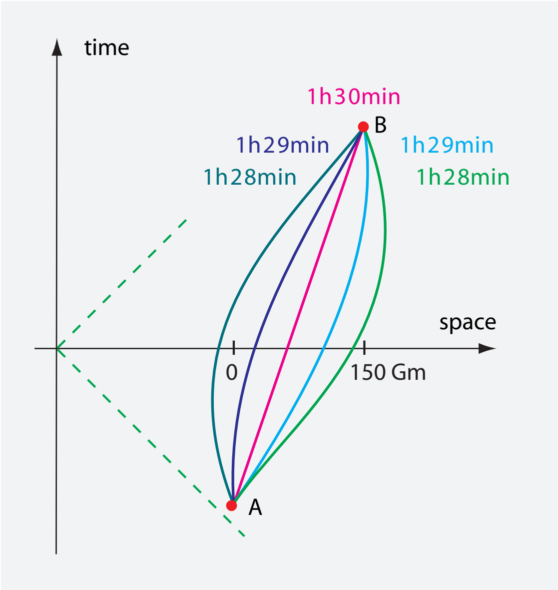


FIGURE 51 The straight motion between two points A and B is the motion that requires the longest proper time.

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

Ref. 25
Challenge 126 ny

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 principle, we need a definition of the *action*. We already know that physical action is a 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

Vol. I, page 248

$$S = -c^2 m \int_{\tau_1}^{\tau_2} \mathrm{d}\tau \;, \tag{69}$$

where τ is the proper time along its path. This is indeed the correct expression.

In short, in nature,

▷ *All* particles move in such a way that the elapsed proper time – or wristwatch 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 this the *'law' of cosmic laziness*.

Vol. I, page 253

Using the invariance of the speed of light, the principle of least action can thus be rephrased:

▷ Bodies idle as much as they can.

Figure 51 shows some examples of values of proper times for a body moving from one point to another in free space. The straight motion, the one that nature chooses, is the Page 49 motion with the *longest* proper time. (Recall the result given above: travelling more keeps you younger.) However, this difference in proper time is noticeable only for relativistic speeds and large distances – such as those shown in the figure – and therefore we do not experience any such effect in everyday, non-relativistic life.

We note that maximum proper time is equivalent to minimum action. Both statements have the same content. Both statements express the principle of least action. For a free body, the change in proper time is maximal, and the action minimal, for straight-line motion with *constant* velocity. The principle of least action thus implies conservation of Challenge 127 e (relativistic) energy and momentum. Can you confirm this?

The expression (69) 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 k, is a relativistic invariant (like the Boltzmann constant). Can you Challenge 128 s imagine how he did this?

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:

$$S = \int L \,\mathrm{d}t = -c^2 m \int_{t_1}^{t_2} \frac{1}{\gamma}\,\mathrm{d}t = -mc \int_{\tau_1}^{\tau_2} \sqrt{u_a u^a}\,\mathrm{d}\tau = -mc \int_{s_1}^{s_2} \sqrt{\eta^{ab} \frac{\mathrm{d}x_a}{\mathrm{d}s} \frac{\mathrm{d}x_b}{\mathrm{d}s}}\,\mathrm{d}s\,, \quad (70)$$

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

$$\eta^{ab} = \eta_{ab} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (71)$$

You can easily confirm the form of the action (70) by deducing the equation of motion Challenge 129 e in the usual way.

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

Vol. I, page 248
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 131 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 you stay younger, as we will see. The actual path is thus indeed the path of maximum aging.

Page 149

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

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* λ. The transformations among these observers or points of view are given by

$$x_a \mapsto \lambda x_a \tag{72}$$

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*

$$x_a \mapsto \frac{x_a}{x^2} \tag{73}$$

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

$$x_a \mapsto x_a + b_a \ , \tag{74}$$

Challenge 132 e
and a second inversion. Therefore the special conformal transformations are

$$x_a \mapsto \frac{x_a + b_a x^2}{1 + 2b_a x^a + b^2 x^2} \ . \tag{75}$$

These transformations are called *conformal* because they do not change angles of (infin-

Challenge 133 e itesimally) 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.*

Note that the way in which special conformal transformations leave light cones invariant is rather subtle. Explore the issue! Challenge 135 e

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 137 e

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 first glimpse of the adventures that we will encounter in the final part of our adventure.

FIGURE 52 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.)

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. 96 of an inertial, Roman, one, and vice versa. We assume that the Greek observer, shown in

Challenge 134 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 Vol. V, page 358 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.

** 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 Challenge 136 e invariant?

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.

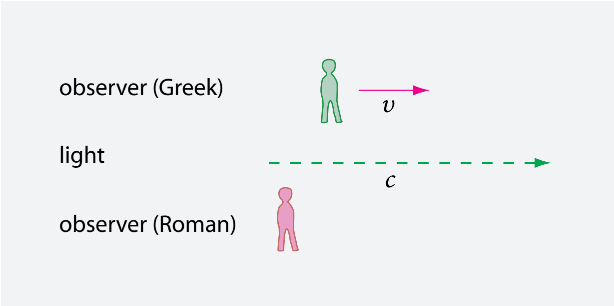


FIGURE 53 The simplest situation for an inertial and an accelerated observer.

Figure 53, 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 τ_1 and τ_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 by

$$\frac{\Delta\tau}{\Delta t} = \frac{d\tau}{dt} = \sqrt{1 - v^2/c^2} = \frac{1}{\gamma_v} , \qquad (76)$$

Challenge 138 e

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 reasoning is necessary. The Greek/Roman clock rate ratio is $d\tau/dt$, and τ 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

Ref. 96

$$\frac{\tau}{\gamma_v} = t - \boldsymbol{x}(t)\boldsymbol{v}(t)/c^2 \qquad (77)$$

and thus

$$\frac{\tau + d\tau}{\gamma_v} = (t + dt) - [\boldsymbol{x}(t) + dt\boldsymbol{v}(t)][\boldsymbol{v}(t) + dt\boldsymbol{a}(t)]/c^2 . \qquad (78)$$

Together, and to first order, these equations yield

$$\text{`}d\tau/dt\text{'} = \gamma_v(1 - \boldsymbol{v}\boldsymbol{v}/c^2 - \boldsymbol{x}\boldsymbol{a}/c^2) . \qquad (79)$$

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

$$\text{`}dt/d\tau\text{'} = \gamma_v , \qquad (80)$$

* These sets form what mathematicians call *hypersurfaces*.

FIGURE 54 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.)

which is *not* the inverse of equation (79). This difference becomes most apparent in the simple case of two clocks with the same velocity, one of which has a constant acceleration
Ref. 96 g towards the origin, whereas the other moves inertially. We then have

$$`\mathrm{d}\tau/\mathrm{d}t' = 1 + gx/c^2 \qquad (81)$$

and

$$`\mathrm{d}t/\mathrm{d}\tau' = 1 \ . \qquad (82)$$

Page 99 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 – 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 an 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 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. 97

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.

Ref. 97 Challenge 139 e Page 54

Can you state how the acceleration ratio enters into the definition of mass in special relativity?

Page 66 Challenge 140 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

$$\boldsymbol{B} \cdot \boldsymbol{B} = -g^2 \,, \tag{83}$$

Ref. 99

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 141 e

$$\gamma^3 \boldsymbol{a} = \boldsymbol{g} \quad \text{or} \quad \frac{\mathrm{d}\gamma \boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{g} \,. \tag{84}$$

Ref. 98

* 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 $r^2 = x^2 + y^2$.

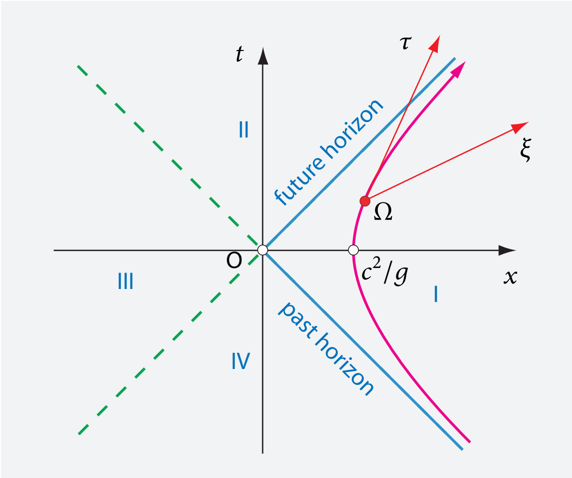


FIGURE 55 The hyperbolic motion of an observer Ω that accelerates rectilinearly and uniformly with acceleration g.

Challenge 142 e Taking the direction we are talking about to be the x-axis, and solving for $v(t)$, we get

$$v = \frac{gt}{\sqrt{1 + \frac{g^2t^2}{c^2}}} , \tag{85}$$

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 143 e

$$x(t) = \frac{c^2}{g}\sqrt{1 + \frac{g^2t^2}{c^2}} , \tag{86}$$

where we assumed that $x(0) = c^2/g$, in order to keep the expression simple. Because of this result, visualized in Figure 55, 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 τ 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 (85) we get*

Ref. 99, Ref. 100

$$t = \frac{c}{g}\sinh\frac{g\tau}{c} \quad \text{and} \quad x = \frac{c^2}{g}\cosh\frac{g\tau}{c} \tag{87}$$

Ref. 101 * 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 $\int dy/\sqrt{y^2 + a^2} = \text{arsinh}\ y/a = \text{Arsh}\ y/a = \ln(y + \sqrt{y^2 + a^2})$.

for the relationship between proper time τ 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 $g = 10\,\mathrm{m/s^2}$ for the proper time τ 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 144 s

For uniform rectilinear acceleration, the coordinates transform as

$$\begin{aligned} t &= \left(\frac{c}{g} + \frac{\xi}{c}\right) \sinh \frac{g\tau}{c} \\ x &= \left(\frac{c^2}{g} + \xi\right) \cosh \frac{g\tau}{c} \\ y &= \upsilon \\ z &= \zeta \,, \end{aligned} \tag{88}$$

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

$$\mathrm{d}\sigma^2 = (1 + g\xi/c^2)^2 c^2 \mathrm{d}\tau^2 - \mathrm{d}\xi^2 - \mathrm{d}\upsilon^2 - \mathrm{d}\zeta^2 = c^2 \mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \,, \tag{89}$$

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

Ref. 102

After this forest of formulae, let's tackle a simple question, shown in Figure 55. The inertial, Roman observer O sees the Greek observer Ω departing under continuous acceleration, moving further and further away, following equation (86). 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 Ω 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 OΩ. Using expression (86), we see that this is

Challenge 145 e

Ref. 103

$$d_{\mathrm{O}\Omega} = \sqrt{\xi^2} = \sqrt{x^2 - c^2 t^2} = c^2/g \,, \tag{90}$$

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 146 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 ξ and τ of the rigidly accelerated frame, of an object located at the departure

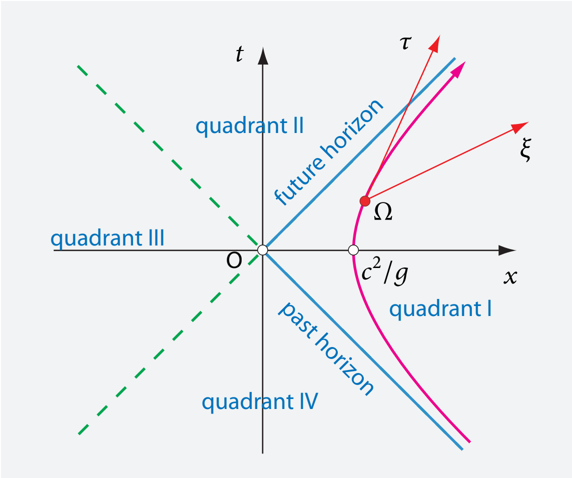


FIGURE 56 Hyperbolic motion and event horizons.

Challenge 147 ny

point $x = x_0 = c^2/g$ at all times t. We get the two relations*

$$\xi = -\frac{c^2}{g}\left(1 - \text{sech}\frac{g\tau}{c}\right)$$
$$\text{d}\xi/\text{d}\tau = -c\ \text{sech}\ \frac{g\tau}{c}\ \tanh\ \frac{g\tau}{c}\ . \qquad (92)$$

These equations are strange. For large times τ the coordinate ξ approaches the limit value $-c^2/g$ and $\text{d}\xi/\text{d}\tau$ 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 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 56 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*, as illustrated in Figure 57?

Challenge 148 e

* The functions appearing above, the *hyperbolic secant* and the *hyperbolic tangent*, are defined using the expressions from the footnote on page 95:

$$\text{sech}\ y = \frac{1}{\cosh\ y} \quad \text{and} \quad \tanh\ y = \frac{\sinh\ y}{\cosh\ y}\ . \qquad (91)$$

Figure 57 How an event horizon looks like according to special (and general) relativity.

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. In particular, 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
Challenge 149 s hyperbolic motion, if the observer has a sufficient head start?

Here is a more advanced challenge, which prepares us for general relativity. What is
Challenge 150 s the two-dimensional *shape* of the horizon seen by a uniformly accelerated observer?
Challenge 151 s Another challenge: what horizon is seen by an observer on a carousel?

The importance of horizons

In special relativity, horizons might seem to play a secondary role. But this impression is wrong. Horizons are frequent and important. In principle, if you want to observe a horizon somewhere, just accelerate in the opposite direction and look back.

In fact, the *absence* of horizons is rare: it implies the lack of acceleration. And we know that uniform, inertial motion is limited in nature: it is limited by gravity and other interactions. Since in everyday life we are not moving inertially, there are horizons everywhere. In other words, *space is not really infinite in everyday life.*

Whenever you accelerate, there is a horizon behind you. Now, gravity and acceleration are equivalent, as they locally just differ by change of reference frame. Therefore, gravity is inextricably linked with horizons.

Horizons are everywhere – because gravity is everywhere. The relativistic description of gravity is called general relativity. We will find that in general relativity, horizons become even more important and frequent: the *night sky* is an example of a horizon. Yes, the sky is dark at night because the universe is not of infinite size. Also the surface of a *black hole* is a horizon. And there are literally billions of black holes in the universe. We will explore these topics below.

But horizons are interesting for a further reason. 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 its discovery of horizons.

Horizons show that *atomism is wrong*. Horizons can be observed and measured. On the one hand, horizons are extended, not localized systems, and they have two spatial dimensions. On the other hand, we will discover that horizons are not completely black, but have a slight colour, and that they can have mass, spin and charge. In short, horizons are neither particles nor space. Horizons are something new.

Later in our adventure, when we combine general relativity and quantum theory, we will discover that horizons are effectively *intermediate* between space and particles. Horizons can also be seen as a *mixture* of space and particles. We will need some time to find out what this means exactly. So far, our exploration of the speed limit in nature only tells us that horizons are a further phenomenon in nature, an unexpected addition to particles and vacuum.

Acceleration changes colours

Page 31 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. 99, Ref. 104 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

$$\mathrm{d}\sigma^2 = \left(1 + \frac{g_0 x}{c^2}\right)^2 c^2 \mathrm{d}t^2 \qquad (93)$$

in which g_0 is the proper acceleration of an observer located at $x = 0$. We can deduce in Challenge 152 e a straightforward way that

$$\frac{f_\mathrm{r}}{f_\mathrm{s}} = 1 - \frac{g_\mathrm{r} h}{c^2} = \frac{1}{1 + \frac{g_\mathrm{s} h}{c^2}} \qquad (94)$$

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 $g_\mathrm{r} = g_0/(1 + g_o x_\mathrm{r}/c^2)$ 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 Challenge 153 s colour of trees along their vertical extension?

The formula usually given, namely

$$\frac{f_\mathrm{r}}{f_\mathrm{s}} = 1 - \frac{gh}{c^2} , \qquad (95)$$

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 154 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 (95) above, an accelerated observer deduces that

$$v_{\text{light}} = c\left(1 + \frac{gh}{c^2}\right) \tag{96}$$

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 163 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,

▷ The value c is the speed of light only relative to *nearby* matter.

In other cases, light can move *faster* than c. Note that this result does not imply that Challenge 155 s signals or energy can be moved faster than c. You may want to check this for yourself.

In practice, non-local effects on the speed of light are negligible for distances l that are much less than c^2/a. For an acceleration of 9.5 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 156 s distant objects, such as stars, not move faster than light, following expression (96)?

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. 105 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 58. 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 $v_{02} = u$.)

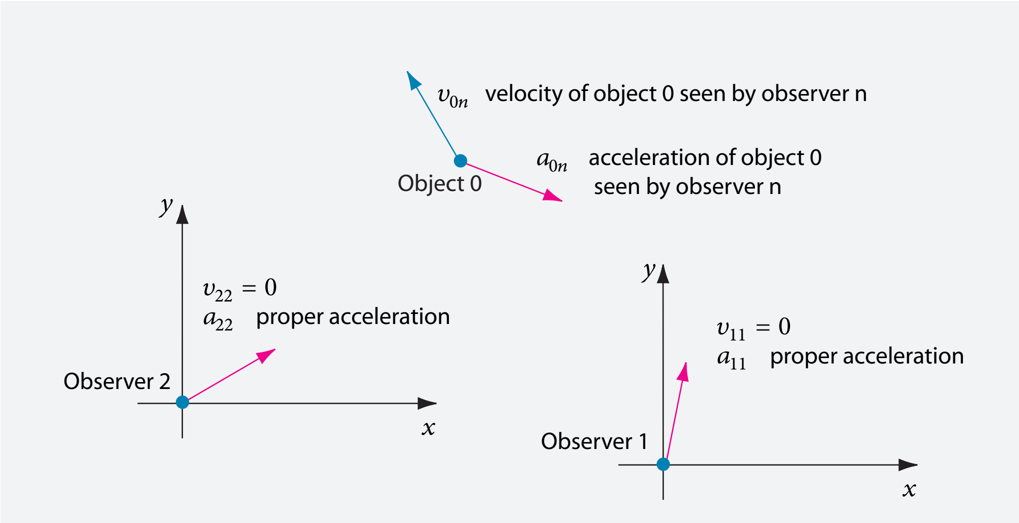


FIGURE 58 The definitions necessary to deduce the composition behaviour of accelerations.

Challenge 157 e

In Galilean physics we have the general connection

$$a_{01} = a_{02} - a_{12} + a_{22} \tag{97}$$

because accelerations behave simply. In special relativity, we get

$$a_{01} = a_{02}\frac{(1-v^2/c^2)^{3/2}}{(1-uv/c^2)^3} - a_{12}\frac{(1-u^2/c^2)(1-v^2/c^2)^{-1/2}}{(1-uv/c^2)^2} + a_{22}\frac{(1-u^2/c^2)(1-v^2/c^2)^{3/2}}{(1-uv/c^2)^3} \tag{98}$$

Challenge 158 e

and you might enjoy checking the expression.

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

$$\frac{v^2}{c^2} \geqslant \frac{2d}{l} . \tag{99}$$

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 acce-

Vol. I, page 294

* 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. More sound speed values were given earlier on.

leration a of a solid body with length l must obey

$$la < c^2 , \tag{100}$$

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 α equal or larger than infinity, or more precisely, 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. 106

Challenge 159 s

$$la < c^2 , \tag{101}$$

where c is now the speed of light.

▷ The speed of light thus limits the size of accelerated solid bodies.

For example, for 9.8 m/s^2, the acceleration of good motorbike, this expression gives a length 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? 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 160 ny

Challenge 161 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'

$$\Delta l \; \Delta a \leqslant c^2 \tag{102}$$

for the length and acceleration indeterminacies?

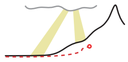
Challenge 162 s

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 163 ny

$$l < \frac{4\pi\varepsilon_0 c^2 d^2 m}{e^2} . \tag{103}$$

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 several times in the rest of our adventure.



Chapter 3

SPECIAL RELATIVITY IN FOUR SENTENCES

The results that we encountered so far can be summarized in four statements:

— All nearby observers observe that there is a unique, maximal and invariant energy speed in nature, the 'perfect' speed $v_{max} = c = 299\,792\,458\,\mathrm{m/s} \approx 0.3\,\mathrm{Gm/s}$. The maximum speed is realized by massless radiation such as light or radio signals, but cannot be achieved by material systems. This observation defines *special relativity*.

— Therefore, even though space-time is the *same* for every observer, measured times and length values – thus also angles and colours – *vary* from one observer to another, as described by the Lorentz transformations (15) and (16), and as confirmed by experiment.

Page 44

— Collisions show that the maximum energy speed implies that *mass is equivalent to energy*, that the total energy of a moving massive body is given by $E = c^2 \gamma m$, and that mass is not conserved.

— 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.

Not only is all motion of radiation and matter is limited in speed, but all speeds are 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 generalized way to predict mirror-inverse motion holds). Above all, also relativistic motion is lazy: it minimizes action.

Vol. I, page 29
Vol. V, page 245

Could the speed of light vary?

The speed of massless light and radiation is the limit speed of energy in nature. Could the limit speed change from place to place, or change as time goes by? This tricky question still makes a fool out of many physicists. The first answer is often 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 by researchers. However, this often-heard answer is *wrong*.

Ref. 107

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.

▷ A change in the speed of light cannot be measured.

Challenge 164 s
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 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. There is no way to detect variations – in time or across space – of a measurement standard. Just try!
Challenge 165 e

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 speed limit – and by the strange properties of *small* speeds.

In summary, there is, in principle, *no* way to falsify 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 the limit speed c from the formulae of everyday motion.

Where does special relativity break down?

The maximum local energy speed is confirmed by all experiments. The speed limit is thus correct: the local energy speed limit is a *fundamental truth* about nature. Indeed, it remains valid throughout the rest of our adventure.

As we approach the speed of light, the Lorentz factor and the quantities in the Lorentz transformation exceed all bounds. However, in nature, *no* observable actually reaches arbitrary large values. For example, no elementary particle with an energy or a momentum above – or even close to – the (corrected) Planck limits

$$E_{\text{Planck}} = \sqrt{\frac{\hbar c^5}{4G}} = 9.8 \cdot 10^8\,\text{J} = 0.60 \cdot 10^{19}\,\text{GeV}$$

$$p_{\text{Planck}} = \sqrt{\frac{\hbar c^3}{4G}} = 3.2\,\text{kg m/s} = 0.60 \cdot 10^{19}\,\text{GeV/c} \qquad (104)$$

has ever been observed. In fact, the record values observed so far are one million times smaller than the Planck limits. The reason is simple: when the speed of light is approached as closely as possible, special relativity *breaks down* as a description of nature.

How can the maximum speed limit remain valid, and special relativity break down nevertheless? At highest energies, special relativity is *not sufficient* to describe nature. There are two reasons.

In the case of extreme Lorentz contractions, we must take into account the curvature of space-time that the moving energy itself generates: *gravitation* needs to be included. Equivalently, we recall that so far, we assumed that point masses are possible in nature. However, point masses would have infinite mass density, which is impossible: gravity, characterized by the gravitational constant G, prevents infinite mass densities through the curvature of space, as we will find out.

In addition, in the case of extreme Lorentz contractions, we must take into account the fluctuations in speed and position of the moving particles: *quantum theory* needs to be included. We recall that so far, we assumed that measurements can have infinite precision in nature. However, this is not the case: quantum theory, characterized by the smallest action value $\hbar$, prevents infinite measurement precision, as we will find out.

In summary, the two fundamental constants G, the gravitational constant, and $\hbar$, the quantum of action, limit the validity of special relativity. Both constants appear in the Planck limits. The gravitational constant G modifies the description of motion for powerful and large movements. The quantum of action $\hbar$ modifies the description of motion for tiny movements. The exploration of these two kinds of motions define the next two stages of our adventure. We start with gravitation.

CHAPTER 4

SIMPLE GENERAL RELATIVITY: GRAVITATION, MAXIMUM SPEED AND MAXIMUM FORCE

General relativity is *easy*. Nowadays, it can be made as intuitive as universal gravity and its inverse square law, so that the important ideas of general 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 twins paradox.

In the following pages we will discover that, just as special relativity is based on and derives from a *maximum speed* c,

▷ General relativity is based on and derives from a *maximum momentum change* or *maximum force* $c^4/4G$ – equivalently, from 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 – or maximum momentum flow – and maximum power – or maximum energy flow – are called ***horizons***.

Horizons are simple generalizations of those horizons that we encountered in special relativity. We will find out shortly why the maximum values are related to them. 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.

Page 96

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.

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.

Challenge 166 s

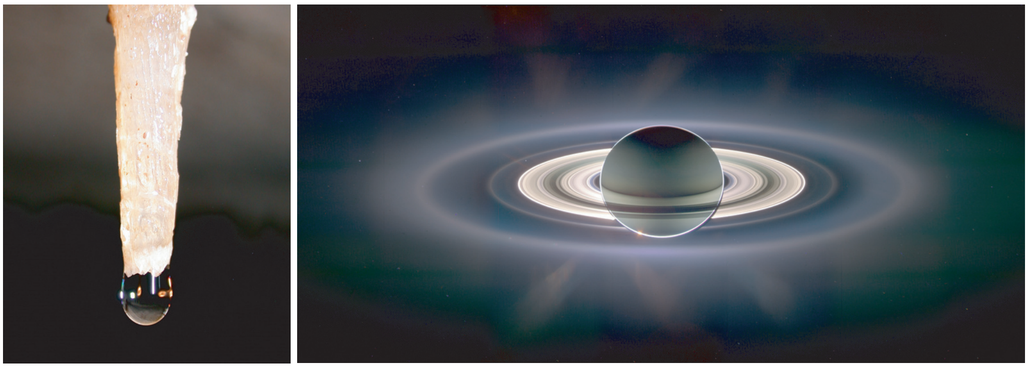


FIGURE 59 Effects of gravity: a dripping stalactite (© Richard Cindric) and the rings of Saturn, photographed when the Sun is hidden behind the planet (courtesy CICLOPS, JPL, ESA, NASA).

Maximum force – general relativity in one statement

Ref. 108

> 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

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
Ref. 109 century it was shown that *general* relativity can be approached by using a similar basic principle:

▷ There is in nature a *maximum force*, or *maximum momentum change per time:*

$$F \leqslant \frac{c^4}{4G} = 3.0258(4) \cdot 10^{43}\,\mathrm{N}\,. \tag{105}$$

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*, or energy change per time, in nature:

$$P \leqslant \frac{c^5}{4G} = 9.071(1) \cdot 10^{51}\,\mathrm{W}\,. \tag{106}$$

No power of any lamp, engine or explosion can exceed this value. It is equivalent to $1.2 \cdot 10^{49}$ horsepower. Another way to visualize the value is the following: the maximum power corresponds to converting 50 solar masses into massless radiation within a millisecond. The maximum power is realized when a (Schwarzschild) black hole is radiated

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 energy speed, on page 26 above, and with the table about a smallest action, on page 19 in volume IV.

Statement	Test
The maximum force value $c^4/4G$ is observer-invariant.	Check all observations.
Force values $> c^4/4G$ are not observed.	Check all observations.
Force values $> c^4/4G$ cannot be produced.	Check all attempts.
Force values $> c^4/4G$ cannot even be imagined.	Solve all paradoxes.
The maximum force value $c^4/4G$ is a principle of nature.	Deduce the theory of general relativity from it. Show that all consequences, however weird, are confirmed by observation.

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.

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

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

$$\frac{dm}{dt} \leqslant \frac{c^3}{4G} = 1.000\,93(1) \cdot 10^{35}\ \text{kg/s}\ . \tag{107}$$

This bound on mass flow 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 existence of a maximum force, power or mass flow implies the full theory of general relativity. In order to prove the correctness and usefulness of this approach, a sequence of arguments is required. This sequence of arguments, listed in Table 3, is the
Page 26
same as the sequence that we used for the establishment of the limit speed in special relativity. The basis is to recognize that the maximum force value is *invariant*. This follows from the invariance of c and G. For the first argument, we need to gather all *observational evidence* for the claimed limit and show that it holds in all cases. 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.

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 meaning of 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, Ref. 110

▷ Change of momentum, and thus force, always takes place through some *boundary surface*.

This connection is of central importance. Whenever we think about force at a point, we *really* mean the momentum 'flowing' through a surface at that point. And that amount is limited.

▷ Force is a *relative* concept.

Any force measurement is relative to a surface. Any momentum flow measurement is relative. In special relativity, speed is relative; nevertheless, speed is limited. In general relativity, force is relative; nevertheless, force is limited. That is the fascination of the force limit.

General relativity usually explains the concept of force as follows: a force keeps 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 or momentum flow is limited, relative to any surface. It is not important whether the surface, i.e., the observer, or the body does not follow geodesics.

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:

▷ A surface is *physical* as long as we can fix observers onto it.

We stress that 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.

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: the *Planck force*.

The force unit fits into the system of natural units that Max Planck derived from the speed of light c, the gravitational constant G and the quantum of action h (nowadays $\hbar = h/2\pi$ is preferred).* The maximum force thus provides a *measurement standard* for force that is valid in every place and at every instant of time.

The maximum force value $c^4/4G$ differs from Planck's originally 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. 111
Vol. VI, page 27
Page 129

The expression $c^4/4G$ 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. 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, it does matter 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.

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.

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

Page 83

* When Planck discovered the quantum of action, he noticed at once the possibility to define *natural* units for all observable quantities. Indeed, on a walk with his seven-year-old son Erwin 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 20

is valid for every component of the 3-force, as well as for its magnitude. (As we will see
Page 122 below, a boost to an observer with high γ 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. The maximum power principle states that there is no way to move or get rid of energy more quickly than that.

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 there is a built-in *braking* mechanism in nature; this braking mechanism is gravity.

The claim of a maximum force, a maximum power or a maximum mass flow 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. We recall that 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. In the past, many people have claimed to have produced energy speeds higher than that of light. So far, nobody has
Challenge 167 s ever claimed to have produced or observed a force higher than the limit value.

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
Page 127 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 construc-
Challenge 168 s ted. You might enjoy searching for such an exception.)

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.

Also the power limit can 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. Ref. 113 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. For example, the black hole merger published in 2016, Ref. 112 possibly the most powerful event observed so far, transformed about 3 solar masses into radiation in 0.2 s. Its power was therefore about three thousand times lower than the power limit $c^5/4G$; the peak power possibly was around three hundred times lower than the limit. It might well be that only the brightness of evaporating black holes in their final phase can equal the power limit. However, no such event has ever been observed yet. (Given that several nearby localised sources can each approach the power limit, the so-called *power paradox* arises, which will be discussed below.) Page 127

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. Ref. 114

In summary, the experimental situation is somewhat disappointing. Experiments do not contradict the limit values. But neither do the data do much to confirm the limits. 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 near future. Page 133

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 readers may skip to the next section.) Page 117

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. These unattainable surfaces are basic to general relativity; they are called *horizons*. Ref. 109, Ref. 111

* This section can be skipped at first reading. The proof mentioned in it 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.

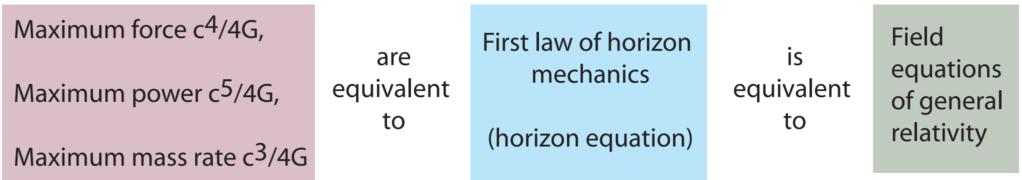


FIGURE 60 Showing the equivalence of the maximum force or power with the field equations of general relativity.

▷ Maximum force and power only appear on horizons.

Page 97

We have encountered horizons in special 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 any 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 used the limit property of the speed of light to deduce the Lorentz transformations. In general relativity, we will now prove that the maximum force in nature, which we can also call the *horizon force*, implies the field equations of general relativity. To achieve this aim, we start by recognizing 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 60. 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 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

$$F = \frac{E}{L} \,. \tag{108}$$

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 (108) 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

$$\frac{c^4}{4G} = \frac{E}{LA} 4\pi R^2 . \qquad (109)$$

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 object accelerated by a is limited, by special relativity, by $L \leqslant c^2/2a$. Already special relativity shows that this limit is related to the appearance of a horizon. Together with relation (109), the statement that horizons are surfaces of maximum force leads to the following important relation for static, spherical horizons:

Ref. 115

$$E = \frac{c^2}{8\pi G} a A . \qquad (110)$$

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 mechanics* or the *first law of horizon mechanics*.

Ref. 116

The above derivation also yields the intermediate result

$$E \leqslant \frac{c^4}{16\pi G} \frac{A}{L} . \qquad (111)$$

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. *Gravity limits energy flow.* 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 equation in its forms (110) and (111) follows 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.

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

$$\delta E = \frac{c^2}{8\pi G} a\, \delta A . \qquad (112)$$

This differential relation – it might be called the *general horizon equation* – is valid for any kind of horizon. It can be applied separately for every piece δA of a dynamic or spatially changing horizon.

The general horizon equation (112) has been known to be equivalent to general relativity at least since 1995, when this equivalence was (implicitly) shown 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 (112) of the relation.

Ref. 117

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 the energy limit

$$\frac{\delta E}{\delta A} \leqslant \frac{c^4}{16\pi G}\frac{1}{L} . \tag{113}$$

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. We also note that the horizon equation – or, equivalently, the force or power limit – implies a maximum mass change rate in nature given by $\mathrm{d}m/\mathrm{d}t \leqslant c^3/4G$.

The remaining, second step of the argument is the derivation of general relativity 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 (112) 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.

Ref. 117

To generalize the general horizon equation, we introduce the general surface element $\mathrm{d}\Sigma$ 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 (112) as

$$\delta E = \int T_{ab} k^a \mathrm{d}\Sigma^b , \tag{114}$$

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 (112) can be rewritten, making use of the (purely geometric) Raychaudhuri equation, as

$$a\,\delta A = c^2 \int R_{ab} k^a \mathrm{d}\Sigma^b , \tag{115}$$

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 (112) becomes

$$\int T_{ab} k^a \mathrm{d}\Sigma^b = \frac{c^4}{8\pi G} \int R_{ab} k^a \mathrm{d}\Sigma^b \,. \tag{116}$$

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

$$T_{ab} = \frac{c^4}{8\pi G} \left(R_{ab} - (\frac{R}{2} + \Lambda) g_{ab} \right) \,, \tag{117}$$

where R is the Ricci scalar and Λ 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 Λ. 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 (just accelerate away, as explained above), 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.

Page 98

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. *In summary, maximum force (or power), the horizon equation, and general relativity are equivalent.*

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. And indeed, 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 the existence of the horizon limits.

Gravity, space-time curvature, horizons and maximum force

Let us repeat the results of the previous section in simple terms. Imagine two observers who start moving freely and parallel to each other. 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. Such deviations from parallel free motion are observed near masses and other localized en-

Challenge 169 s

ergy. We conclude that space-time is curved near masses. Or, simply put: gravity curves space.

Gravitation leads to acceleration. And acceleration leads to a horizon at distance c^2/a. No horizon occurs in everyday life, because the resulting distances are not noticeable; but horizons do occur around bodies whose mass is concentrated in a sphere of radius $r = 2Gm/c^2$. Such bodies are called (Schwarzschild) *black holes*. The spatial curvature around a black hole of mass m is the maximum curvature possible around a body of that mass.

Black holes can be seen as matter in permanent free fall. We will study black holes in Page 262 detail below. In case of a black hole, like for any horizon, it is impossible to detect what is 'behind' the boundary.*

Black holes are characterized by a surface gravity a and an energy flow E.

▷ 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.

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 Page 265 or mass are those falling into the black hole.

The analogy between special and general relativity can be carried further. In special relativity, maximum speed implies $\mathrm{d}x = c\ \mathrm{d}t$, and that 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 horizon equation implies the field equations of general relativity. In short:

▷ The existence of a maximum force implies that space-time is curved near masses, and it implies *how* it 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 2π) 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 Ref. 109

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

a material constant as proportionality factor and as limit value is well known in materials science.

Why is the maximum force also the proportionality factor between curvature and energy? Imagine space 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 μ) 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 Δl, we define the *shear modulus* G by

$$\frac{F}{A} = G\frac{\Delta l}{l} . \tag{118}$$

The value of the shear modulus G 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

$$G_{\text{tss}} = \frac{G}{2\pi} . \tag{119}$$

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 that can be bent is a helpful way to understand general relativity. We will use it regularly in the following. Ref. 118

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 tearing effect in the last part of our adventure. Vol. VI, page 303

Conditions of validity for the force and power limits

The maximum force value is valid only under three conditions. To clarify this point, we can compare the situation to the maximum speed. There are three conditions for the validity 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 light. The limit speed is valid for *motion of energy* only. Page 58

* Does this analogy make you think about aether? Do not worry: physics has no need for the concept of 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. Vol. III, page 136

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. The limit speed is only a *local* limit.

Thirdly, the observer measuring speeds must be *physical*: also the observer must be made of matter and energy, thus must move more slowly than light, and must be able to observe the system. No system moving at or above the speed of light can be an observer. The limit speed is only for *physical* observers.

Ref. 119

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

Challenge 170 s

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 larger than 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'.

A limit to speed is surprising at first, because speed is relative, and therefore it should be possible to let speed take any imaginable value. The situation for force is similar: force is relative, and therefore it should be possible to let force take any imaginable value.

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

* '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), philosopher and scholar. This is aphorism 207 – *Sonnenbahn der Idee* – from his text *Menschliches Allzumenschliches – Der Wanderer und sein Schatten*.

the transfer of energy. This transfer is limited by the horizon equation (112) or the energy limit (113). For any attempt to exceed the force limit, the flowing energy results in the appearance of a horizon. The horizon then prevents the force from exceeding the limit, because it imposes a limit on interaction.

Page 101 Let us explore the interaction limit. 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 breaks, at the latest, at that point. The force F on a body of mass M and radius R is thus limited by

$$F \leqslant \frac{M}{2R} c^2 \,. \tag{120}$$

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 (105). *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*. Also 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. In short, *static* forces cannot exceed the maximum force.

* *

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 Page 101 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.

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 momentum 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 through reflection.

* *

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 impossible to use light and mirrors to create a momentum change larger than $c^4/4G$.

It is even 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 (113) 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 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
Ref. 120 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.
Page 83 In all reference frames, the measured 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.)

* *

The divergence attempt. The force on a test mass m at a radial distance d from a Schwarz-

Ref. 113 schild black hole (for $\Lambda = 0$) is given by

$$F = \frac{GMm}{d^2\sqrt{1 - \frac{2GM}{dc^2}}} . \quad (121)$$

Similarly, the inverse square expression of universal gravitation states that the force between two masses m and M is

$$F = \frac{GMm}{d^2} . \quad (122)$$

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

However, gravitational force can diverge only for non-physical, *point-like* masses. However, there is a minimum approach distance to a mass m given by

$$d_{\text{min}} = \frac{2Gm}{c^2} . \quad (123)$$

The minimum approach distance is the corresponding black hole radius. Black hole formation makes it impossible to achieve zero distance between two masses. Black hole formation also makes it impossible to realize point-like masses. Point-like masses are unphysical. As a result, in nature there is a (real) minimum approach distance, proportional to the mass. If this minimum approach distance is introduced in equations (121) and (122), we get

$$F = \frac{c^4}{4G} \frac{Mm}{(M+m)^2} \frac{1}{\sqrt{1 - \frac{M}{M+m}}} \leqslant \frac{c^4}{4G} . \quad (124)$$

The approximation of universal gravitation yields

$$F = \frac{c^4}{4G} \frac{Mm}{(M+m)^2} \leqslant \frac{c^4}{4G}. \quad (125)$$

In both cases, the maximum force value is never exceeded, as long as we take into account the physical size of masses or 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 in general relativity, as long as the typical dimensions in the situation are much larger than the Planck value.

* *

The quantum problem. If quantum effects are neglected, it is possible to construct sur-
Challenge 171 e
faces 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.

Ref. 109, Ref. 111

* *

The relativistically extreme observer attempt. Any extreme observer, whether in rapid inertial or in accelerated motion, has no chance to beat the force 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.

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

Ref. 121

$$a \leqslant \frac{2mc^3}{\hbar} \,. \tag{126}$$

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 (126). Inserting the maximum possible mass for an elementary particle, namely the (corrected) Planck mass, we find that equation (126) then states that the horizon force is the upper bound for elementary particles.

Vol. VI, page 40

* *

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 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.

Ref. 113

* *

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 nature, large forces do not add up.

* *

Challenge 172 s 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 of reference is also smaller in any other frame?

* *

Ref. 122 We could also try to use the cosmological constant to produce forces that exceed the maximum force. But also this method does not succeed, as discussed by John Barrow and Gary Gibbons.

* *

Challenge 173 r Can you propose and then resolve an additional attempt to exceed the force limit?

Gedanken experiments with the power and the mass flow limits

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 61. 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.

The surface bent around atoms 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

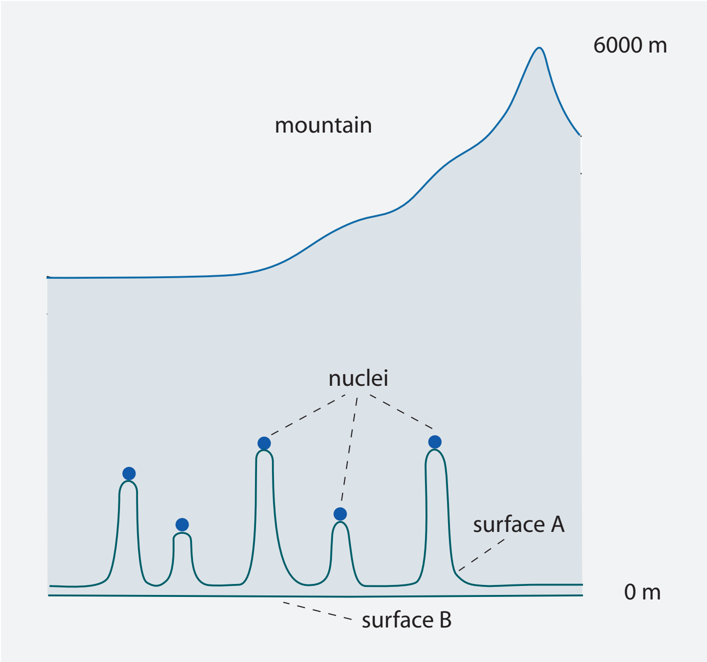


FIGURE 61 The mountain attempt to exceed the maximum mass flow value.

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 61. 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. Again, also in this case the indeterminacy in the atomic positions makes it impossible to observe or state 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. Ref. 113 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.

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; Vol. VI, page 67 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. Challenge 174 s However, these unphysical counter-examples make no statements about nature. (*Ex falso quodlibet.***)

* *

The many lamps attempt, or *power paradox*. An absolute power limit imposes a limit on the rate of energy transport through any imaginable, physical 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 a total of 3/2 times the maximum value, and thus allow us 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. Challenge 175 e Again, since the horizon limit (113) is achieved, the arising horizon swallows parts of the radiation and prevents the force or power limit from being exceeded. Exploring a numerical simulation of this situation would be instructive. Challenge 176 r Can you provide one? In short, we can say that large power values *do not add up* in nature.

* *

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.

* It can also be called *physically sensible*.

** 'Anything can be deduced from a falsehood.'

* *

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, any new power source would break it, or at least imply that another elsewhere must close down. Can you find the fallacy in this argument?

Challenge 177 s

* *

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.

* *

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 limits are 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 then 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 think of a good one, let the author know.)

Challenge 178 r

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 probably 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 notable exceptions – and thus the concept is often avoided.

Ref. 123

In summary, the principle of maximum force – or of maximum power – has 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

$$a = f\frac{GM}{R^2} , \qquad (127)$$

where the numerical factor f must still be determined. To determine it, we study the

* 'We all live under the same sky, but we do not have the same horizon.' Konrad Adenauer (1876–1967), West German Chancellor.

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 (127), 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 174

* *

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 maximum speed to be their propagation speed. This is indeed the case, as we will see. Page 174

* *

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:

$$L \geqslant \frac{4Gm}{c^2} . \tag{128}$$

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 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.) Ref. 124, Ref. 125

* *

There is a power limit for all energy sources. In particular, the value $c^5/4G$ limits the lu-

* The maximum value for the mass to size limit is obviously equivalent to the maximum mass change given above.

minosity of all gravitational sources. Indeed, all formulae for gravitational wave emission Ref. 113 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 nor dressed) singularities in nature. Horizons prevent the appearance of singularities. In particular, the big bang was *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. In a sense, this is indeed the case, even though horizons usually prevent it. However, the way that this tearing 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 Ref. 126 cannot be detected. Present data support this claim to a high degree of precision.

* *

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, Ref. 113 the maximum force principle predicts that the Nordtvedt effect vanishes. 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. We will explore this question Page 262 later on.

* *

Vol. VI, page 37 Later on we will find that quantum effects cannot be used to exceed the force or power

Challenge 179 e 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.

* *

Given that the speed c and the force value $c^4/4G$ are limit values, what can be said about G itself? The gravitational constant G describes the strength of the gravitational interaction. In fact, G characterizes the strength of the *weakest possible* interaction. In other words, given a central body of mass M, and given the acceleration a of a test body at a distance r due to any interaction whatsoever with the central body, then the ratio ar^2/M is *at least equal to* G. (Can you show that geostationary satellites or atoms in crystals are not counterexamples?) In summary, also the gravitational constant G is a limit value in nature. Challenge 180 e

An intuitive understanding of cosmology

Page 240 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 luminosity values 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 – and most of the emitted power 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, the much simpler explanation with the power limit 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 of density ρ_0. But in this 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 of radius R_0 whose centre lies at the Earth, which encompasses all the universe, and whose radius changes with time (almost) as rapidly as the speed of light. The mass flow $\mathrm{d}m/\mathrm{d}t = \rho_0 A_0 c$ through that sphere is predicted to reach the mass flow limit $c^3/4G$; thus we have

$$\rho_0 4\pi R_0^2 c \leqslant \frac{c^3}{4G} . \qquad (129)$$

We can compare this with the Friedmann models, who predict, in a suitable limit, that *one third* of the left hand side saturates the mass flow limit. The precision measurements Ref. 127

of the cosmic background radiation by the WMAP satellite confirm that the present-day total energy density ρ_0 – including dark matter and dark energy – and the horizon radius R_0 just reach the Friedmann value. The above argument using the maximum force or mass flow thus still needs a slight correction.

In summary, the maximum force limit predicts, within a factor of 6, the observed relation between the size and density of the universe. In particular, the maximum force principle predicts that the universe is of finite size. By the way, a finite limit for power also suggests that a finite age for the universe can be deduced. Can you find an argument?

Challenge 181 s

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 vicinity of researchers. Nevertheless, the limit values are observable and falsifiable.

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 error Δx and the energy error ΔE. For all systems we have

Ref. 109, Ref. 111

$$\frac{\Delta E}{\Delta x} \leqslant \frac{c^4}{4G} . \qquad (130)$$

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. Pulsar experiments and gravitational wave detectors therefore can test the power limit in the coming years.

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 $P = c^2 M/(R/v)$ 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. Gravitational wave detectors therefore can test the power limit in the coming years.

A candidate surface that reaches the power 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:

$$4\pi R^2 I \leqslant \frac{c^5}{4G} . \tag{131}$$

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 (131) with precision.

It might well be that the amount of matter falling into some black hole, such as the one at the centre of the Milky Way, might be measurable one day. The limit $\mathrm{d}m/\mathrm{d}t \leqslant c^3/4G$ could then be tested directly.

In short, direct tests of the limits are possible, but not easy. In fact, you might want to predict which of these experiments will confirm the limit first. The scarcity of direct experimental tests of the force, power and mass flow 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 limits: the field equations of general relativity. This will be our next topic.

Challenge 182 e

A summary of general relativity – and minimum force

> “Non statim pusillum est si quid maximo minus est.*”
>
> Seneca

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 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 leads us to an additional aspect of gravitation. The cosmological constant Λ 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 cor-

Challenge 183 e
Page 243

* ‘Nothing is negligible only because it is smaller than the maximum.’ Lucius Annaeus Seneca (*c.* 4 bce –65), *Epistolae* 16, 100.

Vol. VI, page 37 rected) Planck area $4G\hbar/c^3$, the smallest area in nature, gives a force value

$$F = 4\Lambda\hbar c = 1.20 \cdot 10^{-77}\,\mathrm{N}\,. \tag{132}$$

This is also the gravitational force between two (numerically corrected) Planck masses $\sqrt{\hbar c/4G}$ located at the cosmological distance $1/4\sqrt{\Lambda}$.

We conjecture that expression (132) is the *minimum force* in nature. Proving this conjecture is more involved than for the case of maximum force. So far, only some hints are possible. Like the maximum force, also the minimum force must be compatible with gravitation, must not be contradicted by any experiment, and must withstand any thought experiment. A quick check shows that the minimum force allows us to deduce the cosmological constant of gravitation; minimum force is an invariant and is not contradicted by any experiment. There are also hints that there may be no way to generate or measure any smaller value. For example, the gravitational force between any two neutral particles at cosmological distance, such as between two atoms or two neutrinos, is much smaller than the minimum force; however, it seems impossible to detect experimentally whether two such particles interact at all: the acceleration is too small to be measured. As another 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 measurement of a still smaller force. (Can you do so?)

Challenge 184 e

Challenge 185 d

If we leap to the – not completely proven – conclusion that expression (132) 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.

We have seen that both 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 as general relativity. 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 confirmation of the minimum force are two important challenges for our ascent beyond general relativity. We come back to the issue in the last part of our adventure.

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 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, in universal gravity, 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. 128, Ref. 129

The results of combining maximum speed with gravity are 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 inertial 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 186 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 174 and page 204.

Ref. 130 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 in our adventure:

▷ *Rest* is free fall.

Rest, like free fall, is the lack of disturbance.

The set of all possible 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 on rest in other words:

▷ True *motion* is the opposite of free fall.

This statement immediately rises a number of questions: Most trees or mountains are Challenge 187 s not 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 129 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. 131 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.**

* Nowadays it is possible to book such flights at 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 Challenge 188 s larger than the length effect, which can usually be neglected. Can you see why?

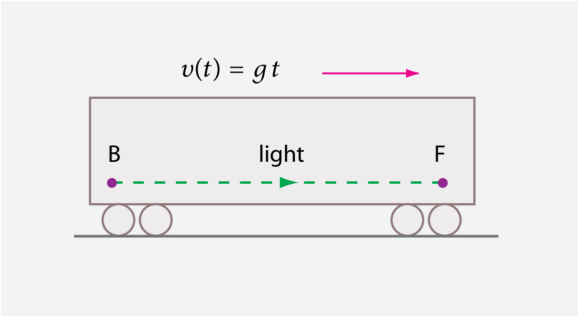


FIGURE 62 Inside an accelerating train or bus.

Challenge 189 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 on the observation 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.

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 or bus of length Δh that is accelerating forward with acceleration g, as shown in Figure 62. The light arrives at the front of the train or bus after a time $t = \Delta h/c$. However, during this time the accelerating train or bus 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 get*

Page 55
Challenge 190 e

$$\frac{\Delta f}{f} = \frac{g\Delta h}{c^2} . \tag{133}$$

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.

Challenge 192 s

▷ Acceleration induces frequency changes in light.

Let us compare this first effect of acceleration with the effects of gravity.

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 63. The descending, grey masses are

Ref. 132

* The expression $v = gt$ is valid only for non-relativistic speeds; nevertheless, the conclusion of this section is not affected by this approximation.

Challenge 191 e

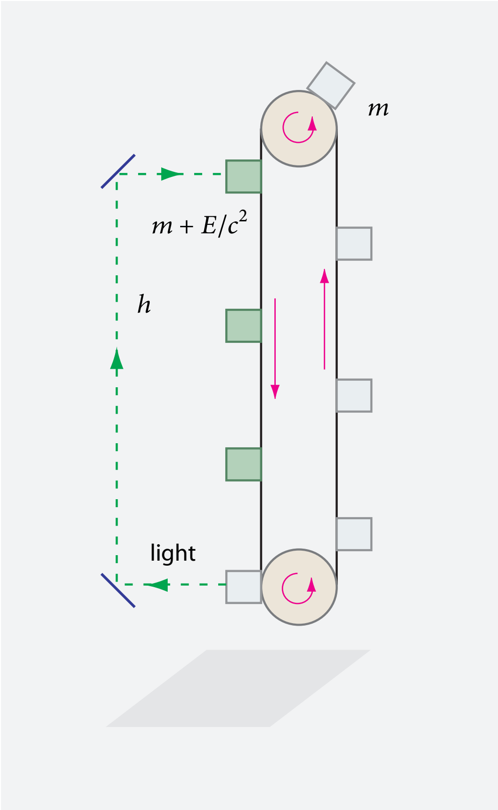


FIGURE 63 The necessity of blue- and red-shift of light: why trees are greener at the bottom.

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 $E = c^2 m$, 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.**

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 *different* from the one at which it is emitted from the bottom.***

Vol. I, page 280

In short, it turns out that

▷ Rising light is gravitationally red-shifted.

* As in special relativity, here and in the rest of our adventure, the term 'mass' always refers to rest mass.
Challenge 193 s
** Can this process be performed with 100 % efficiency?
*** The precise relation between energy and frequency of light is described and explained in the discussion on quantum theory. But we know already from classical electrodynamics that the energy of light depends on its intensity and on its frequency.
Vol. IV, page 44

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 (133). That is what we would expect, 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 195 e
Challenge 196 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 Mössbauer effect, Pound and Rebka confirmed the gravitational red-shift in their university tower using γ radiation.

Ref. 133
Vol. IV, page 380
Ref. 134
Ref. 135

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 Δh and the acceleration of gravity g according to

$$\frac{\Delta\tau}{\tau} = \frac{\Delta f}{f} = \frac{g\Delta h}{c^2} . \qquad (134)$$

In simple words,

▷ In gravity, time is height-dependent.

In other words, *height makes old.* Can you confirm this conclusion?

Challenge 197 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 (134). 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. 55
Ref. 136
Ref. 137
Challenge 198 e

* How does this argument change if you include the illumination by the Sun?

Challenge 194 ny

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 199 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:

$$\mathrm{d}s^2 \neq c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \ . \tag{135}$$

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 200 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 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, or two parts of one observer.

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 201 e

Challenge 202 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 203 e

Gravitation leads to changes of relative distance. These changes evince another effect,

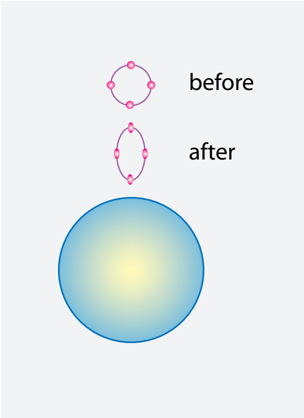


FIGURE 64 Tidal effects: the only effect bodies feel when falling.

shown in Figure 64: 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.

Vol. I, page 197

Ref. 138

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. 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. From the equations of general relativity, 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. 139
Vol. I, page 201
Ref. 140

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

Challenge 204 e

* '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.'

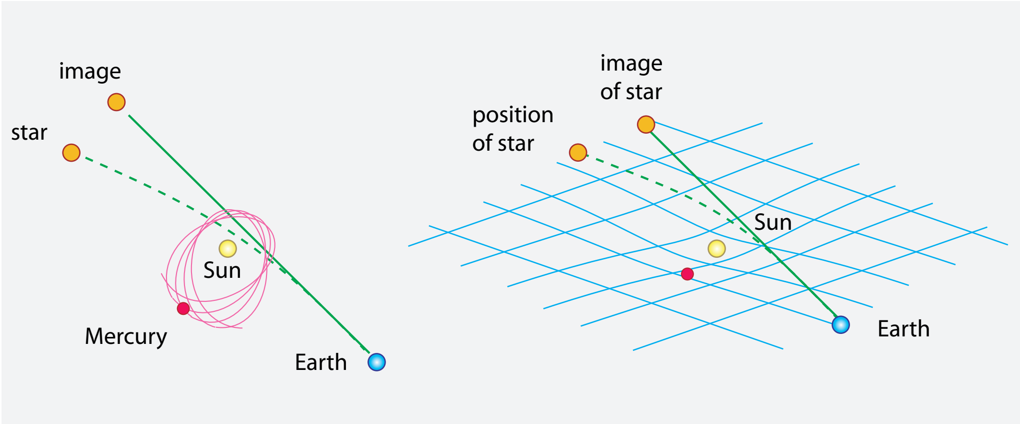


FIGURE 65 The mattress model of space: the path of a light beam and of a satellite near a spherical mass.

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 65 gives an impression of the situation.

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. 141

Challenge 205 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.)

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.)

Vol. III, page 136

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

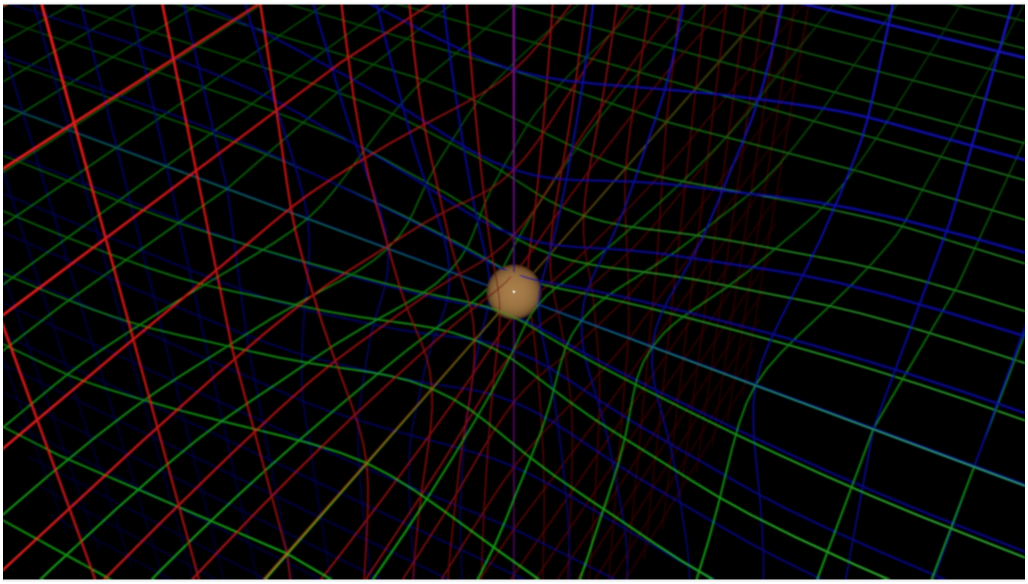


FIGURE 66 A three-dimensional illustration of the curvature of space around a mass (© Farooq Ahmad Bhat.

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 206 s

Challenge 207 ny You might be led to ask: if the flat space containing a macroscopic body is bent by a gravitational field, what happens to the body contained in it? (For simplicity, we can imagine that the body is suspended and kept in place by massless ropes.) The gravitational field will also affected the body, but its bending is *not* related in a simple way to the bending of the underlying space. For example, bodies have higher inertia than empty space. And in static situations, the bending of the body depends on its own elastic properties, which differ markedly from those of empty space, which is much stiffer.

Curved space-time

Figure 65 and Figure 66 shows the curvature of space only, but in fact the whole of 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 65, 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 2π. 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 $(t + \mathrm{d}t, r + \mathrm{d}r)$. Page 45

Page 141 As we saw above, gravity means that in spherical coordinates we have

$$d\tau^2 = \frac{ds^2}{c^2} \neq dt^2 - dr^2/c^2 - r^2 d\varphi^2/c^2 \ . \qquad (136)$$

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

$$d\tau^2 = \frac{ds^2}{c^2} = \left(1 - \frac{2GM}{rc^2}\right) dt^2 - \frac{dr^2}{c^2 - \frac{2GM}{r}} - \frac{r^2}{c^2} d\varphi^2 \ . \qquad (137)$$

This expression is called the *Schwarzschild metric* after one of its discoverers.* The metric (137) 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?) Challenge 208 s Expression (137) also shows that gravity's strength around a body of mass M and radius R is measured by a dimensionless number h defined as

$$h = \frac{2G}{c^2} \frac{M}{R} \ . \qquad (138)$$

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*

$$R_S = \frac{2GM}{c^2} \ , \qquad (139)$$

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 Page 266 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 Challenge 209 e every system in everyday life is larger than its Schwarzschild radius. Physical systems Ref. 143 which realize the Schwarzschild radius are called *black holes*; we will study them in de- Page 262 tail shortly. In fact, general relativity states that *no* system in nature is smaller than its

* Karl Schwarzschild (1873–1916), influential 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 Ref. 142 the metric, unknown to Einstein, was Johannes Droste, a student of Lorentz.

Schwarzschild size, in other words that the ratio h defined by expression (138) is never above unity.

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. 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. 144

It is straightforward to deduce that the temporal equivalent of spatial curvature is the uneven running of clocks. Taking the two curvatures together, we conclude that when gravity is present, *space-time* is curved. Challenge 210 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. 145 Challenge 211 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:

$$v \leqslant c \quad (140)$$

* ‘If I rest, I die.’ This is the motto of the bird of paradise.

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.

▷ For all observers, the force F on a system is limited by

$$F \leqslant \frac{c^4}{4G} \, , \qquad (141)$$

where G is the universal constant of gravitation.

In short, there is a maximum force in nature. Gravitation leads to attraction of masses. However, this force of attraction is limited. An equivalent statement is:

Challenge 212 e

▷ For all observers, the size L of a system of mass M is limited by

$$\frac{L}{M} \geqslant \frac{4G}{c^2} \, . \qquad (142)$$

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

$$P \leqslant \frac{c^5}{4G} \, . \qquad (143)$$

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 have seen above, the speed limit together with any one of the last three principles imply all of general relativity.*

Page 113

For example, can you show that the formula describing gravitational red-shift complies with the general limit (142) on length-to-mass ratios?

Challenge 213 ny

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.

* This didactic approach is unconventional. It is possible that is has been pioneered by the present author, though several researchers developed similar ideas before, among them Venzo de Sabbata and C. Sivaram.

Ref. 109

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 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
Page 87 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
Vol. I, page 59 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.

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 *time-like geodesics*. For light, such paths are called *light-like* 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 consistent.

If falling is seen as a consequence of the Earth's surface approaching – as we will
Page 158 argue below – we can deduce directly that falling implies a proper time that is as long as
Challenge 214 e possible. Free fall indeed is motion along geodesics.

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;

* Or it would be, were it not for a small deviation called quantum theory.

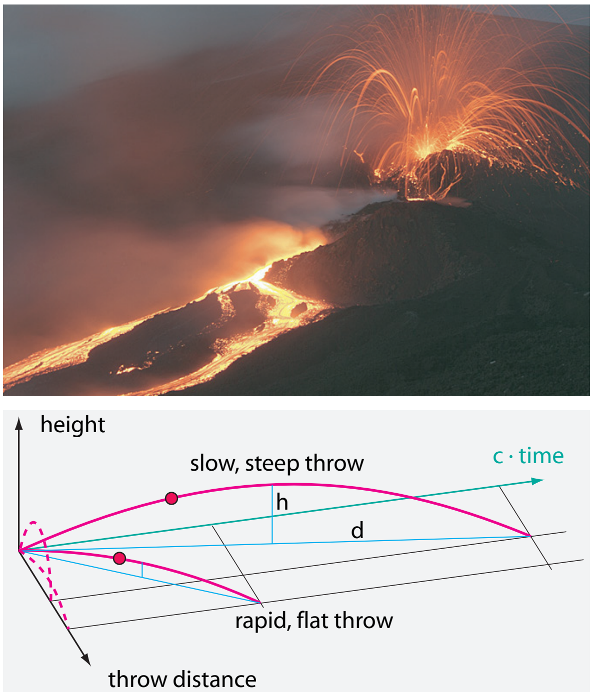


FIGURE 67 All paths of flying stones, independently of their speed and angle, have the same curvature in space-time (photograph © Marco Fulle).

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 215 ny

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 67. All paths have the *same* curvature radius r, given by Challenge 216 ny

$$r = \frac{c^2}{g} \approx 9.2 \cdot 10^{15}\ \mathrm{m}\ . \qquad (144)$$

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 217 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 $\int d\tau$. We rephrase this by saying that any particle in free fall from space-time point A to space-time point B minimizes the action S given by

$$S = -c^2 m \int_A^B d\tau \,. \tag{145}$$

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 289

Ref. 146

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.

Challenge 218 s

Ref. 147

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?) For example, on the Moon, where there is no air, David Scott from Apollo 15 dropped a hammer and a feather and found that they fell together, alongside each other. The independence on material composition has been checked and confirmed over and over again.

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 137 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 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 depends on the mass of the body and the distance of passage. However, nobody in the nineteenth century was able to check the result experimentally. Ref. 148 Vol. I, page 201

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 details will be given shortly. Page 161

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 light near the Earth must be the same as that of stones, given by expression (144)? Challenge 219 ny

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

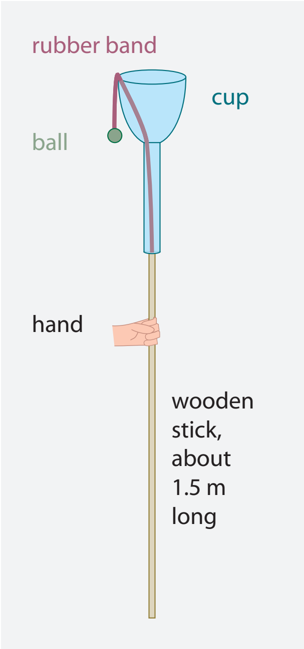


FIGURE 68 A puzzle: what is the simplest way to get the ball attached to the rubber band into the cup?

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 220 s

* *

On his seventy-sixth birthday, Einstein received a birthday present specially made for him, shown in Figure 68. 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 221 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 222 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.'

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 223 e

* *

Challenge 224 s
A piece of wood floats on water. Does it stick out more or less in a lift accelerating upwards?

* *

Page 55
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 225 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 226 s

* *

The experiments of Figure 62 and 63 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 227 s

* *

Challenge 228 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 229 s

* *

Vol. I, page 106
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.
Vol. V, page 140

* *

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
Ref. 149
of negative mass. The statement that this does not happen in nature is also called the *weak*
Challenge 230 ny
energy condition. Is it implied by the limit on length-to-mass ratios?

* *

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
Challenge 231 ny
that the body's length contraction reaches the limit?

* *

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 explored in the last part of our adventure.

* *

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 232 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
Challenge 233 s
would the size of this 'molecule' be?

* *

Challenge 234 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
Vol. I, page 156
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
Vol. I, page 456
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
Ref. 150
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 where the law follows science. (Only the BBC continues to use it.)

* *

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.*

* *

General relativity is also used by space agencies around the world to calculate the exact Ref. 151 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 (137) as

$$\left(\frac{\mathrm{d}\tau}{\mathrm{d}t}\right)^2 = 1 - \frac{2GM}{rc^2} - \frac{r^2}{c^2}\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2 = 1 - \frac{2GM}{rc^2} - \frac{v^2}{c^2} . \tag{147}$$

Challenge 235 e For the relation between satellite time and Earth time we then get

$$\left(\frac{\mathrm{d}t_{\mathrm{sat}}}{\mathrm{d}t_{\mathrm{Earth}}}\right)^2 = \frac{1 - \frac{2GM}{r_{\mathrm{sat}}c^2} - \frac{v^2_{\mathrm{sat}}}{c^2}}{1 - \frac{2GM}{r_{\mathrm{Earth}}c^2} - \frac{v^2_{\mathrm{Earth}}}{c^2}} . \tag{148}$$

Can you deduce how many microseconds a satellite clock gains every day, given that the Challenge 236 s GPS satellites orbit the Earth once every twelve hours? Since only three microseconds

* 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:

$$a = \frac{GM}{r^2} + f_2\frac{GM}{r^2}\frac{v^2}{c^2} + f_4\frac{GM}{r^2}\frac{v^4}{c^4} + f_5\frac{Gm}{r^2}\frac{v^5}{c^5} + \cdots \tag{146}$$

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. Page 180 Higher-order effects, up to f_5, have been measured in the binary pulsars, as discussed below.

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.

** For more information, see the www.gpsworld.com website.

Ref. 152 would give a position error of one kilometre after a single day, the clocks in the satellites 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. 153 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 be checked? Challenge 237 d

* *

Could our experience that we live in only three spatial dimensions be due to a limitation of our senses? How? Challenge 238 s

* *

Challenge 239 s Can you estimate the effect of the tides on the colour of the light emitted by an atom?

* *

Ref. 154 The strongest possible gravitational field is that of a small black hole. The strongest gravitational 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 240 ny

* *

Ref. 155 Challenge 241 e Light deflection changes the angular size δ of a mass M with radius r when observed at distance d. The effect leads to the pretty expression

$$\delta = \arcsin\left(\frac{r\sqrt{1-R_S/d}}{d\sqrt{1-R_S/r}}\right) \quad \text{where} \quad R_S = \frac{2GM}{c^2} \,. \qquad (149)$$

Challenge 242 e What percentage of the surface of the Sun can an observer at infinity see? We will examine this issue in more detail shortly. Page 276

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 Ref. 156 is indeed observed in experiments, and to the highest precision achievable. Roland von Eötvös* performed many such high-precision experiments throughout his life, without finding any discrepancy. In these experiments, he used the connection that the inertial mass determines centrifugal effects and the gravitational mass determines free fall. (Can Challenge 243 ny you imagine how he tested the equality?) Recent experiments showed that the two masses Ref. 156 agree to one part in 10^{-12}.

However, the mass equality is not a surprise. Remembering the definition of mass Vol. I, page 101 ratio as negative inverse acceleration ratio, 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 Vol. I, page 202 in Galilean physics, 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 building; 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

* Roland von Eötvös (b. 1848 Budapest, d. 1919 Budapest), physicist. He performed many high-precision gravity experiments; among other discoveries, he discovered the effect named for him. The university of Budapest bears his name.

** ‘Going it acquires strength.’ Publius Vergilius Maro (b. 70 bce Andes, d. 19 bce Brundisium), from the *Aeneid* 4, 175.

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 244 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

In situations with gravity, time depends on height. The invariance of the speed of light implies that space and space-time are *curved* in all regions where gravity plays a role. Curvature of space can be visualized by threading space with lines of equal distance or by imagining space as a mattress. In situations with gravity, these lines are curved. Masses thus curve space, especially large ones. Curved space influences and determines the motion of test masses and of light.

Chapter 6

OPEN ORBITS, BENT LIGHT AND WOBBLING VACUUM

> 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.

Ref. 157

Before we tackle the details of general relativity, we first 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 check how weak curvature influences motion.

Gravity is strong only near horizons. Strong gravity occurs when the mass M and the distance scale R obey

$$\frac{2GM}{Rc^2} \approx 1 . \qquad (150)$$

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 adventure. In contrast, in most regions of the universe, including our own planet and our solar system, there are *no* nearby horizons; in these cases, gravity is a *weak* effect. This is the topic of the present chapter.

Weak fields

In everyday life, despite the violence of avalanches or of falling asteroids, forces due to gravity are much weaker than the maximum force. On the Earth, the ratio $2GM/Rc^2$ is only about 10^{-9}. Therefore, all cases of everyday life, relativistic gravitation can still be approximated by a field, i.e., with a potential added to flat space-time, despite all what was said above about curvature of space.

Weak gravity situations are interesting because they are simple to understand and to describe; 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 effects that do not occur in universal gravity. To explore these interesting effects, we just need a consistent relativistic treatment.