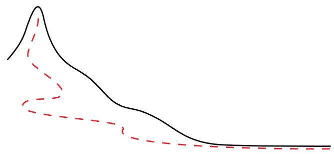
Christoph Schiller

Motion Mountain



Hiking beyond space and time

along the concepts of modern physics

available at

www.motionmountain.org

Picture credits are listen on page 772.

cs@motionmountain.org

To my brothers Stephan and Philipp

Alla zia sconosciuta

τῷ ἐμοὶ δαίμονι

Das Schaudern ist der Menschheit bestes Teil.
Wie auch die Welt ihm das Gefühl verteure,
Ergriffen fühlt er tief das Ungeheure.

Goethe, *Faust*.

Die Menschen stärken, die Sachen klären.

Content Overview

Third Part : Motion Without Motion What are Space, Time, and Particles?

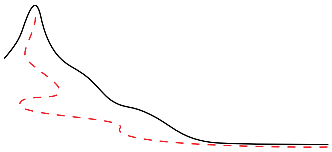
Appendices

Contents

Appendices

This is a section of the freely downloadable e-textbook

Motion Mountain



Hiking beyond space and time along the concepts of modern physics

available at
www.motionmountain.org

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Challenge - Did you find any mistakes?

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

Foreword

Primum movere, deinde docere.

When watching the intensity with which small children explore their environment, one cannot avoid coming to the conclusion that there is a drive to grasp the way the world works, a 'physics instinct', built into each of us. What would happen if this drive, instead of dying out when the limits of the usual education are reached, would be allowed to thrive in an environment without bounds, reaching from the atoms to the stars? Probably each adolescent would know more about nature than the average physics professor today. This text tries to provide this possibility to the reader. It acts as a guide for such an exploration, free of all limitations, of the exotic 'island of motion.'

Every text on physics is a gamble. A gamble on the selection of the topics as well as a gamble on the depth with which they are treated. The present project is the result of a threefold aim I have pursued since 1990: to present the basics of *motion* in a way that is simple, up to date, and vivid. Being *simple* implies to focus on concepts and understanding, and to reduce the mathematics to the necessary minimum. It was my aim to focus on the essence of problems, and to avoid unnecessary detail. Learning the language of physics was given precedence over using formulas in calculations.

Being *up to date* implies the inclusion of quantum gravity, string theory, and M theory. These domains have lead to numerous results not yet found in books accessible to undergraduates. But also the standard topics of mechanics, light, quantum theory, and general relativity are greatly enriched by a systematic collection of results scattered around the scientific literature and not commonly found in textbooks. However, the topic of this text being the fundamentals of motion, domains like material physics, biophysics, statistical physics, hydrodynamics, selforganisation, and most of elementary particle physics are skipped, despite their interest.

Being *vivid* means to challenge, to question, and to dare. The text is everywhere as provocative as possible without being wrong. I avoided boredom as much as possible. In many texts the intellectual interest of the physical description of nature is not made tangible at all. Often they tend to be commented formula collections and to drown the beauty of the topic in pages and pages of formalism. My experience with colleagues, physics students and laymen showed that the best way to produce interest for a topic is to make a simple, correct, but surprising statement. Surprises perplex, confuse, make angry, open up the mind, and usually reveal something about oneself and the world. Following this connection, this text is built around a collection of the most astonishing facts known on the topic of motion. In my view, reading a book on general physics should be similar to a visit to a magician's show. One watches, is astonished, doesn't believe one's eyes, thinks, understands the trick, goes on watching, is astonished again, and so on. Nature is similar: many things are different from the way they seem to be. The text tries to follow a simple rule: *On each page, there is at least one surprise or one provocation to think.*

The surprises are organized to lead in a natural way to the most astonishing of all, namely that space and time do not exist, and that these concepts, so useful for everyday life, are only approximations which are not valid in all cases. Time and space turn out to be mental crutches which hinder the complete exploration of the world. This text is an introduction on how one can think without using these so familiar concepts. The flexibility necessary to achieve this result is trained by the collection of the preceding surprises.

Surprises have the strongest effect whenever they question everyday observations. Therefore the examples in this text are taken as much as possible from daily life; most are taken from the experiences one makes when climbing a mountain. Observations about trees, stones, the moon, the sky, and people are used wherever possible; complicated laboratory experiments are mentioned only where necessary, to avoid cluttering the mind.

Books are old-fashioned means of communication: their possibilities for interactivity are limited. Nevertheless, this text wants to make it clear that the study of motion, like the rest of science, is also a part of the entertainment industry. To achieve this, I tried to make the reader discover and enjoy conceptual physics in the same way that children discover and enjoy the world around them. Numerous challenges are proposed, some hard, some less hard. Most, but not all, are answered later in the text. In addition, the walk is as structured and as complete as possible, so that every topic can be enjoyed by itself. All children discover the world starting from scratch and learning everyday life, step after step. This happens here as well.

Children are often refreshingly direct. In this adventure, a clear and consistent line of thought is presented, even though it may often be controversial. On the other hand, children are flexible; they like to change situations around and explore them from new and fresh viewpoints. Also in physics, viewpoints have to be changed regularly to get to the bottom of things. The definitions of terms such as 'object', 'particle', 'state', 'mass', 'space', 'vacuum', or 'time' are changed several times, as indeed happened in the history of the subject.

Children also like to dare. One needs courage to drop space and time as tools for the description for the world and to approach nature with the openness resulting from complete freedom of preconceptions. Just ask a physicist whether the world is deterministic or space continuous, and whether he would put his hands into fire about the answer. Emotions will quickly run high. But after all, nothing is more challenging, intense and satisfying than overcoming one's own fears. Facing the dark sides of the world and of oneself is an opportunity one should never miss.

Challenge

In the literature, in lectures, and in the mass media one finds a number of regularly repeated statements which do not hold water when subjected to close scrutiny. The uncovering of such beliefs forms an important part of the surprises mentioned above. The persistence of many of these statements is amazing, sometimes saddening and sometimes hilarious; but in most cases it blocks the advance of understanding. In this sense, this text does its best to cut down the number of lies, simply because lies reduce the intensity of life and pleasure. With the freedom of thought thus achieved, one is ready to enjoy one's curiosity much more intensely and to be entertained by its discoveries.

Many people who have kept the flame of their childhood curiosity alive have helped to make this project come true. Fernand Mayné, Saverio Pascazio, Anna Koolen, Ata Masafumi, Roberto Crespi, Serge Pahaut, Valentin Altarez Menendez, Frank van Heyningen and Maria Scali, Herman Elswijk, Marcel Krijn, Marc de Jong, Martin van der Mark, Kim Jalink, my parents Peter and Isabella Schiller, Mike van Wijk, Renate Georgi, Paul Tegelaar, Ron Murdock, and especially my wife Britta have given me encouragement, maybe the most precious present one can get. The project and the collection of material owes much to the numerous acquaintances from my work environment and from the internet, amongst them

Bert Peeters, Anna Wierzbicka, William Beaty, Jim Carr, John Merrit, John Baez, Frank DiFilippo, Jonathan Scott, Luca Bombelli, Douglas Singleton, George McQuarry, Tilman Hausherr, Brian Oberquell, Peer Zalm, Martin van der Mark, Vladimir Surdin, Julia Simon, Antonio Fermani, Don Page, Stephen Haley Peter Mayr, Allan Hayes, Norbert Dragon, Igor Ivanov, Doug Renselle, Wim de Muynck, Steve Carlip, Tom Bruce, Ryan Budney, Gary Ruben, Chris Hillman, Jochen Greiner, squark and Martin Hardcastle. The software tools were refined with help from Donald Arsenau, Sebastian Rahtz, Don Story, Vincent Darley, Johan Linde, Joseph Hertzlinger, Rick Zaccone, John Warkentin.

In a simple view of the world, one can talk about three types of adventures: of the body, of the mind, and of emotions. Achieving a description of the world without the use of space and time is one of the most beautiful of all the possible adventures of the mind. Telling its story allowed me to include the other two aspects. The gamble paid off if the text fills a need.

Eindhoven, 29 April 2001

1. An appetizer

Die Lösung des Rätsels des Lebens in Raum und Zeit
liegt *außerhalb* von Raum und Zeit.*
Ludwig Wittgenstein, *Tractatus*, 6.4312

What is the most daring and most exciting journey one can have in a lifetime? Which is the most interesting place to visit? One can travel to places as remote as possible, like explorers or astronauts do, one can look into places as far as imaginable, like astronomers do, one can visit the past, like historians or archaeologists do, or one can delve as deeply as possible into the human soul, like artists or psychologists do. All these voyages lead either to other places or to other times (or nowadays, to other servers on the internet). However, one can do much better: the most daring trip is not the one leading to the most inaccessible place, but the one leading to where there is no place at all. Such a journey implies leaving the prison of space and time and venturing beyond it, into a domain where there is no position, no present, no future, and no past, where one is free of all restrictions, but also of any security of thought. There, discoveries still are to be made, adventures to be fought, and stories to be told. Almost nobody has ever been there; humanity took 2500 years to complete the trip, and she achieved it only recently.

For example, did you know that when two points are too near to each other it is not always possible to find room for a third one in between? To venture into this part of nature, one needs to be curious about the essence of travel itself, with ever-increasing precision. Obviously, the essence of any travel is *motion*. In principle, the search to understand motion can be performed behind a desk, with a few books, some paper and a pen, without moving much at all, all by oneself.

In contrast, this text tells the story of this search as the escalation of a mountain. During this escalation every step towards the top will correspond to a step towards higher precision in the description of motion. At the same time, each step will increase the pleasure and the delights. At the top of the mountain, one will arrive in the domain one was looking for, where 'space' and 'time' are words which have lost all content, and where the sight on the beauty of the world is overwhelming and unforgettable.

The challenges one has to face when trying to describe the world without space or time can be experienced with the following self-test: Challenge

- Can you show whether between two points extremely near to each other there is always room for a third one in between?
- Have you ever tried to describe the shape of a knot through the telephone?
- Have you ever tried to make a telephone appointment with a friend without using one of the following words: clock, hour, place, at, near, before, after, near, upon, under, above, below?
- Can you explain on the telephone what 'right' and 'left' mean, or what a mirror is?
- Can you describe the fall of a stone without using space or time?
- Do you know *any observation at all* which you can describe without concepts from the domains 'space', 'time' or 'object'?

* The solution of the riddle of life in space and time lies *outside* space and time.

• Can you imagine a domain of nature where matter and vacuum are indistinguishable?
• Can you imagine a finite history of the universe, but without a ‘first instant of time’?
• Have you ever tried to talk to a theoretical particle physicists with the hope to understand what he does and did you hear him – falsely – saying that he could not explain his work to somebody who has only a high school physics background?
• Can you explain what time is? And what clocks are?
• Have you ever tried to understand why things move? Or why motion exists at all?

This book tells how to achieve these feats, bringing to completion an old adventure of the human spirit, namely the quest to describe *every* possible aspect of motion.

Why do shoestrings usually remain tied? Why does space have three dimensions? Why not another number? It took people thousands of years to uncover the answer. In order to find it, several other simple but disturbing questions had to be answered, such as: What is space? Well, everybody knows that it somehow describes the possibilities one has to move things around. Therefore space has something to do with motion. What is motion precisely? Well, motion is change of position with time. But what is time? One is used to hearing: “nobody knows!” That would imply that nobody knows what space, time and motion really are. However, this is not the case any more. Even though simple questions such as these are among the most difficult known, results from the last twenty years of research finally allow them to be answered.

Why does everything fall downwards? Why is grass green and not blue, milk white and not black, gold yellow, blood red, and hair sometimes blond and sometimes black, but never mauve? Why does the sun shine? Why does the moon not fall from the sky? Why is the sky dark at night? Why is water liquid and fire hot? Why is the universe so big? Why does the weather change so often? Why can birds fly but men can’t? Why is there so much to learn? Why is lightning not straight? Why is there no purple in a rainbow? Why are atoms not square, nor the size of cherries? Why does the floor not fall? Why are computers not faster? All these questions seem to have little in common. But this impression is wrong. They are all about bodies, about interactions, and most of all, about motion. Indeed, they will all appear and be answered in what follows.

In the course of this promenade, one learns that in contrast to personal experience, motion never stops. One learns that there are more cells in the brain than stars in the galaxy. (People almost literally have a whole universe in their head.) One learns that perfect memory cannot exist. One learns that every clock has a certain probability to go backwards. One learns, literally, that time does not exist. One finds that all objects in the world are connected. One learns that matter and empty space cannot be distinguished precisely. One learns that accelerated mirrors emit light. One learns that one can measure gravity with a thermometer. One learns that we literally are made of nothing.

What type of bodies exist? What type of interactions? Why do they produce motion? What is motion anyway? People went on asking and asking, until they were able to show that bodies, motion and forces are terms which cannot be defined precisely, cannot even be distinguished exactly, and that they are only approximations of a single, deeper layer of nature, of which they in fact are three different manifestations. Which layer? Well, that is the story told in this text.

For children and physicists alike, delving into these connections is the way to have fun; curiosity always leads to strong emotions. This adventure into the unknown, with its fascinating, its frightening and its mysterious steps, is divided into three parts. Don't panic. All topics will be introduced step by step, in such a way that one can understand and enjoy them all.

How do things move? The standard general answer is that all motion is the change of position with time of some interacting thing. This seems a boring statement; however, it contains in it general relativity, one of the most incredible descriptions of nature one can imagine. One finds that space is warped, that light does not usually travel straight, and that time is not the same for everybody; one also discovers that gravity is not an interaction, but that it is the change of time with position in space, that the surface of the earth is continually accelerating upwards, and that the blackness of the sky at night proves that the universe has a finite age. These and other strange properties of motion are summarized in the first part of this text, on classical physics. They lead directly to the next question:

What are things? The first answer is that all things are made from a few types of particles. This is common knowledge today; it is complemented by the result that all interactions and forces, those of the muscles, those which make the sun burn, those which make the earth turn, those which decide over attraction, repulsion, indifference, friction, creation and annihilation, are also made of particles. The growth of trees, the colours of the sky, the burning of fire, the warmth of a human body, the waves of the sea, and the mood changes of people are variations of motion of particles. This story is told in more detail in the second part, that on quantum mechanics. One learns that in principle, watches cannot work properly, that it is impossible to completely fill a glass of wine, and that some people are able to transform light into matter. You still think it's boring? Just read about the substantial dangers one incurs when buying a can of beans. At this point the path is prepared for the central theme of this escalation:

What are particles, position, and time? The recent results of an age-long search will make it possible to answer this question in the near future. This third part is *not complete yet*, because the final research results are not yet available. But there are good reasons to continue the adventure:

- It is *known already* that space and time are not continuous, that – to be precise – neither points nor particles exist, and that there is no way to distinguish space from time, nor vacuum from matter, nor matter from radiation. It even turns out that nature is *not* made of particles and vacuum, in contrast to what one is used to think. See chapter XII

- It *seems* that position, time, and every particle are aspects of a complex, extended entity incessantly varying in shape. The precise meaning of this statement will become clear during the third leg.

- Mysteries which should be cleared up *in the coming years* are the origin of the three space dimensions, the origin of time, and the details of the big bang.

- Research is *presently uncovering* that motion is an intrinsic property of matter and radiation and that it appears as soon as one introduces these two concepts in the description of nature. On the other hand, it is impossible *not* to introduce them, because they automatically appear when one divides nature into parts, an act we cannot avoid due to the mechanisms of our thinking.

- Research is also *presently uncovering* that the final description of nature, with complete precision, does not use any form of infinity. One finds, step by step, that all infinities appearing in the human description of nature, both the infinitely large as well as the infinitely small, result from approximations. "Infinity" turns out to be an exaggeration which does not apply to nature at all. At the same time, one finds that the precise description does not include any finite quantities either! These and many other astonishing results of modern physics will form the third part of this text.

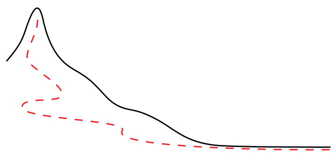
This final part develops the present state of the search for a unified description of general relativity and that of quantum mechanics, overcoming their mutual contradictions. Wiping out the inconsistencies between these two descriptions will be one of the most astonishing successes of physics; it will complete the description of motion in all its aspects, from the motion of electrons in the brain when thinking to the motion of the stars on the other end of the universe. The secrets of space, time, matter and forces have to be unraveled to achieve it. It is a fascinating story, put together piece by piece by thousands of researchers.

In any escalation, every now and then one finds something particularly interesting. Often, one then takes a small detour in order to have a closer look. Sometimes one discovers an interesting route to be tried in a following trip. In this text, the 'intermezzi', the sections entitled 'curiosities', and the footnotes correspond to such detours. The footnotes give a selection of interesting literature into nearby fields of inquiry, to satisfy any strong curiosity in directions different from the one chosen here; books telling how to build telescopes, how to fool one's senses of sight, how to move without tension in one's body, how to understand colours, how to talk, how order and beauty appear in nature, how elementary particles were discovered, and many others are mentioned and recommended, selected for quality in their exposition. In contrast, the references at the end of each chapter list sources for material used or mentioned in the text. In addition, a large number of world-wide web sites is cited, with the same distinction. In the electronic version of this text, clicking allows to access them directly.

The text is completed by a number of appendices which list the symbols used in the notation, give the definitions of physical units, provide an overview of physical constants and particles, present intuitive definitions of some mathematical concepts, and list general sources for literature and internet information. Lists of all tables and figures are given, as well as an index referring to all used concepts and to all the mentioned names and cited authors.

At the end of the escalation, on the top of the mountain, the idea of motion will have undergone a deep transformation. Without space and time, the world looks magical, incredibly simple and astonishingly fascinating at the same time – pure beauty.

First Part



Classical Physics
How do Things and Images Move?

In which the description of hiking and other everyday motion
leads us to introduce for its description
the concepts of time, length, mass, charge,
field, manifold, and lagrangian,
which allow us to understand among others
why we have legs instead of wheels,
how empty space can wobble and move,
what sex has to do with magnets and amber,
and why we can see the stars.

Typeset in April 2001

Chapter II Classical Motion *

2. Why care about motion?

All motion is an illusion.
Zeno of Elea, ca. 450BCE

Wham! The lightning crashing in the tree nearby violently disrupts the quietness of our walk through the forest. Our hearts suddenly beat faster. But the fire that started in the tree quickly fades away. The gentle wind moving the leaves around us helps to restore the calmness of the place. Nearby, the water in a small river follows its complicated way down the valley, reflecting on its surface the everchanging shapes of the clouds.

Motion is everywhere, friendly and threatening, horrible and beautiful. It is fundamental to our human existence; we need motion for learning, for thinking, for growing, and for enjoying life. We need it for walking through a forest, for listening to it with our eardrums and for talking about it with our vocal chords. Like all animals, we rely on motion to get food, to survive dangers,** and to reproduce; like all living beings, we need motion to breathe and to digest; like all objects, motion keeps us warm. Motion is such a basic part of our observations that even the origin of the word is lost in the darkness of indo-european linguistic history.

Figure will be added in the future

Figure 1 An example of motion observed in nature

Who are we? Where do we come from? What will we do? What will future bring? Where do people come from? Where do they go to? What is death? Where does the world come from? Where does life lead to? All these questions are questions about motion. (And the answers are often surprising.)

Motion is the most fundamental observation about nature at large. It turns out that *everything* which happens in the world is some type of motion. There are no exceptions.

∗ Since the final chapter is also the first, this is the second chapter.

∗∗ Plants for example cannot move; for their self-defense, they developed *poisons*. Examples of such plants are the stinging nettle, the tobacco plant, digitalis, belladonna, and poppy; poisons include caffeine, nicotine, curare, and many others. Poisons such as these are at the basis of most medicines. Therefore, most medicines exist essentially because plants have no legs.

Therefore it has always been a favorite object of curiosity. Already at the beginning of written thought, in the sixth century BCE in ancient Greece, its study had been given a name: *physics*. Motion is mysterious. Even though motion is everywhere, in the stars, in the tides, in our eyelids, neither these ancient thinkers nor myriads of others in the following twenty-five centuries were able to shed some light on the central mystery: *what is motion?* Until a few years ago the only answer was "motion is the change of place in time", or some variation of it. But we will soon discover that these words have almost no meaning! Only recently an answer has finally been found; this is the story of the way to reach it. Ref. 1

Apart from being fascinating and important, the subject of motion is first of all vast. Studying it resembles the exploration of a large unknown island, after being carried to the shore by the waves. The size of the unknown jungle on the island is overwhelming. We wonder where to start, knowing that a whole life does not suffice to explore it all.

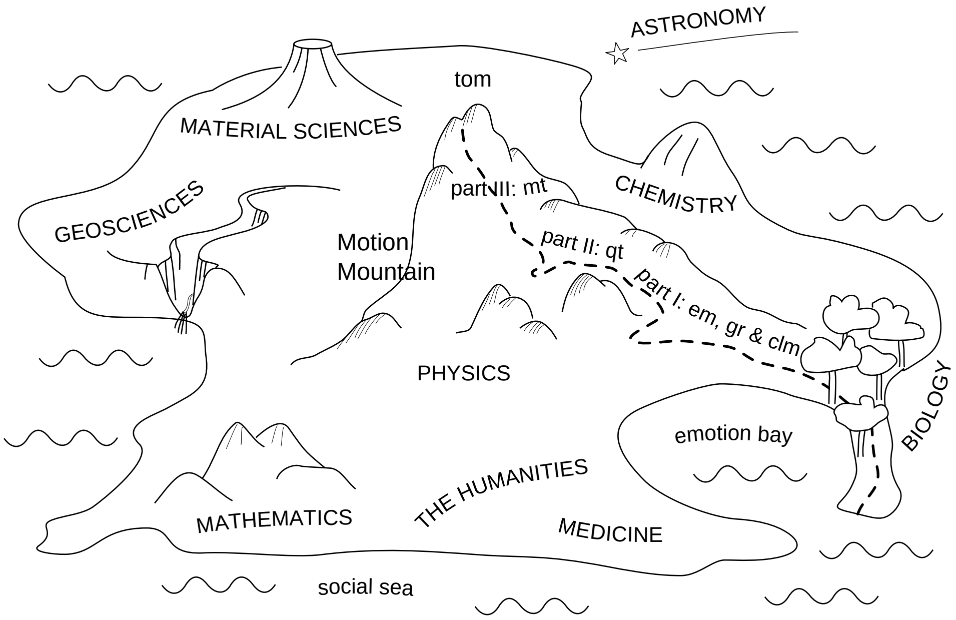


Figure 2 Experience island, or Motion island, and the trail to be followed (clm: classical mechanics, gr: general relativity, em: electromagnetism, qt: quantum theory, mt: M-theory, tom: the theory of motion)

However, near the center of the island an especially high mountain stands out. From its top we can oversee the whole landscape, and we can get an impression of the relations between the various regions, i.e. between the various examples of motion. This is a guide to the top of this mountain, and to one of the most beautiful adventures of the human mind. Obviously, the first question to ask is:

Das Rätsel gibt es nicht. Wenn sich
eine Frage überhaupt stellen läßt,
so *kann* sie beantwortet werden.*
Ludwig Wittgenstein, *Tractatus*, 6.5

* *The riddle* does not exist. If a question can be put at all, it *can* be answered.

Does motion exist?

> Any fool can ask more questions
> than seven sages can answer.

Ref. 2

No, was the answer given by many. Their arguments are important: they deeply influenced the investigation of motion in the past and they still continue do so. The issues raised by them will accompany us throughout our escalation.

Challenge

To sharpen the mind, one only has to look at figure 3, and slightly move the page. The figure seems to rotate. How can one make sure that real motion is actually different from this or some similar type of illusion? *

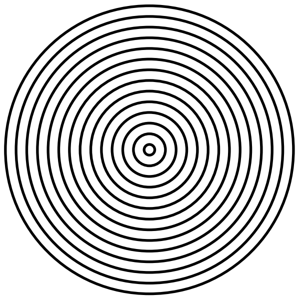


Figure 3 Illusion of motion: fake rotation.

Ref. 3

One answer was put forward as main thesis by the greek philosopher Parmenides of Elea (born ca. 515 BCE). He pondered the following question: given that nothing comes from nothing, how can change be? He underlined the *permanence* of nature and maintained that all change and thus all motion is an illusion.

Challenge

The opposite view was held by Heraclitos (ca. 540–ca. 480 BCE) and expressed in his sentence πάντα ῥεῖ ‘everything flows’. These two equally famous opinions induced many to investigate in more detail whether in nature there are *conserved* quantities, or whether *creation* is possible. Can you guess the result of of these efforts?

Ref. 4

Parmenides’ collaborator Zeno of Elea (born ca. 500 BCE) argued so intensely against motion that some people still worry about it today. In one of his arguments he claims – in simple language – that it is impossible to slap somebody, since the hand has first to travel halfway to the face, then travel through the remaining half, then again so, and so on; it would therefore never reach the face. This argument induces one to think carefully about the relation between *infinity* and its opposite, finitude, in the description of motion. In modern quantum theory, the same issue still troubles many scientists up to this day.

Another of Zeno’s provocations was to say that by looking at a moving object at a *single* instant of time, one cannot maintain that it moves. This raises the question whether motion can clearly be distinguished from its opposite, namely *rest*. In our walk, like in the history of physics, we will change from a positive to a negative answer a few times! Thinking about this question led Albert Einstein to the development of general relativity, the high point of the first part of our escalation. Additionally, one is led to ask whether single instants do exist at all. This far-reaching question is central to the third part of our walk.

The second leg of the escalation, that on quantum theory, will show that motion is in many ways an illusion, as Parmenides claimed. More precisely, we will show that the observation of motion is due to the limitations of our human condition. We will find that we experience motion only because as human beings we live on the earth, only because we are of finite size, only because we are made of a large but finite number of atoms, only because we have a finite but moderate temperature, only because we are electrically neutral, only because we are large compared to a black hole of our same mass, only because we are large compared to our quantum mechanical wavelength, only because we have a limited memory, only because

∗ Solution to *challenges* are usually given later on in the walk.

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Did you find any mistakes? Challenge

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

our brain forces us to approximate space and time by continuous entities, and only because our brain cannot avoid describing nature as made of different parts. If any of these conditions were not fulfilled, we would not observe motion; motion would not exist.

In order to get through all these topics in an efficient manner, one usually starts with the following question:

How should one talk about motion?

[...]
Je hais le mouvement, qui déplace les lignes,
Et jamais je ne pleure et jamais je ne ris.
[...]
Charles Baudelaire, *La Beauté*.*

Simple: with *precision*; and with *curiosity*. Precision makes meaningful communication possible, and curiosity makes it worthwhile.** Whenever one is talking about motion and one is aiming for increasing precision or for more detailed knowledge, one is engaged, whether one knows it or not, in the ascent of motion mountain. Every time one increases the precision of description, one gains some height towards the top.

High precision means going into details. In contrast to what one may think, this method *increases* the pleasure of the adventure.*** The higher one gets, the farther one can see and the more curiosity gets rewarded. The views offered are breathtaking, especially at the very top. The path we will follow – one of the many possible ones – starts from the the side of
Ref. 6 biology, and begins in the forest lying at the foot of the mountain.

What are the types of motion?

Every movement is born of
a desire for change.

The best place to get a general answer is a big library. The domains where motion, movements and moves play a role a rather varied. One can find books on motion pictures and on motion therapy, on motion perception and on motion sickness, on motion for fitness and for meditation, on perpetual motion, on motion in dance, in music and in the other arts, on
Ref. 7 motion as proof of existence of various gods, on motion of insects, horses, robots, stars and angels, on the economic efficiency of motion, on emotion, locomotion and commotion, on motions in parliament, on movements in art, in the sciences and in politics, on movements in watches and in the stock market, on movement teaching and learning, on movement development in children, on musical and on troop movements, on religious and on bowel movements, on moves in chess, and many many more.

* Charles Baudelaire (1821, Paris–1867, Paris) *Beauty*: [...] I hate movement, which changes shapes, [...] The full text of this and the other poems from *Les fleurs du mal*, one of the most beautiful books of poetry ever written, can be found at the http://hypermedia.univ-paris8.fr/bibliotheque/Baudelaire/Spleen.html web site.

** For a collection of interesting examples of motion in everyday life, see the excellent and rightly famous book
Ref. 5 by Jearl Walker, *The flying circus of physics*, Wiley, 1975.

Challenge *** Distrust anybody who wants to talk you *out* of investigating details. He is trying to get you. Be vigilant also during *this* walk.

Already a long time ago people had the suspicion that all these types of motion, as well as other types of change, are related. It is commonplace to distinguish at least three categories.

The first category is that of material *transport*, such as a person walking or a leaf falling from a tree. Transport is the change of position and of orientation of objects. For example, the behaviour of people falls into this category.

Another category groups observations such as the dissolution of salt in water, the freezing of water, the putrefaction of wood, the cooking of food, the cicatrization of blood and the melting and alloying of metals. These changes of colour, of brightness, of hardness, of temperature and of other material properties are all *transformations*. Transformations are changes not visibly connected with transport. To this category, a few ancient thinkers already added the emission and absorption of light. In the twentieth century, these two effects were proven to be special cases of transformations, as were the newly discovered appearance and disappearance of matter, as observed in the sun and in radioactivity. Also *emotion change*, such as change of mood, of expression, of health, of education, and of character is a type of transformation. Ref. 8

An especially important category of change, is *growth*; it is observed for animals, plants, bacteria, crystals, mountains, etc. In the nineteenth century, changes in the population of systems, *biological evolution*, and in the twentieth century, changes in the size of the universe, the *cosmic evolution*, were added to this category. Traditionally, these phenomena were studied by separate sciences. Independently they all arrived at the conclusion that growth is a combination of transport and of transformation. The difference lies in the complexity and the time scale. Ref. 9

At the beginning of modern science during the renaissance, only the study of transport was seen as the topic of physics. Motion was equated to transport. Despite this restriction, one is still left with a large field of enquiry, covering a large part of experience island. The obvious way to structure the field is to distinguish transport by its origin. Movements such as those of the legs when walking are *volitional*, because they are controlled by one's will, whereas movements of external objects, such as the fall of a snowflake, which one cannot influence by one's will-power, are called *passive*. This distinction is completed by children around the age of six, and marks a central step in the development of every human towards a precise description of its environment.* From this distinction stems the historical, but now outdated definition of physics as the science of motion of non-living things.

Then, one day, machines appeared. From that moment, the distinction between the volitional and passive motion was put into question. Machines, like living beings, are self-moving, and thus mimic volitional motion. But careful observation shows that every part in a machine is moved by another, so that their motion is in fact passive. Are living beings also machines? Are human actions examples of passive motion as well? The accumulation of observations in the past hundred years made it clear that the motion of living beings in

* Failure to pass this stage completely can result in various strange beliefs, such as in the ability to influence roulette balls, as found in compulsive players, or in the ability to move other bodies by thought, as found in numerous otherwise healthy looking people. An entertaining and informative account of the deception and self-deception involved in these beliefs is given by James Randi, a professional magician, in *The faith healers*, Prometheus Books, 1989, as well as in several of his other books. See also the http://www.randi.org web site.

general, including their volitional movements,* has indeed the same physical properties as passive motion in non-living systems. (Of course, from the emotional viewpoint, there are many differences; for example, *grace* can only be ascribed to volitional movements.) The distinction between the two types is thus not necessary and is dropped in the following. This means that since the two types of motion have the same properties, through the study of motion of non-living objects one can learn something about the human condition, as is most evident when one touches the topics of determinism, of causality, of probability, of infinity, of time, and of sex, to name but a few of the themes one encounters on the way.

Ref. 10

With the accumulation of observations in the nineteenth and the twentieth century, more and more restrictions on the study of motion were put into question. Extensive observations showed that also all transformations and all growth phenomena, including behaviour change and evolution, are types of transport. In the middle of the twentieth century this culminated in the confirmation of an idea formulated already in ancient Greece: every type of change is a form of transport, and in particular, *every* type of change is due to motion of particles. It takes time and work to reach this conclusion, which appears only when one relentlessly pursues higher and higher precision in the description of nature. The first two parts of this walk retrace the way to reach this result.

Then, in the third part, it is shown to be plain wrong. But until then we still have some way to go. At the present point, in the beginning of our walk, the large number of manifestations of motion and of change tell us only one thing: classifying the various appearances of motion is not productive. It does not allow to talk about motion with precision. To achieve precision, we need to select a few specific examples of motion, and to study them in full detail. Only by trying to achieve maximum precision we can hope to arrive at the fundamental properties of motion.

It is intuitively obvious that the most precise description is possible for the *simplest* possible examples. In everyday life, this is for example the motion of a non-living, solid, rigid body near us, such as a stone thrown through the air. Indeed, like all humans, each of us has learned to throw objects long before learning to walk. Throwing was the first act we performed ourselves in a chain of events that lead us here, walking in the forest at the foot of this mountain. During our early childhood, by throwing stones and similar objects until our parents feared for every piece of the household, we learned among other things that in order to describe and to understand motion we needed first of all to distinguish *permanent* aspects, such as objects and images, and *variable* aspects, such as dimensions, position and instants.

> Die Welt ist unabhängig von meinem Willen.**
> Ludwig Wittgenstein, *Tractatus*, 6.373

Do you dislike formulas?

Ref. 8

If you dislike formulas, use the following three minute method to change the situation. It's

* The word 'movement' is rather modern; it was imported from the old french and became popular only at the end of the eighteenth century. It is never used by Shakespeare.

** The world is independent of my will.

worth trying it; it will make you enjoy this book much more. Life is short, and reading this text should be a pleasure, or shouldn't it?

• Close your eyes and think of an experience which you had which was *absolutely marvelous*, when you felt excited, curious, and positive.

• Open your eyes for a second or two and look at page 176 or any other page of your choice which contains many formulas. Then close your eyes and return to your marvelous experience.

• Open your eyes again and look at page 176, then close them again and return to that marvelous experience.

• Redo this three more times.

Then leave the memory, look around yourself to get back into the here and now, and test yourself. Have a look at page 176. How do you feel about formulas now?

Perception, permanence and change

> Only wimps specialize in the general case;
> real scientists pursue examples.
> Beresford Parlett

Human beings enjoy perceiving. We start already before our birth, and continue enjoying it as long as we can. That is why television, even when devoid of content, is so successful. Also during our walk through a forest, as concentrated or absentminded we might be, we cannot avoid perceiving. Perception is first of all the ability to *distinguish*. For example, during childhood we first learned to distinguish *familiar* from *unfamiliar* observations. We use the basic mental act of distinguishing in almost every instant of life, usually together with another basic ability, namely the capacity to *memorize* experiences. Without memory, we would loose the ability to experience, to talk and thus also to study nature. Together, the three activities of perceiving, classifying and memorizing form what is called *learning*. If humans would loose any of the three abilities, they could not study motion.

During early childhood every human rapidly learns that in all examples of motion it is possible to distinguish *permanent* from *changing* aspects. One starts to learn this with human faces, which one learns to *recognize*, even though a face never looks exactly the same as in previous observations. Recognition works pretty well in everyday life – one even recognizes friends at night, after a many beers. Anyway, when growing up, children extend recognition to all observations they make.

Sitting on the grass in a clearing of the forest at the feet of motion mountain, surrounded by the trees and the silence typical for such places, a feeling of calmness and tranquillity pervades us. But suddenly, something moves in the bushes; immediately our eyes turn, the attention focuses. All these reactions are built into our organism. The nerve cells in our brain detecting motion are part of the most ancient piece of our brain, the one we share with birds and reptiles: the brain stem. (The brain stem also controls all our involuntary motions, such as the blinking of the eyes.) Only then the cortex, the modern brain, takes over to analyze the type of motion, i.e. to try to identify its origin. Watching the motion across our field of vision, we observe two invariant entities: the fixed landscape and the moving animal. After we recognized it as a deer, we relax again. But how did we distinguish between landscape and deer? Ref. 11

Several steps in the eye and in the brain are involved. Motion plays an essential part in them, as is best deduced from the flip movie shown in the lower left corner of the following Ref. 12 pages. Each image only shows a rectangle filled with a mathematically random pattern. But when the pages are scanned, one discerns a shape moving against a fixed background. At any given instant, the shape cannot be distinguished from the background; there is no visible object at any given fixed instant of time. Nevertheless it is easy to perceive its motion.* Perception experiments such as this one have been performed in many variations; among others it was found that detecting such a window is nothing special: flies have the same ability, as do, in fact, all animals which have eyes.

The flip movie in the lower left corner, as well as many similar experiments show two connections. First, we perceive motion only if we are able to distinguish an *object* and a Ref. 12 *background* or *environment*. Second, we need motion to define objects and environments, and to distinguish them from each other. In fact, our concept of space is – among others – an abstraction of the idea of background. The background is extended; the moving entity is localized.**

One calls the set of localized aspects which remain invariant during motion, such as size, shape, colour etc., taken together, a (physical) *object* or a (physical) *body*. (We will tighten the definition shortly, since otherwise images would be objects as well.) In other words, right from the start we experience motion as a *relative* process; it is perceived in relation and in opposition to the environment. The same is valid for the concept of object: it is a relative concept as well. The introduction of a basic conceptual distinction between object and environment seems trivial and unimportant. We are so used to the possibility of isolating local systems that we take it for granted. However, as we will see in the third part of our walk, this distinction turns out to be logically and experimentally impossible!*** Our walk will lead us to discover the reason for this impossibility and its important consequences.

Does the world need states?

> “Wisdom is one thing: to understand the thought
> which steers all things through all things.”
> Ref. 14 Heraclitos of Ephesos

> Das Feste, das Bestehende und der Gegenstand sind Eins.
> Der Gegenstand ist das Feste, Bestehende;
> die Konfiguration ist das Wechselnde, Unbeständige.****
> Ludwig Wittgenstein, *Tractatus*, 2.027 - 2.0271

* The human eye is rather good at detecting motion. For example, the eye can detect motion of a point of light even if the change of angle is smaller than what can be distinguished for fixed images. Details of this and similar Ref. 13 topics for the other senses are the domain of perception research.

** There are many more aspects around the perception of motion. A good introduction is chapter 6 of the beautiful text by Donald D. Hoffman, *Visual intelligence – how we create what we see*, W.W. Norton & Co., 1998. Many motion illusions can be experienced and explored on the associated ari.ss.uci.edu/cogsci/personnel/hoffman.html web site.

*** However, the distinction *is* possible in quantum theory, contrary to what is often read in popular literature; the distinction becomes impossible only when quantum theory is unified with general relativity.

**** Objects, the unalterable, and the subsistent are one and the same. Objects are what is unalterable and subsistent; their configuration is what is changing and unstable.

motion
the basic type of change

parts	*relations*	*background*
permanent bounded shaped	variable unbounded unshaped	permanent extended measurable

images	*objects*	*states*	*interactions*	*phase space*	*space-time*
penetrable	impenetrable	global	local	composed	simple
aspects:					
intensity	mass	instant	source	dimension	curvature
colour	size	position	domain	distance	topology
appearance	charge	momentum	strength	volume	distance
disappearance	spin	energy	direction	subspaces	area
etc.	etc.	etc.	etc.	etc.	etc.

world – nature – universe – cosmos
the collection of all parts, relations, and backgrounds

Table 1 Family tree of the basic physical concepts

What distinguishes the various patterns in the lower left corners of this text? In everyday life we would say: the *situation* or *configuration* of the involved entities. The situation somehow describes all those aspects which can differ from case to case. It is customary to call the list of all non-permanent or *variable* aspects of a set of objects their *(physical) state of motion*, or simply their *state*.

The situations in the lower left corners differ first of all in *time*. Time is what makes opposites possible: a child is in a house and the same child is outside the house; time describes and resolves this type of contradictions. But the state not only distinguishes situations in time. The state also contains all those aspects of a system (a group of objects) which sets it apart from all *similar* systems. Two objects can have the same mass, the same shape, the same colour, the same composition and can be indistinguishable in all other intrinsic properties; but at least they will differ in their position, or their velocity, or their orientation. The state pinpoints the individuality of a physical system, and allows one to distinguish it from exact copies of itself. Therefore, the state also describes the relation of an object or a system with respect to its environment. Or in short: *the state describes all aspects of a system which depend on the observer.* All this seems too boring? Then just ponder this: does the universe have a state? Challenge

Describing nature as a collection of permanent entities and changing states is the starting point of the study of motion. The various aspects of objects and of their states are called *observables*. All these rough, preliminary definitions will be refined step by step in the following. Using the terms just introduced, one can say: *motion is the change of state of* *objects*.*

In order to proceed and to achieve a *complete* description of motion, we thus need a complete description of objects and a complete description of their possible states. This program can indeed be realized: several *consistent* descriptions of motion are possible, depending on the level of precision required. Each description is an approximation of another, finally leading to the complete description of motion. At each of these levels, the characteristic ingredient is the description of the state of objects, and thus in particular the description of their velocity, their position, and their interactions with time.

3. Galilean physics – motion for kids

Die Maxime, jederzeit selbst zu denken, ist die Aufklärung.**
Immanuel Kant

The simplest description of motion is the one we all unconsciously use in everyday life. It is characterized by a simple general idea: *only one thing at a time can be at one given spot*. This general idea can be separated into three parts that characterize its general approach: matter *moves* and is *impenetrable*, time is made of *instants*, and space is made of spots or *points*. This description is called *galilean physics*, or also newtonian physics. The name is derived from Galileo Galilei (1564–1642), tuscan professor of mechanics, one of the founders of modern physics and a famous advocate of the importance of observations as checks of statements about nature. By requiring and performing these checks throughout his life, he was one of the first researchers on motion to insist on accuracy in its description. This approach changed the speculative description of ancient Greece into the experimental physics in renaissance Italy. He is therefore regarded as the founder of modern physics. The english alchemist, theologian, physicist, and politician Isaac Newton (1643–1727) was one of the first to pursue with vigour the idea that different types of motion have the same properties, and made important steps in constructing the concepts necessary to achieve this program.*** The way that motion, points, instants, and impenetrability are described with precision only needs a short overview.

What is velocity?

There is nothing else like it.
Jochen Rindt****

* Investigating more closely, one finds that the exact separation between those aspects belonging to the object and those belonging to the state depends on the precision of observation. For example, the length of a piece of wood is usually taken to be permanent when one constructs a house; however, more precise observations show that it shrinks and bends with time, due to processes at the molecular scale. To be precise, the length of a piece of wood is thus not an aspect of the object, but an aspect of its state. Precise observations thus *shift* the distinction between the object and its state; the distinction itself does not disappear. Only in the third part of this walk there is a surprising twist to this topic.

** The maxim to think at all times by oneself is the enlightenment.

*** The best and most informative book on the life of Galileo and his times is by Pietro Redondi (see the footnote on page 416). On the http://www.mpiwg-berlin.mpg.de web site one can read an original manuscript
Ref. 15 by Galileo. About Newton and his importance for classical mechanics, see the text by Clifford Truesdell. About
Ref. 16 Newton's infatuation with alchemy, see the books by Dobbs.

**** Jochen Rindt, (1942–1970), austrian formula one driver.

Velocities can	Physical property	Mathematical name (see later for definitions)
be distinguished	distinguishability	element of set
change gradually	continuity	completeness
point somewhere	direction	dimensionality
be compared	measurability	metricity
be added	additivity	euclidean
beat any limit	infinity	unboundedness, openness

Table 2 Properties of galilean velocity

On certain mornings it can be dangerous to walk in forests. When the hunting season opens, men armed with rifles ramble through the landscape and, eager to use again – at last – their beloved weapon, they shoot, as malicious tongues pretend, on everything which moves, including falling leaves or fellow hunters. Here, we are going to do something similar. Like a hunter, we will concentrate on everything that moves; unlike them however, we will not be filled with the desire to stop it moving, but with the desire to follow it and to understand its motion in detail.

We observe that objects can move differently against each other: they can overtake each other. We also observe that they can move in different directions. We observe that velocities can be composed. More properties of velocity can be found in table . For the sum of all these properties together, mathematicians have invented a special term; they say that velocities form a *euclidean vector space*. This is an example of a general connection: every time one aims for the highest precision in describing nature, mathematical concepts are adopted. (For more details about this strange term, see page 50.)

When velocity is said to have these properties, it is called *galilean*. Velocity seems to be a simple and almost boring subject. Well, it is not. The first mistake: one is usually brought up with the idea that velocity needs space and time measurements to be defined first. But this is utterly wrong. Are you able to define a means to measure velocities without measuring space and time? If you are able to do so, you probably want to continue reading on page 145, jumping 2000 years of inquiries. If you cannot do so, here are some hints. Whenever one measures a quantity, one assumes that this is possible, that everybody can do it, and that everybody will get the same result. This means that *measurement* is the comparison with a standard. One thus implicitly assumes that such a standard exists, i.e. that an example of a “perfect” velocity can be found. Historically, the study of motion has not investigated this question first, because nobody could find this velocity, and nobody discovered this measurement method. You are thus in good company.

Ref. 17

Challenge

There is a second mistake in thinking that velocity is a boring subject: the latter stages of our walk will show that every single property mentioned in table is only approximate; *none* is actually correct. That is one reason that makes our hike so exciting. But for the moment, we continue with the next aspect of galilean states.

What is time?

Time does not exist in itself, but only through the perceived objects,

Observation	Velocity
Typical motion of continents	10 mm/a=0.3 nm/s
Hair growth	ca. 5 nm/s
Stalactite growth	ca. 10 nm/s
Tree growth	up to 30 nm/s
Ketchup motion	1 mm/s
Electron speed in metals	1 mm/s
Speed of snail	5 mm/s
Slowest measured speed of light in matter	0.3 m/s Ref. 18
Signal speed in human nerve cells	0.5 m/s to 120 m/s
Speed of air in throat when sneezing	ca. 42 m/s
Record car speed	ca. 340 m/s
Speed of a rifle bullet	ca. 3000 m/s
Speed of cracks in breaking silicon	ca. 5 km/s
Highest macroscopic speed ever achieved by man (the Voyager space probes)	14 km/s
Speed of lightning tip	ca. 100 km/s
Speed of earth through universe	ca. 370 km/s
Highest macroscopic speed measured in our galaxy	ca. $0.97 \cdot 10^{8}$ m/s Ref. 19
Speed of electrons inside a color tv	ca. $1 \cdot 10^{8}$ m/s
Speed of telephone messages	up to $3 \cdot 10^{8}$ m/s
Highest ever measured group velocity of light	ca. $10 \cdot 10^{8}$ m/s
Speed of light spot from a light tower when passing over the moon	ca. $2 \cdot 10^{9}$ m/s
Highest proper velocity ever achieved for electrons	ca. $7 \cdot 10^{13}$ m/s
Highest velocity of a light spot	infinite

Table 3 Some velocity measurements

Observation	Duration
Shortest measurable time	ca. 10^{-44} s
Shortest time ever measured	ca. 10^{-23} s
Time for light to cross an atom	ca. 10^{-18} s
Wingbeat of fly	ca. 1 ms
Human instant	ca. 20 ms
Shortest lifetime of living being	ca. 0.3 d
Your 1000 million seconds anniversary	31.7 a
Time since human language is used	ca. 200000 a
Age of earth	$4.6 \cdot 10^9$ a
Age of oldest stars	ca. $12 \cdot 10^9$ a
Age of most protons in your body	ca. $12 \cdot 10^9$ a
Lifetime of ^{180}Ta	ca. 10^{15} a

Table 4 Some time measurements

from which the concepts of past, of present and of future ensue.
Lucretius (ca. 95–ca.55 BCE), *De natura rerum*, lib. 1, v. 460 ss.

In their first years of life, children spend a lot of time throwing objects around them. The term ‘object’ in fact *means* ‘that which has been thrown in front.’ Developmental psychology has even shown experimentally that from this experience children actually extract the concepts of time and space. Adults do the same when studying motion, with the difference that they repeat it consciously, using language. Ref. 20

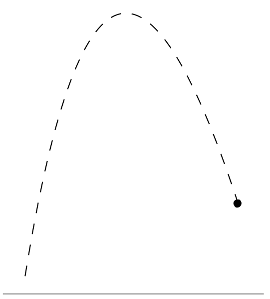


Figure 4 A typical path followed by a stone thrown through the air

Watching a stone flying through the air, one experiences the ability to define a *sequence* among observations. This ability results from the properties of our memory and of our senses. Most notably, one does this regularly with the sense of hearing, which allows to collect the various sounds during the rise, the fall and the hitting of the target. One finds that all observations one makes have other observations preceding them, observations simultaneous to them, and still others succeeding them. One says that observations happen at various *instants*, and one calls the sequence of all instants together *time*. If one wants to stress that an observation is the smallest part of a sequence, i.e. not itself a sequence, one calls it an *event*. Events are central to the definition of time, as starting or stopping a stopwatch is an event. But do events exist? You might want to ponder the question and keep it in the back of your head. We will find the definitive answer only later on.

Challenge

See page 584

Apart from detecting sequences, one also discovers that phenomena have an additional aspect that one calls indifferently stretch, extension, or duration. *Duration* expresses the fact that sequences “take” time, i.e. that they cannot be accelerated or slowed down.

How exactly do we deduce the concept of time, including sequence and duration, from observation? Many people have looked into this question: astronomers, physicists, watchmakers, psychologists, and philosophers. All find the same answer: *time is deduced by com-*

paring motions. Children, beginning at a very young age, develop the concept of 'time' from the comparison of motions in their surroundings. As grown-ups, if one takes for comparison the motion of the sun one calls this type of time *local time*; from the moon one gets a *lunar calendar*; if one takes a particular village clock on a European island one calls it *universal time coordinate* (UTC), once known as "Greenwich mean time."* Astronomers take for comparison the movements of the stars and call the result *ephemeris time*. An observer which uses his personal clock calls the reading his *proper time*; it is often used in the theory of relativity.

Ref. 20

All these methods have in common that in order to make the concept of time as precise as possible, one defines a standard motion, and with it a standard sequence and a standard duration. The device which performs this task is called a *clock*.** We can thus answer the question of the title: *time is what we read from a clock*. Note that all definitions of time used in the various branches of physics are equivalent to this one; no "deeper" or more fundamental definition is possible. We will need to recall this several times in our escalation. Note as a curiosity that the word 'moment' is even derived from the word 'movement'. Astonishingly, the definition of time is final; it will never be changed, not even at the top of motion mountain. This is surprising at first sight, because many books have been written on the nature of time. In fact, they should investigate the nature of motion instead! But this is the aim of our walk anyhow.

See page 730

Note that in the year 2000 an earth rotation does not take 86400 seconds any more, as it did in the year 1900, but 86400.002 seconds. Can you deduce in which year your birthday will have shifted by by a whole day?

Challenge

A *clock* is a moving system whose position can be read out. Of course, a *precise* clock is a system moving as regularly as possible, with as little outside disturbances as possible. Is there a perfect clock in nature? Do clocks exist at all? If one goes into the details, these turn out to be tricky questions. We will continue to study them throughout our walk, and come to a conclusion only towards the end. Every clock reminds us that in order to understand time, we need to understand motion. Cheap literature often suggests the opposite, in contrast to the facts. Our first thought about clocks is that they exist; this means that in nature there somehow is an intrinsic, natural and *ideal* way to measure time. Can you see which one?

Challenge

Time is not only an aspect of observations, it is also a facet of personal experience. Even in the innermost private life, in one's thoughts, feelings and dreams, one experiences sequences and durations. During childhood one learns to relate this internal experience of time with the external observations, and one learns to make use of the sequential property of events in one's actions. But when one studies the origin of psychological time, one finds that it coincides – apart from its lack of accuracy – with clock time.*** Every living human

* Official time is used to determine power grid's phase, phone companies' bit streams and the signal to the GPS system which is used by many navigation systems around the world, especially in ships, airplanes and trucks. For more information, see the http://www.gpsworld.com web site. The time keeping infrastructure is important for many more parts of the modern economy.

** The oldest clocks are sundials. The science of making them is called *gnomonics*. An excellent and complete introduction into this somewhat strange world can be found at the http://www.sundials.co.uk web site.

*** This internal clock is more accurate than often imagined, especially when trained. For times between a few tenths of a second, as necessary for music, and a few minutes, people can achieve accuracies, when trained, of a few percents. Only recently it became clear what type of clock lies at the basis of this personal time. It seems that macroscopic currents flowing around the brain in loops of about 10 cm size are at the basis of our own,

Ref. 21

Instants:	Physical property	Mathematical name (see later for definitions)
can be distinguished	distinguishability	set
can be lined up	sequence	order
define duration	measurability	metricity
can have vanishing distance	continuity	completeness
allow to add distances	additivity	metricity
don't bring surprises	translation invariance	homogeneity
don't end	infinity	openness
can beat any limit	infinity	unboundedness
can be defined for all	absoluteness	uniqueness

Table 5 Properties of galilean time

necessarily uses in his daily life the concept of time as combination of sequence and duration; this fact has been checked and confirmed in numerous investigations. For example, the term 'when' is present in all human languages. Ref. 22

Time is also a *necessary* concept of our thinking: we introduce it automatically when we distinguish between observations which are part of a sequence. There is no way to avoid time when talking about life. This seems to contradict the title of this adventure. But in fact it doesn't, as will be seen later on.

All experiences collected in everyday life with help of clocks can be summarized in a few sentences. One observes that events succeed each other smoothly, apparently without end. In this context, 'smoothly' means that observations not too distant tend to be not too different. One also observes that between two instants, as close as one can observe them, there is always room for other events. One further finds that durations, called *time intervals*, measured by different people with different clocks agree in everyday life; moreover, the order in the sequence of events is unique; everybody agrees on it.

These properties form what is called *galilean time*; it corresponds to the precise version of our everyday experience of time, as listed in table 5. All these properties can be expressed simultaneously by describing time with the real numbers; they have been constructed to have exactly the same properties as galilean time has, as explained in the intermezzo. Every instant of time is then described by a real number, often abbreviated t, and the duration of a sequence of events is then given by the difference between the values for the starting and the final event. See page 391

Note that hundreds of years of close scrutiny have shown in the meantime that *every single* property of time just listed is approximate, and none is strictly correct. These discoveries, with all the surprises that follow, are part of our journey.

By the way, when Galileo studied motion in the 17th century, there were no stopwatches in the shops yet. He thus had to build one himself, in order to measure times in the range between a fraction and a few seconds. Can you guess how he did it? Ref. 23 Challenge

What time is it at the *north pole*?

conscious feeling of time. But other clocks are also part of the human body; the time keepers for shorter times are electrical oscillators at cellular level, and for longer times chemical reactions.

Does time flow?

Wir können keinen Vorgang mit dem "Ablauf der Zeit"
vergleichen – diesen gibt es nicht –, sondern nur mit
einem anderen Vorgang (etwa dem Gang des Chronometers).*
Ludwig Wittgenstein, *Tractatus*, 6.3611

The "flow of time" is an often heard expression. With it one means to say that events flow, and that in nature change follows after change, in a continuous manner. But even though the hands of a clock "flow",** time itself does not. Time is a concept introduced specially to describe the flow of events around us; it does not itself flow, it *describes* the flow. Time does not advance. Time is neither linear nor cyclic. The idea that time flows is as hindering to understanding nature as is the idea that mirrors exchange right and left.

See page 339

The confusion at the origin of the expression "flow of time", already propagated by some greek thinkers and then by Newton, still continues. Aristoteles (384–322 BCE), always careful to think logically, already points out its misconception. Nevertheless, one continues to hear expressions such as "time reversal", the "irreversibility of time", and, most abused of all, "time's arrow". Time cannot be reversed, only motion can, or more precisely, only velocities of objects; time has no arrow, only motion has; it is not the flow of time which humans are unable to stop, but the motion of all the objects around. Incredibly, there are even books written by respected physicists which study different types of "time's arrow" and compare them with each other. Have a look at some of these horrible texts! Predictably, no tangible or new result is extracted.

Ref. 24
Ref. 25
Challenge

However, the mentioned expressions can lead reason astray in many ways, and we must avoid them because they render the ascent of motion mountain unnecessarily and increasingly difficult. They even prevent it beyond a certain stage, located about halfway to the top. We can now continue with the following aspect of motion states.

What is space?

The introduction of numbers as coordinates is an act of violence.
Hermann Weyl***

Why can we distinguish a tree from another? We see that they are in different positions. In fact, the capacity of distinguishing positions is the main ability of our sense of sight. When-

* We cannot compare a process with 'the passage of time' – there is no such thing – but only with another process (such as the working of a chronometer).

** Why do *clocks* go *clockwise*, even though all other rotational motions in our society, such as athletic races, horse races, bicycle races, ice skaters etc. go the other way? Most people are right-handers, and the right hand has more freedom at the outside of a circle. Therefore races go anticlockwise, as chariot races did already thousands of years ago. (For the same reason, helical stairs in castle are built in such a way that defending right-handers, usually from above, have their hand on the outside.) On the other hand, the clock imitates the shadow of sundials; obviously, this is true on the northern hemisphere only. (The old trick to determine south by pointing the hour hand of an horizontal watch to the sun and halving the angle between it and the direction of 12 o'clock does not work on the southern hemisphere.) So each clock implicitly continues to tell on which hemisphere it has been invented.

*** Hermann Weyl (1885–1955), one of the most important mathematicians of his time, also important theoretical physicist, one of the last universalists in both fields, contributor to quantum theory and relativity, father of the term 'gauge' theory, and author of many popular texts.

ever we distinguish two objects from each other, such as two stars, we first of all distinguish their positions. Position is therefore an important aspect of the state of an object. Positions are taken by only one object at a time. They are limited. The set of all available positions, called *(physical) space*, acts like a container and at the same time acts as a background.

A second aspect of observations is connected with space and position: *size*. The size of an object describes the set of positions it occupies. Small objects occupy only subsets of the positions occupied by large ones. We discuss size shortly.

How do we deduce space from observations? During childhood, humans (and most higher animals) learn to forge together the various perceptions of space, namely the visual, the tactile, the auditory, the kinesthetic, the vestibular etc., into one coherent set of experiences and relations. The result of this learning process is a certain "image" of space in the brain. Among others, every child also learns the details of the shape and size of its own body. During this development, which takes place mainly before school age, every human learns how to use the properties of size and of space in his actions. Everybody who lives uses these properties of space. The question 'where?' can be asked and answered in all world languages.

Adults derive space, i.e. position and size, from distance measurements. The concepts of length, area, volume, angle, and solid angle are all deduced from such measurements. Geometers, surveyors, astronomers, and producers of cloth and meter bars base their trade on distance measurements. Space is thus a concept formed to describe observations by summarizing all the distance relations between objects.

To measure lengths one needs meter bars. Their main property is that they are straight. That is astonishing, if one thinks about it. When humans lived in the jungle, there was not a single straight object around them. No straight rulers, no straight tools, nothing. Today, a cityscape is essentially a collection of straight lines. How did humans do this? Challenge

Once humans came out of the jungle with their newly built meter bars, they collected a wealth of results which are easily confirmed by personal experience. One observes first of all that objects can take positions in an apparently *continuous* manner: there are more positions that can be counted.* One further finds that size is captured by defining the distance between various positions, which is called *length*, or by using the field of view an object takes when one touches it, which is called its *surface*. Length and surface can be measured with help of a meter bar. In daily life, all length measurements performed by different people coincide, they are unique. One observes that the length of objects is independent of the person measuring it, of the position of the objects, and of their orientation. One also observes that in daily life the sum of angles in any triangle is equal to two right angles. Finally, one does not observe any limits in space.

One also observes that space has three dimensions, i.e. that one can define sequences of positions in precisely three independent ways. This can also be seen from the inner ear of man, of all other mammals, and even of practically all vertebrates: the ear has three semicircular canals which help the body to sense its position in the three dimensions of space. Another proof that space has three dimensions is given by the problems posed and solved regularly by shoe-lace: if space had more than three dimensions, shoe-lace would never remain tangled, because knots exist only in three-dimensional space. (Can you confirm that

* For a definition of uncountability, see page 385.

Points:	Physical property	Mathematical name (see later for definitions)
can be distinguished	distinguishability	set
can be lined up	sequence	order
can form shapes	shape	topology
can be lined up to form knots	possibility of knots	dimensionality
define distances	measurability	metricity
can have vanishing distance	continuity	completeness
allow to add distances	additivity	linearity
don’t hide surprises	translation invariance	homogeneity
don’t end	infinity	openness
can beat any limit	infinity	unboundedness
can be defined for all	absoluteness	uniqueness

Table 6 Properties of galilean space

Challenge

in four dimensions knots are impossible?) Why three? This is perhaps the most difficult question of physics; it will be answered only in the last part of our walk.

Like time intervals, length intervals can be described most precisely with the help of *real numbers*. In order to simplify communications, one uses standard *units*, so that everybody uses the same numbers for the same length or time interval. We then can study experimentally the general properties of *galilean space*: space contains objects; it is continuous, i.e. mathematically complete, three-dimensional, isotropic, homogeneous, infinite, euclidean, and unique, i.e. “absolute”. In mathematics, a structure, i.e. a mathematical concept, with all the properties just mentioned is called a three-dimensional *euclidean space*, and its elements, the (mathematical) points, are described by three real parameters, usually written as

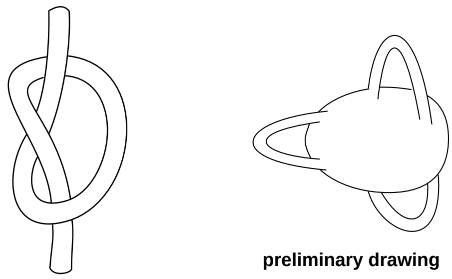


Figure 5 Two proofs of the three-dimensionality of space: a knot and the inner ear of a mammal

$$(x, y, z) \tag{1}$$

and called *coordinates*, which specify the location of a point in space. (For the precise definition of euclidean spaces, see page 50.)

This description is mentioned here in just half a page; in fact it took two thousand years to work it out, mainly because the concepts of ‘real number’ and of ‘coordinate’ had to be discovered first. The first person to describe points of space in this way was the famous french-born mathematician and philosopher René Descartes (1596–1650), after whom the coordinates (1) are named *cartesian*.

Like time, space is a *necessary* concept to describe the world. Indeed, it is automatically introduced when one distinguishes between observations, e.g. between a tree and a person.

Observation	Distance
Galaxy Compton length	ca. 10^{-85} m (prediction only)
Planck length	ca. 10^{-35} m
Shortest measurable length	ca. 10^{-32} m
Proton size	ca. 1 fm
Electron Compton wavelength	2.4 pm
Wavelength of visible light	ca. 0.4 to 0.8 μm
Size of small bacterium	ca. 5 μm
Point: diameter of smallest object visible with naked eye	ca. 20 μm
Total length of human DNA	ca. 2 m
Size of largest living being	ca. 100 m
Total length of human nerve cells	$8 \cdot 10^5$ km
Light year	9.5 Pm
Distance of typical star at night	ca. 10^{19} m
Size of galaxy	ca. 10^{229} m
Most distant visible object	ca. $1 \cdot 10^{26}$ m

Table 7 Some distance measurements

Since length measurement methods are possible, there must be a *natural* or *ideal* way to measure distances, sizes and straightness. Can you find it? Challenge

Like in the case of time, each of the properties of space just listed has to be checked by careful observation. And again, more precise observation later on will show that each of them is an approximation, or, in more simple and drastic words, that *each* of them is wrong. This confirms the quote at the beginning of this section; we will discover that this story is told by every forest, such as the one we are crossing now, at the feet of motion mountain. We only need to listen carefully to what the trees have to tell.

Are space and time absolute or relative?

In everyday life, the concepts of of galilean space and time include two opposing aspects which have coloured every discussion about them for several centuries. On one hand, space and time express something invariant and permanent; both act like big containers for all the objects and events found in nature. Seen this way, space and time have an existence of their own. In this sense one can say that they are *fundamental* or *absolute*. On the other hand, space and time are tools of description which allow to talk about relations between objects. In this view, they do not have any meaning when separated from objects, and only result from the relations between objects; they are *relational* or *relative*. In this opposition of standpoints, (what do you favour?) the results of physics have alternatively favoured one Challenge view over the other; we will follow this alternation throughout our walk. Only the last part will provide the solution of the puzzle.

Size: why area exists, but length and volume do not

We saw that a central aspect of objects was their size. It seems obvious that with the definition of *distance* as difference between coordinates is it possible to define *length* in a reliable way. It took hundreds of years to discover that this is not the case. In fact, several investigations both in physics and in mathematics lead to complications.

The physical issues started with the astonishingly simple question asked by the english physicist and psychologist Lewis Fray Richardson (1881–1953): how long is the coastline of Britain?

Following the coastline on a map with a odometer, a device such as the one shown in the figure 6, one finds that the coastline l depends on the scale s (say 1/10 000 or 1/500 000) of the map used:

$$l = l_o \, s^{0.36} \qquad (2)$$

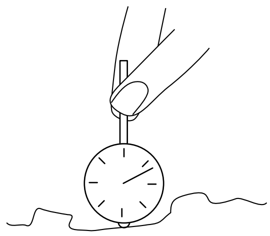


Figure 6 A curvemeter or odometer

The larger the map, the longer the coastline. What would happen if the scale of the map is increased even beyond the size of the original? Could a coastline really have infinite length? It is not a secret that there indeed are such curves, and that there are actually an infinite
Challenge number of them.* Can you construct another example?

Length has other strange properties. The great mathematician Felix Hausdorff discovered that it is possible to cut the a line segment of length 1 into pieces which can be reassembled, only by shifting them in direction of the segment, into a line segment of length 2. Are you able to find such a division using the hint that it
Challenge only possible using infinitely many pieces?

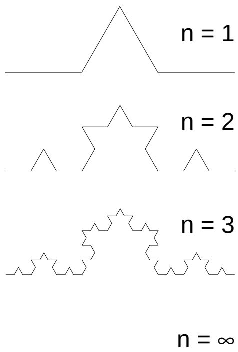


Figure 7 A fractal: a self-similar curve of infinite length, and its construction

In summary, length is well defined for straight and nicely curved lines, but not defined for intricate lines or lines made of infinitely many pieces. We thus have to avoid fractals and other strangely shaped curves in the following, and also have to be very careful when talking about infinitely small segments. These are central but often hidden assumptions in the first two parts of this walk, and should never be forgotten. We will return to them in the third part. But the mentioned problems pale when compared to the following.

Using length, one commonly defines area and volume. You think it's easy? You're wrong, as well as a victim of prejudices spread in schools all around the world. To define area and volume with precision, one needs definitions with two properties: the values have to be *additive*, i.e. for finite and infinite sets of objects, the total area and volume have to be the sum of the areas and volumes of each element of the set; and they have to be *rigid*, i.e. if one cuts an area or a volume into pieces and then rearranges them, the value remains the same. Do such concepts exist?

For areas, one proceeds in the following standard way: one defines the area A of a rectangle of sides a and b as $A = ab$; since any polygon can be rearranged into a rectangle with a
Challenge finite number of straight cuts, one can define an area for all polygons. One can then define

* Most of these curves are selfsimilar, i.e. they follow scaling laws similar to the mentioned one, and are often called *fractals*. The term is due to the polish mathematician Benoit Mandelbrodt. Coastlines and other fractals are beautifully presented in Heinz-Otto PEITGEN, Hartmut JÜRGENS & Dietmar SAUPE, *Fractals for the classroom*, Springer Verlag, 1992, on pages 232-245. Their book has also been translated into several other languages.

area for nicely curved shapes as limit of the sum of infinitely many polygons. (This method is called integration, and is introduced in detail on page 761.)

But integration does not allow to define area for arbitrarily bounded regions. (Can you imagine such a strangely bounded region?) One needs other tricks. But in 1923 the famous polish mathematician Stefan Banach (Krakow, 1892–Lvov, 1945) showed that one can indeed define an area for any set of points whatsoever, even if the border is not nicely curved but extremely complicated, such as a fractal curve just mentioned. Today this concept of area, technically a 'finitely additive isometrically invariant measure', is called a *Banach measure* in his honor.* Challenge

Mathematicians sum up this discussion by saying that since in two dimensions there is a Banach measure, there is a way to define the concept of area, an additive and rigid property, for any set of points whatsoever.**

What is the situation for volume? One can start in the same way by defining the volume V of a rectangular polyhedron with sides a, b, c as $V = abc$. But then one encounters the first problem: a polyhedron cannot be cut into a cube by straight cuts: this is possible only if they have the same Dehn invariant! In 1900 and 1902 the german mathematician Max Dehn discovered that if one ascribes to every edge of a general polyhedron a number given by its length l times a special function $g(\alpha)$ of its dihedral angle α, then the sum of all the numbers for all the edges of a solid does not change under dissection, provided that $g(\alpha+\beta) = g(\alpha) + g(\beta)$ and $g(\pi) = 0$. An example of such a strange function g is the one giving the value zero to any rational multiple of π and the value one to any irrational multiple of π. You may then deduce for yourself that a cube cannot be dissected into a regular tetrahedron because their Dehn invariants are different.*** Nevertheless, one can define a rigid and additive concept of volume for polyhedra, since for them and in general for all "nicely curved" shapes, one can again use integration for the definition of volume. Challenge Challenge See page 761

Now let us consider general shapes and general cuts in *three* dimensions. One then gets the famous *Banach-Tarski theorem* (or paradox). In 1924, Stefan Banach and Alfred Tarski (1902, Warsaw– 1983, Berkeley) proved that it is possible to cut one sphere into five pieces which can be recombined to give two spheres, each of the size of the original. Even worse, another version of the theorem says: take any two sets not extending to infinity and containing a solid sphere each, it is always possible to dissect one into the other with a finite number of cuts. In particular it is possible to dissect a pea into the earth, or vice versa. Size does not count!**** Ref. 26

The Banach-Tarski theorem raises two questions: Can one do this with gold, i.e. is matter continuous? And can one do this with empty space, i.e. is empty space continuous? Both

* Defining a Banach measure means to be able to assign a finite positive value to any set of points, however weird, with the properties of being *rigid*, i.e. invariant under translations, and *additive* for disjunct sets.

** Actually, this is strictly true only for the plane. For curved surfaces, such as the surface of a sphere, there are complications, but they will not be discussed here. Note also that the problems of the definition of length reappear for area if the surface to be measured is not flat but full of hills and valleys. A typical example is the area of the human lung: depending on the details one looks, one finds area values from a few square metres up to over100 m^2.

*** See also the beautiful book by M. AIGLER & G.M. ZIEGLER, *Proofs from the book*, Springer Verlag, 1999.

**** The proof of the result is explained beautifully by Ian Stewart in *Paradox of the spheres*, New Scientist, 14 January 1995, pp. 28–31.

topics will be studied in the rest of our walk. For the moment, we eliminate the troubling issue altogether by restricting our interest to nicely, i.e. smoothly curved shapes.* With this restriction, volumes of matter and space behave nicely: they are additive and rigid. Nevertheless, we need to keep in the back of our mind that the size of an object is a tricky quantity and that we need to be careful whenever we talk about it.

What is straight?

When one sees a solid object with a straight edge, it is a 99% safe bet to take the following conclusion: it is human made.** The contrast between the objects seen in a city, houses, furniture, cars, boxes, books, and the objects seen in a forest, trees, plants, mountains, clouds, is evident: in the forest nothing is straight or plane, in the city most objects are. How is it possible for us to make straight objects if in nature there are none to be found? How can man make objects which are more straight than the machine tools one finds in nature?

Traditionally one calls any line *straight* which touches a plumb-line all along, or which
Challenge touches a light ray all along. Can you find another definition? In fact, the definition with the
Challenge plumb line and the definition with light are equivalent. Can you confirm this? (This is not an easy question.) Obviously, we call a surface *flat* if it can be rotated by any angle and still touches a plumb-line or a light ray all along.

There are people who maintain that we do not live on the outside of a spherical planet, but on the *inside* of a sphere. People who defend this so-called *hollow earth theory* say that the sun and the stars are located near the center of the hollow earth. One can indeed build a complete model of the universe if, as they say, light does not travel in straight lines. Is this
See page 277 correct? We will come back to this problem in the section on general relativity.

How to describe motion: kinematics

As experiments show, the mentioned properties of galilean time and space are extracted from the environment by most higher animals as well as by young children. Later, when children learn to speak, they can put these experiences and concepts into words, as was just done here. With help of the concepts just introduced, one can say that *motion is change of position with time*. Scanning rapidly the lower right corners of the pages of this book illustrates this description. Each page simulates an instant of time, and the only change taking place during motion is the position of the object, represented by the dark spot. The variations from one movie picture to the next, due to the imperfections of printing techniques, even simulate the inevitable measurement errors.

It is evident that calling motion the change of position with time is *not* an explanation of motion *nor* a definition, since both the concepts of time and position are deduced from motion itself. It is only a *description* of motion. Nevertheless, the rephrasing is useful because, as we will see, it allows for high *precision*, and that is after all, our guide during this promenade. The study of this description is traditionally called *kinematics*.

* In 4 dimensions, the Banach-Tarski paradox exists as well, as it does in any higher dimension. More mathematical detail can be found in the book by Steve Wagon, *The Banach Tarski Paradox*, Cambridge University Press, 1993.

** Exceptions are some crystalline minerals. Other examples which come to mind, such as some bacteria which
Ref. 27 can have (almost) square and triangular shapes are not exceptions.

The set of all positions taken by an object over time form the *path* or *trajectory*. The origin of the concept is evident when one watches fireworks* or the flip movie on the lower right corners of this text. With the description of space and time by real numbers, the trajectory can be described by specifying its three coordinates (x,y,z), one for each dimension, as continuous functions of time t. (Functions are defined in detail on page 386.) This is usually written as $\mathbf{x} = \mathbf{x}(t) = (x(t), y(t), z(t))$. For example, observation shows that the height z of any thrown or falling stone changes as

$$z(t) = z_o + v_o(t - t_o) - \tfrac{1}{2} g (t - t_o)^2 \qquad (3)$$

where t_o is the time one starts the experiment, z_o is the initial position, the quantity v_o is the initial velocity in the vertical direction and $g = 9.8\,\mathrm{m/s^2}$ is a constant which is found to be the same, within one part in about 300, for all falling bodies on all points of the surface of the earth. Where does the value $9.8\,\mathrm{m/s^2}$ and its slight variations come from? A preliminary answer will be given shortly; but the complete elucidation of the origins will occupy us during the largest part of this hike. Ref. 28

Equation (3) allows e.g. to determine the depth of a well given the time a stone takes to reach is bottom. It also gives the speed v with which one hits the ground after jumping from a tree, namely $v = \sqrt{2gh}$; a height of 3 m gives a velocity of 27 km/h. The velocity is thus only proportional to the square root of the height – perhaps the fear of falling results from a overestimate of its actual effects? Challenge

If one completes the description of equation (3) with the two expressions for the horizontal coordinates x and y, namely

$$\begin{aligned} x(t) &= x_o + v_{xo}(t - t_o) \\ y(t) &= y_o + v_{yo}(t - t_o) \quad , \end{aligned} \qquad (4)$$

one has a complete description for the path followed by thrown stones. A path of this shape is called a *parabola*. It is the same shape as used for the reflector of a pocket lamp. Can you show why? Challenge

The kinematic description of motion** is useful to answer questions such as:

- What is the distance one can reach with a stone, given the speed and the angle with which one shoots? Challenge

- How can one measure the speed of falling rain with an umbrella? Challenge

- What is the maximum numbers of balls one can expect to juggle at the same time? Challenge

- What is an upper limit for the long jump record? One can use as input that the running speed world record in 1997 is $12\,\mathrm{m/s} \approx 43\,\mathrm{km/h}$ by Ben Johnson, and the women's record is $11\,\mathrm{m/s} \approx 40\,\mathrm{km/h}$. Challenge Ref. 29

- Are gun bullets falling back after being fired in the air dangerous? Challenge

- Is it true that rain drops would kill if it weren't for the air resistance of the atmosphere? Challenge

∗ On the world of fireworks, see the frequently asked questions list of the usenet group rec.pyrotechnics, or search the web. A simple introduction is the article by J.A. Conkling, *Pyrotechnics*, Scientific American pp. 66–73, July 1990.

∗∗ Apart from the graphs shown in figure 8, there is also the *configuration space* spanned by the coordinates of all particles of a system; only for a single particle it is equal to the real space.

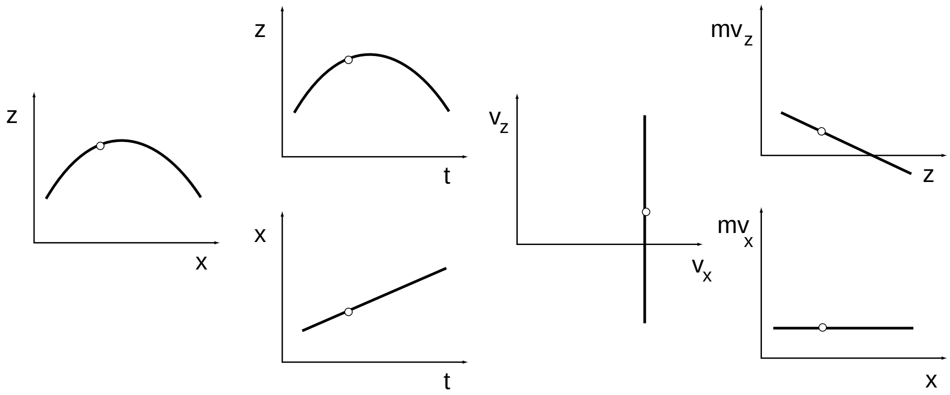


Figure 8 Various types of graphs describing the same flying stone: the path in configuration space, the space-time diagrams, the hodograph, and the phase space or state space diagrams

The last two questions derive from the fact that equation (3) does not hold in all cases. For example, leaves or potato chips do not follow it. That is a consequence of air resistance. We will discuss it shortly.

What is rest?

This question seems to have an obvious answer. A body is at rest when its position, i.e. its coordinates do not change with time. In other words, *rest* is

$$\mathbf{x}(t) = \text{const} \quad . \tag{5}$$

Later we will see that this definition, contrary to first impression, is *not* of much use and will have to be modified. In any case, non-resting objects can be distinguished by comparing the rapidity of their displacement. One thus defines the *velocity* **v** of an object as the change of their position **x** with time t, which in symbolic notation is usually written as

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} \quad . \tag{6}$$

The *speed* v is the name given to the magnitude of the velocity **v**. In this expression, valid for each coordinate separately, d/dt means 'change with time'; one can thus say that velocity is the *derivative* of space with respect to time. Derivatives are written as fractions to remind that they are derived from the idea of slope; the expression

$$\frac{dy}{dt} \quad \text{is meant as an abbreviation of} \quad \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} \quad , \tag{7}$$

a shorthand for saying that the *derivative at a point* is the limit of the slopes in the neighbourhood of the point, as shown in figure 9. From this definition follow the working rules (c being any number)

$$\frac{d(y+z)}{dt} = \frac{dy}{dt} + \frac{dz}{dt} \quad , \quad \frac{d(cy)}{dt} = c\frac{dy}{dt} \quad , \quad \frac{d}{dt}\frac{dy}{dt} = \frac{d^2y}{dt^2} \quad , \quad \frac{d(yz)}{dt} = \frac{dy}{dt}z + y\frac{dz}{dt} \tag{8}$$

which is all one needs to know about derivatives. The quantities dt and dy, sometimes useful by themselves, are called *differentials*. These concepts are due to the saxon thinker – he was physicist, mathematician, philosopher, diplomat, jurist, historian – Gottfried Wilhelm Leibniz (1646, Leipzig–1716, Hannover), and are at the basis of all calculations based on the continuity of space and time.

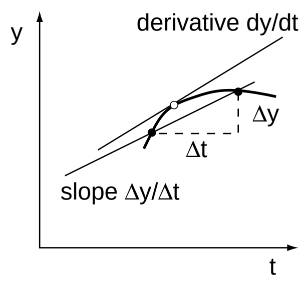


Figure 9 Derivatives

Indeed, the definition of velocity assumes that it makes sense to take the limit $\Delta t \rightarrow 0$, in other words, that infinitely small time intervals do exist in nature. The concept of velocity, being based on derivatives, can be defined only because both space and time are described by sets which are *continuous*, or in mathematical language, *complete*. In the rest of our walk we should never forget that right from the beginning of physics, *infinities* are present in its description of nature. In fact, differential calculus *is* the study of infinities and of their use. We thus learn straight away that the appearance of infinity does not automatically render a description impossible or imprecise. (The definition of the various types of infinities are presented in the intermezzo following this chapter.)

The appearance of infinity in the usual description of motion was first criticized with ironic arguments by Zeno of Elea (around 445 BCE), a disciple of Parmenides. In his famous third argument, Zeno explains that since at every instant a given object occupies a part of space corresponding to its size, the notion of velocity at a given instant makes no sense; he provokingly concludes that therefore motion does not exist. Nowadays we would not call this an argument against the *existence* of motion, but against its usual *description*, in particular against the use of infinitely divisible space and time. (Do you agree?) However, the description so criticized by Zeno actually works quite well in everyday life. On the other hand, later in our walk Zeno will be partly rehabilitated, and the more so the more we proceed.

Ref. 30

Challenge

See page 584

Why is velocity necessary as a concept? Aiming for precision in the description of motion, we need to find the complete list of aspects necessary to specify the state of an object. And the concept of velocity is obviously a member of this list. Continuing in the same way, one calls *acceleration* **a** of a body the change of velocity with time, or

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{x}}{dt^2} \quad . \tag{9}$$

Higher derivatives can also be defined in the same manner but add little to the description of nature. In fact, it turns out that neither they nor even acceleration itself are useful for the description of the state of motion of a system, as we will show shortly.*

Ref. 31

* Both velocity and acceleration have a magnitude and a direction, properties indicated by the use of bold letters for their abbreviations. Such physical quantities are called *vectors*. In more precise, mathematical language, a vector is an element of a set, called *vector space* V, in which the following properties hold for all vectors **a** and **b** and for all numbers c and d:

$$c(\mathbf{a}+\mathbf{b}) = c\mathbf{a}+c\mathbf{b} \quad , \quad (c+d)\mathbf{a} = c\mathbf{a}+d\mathbf{a} \quad , \quad (cd)\mathbf{a} = c(d\mathbf{a}) \quad \text{and} \quad 1\mathbf{a} = \mathbf{a} \quad . \tag{10}$$

Another example of vector space is the set of all *positions* of an object. Does the set of all rotations form a vector space? All vector spaces allow to define a unique null vector and for each vector a unique negative vector.

Challenge

Observation	acceleration
Centrifugal acceleration due to earth's rotation	$33\,\mathrm{mm/s^2}$
Electron acceleration in household wire	ca. $50\,\mathrm{mm/s^2}$
Gravitational acceleration on the moon	$1.6\,\mathrm{m/s^2}$
Gravitational acceleration on earth's surface, depending on place	$9.8 \pm 0.1\,\mathrm{m/s^2}$
Fastest wheel accelerated car	ca. $15\,\mathrm{m/s^2}$
Gravitational acceleration on jupiter's surface	$240\,\mathrm{m/s^2}$
Fastest leg acceleration (insects)	ca. $2000\,\mathrm{m/s^2}$
Tennisball on wall	ca. $10^5\,\mathrm{m/s^2}$
Bullet acceleration in rifle	ca. $5 \cdot 10^6\,\mathrm{m/s^2}$
Fastes centrifuges	ca. $10^8\,\mathrm{m/s^2}$
Acceleration of protons inside nucleus	ca. $10^{31}\,\mathrm{m/s^2}$
Highest possible acceleration	ca. $10^{51}\,\mathrm{m/s^2}$

Table 8 Some acceleration values

Objects and point particles

Wenn ich den Gegenstand kenne, so kenne ich auch sämtliche
Möglichkeiten seines Vorkommens in Sachverhalten.*
Ludwig Wittgenstein, *Tractatus*, 2.0123

One aim of the study of motion is to find a complete and precise description of both states and objects. With help of the concept of space, the description of objects can be refined considerably. In particular, one knows from experience that all objects seen in daily life have an important property: they can be divided into *parts*. Often this observation is expressed by saying that all objects, or bodies, have two properties. First, they are made out of *matter*,** defined as that aspect of an object which is responsible for its impenetrability, i.e. the property preventing two objects to be in the same place. Secondly, bodies have a certain form or *shape*, defined as the precise way in which this impenetrability is distributed in space.

In order to describe motion as accurately as possible, it is convenient to start with those bodies which are as simple as possible. In general, the smaller a body, the simpler it is. A body that is so small that one does not need any more to take into account its parts is called a *particle*. (The older term *corpuscule* has come out of fashion.) Particles are thus idealized little stones. The extreme case, a particle whose size is *negligible* compared to the dimensions of its motion, so that its position is described completely by a *single* triplet of

In many vector spaces the concept of *length* can be introduced, usually via an intermediate step. A vector space is called *euclidean* if one can define for it a *scalar product* between two vectors, a number satisfying

$$\mathbf{aa} \geqslant 0 \quad , \quad \mathbf{ab} = \mathbf{ba} \quad , \quad (\mathbf{a}+\mathbf{a}')\mathbf{b} = \mathbf{ab}+\mathbf{a}'\mathbf{b} \quad , \quad \mathbf{a}(\mathbf{b}+\mathbf{b}') = \mathbf{ab}+\mathbf{ab}' \quad \text{and} \quad (c\mathbf{a})\mathbf{b} = \mathbf{a}(c\mathbf{b}) = c(\mathbf{ab}) \quad . \tag{11}$$

In coordinate notation, the standard scalar product is given by the number $a_x b_x + a_y b_y + a_z b_z$. When it vanishes the two vectors are orthogonal. The *length* or *norm* of a vector can then be defined as the square root of the scalar product of a vector with itself: $a = \sqrt{\mathbf{aa}}$.

* If I know an object I also know all its possible occurrences in states of affairs.

Ref. 32 ** Matter is a word derived from the latin 'materia', which originally meant 'wood' and was derived via some intermediate steps from 'mater', meaning 'mother'.

coordinates, is called a *point particle* or a *mass point*. In equation (3), the stone was assumed to be such a point particle.

Do pointlike objects, i.e. objects smaller than anything one can measure, exist in daily life? Yes, they do. The most notable examples are the stars. At present one is able to measure angular sizes as small as 2 μrad, a limit given by the fluctuations of the air in the atmosphere; without atmosphere, in space, e.g. for the Hubble telescope orbiting the earth, the limit is due to the diameter of the telescope and is of the order of 10 nrad. Practically all stars seen from earth are smaller than that, and are thus effectively "pointlike", even when seen with the most powerful telescopes.

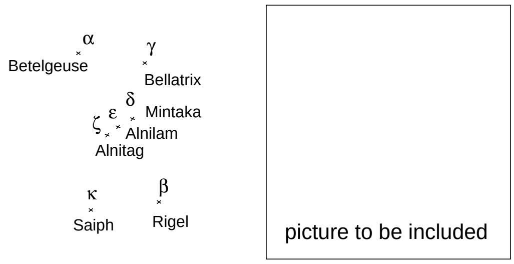


Figure 10 Betelgeuse

One can even see the difference between "pointlike" sources and finite size ones with the naked eye alone: stars twinkle, planets do not. This is due to the fact that the turbulence of air, which is responsible for the twinkling of stars by deflecting light rays by very small amounts, is too small to lead to twinkling of sources of larger size, like planets or satellites. In reality, the size of a few large and nearby stars, of red giant type, can be measured with special instruments.* Mira in Cetus, Betelgeuse, the higher of the two shoulders of Orion, Antares in Scorpio, Aldebaran in Taurus, and Sirius in Canis Major are examples of stars whose size has been measured; they are all only a few light years from earth. Of course, stars have a finite size, like the sun has, but there is no way to prove this by taking pictures.** Ref. 33

An object is *pointlike for the naked eye* if its angular size is smaller than about $2' = 0.6$ mrad. Can you guess the size of a pointlike dust particle? By the way, an object is *invisible* to the naked eye if it is pointlike and if its luminosity, i.e. the intensity of the light from the object reaching the eye, is below some critical value. Can you estimate whether there are any man-made objects visible form the moon, or form the space shuttle? Challenge Challenge

The above definition of 'point-like' in everyday life is a fake one. Do proper point particles exist? In fact, is it possible at all to show that a particle has vanishing size? This question will be central in the last two parts of our walk. In fact we have even forgotten to check whether points in space exist. Our walk will lead us to the astonishing result that all the answers to these questions are negative. Can you imagine how this is proven? Challenge

Once one knows how to describe the motion of point particles, the motion of extended bodies, rigid or deformable, can be described by assuming that they are made of parts, in the same way as the motion of an animal as a whole results from the motion of its various parts.

* An introduction to the different types of stars can be found at the http://www.astro.uiuc.edu/~kaler/sow/sowlist.html web site. About constellations, see the http://www.astro.uiuc.edu/~dolan/constellations/constellations.html web site.

For an overview of the planets, see the beautiful book by K.R. LANG, C.A. WHITNEY, *Vagabonds de l'espace – Exploration et découverte dans la système solaire*, Springer Verlag, 1993. The most beautiful pictures of the stars in the sky can be found in D. MALIN, *A view of the universe*, Sky Publishing and Cambridge University Press, 1993. To learn more about what people do to take these pictures, a fascinating book is also P. MANLY, *Unusual telescopes*, Cambridge University Press, 1991.

Challenge ** Do you agree?

The simplest description, the *continuum approximation*, is to describe extended bodies as an infinite collection of point particles. One can thus describe the motion of milk and honey, the motion of the air in hurricanes, of perfume in rooms, of fire and of all other gaseous bodies, the bending of bamboo in the wind, the shape changes of chewing gum and of all other deformable solids, and the growth of plants and animals. Ref. 34

A better approximation than the continuum approximation is described shortly. All observations have confirmed that the motion of large bodies can be described as the result of the motion of their parts. Interestingly, in the third part of our escalation we will discover that at a fundamental scale, this decomposition is *not* possible.

Legs and wheels

Shape is an important aspect of bodies: it allows to count them. For example, one finds that that living beings are always made of a single body. This is not an empty statement: from this fact one can deduce that animals cannot have wheels nor propellers, but only legs, fins, or wings.

Why? Living beings have only one surface; simply put, they have only one piece of skin. Mathematically speaking, animals are *connected*. Thus in a first reaction one tends to imagine that the blood supply to a rotating part would get tangled up. But this is argument is not correct, as the figure 11 shows. Can you find an example for this kind of motion in your own body? (See appendix D; Ref. 35; Challenge)

However, such a rotating part still cannot make a wheel. Can you see why? Could it make a propeller? By the way, can you see how many cables may be attached to the rotating body of the figure without hindering the rotation? Challenge

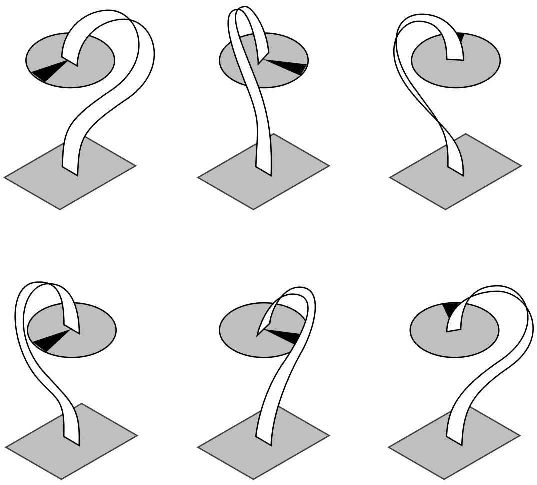


Figure 11 How an object can rotate without tangling up the connection to a second one

In summary, whenever one observes a construction of which some part is turning continuously (and without the “wiring” of the figure), one knows immediately that it is an artifact, a machine, not a living being, but built by one. Of course this does not rule out living bodies which move by rotation as a whole: the tumbleweed so common in Hollywood westerns, seeds from various trees, some animals, as well as children or dancers sometimes move by rotating as a whole. Ref. 36

Single bodies, and thus all living beings, can only move through *deformation* of their shape: therefore they are limited to walking, crawling, running, and flapping wings or fins. In contrast, systems of several bodies, such as bicycles, pedal boats or any other machine, can move without any change of shape of their components, thus enabling the use of wheels, propellers, or other rotating devices.*

Figure still to be added

Figure 12 Legs and wheels in living beings

However, like so many statements about living creatures, also this one has exceptions. The distinction between one and two bodies is impossible to make sharp if the two bodies are each made of a single molecule. Organisms such as Escherichia coli, the well-known bacterium found in the human gut, or bacteria from the Salmolella family, all swim using flagella, i.e. thin filaments, similar to tiny hair sticking out of the cell membrane. In the seventies it was shown that each flagellum, which is made of one or a few long molecule with a diameter of a few tens of nanometers, does in fact turn around its axis. A bacterium is able to turn its flagella in the clockwise and anticlockwise direction, can achieve more than thousand turns per second, and can turn all its flagella in perfect synchronism. Therefore wheels actually do exist in nature. But let us now continue with our study of simple objects. See page 570 Ref. 37

Objects and images

In our walk through the forest here at the base of motion mountain, we observe two rather different types of motion: the breeze moves the leaves, and at the same time their shadows move on the ground. In general, both objects and images can move. Running tigers, falling snowflakes, material ejected by volcanos, but also the shadow following our body, the beam of light circling the tower of a lighthouse in a misty night, and the rainbow that constantly keeps the same distance from the hiker are examples of motion.

Everybody who has seen an animated cartoon in the cinema knows that images can move in more surprising ways than objects. Images can change their size and shape, and they can change their colour, a feat only few objects are able to perform.** Images can appear and disappear without trace, can multiply, can interpenetrate, can go backwards in time, can defy gravity and any other force; images, even usual shadows, can move faster than light. Images can seem to float in space and stay always at the same distance to approaching objects. Objects can do almost none of these things. In general, the "laws of cartoon physics" are rather different from those in nature. In fact, motion of images do not seem to follow any Ref. 39

* Despite the disadvantage of not being able to use rotating parts and being restricted to one piece only, nature's moving constructions, usually called animals, are often better than human ones. This has two reasons. First of all, nature's systems have integrated repair and maintenance systems. But on top of that, nature can build large structures inside containers with small openings. Nature is very good at building sailing ships inside glass bottles. The human body is full of such examples; can you name a few? Challenge

** Excluding very slow changes such as the change of colour of leaves in the fall, in nature only certain crystals, the octopus, the chameleon and a few other animals achieve this. Of human made objects, only television, computer displays, heated objects, and certain lasers can do it. Do you know more examples? Challenge An excellent source of information on the topic of colour is the book by K. Nassau, *The physics and chemistry of colour – the fifteen causes of color*, J. Wiley & Sons, 1983. In the popular science domain, the most beautiful book is the classic work by the flemish astronomer Marcel G.J. Minnaert, *Light and color in the outdoors*, Springer, 1993, an updated version extracted from his wonderful book dating from 1937, *De natuurkunde van 't vrije veld*, Thieme & Cie, Zutphen. Ref. 38

rule at all, in contrast to motion of objects. Together, both objects and images differ from their environment in that they have *boundaries* which allow to define size and shape. One feels the need for a precise criterion allowing to distinguish the two cases.

The clearest distinction between images and objects is performed by the same method that children use when they stand for the first time in front of a mirror: they try to *touch* what they see. Indeed, if one is able to touch what one sees, or more precisely, if one is able to move it, one calls it an *object*, otherwise an *image*.* One cannot touch images, but one can touch objects. And as everybody knows, touching something means to feel that it resists being moved. Certain bodies, such as butterflies, are moved with ease, others, such as ships, with more difficulty. This resistance to motion – more precisely, to change of motion – is called *inertia*, and the difficulty with which a body can be moved is called its *(inertial) mass*. Images have neither inertia nor mass.

See page 468

Summing up, for the description of motion one has to distinguish bodies, which can be touched and are impenetrable, and images which cannot and are not. Note that everything visible is either an object or an image; there is no third possibility. If the object is so far away that it cannot be touched, such as a star or a comet, it can be difficult to decide whether one is dealing with an image or an object; we will repeatedly come back to the question.

Moving images are made of *radiation* in the same way as objects are made of *matter*. Images are the domain of shadow theatre, of cinema, of television, of computer graphics, of belief systems and of drug experts: photographs, motion pictures, ghosts, angels, spirits and other illusions and hallucinations usually are images, or sounds. Due to the importance of objects – after all we are objects ourselves – we study them first.

Ref. 40

Motion and contact

When a child learns to ride a monocycle, it makes use of a general relation in our world: a body acting on another puts it in motion. In about six hours, anybody can learn and enjoy riding it. In all what is fun in life, such as toys, animals, women, machines, children, men, the sea, wind, cinema, juggling, rambling, loving, something pushes something else. Thus, the first challenge is to describe this transfer of motion in more precise terms.

Contact is not the only way to put something into motion; a counterexample is an apple falling from a tree. Non-contact influences are more fascinating and mysterious: nothing seems hidden, but nevertheless something happens. But contact motion seems to be more easy to grasp, and that is why one usually starts with it. However, we will soon find out that taking this choice one makes the same experience as when riding a bicycle at sustained speed and trying to turn right by turning the steering bar to the right: one takes a left turn.** The rest of our walk will rapidly force us to study non-contact interactions as well.

What is mass?

Δός μοι ποῦ στω καὶ κινῶ τὴν γῆν.
Da ubi consistam, et terram movebo.***

* One could imagine to include the requirement that objects may be rotated; however, it gives difficulties in the case of atoms, as explained on page 456, and with elementary particles.

Challenge ** This surprising effect obviously works only above a certain minimal speed. Can you determine which one?

*** Give me where to stand, and I'll move the earth.

Archimedes (ca. 283–212), as cited by Pappus. Ref. 41

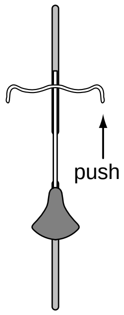


Figure 13 In which direction does the bicycle turn?

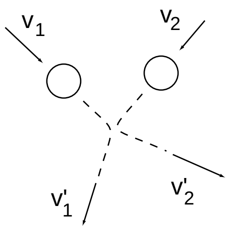


Figure 14 Collisions define mass

When we touch something we do not know, such as a when we kick some object on the street, we automatically pay attention to two aspects: how much we push, and how much the unknown object moves. If one want to be more precise, one performs experiments like the one shown in figure 14. Repeating the experiment with various pairs of objects, one notes that one can ascribe a fixed quantity m_i to every object i. These quantities are determined by the relation

$$\frac{m_2}{m_1} = -\frac{\Delta v_1}{\Delta v_2} \tag{12}$$

where Δv is the velocity produced by the collision. This particular number is called the *mass* of the object. In order to get a mass values common to everybody, the mass of one particular, selected body has to be fixed in advance. This special body is called the *standard kilogram* and is kept with great care under vacuum in a glass container near Paris. It is touched only once every few years because otherwise dust, humidity or scratches would change its mass. Through the standard kilogram the value of the mass of every other body is determined.

The mass thus measures the difficulty to get something moving. High masses are harder to move than low masses. Obviously, only objects, i.e. entities showing solidity, have mass; images don't. (By the way, the word 'mass' is derived, via latin, from the greek μαζα, bread, or the hebrew 'mazza', unleavened bread – quite a change in meaning.) Ref. 32

One also finds the important result that throughout any collision, the sum of all masses is conserved:

$$\sum_\mathrm{i} m_\mathrm{i} = \mathrm{const} \quad . \tag{13}$$

Therefore the mass of a composite systems is the sum of the mass of the components. Galilean mass is thus a measure for the quantity of matter.

Another way to formulate the same definition of mass is to say that one can ascribe a mass m_i to every object i such that for collisions free of outside interference, the sum of all products between the body's masses and their velocities, their momenta, is unchanged *throughout* the collision:

$$\sum_\mathrm{i} m_\mathrm{i}\mathbf{v}_\mathrm{i} = \mathrm{const} \quad . \tag{14}$$

In particular, the sum is the same before and after the collision, and is thus a conserved quantity. The two conservation "laws" (13) and (14) were first stated in this way by the important dutch physicist Christiaan Huygens (1629, 's Gravenhage –1695, Hofwyck).

By the way, if a moving sphere hits a resting one, after the collision there is a simple rule determining the angle between the directions the two spheres take. Can you find this rule which is so useful when playing billiard?

Challenge

In 1994, a cover photograph of the CERN Courier showed a man lying on a bed of nails with two large blocks of concrete on his stomach. Another man is hitting the concrete with a heavy sledgehammer. Most energy is absorbed by the concrete, thus there is no pain and no danger.

figure missing

Figure 15 Is this dangerous?

The definition of mass has been generalized by the austrian physicist and philosopher Ernst Mach (1838, Chrlice–1916)* in such a way that it is valid even if the two objects interact without contact, as long as they do so along the directions of the line connecting their positions. The mass ratio between two bodies is defined in general via the relation

$$\frac{m_2}{m_1} = -\frac{a_1}{a_2} \tag{15}$$

where a is the acceleration of each body during the interaction. Mass ratio is negative inverse acceleration ratio. This definition has been studied in much detail in the physics community, mainly in the nineteenth century. A few points sum it up:

- The definition of mass *implies* the conservation of momentum $\sum mv$ as elaborated below; the latter is not a separate "law"; the conservation of momentum cannot be checked experimentally, because mass is defined in such a way that it holds.

- The definition of mass *implies* the equality of *action* m_1a_1 and *reaction* $-m_2a_2$; the equality of action and reaction is not a separate "law" – mass is defined in such a way that it holds.

* The *mach* unit for airplane speed as multiple of the speed of sound in air (about 0.3 km/s) is named after him. He developed the so-called Mach-Zehnder interferometer; he also studied the basis of mechanics. His thoughts about mass and inertia influenced the development of general relativity, and let to Mach's principle, to be discussed later on. He also was proud to be the last scientist denying – humorously, but against all evidence – the existence of atoms.

Masses	Physical property	Mathematical name (see later for definitions)
can be distinguished	distinguishability	set
can be ordered	sequence	order
can change gradually	continuity	completeness
can be added	*quantity of matter*	additivity
don't change	conservation	invariance
do not disappear	impenetrability	positivity

Table 9 Properties of galilean mass

- The definition of mass is *independent* on whether contact is involved or not, and on whether the origin of the accelerations is due to electricity, gravitation, or other interactions.*

- The definition is valid only for observers at rest or in inertial motion. More about this issue later on.

By studying the measurements of mass with the bodies around, one discovers its main properties: mass, as defined and used in everyday life, is additive – not addictive – i.e. the mass of two bodies combined is equal to the sum of the two masses; furthermore, mass is continuous; it seemingly can take any positive value; finally, mass is conserved. One finds that the mass of a system, defined as the sum of the mass of all constituents, does not change when the system is kept isolated from the rest of the world. In fact it is found that mass in conserved in more processes than in collisions alone: also during melting, evaporation, digestion, etc., mass does not seem to disappear or to appear.

Later on we will find that also in the case of mass all these properties are only approximate; precise experiments show that none of them is correct. Note also that in order to define mass one has to be able to distinguish the two bodies from each other. This seems a trivial requirement, but later on we will see that this is not always possible in nature.

The definition of mass implies that during the fall of an object, the earth is accelerated upwards by a tiny amount. If one could measure this tiny amount, one could determine the mass of the earth. Unfortunately, this is impossible. One needs a smarter way to do this. Can you find one? Challenge

The mass of a body is thus most precisely described by a *positive* real number, often abbreviated m or M. This is a direct consequence of the impenetrability of matter. A negative (inertial) mass would mean that such a body would move in opposite direction to any applied momentum change. Such a body could not be kept in a box; it would break through any wall trying to stop it. Strangely enough, negative mass bodies still would fall downwards in the field of a large positive mass. However, a small positive mass object would float away from a large negative mass body, as you can easily deduce by comparing the various accelerations involved. A positive and a negative mass of the same size would stay at constant distance and spontaneously accelerate away along the line connecting the two masses. Note that Challenge

* As mentioned above, only *central* forces obey equation (15), i.e. forces acting between the center of mass of bodies. But since all fundamental forces are central, this is not a restriction in this discussion. There is one notable exception: magnetism. Is the definition of mass valid in this case? Challenge

Observation	Mass
Mass increase through absorption of one green photon	ca. $3.7 \cdot 10^{-36}$ kg
Lightest known object: electron	ca. $9.1 \cdot 10^{-31}$ kg
Mass of human at early age	ca. 10^{-11} kg
Mass of water adsorbed on a kilogram mass	10^{-8} kg
Planck mass	ca. $2.2 \cdot 10^{-8}$ kg
Mass of fingerprint	ca. 10^{-7} kg
Mass of typical ant	ca. 10^{-7} kg
Mass of water droplet	ca. 10^{-6} kg
Mass of largest living being	ca. 10^{6} kg
Mass of largest ship	ca. $400 \cdot 10^{6}$ kg
Largest object moved by man (Troll gas rig)	ca. $687.5 \cdot 10^{6}$ kg
Water on earth	10^{21} kg
Solar mass	$2.0 \cdot 10^{30}$ kg
Mass of galaxy	ca. 10^{41} kg
Total mass visible in universe	ca. 10^{54} kg

Table 10 Some mass values

both energy and momentum is conserved in all these situations.* Negative mass bodies have never been observed. Antimatter, which will be discussed later, also has positive mass.

See page 169 and 453

Is motion eternal?

> Every body continues in the state of rest or of uniform motion
> in a straight line except in so far as it doesn't.
> Arthur Eddington (1882–1944), british astrophysicist.

Using the definition of mass, the product $\mathbf{p} = m\mathbf{v}$ is called the *momentum* of a particle; it describes the tendency of an object to keep moving in collisions. The bigger it is, the harder it is to stop the object. Like velocity, momentum has a direction and a magnitude: it is a vector. (In french, momentum is called 'quantity of motion', a more appropriate term. In the old days, the term 'motion' was used instead of 'momentum', by Newton, for example.) Relation (14), the conservation of momentum, therefore expresses the conservation of motion during interactions.

Momentum and energy are *extensive quantities*. That means that one can say for each of them that they 'flow' from one body to the other, and that they can be 'accumulated' in bodies, in the same way that water flows and can be accumulated in containers. Imagining momentum as something which can be exchanged between bodies, e.g. in collisions, is a good help when thinking about the description of moving objects.

* For more curiosities, see R.H. PRICE, *Negative mass can be positively amusing*, American Journal of Physics **61**, pp. 216–217, 1993. Negative mass particles in a box would heat up a box made of positive mass while traversing its wall, and accelerating, i.e. losing energy at the same time. They would allow to build a perpetuum mobile of the second kind, circumventing the second "law" of thermodynamics. Moreover, such a system would have no thermodynamical equilibrium, because its energy can decrease for ever. The more one thinks about negative mass, the more one finds strange properties in contrast with observations.

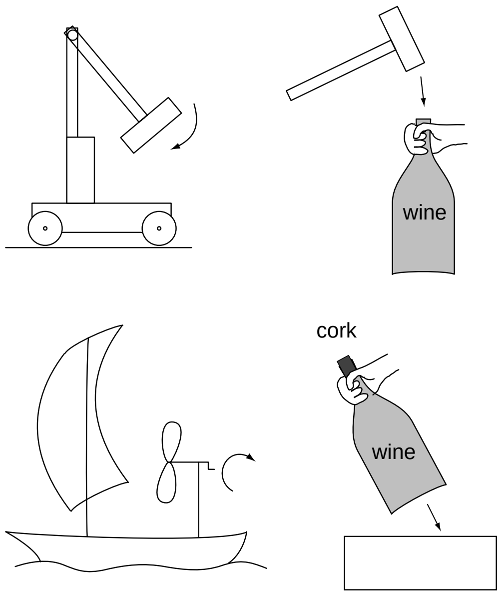


Figure 16 What happens?

An observation included in the conservation of momentum is the limitation one experiences when one is on a perfectly frictionless surface, such as ice or a polished oil covered marble surface: you cannot propel yourself forward by patting yourself on your own back. (Have you ever tried to put a cat on such a marble surface? It is not even able to stand on its four legs. Can you imagine why?) The conservation of momentum – and mass – also means that teleportation ("beam me up") is impossible. Challenge

The conservation of momentum also means that motion, at least in collisions, never stops, but simply is exchanged. However, motion often seems to disappear, as for example in the case of a stone dropped on the ground, or of a ball left rolling on grass. Moreover, in daily life one often observes creation of motion, e.g. every time one opens a hand. How do these examples fit with the conservation of momentum?

It turns out that the answer lies in the microscopic aspects of these systems. In fact a muscle only *transforms* one type of motion, namely that of the electrons in certain chemical compounds (usually adenosinetriphosphate (ATP), the fuel of most processes in animals) Ref. 42 into another, the motion of the fingers. The working of muscles is similar to that of a car engine transforming the motion of electrons in the fuel into motion of the wheel. Both systems need fuel and get warm in the process.

Also when a ball rolls on grass until it stops, one has to study the microscopic behaviour. Motion seems to disappear in this case, and one uses the term *friction*. Studying the situation carefully, one finds that the grass and the ball heat up a little during this process. Friction is thus a process transforming visible motion into heat. Once one knows the structure of matter it becomes clear that heat is the disorganized motion of the microscopic constituents of every material. When these constituents all move in the same direction, the object as a whole moves; when they oscillate randomly, the objects is at rest but is warm. Heat is a form of motion.

In summary, motion is indeed eternal on microscopic scale. On the other hand, *macroscopic* perpetual motion does not exist, since friction cannot be eliminated completely.* Therefore the disappearance and also the spontaneous appearance of motion is an illusion.

* Some funny examples of past attempts to built a perpetuum mobile, i.e. a machine that never stops moving, are described in Stanislav Michel, *Perpetuum mobile*, VDI Verlag, 1976. The conceptual mistake made by the eccentrics behind these attempts is always the same: the hope to overcome friction. If the machine is well constructed, i.e. with little friction, it can take the little energy in needs for the sustention of motion from very subtle environmental effects. For example, in the Victoria and Albert Museum in London one can admire a Ref. 43 beautiful clock powered by the variations of air pressure over time.

Small friction means that motion takes a long time to stop. One directly thinks about motion of the planets. Challenge In fact, there is friction between the earth and the sun. (Can you guess one of the mechanisms?) But the value is so small that the earth circles around the sun already for thousands of millions of years.

For example, the motion proper to every living being exists before its birth, and remains after its death. The same happens with its energy. This is probably the closest one can get to the idea of everlasting life from evidence collected by observation. It is perhaps less than a coincidence that energy used to be called 'vis viva', living force, by Leibniz.

Note that since motion it conserved, it has no origin. Therefore, at this stage of our walk we cannot at all answer the fundamental question: Why does motion exist? What is it origin? The end of our walk is nowhere near.

But the example of the ball rolling on grass also shows that motion cannot be described adequately by momentum alone. Momentum is never lost, only exchanged. One also needs a quantity which distinguishes situations with friction from situations without friction. In daily life, one feels that in a collision without friction, an *elastic* collision, in which the bodies bounce well, little is lost to friction, whereas in a collision where the bodies stick to each other, the losses are greater. What is the quantity we are looking for? Obviously, the quantity must depend on the mass of a body, and on its velocity. Experiments show that in the case elastic collisions, and in that case only, the following quantity, called the *kinetic energy* T of the system, is also conserved:

$$T = \sum_{\mathrm{i}} \frac{1}{2} m_{\mathrm{i}} \mathbf{v}_{\mathrm{i}}{}^2 = \mathrm{const} \quad . \tag{16}$$

The factor 1/2 and the name was introduced by the french engineer and mathematician Gustave-Gaspard Coriolis (Paris, 1792– Paris, 1843) so that the relation $dT/dv = p$ would be obeyed. In non-elastic collisions, part or all of the kinetic energy is lost. In these and in other cases one finds thus the general rule: *friction* leads to the loss of kinetic energy. Energy, a word taken from ancient greek originally used to describe character and meaning "intellectual or moral vigour", was taken over into physics because its literal meaning is 'force within.'

Physical energy measures the ability of a body to generate motion. Kinetic energy is thus not conserved in everyday life: friction destroys kinetic energy. It was one of the important conceptual discoveries of physics that *total* energy *is* conserved if one includes the discovery at heat is a form of energy. Friction is a process transforming kinetic energy, i.e. the energy connected with the motion of a body, into heat. However, on a microscopic scale, energy is conserved.

Do not be surprised if you do not grasp the differences between momentum, energy, and force (to be introduced later) straight away: physicists took about 100 years to figure them out; for some time they insisted in using the same word for all of them, and they didn't know which problem required which concept. So one is allowed to take a few minutes for the topic.

One way to express the difference is to note that if a body is accelerated by a constant force, momentum is what increases with time, and energy is what increases with distance. Is it therefore more difficult to stop a running man with mass m and speed v, or one with mass $m/2$ and speed $2v$, or one with mass $m/2$ and speed $\sqrt{2}v$? You may want to ask a rugby playing friend for confirmation. Challenge

By the way, when a car travelling at 100 m/s runs frontally into a parked car of the same make, which car has the larger damage? Challenge

To get a feeling for energy, here is an additional way. The world use of energy by human machines (coming from solar, geothermal, biomass, wind, nuclear, hydro, gas, oil, coal, or animals sources) in the year 2000 is about 500 EJ, for a world population of about 6000 million people. To see what this energy consumption means, translate it into a personal power consumption; one gets about 2.6 kW. Now a working person can produce mechanical work of about 100 W. In short, the average human energy consumption corresponds to about 26 humans working 24 hours a day. In other words, if one looks at the energy consumption in countries of the first world, such as the one of the reader, the average inhabitant there has machines working for him equivalent to more than 100 'servants'. Can you name a few? Challenge

Rotation

Rotation keeps us alive. Without the change of day and night, we would either be fried or frozen to death. A short summary of rotation is thus appropriate. We saw before that a body is described by its reluctance to move, called its inertial mass; similarly, a body also has a reluctance to turn. This quantity is called its *moment of inertia*, and is often abbreviated Θ. The speed or rate of rotation is described by the *angular velocity*, usually abbreviated ω. Like mass, the moment of inertia is defined in such a way that the sum of *angular momenta* L – the product of moment of inertia and angular velocity – is conserved in systems which do not interact with the outer world:

$$\sum_{\mathrm{i}} \Theta_{\mathrm{i}} \omega_{\mathrm{i}} = \sum_{\mathrm{i}} L_{\mathrm{i}} = \text{const} \qquad (17)$$

The moment of inertia can be related to the mass and shape of a body; it is given by the expression

$$\Theta = \sum_{\mathrm{l}} m_{\mathrm{l}} \rho_{\mathrm{l}}^2 \quad , \qquad (18)$$

where ρ_{l} is the distance form the mass element m_{l} and the axis of rotation. The moment of inertia of a body therefore depends on the chosen axis of rotation.*

Obviously, also the value of the angular momentum depends on the location of the axis used for its definition. One distinguishes *intrinsic* angular momentum, when the axis goes through the center of mass of the body, from *extrinsic* angular momentum, when it does not.** (By the way, the *center of mass* of a body is that imaginary point which moves straight during flight or fall, even if the body is rotating while doing so. Can you find a way to determine it for a specific body?) Challenge

* The set of all possible values is described by a generalization of a vector called a *tensor*. Vectors are quantities with a magnitude and a direction; tensors are quantities with a magnitude, and with a direction depending on a second, chosen direction. They describe simple distributions in space. If vectors can be visualised as arrows, tensors can be visualized as ellipsoids. We will encounter them every now and then in our walk. A vector is described by a list of components; a tensor is described by a *matrix* of components. A vector has the same length and direction for every observer; a tensor (of rank 2) has the same determinant, the same trace, and the same sum of diagonal subdeterminants for all observers.

** For macroscopic bodies, the extrinsic angular momentum is related to the intrinsic one by

$$\Theta_{\mathrm{ext}} = \Theta_{\mathrm{int}} + md^2 \quad , \qquad (19)$$

where d is the distance between the center of mass and the axis of extrinsic rotation. This relation is called *Steiner's parallel axis theorem*. Are you able to deduce it? Challenge

Quantity	Linear motion		Rotation	
State	momentum	p	angular momentum	L
	position	x	angle	φ
Reluctance to move	mass	m	moment of inertia	Θ
Motion	velocity	v	angular velocity	ω
	acceleration	a	angular acceleration	α

Table 11 Correspondence between linear and rotational motion

Observation	Angular velocity
Average sun rotation around axis	ca. $2\pi\, 3.8 \cdot 10^7\,/\mathrm{s} = 1/\,30\mathrm{d}$
Typical lighthouse	ca. $2\pi\, 0.08\,/\mathrm{s}$
Jumping ballet dancer	ca. $2\pi\, 3\,/\mathrm{s}$
Bacterial flagella	ca. $2\pi\, 100\,/\mathrm{s}$
Fastest turbine built	ca. $2\pi\, 10^3\,/\mathrm{s}$
Fastest rotating stars	ca. $2\pi\, 10^3\,/\mathrm{s}$
Proton rotation	ca. $2\pi\, 10^{20}\,/\mathrm{s}$
Highest possible angular velocity	ca. $2\pi\, 10^{35}\,/\mathrm{s}$

Table 12 Some rotation speeds

Every object which has an orientation also has an intrinsic angular momentum. (And a sphere?) Therefore, point particles do not have intrinsic angular momenta – at least in first approximation. (Later, in quantum theory, this conclusion is changed.) The extrinsic angular momentum of a point particle is given by

Challenge

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = \frac{2\mathbf{A}m}{T} \tag{20}$$

where $\mathbf{A}$ is the surface swept by the position vector of the particle during time T.*

One can define a corresponding *rotational energy* as

$$T_{\mathrm{rot}} = \tfrac{1}{2}\Theta\,\omega^2 = \tfrac{1}{2}\frac{L^2}{\Theta} \quad . \tag{22}$$

* The *cross product* or *vector product* $\mathbf{a} \times \mathbf{b}$ between two vectors $\mathbf{a}$ and $\mathbf{b}$ is defined as that vector which is orthogonal to both, whose length is given by $ab \sin\sphericalangle(\mathbf{a},\mathbf{b})$, i.e. by the surface of the parallelogram spanned by the two vectors, and whose orientation is given by the *right hand rule*. The vector product thus has the properties

Challenge

$$\begin{aligned}
&\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a} \quad , \quad \mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c} \quad , \quad \lambda\mathbf{a}\times\mathbf{b} = \lambda(\mathbf{a}\times\mathbf{b}) = \mathbf{a}\times\lambda\mathbf{b} \quad , \quad \mathbf{a}\times\mathbf{a} = \mathbf{0} \quad , \\
&\mathbf{a}(\mathbf{b}\times\mathbf{c}) = \mathbf{b}(\mathbf{c}\times\mathbf{a}) = \mathbf{c}(\mathbf{a}\times\mathbf{b}) \quad , \quad \mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \mathbf{b}(\mathbf{ac}) - \mathbf{c}(\mathbf{ab}) \quad , \\
&(\mathbf{a}\times\mathbf{b})(\mathbf{c}\times\mathbf{d}) = \mathbf{a}(\mathbf{b}\times(\mathbf{c}\times\mathbf{d})) = (\mathbf{ac})(\mathbf{bd}) - (\mathbf{bc})(\mathbf{ad}) \quad , \\
&(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d}) = \mathbf{c}((\mathbf{a}\times\mathbf{b})\mathbf{d}) - \mathbf{d}((\mathbf{a}\times\mathbf{b})\mathbf{c}) \quad , \quad \mathbf{a}\times(\mathbf{b}\times\mathbf{c})+\mathbf{b}\times(\mathbf{c}\times\mathbf{a})+\mathbf{c}\times(\mathbf{a}\times\mathbf{b}) = 0 \quad .
\end{aligned} \tag{21}$$

The vector product exists only in three-dimensional vector spaces. The cross product vanishes if and only if the vectors are parallel. The tetrahedron formed by three vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ has a volume given by $V = \mathbf{c}(\mathbf{a}\times\mathbf{b})/6$. The parallelepiped spanned by the three vectors has 6 times that volume.

As in the case of linear motion, the angular quantities are not always conserved in the macroscopic world, due to friction; but they are always conserved on the microscopic scale. We will study the consequences more in detail in the part on quantum theory.

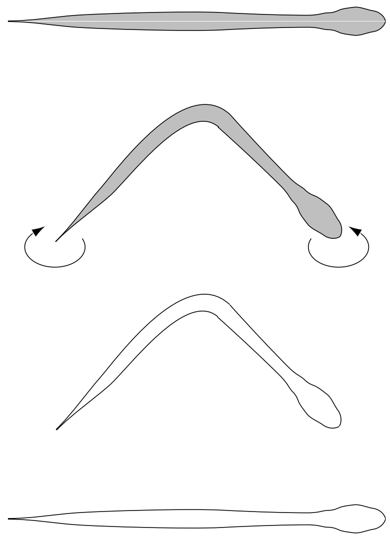


Figure 17 A snake is able to turn itself around its axis.

On a frictionless surface, as approximated by smooth ice or by a marble floor covered by a layer of oil, it is impossible to move forward. In order to move, one needs to push against something. Is this also the case for rotation?

Surprisingly, no. It is possible to *turn* even without pushing against something. One can check this on an (almost) frictionless rotating office chair: one simply rotates an arm above the head. After each turn of the hand, the orientation of the chair has changed by a small amount. Indeed, conservation of angular momentum and of rotational energy do not prevent bodies from changing their orientation. Cats learn this in their youth; after they learned the trick, if they are dropped legs up, they can turn themselves in such a way that they always arrive on the floor on their legs. Snakes also know how to do it. During the Olympic games one can watch board divers perform similar tricks. Rotation is thus different from translation in this aspect. Why? The final reason will Challenge
be unveiled later in our escalation.

Rolling wheels

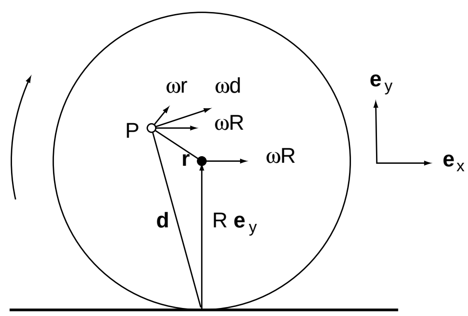


Figure 18 The velocities and unit vectors for a rolling wheel

Rotation is an interesting phenomenon. A rolling wheel does *not* turn around its axis, but around its point of contact, as can easily be shown. *Rolling* means that if the axis of a wheel of radius R moves with v_{axis} then the angular velocity is

$$\omega = v_{\mathrm{axis}}/R \quad . \tag{23}$$

One easily sees that any point P on the wheel, with distance r from the axis, has the velocity

$$\mathbf{v}_{\mathrm{P}} = \omega R \mathbf{e}_{\mathrm{x}} - \omega r \mathbf{e}_{\theta} \quad , \tag{24}$$

where $\mathbf{e}_{\theta}$ in the second term is a unit vector orthogonal to the line connecting the point P and the axis. Now take $\mathbf{e}_{\mathrm{z}}$ as the unit vector along the axis; then one can transform the previous expression into

$$\mathbf{v}_{\mathrm{P}} = (-\omega \mathbf{e}_{\mathrm{z}}) \times (R\mathbf{e}_{\mathrm{y}} + \mathbf{r}) = \omega \times \mathbf{d} \tag{25}$$

which shows that a rolling wheel does indeed rotate around its contact point with the ground.

Some points on a rolling wheel even move towards the axis, some stay fixed, while others move away from it. They lead to interesting pictures when a rolling wheel with spokes, such as a bicycle wheel, is photographed. Can you determine where these points are located?

Ref. 44
Challenge

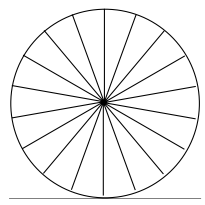


Figure 19 A simulated photograph of a rolling wheel with spokes

How do we walk?

> Golf is a good walk spoiled.
> Mark Twain

Why do we move our arms when walking or running? To conserve energy. In fact, a body movement is natural and graceful when it is performed with as little energy as possible. (This can indeed be taken as the actual definition of grace. It is common knowledge in the world of dance and it is a central aspect also of the methods used by actors to learn how to move their bodies as beautifully as possible, such as the Alexander technique.)

Ref. 10

To be convinced about the energy savings, try walking or running with the arms moving in the opposite direction than usual: the effort is considerably higher. This is also valid if one does not move one's arms at all. When a leg is moved, it produces a moment of rotation around the body axis, which has to be counterbalanced. The method using least energy is the swinging of arms. Since the arms are lighter than the legs, to compensate for the momentum, they must move further from the axis of the body; evolution has therefore moved the attachment of the arms, the shoulders, farther away than that of the legs, the hips. Animals on two legs which have no arms have more difficulties walking and have to move their whole torso, such as penguins or pigeons.

Which muscles do most of the work when walking, i.e. during the motion which the experts call *gait*? In 1980, Serge Gracovetsky found that in gait, most power comes from the spine muscles, not from the leg muscles. (Note that people without legs are also able to walk.) When one makes a step, the lumbar muscles straighten the spine, this automatically makes it turn a bit to one side, and then the knee of the leg on that side automatically comes forward. When the foot is moved, the lumbar muscles can relax, and then straighten again for the next step. The arm swing helps to reduce the necessary energy – one can feel the increase in tension in the back muscles if one walks without moving the arms.

Ref. 45

Is the earth at rest?

> Eppur si muove![*]

The search for answers to this question gives a beautiful cross section through the history of classical physics.

[*] 'And yet she moves' is the sentence falsely attributed to Galileo about the earth; true is that in his trial he was forced to publicly retract the idea of a moving earth to save his life (see also the footnote on page 416).

Around the year 265 BCE, the greek thinker Aristarchos of Samos maintained that the earth rotates. He measured the parallaxis of the moon (today known to be up to 0.95 degrees) and of the sun (today known to be 8.8”). The *parallaxis* is an interesting effect; it is the angle describing the difference between the directions of a body in the sky when seen from an observer on the surface of the earth and when seen form a hypothetical observer at its centre. Did this measurement provide Aristarchos with enough arguments for his conclusion?

Ref. 46

Challenge

In 1802, the hamburger physicist A. Benzenberg (1777–1846) showed that an object falling from a large height does not hit the earth on the spot below, but is deviated to the east because it keeps the larger horizontal velocity it had at the height from where it fell. Using steel balls which he dropped from the tower of the Michaelis church in Hamburg – a height of 76 m – Benzenberg found that the deviation to the east was 9.6 mm. This was the first non-astronomical proof of the earth’s rotation. (There is also a much smaller south deviation, not measured by Benzenberg; it is mentioned here in order to complete the list of the effects of the rotation of the earth.)

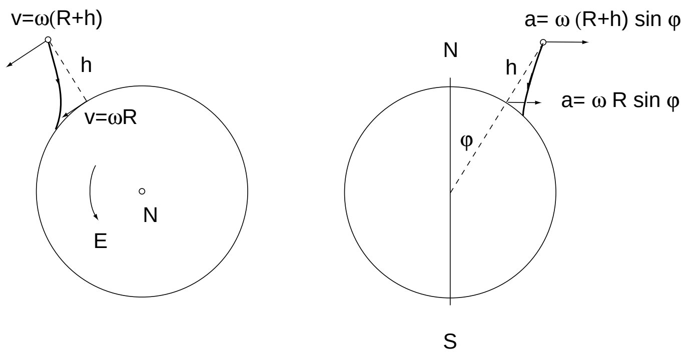


Figure 20 The deviations of free fall from vertical towards the east and towards the equator due to the rotation of the earth.

In 1835, the french engineer and mathematician Gustave-Gaspard Coriolis (1792–1843), the same who also introduced the modern concepts of ‘work’ and of ‘kinetic energy’, described an effect up to then overlooked on the motion of objects traveling on a rotating background. If the rotation is counterclockwise, as is the case for the earth on the northern hemisphere, the direction of the velocity of any object is slightly turned to the right, whereas its magnitude remains constant.

The acceleration is due to the change of distance to the axis of rotation. Can you deduce the analytical expression for it, namely $\mathbf{a}_C = 2\omega \times \mathbf{v}$? In many large scale phenomena with a spiral symmetry, this so-called *Coriolis acceleration* (or Coriolis force) determines the handiness: the directions of cyclones and anticyclones in meteorology, the general wind patterns on earth, the deflection of ocean currents and tides. It also plays a role in the flight of canon balls (that was the original interest of Coriolis), in satellite launches, in the motion of sun spots and even in the motion of electrons in molecules; most beautifully, it explains why icebergs do not follow the direction of the wind as they drift away from the polar caps.

Challenge

Ref. 47

In summary, when one throws a stone, the point below it on the ground does not follow a straight line; the motion of the stone is *not* in a plane, as the moving pictures in the lower right corner seem to suggest. The rotation of the earth makes the real path a curve in all three dimensions.

Only in 1962, A. Shapiro was the first to verify that the Coriolis effect has an influence on the direction of the vortex formed by the water flowing out of a “bathtub”. More than a bathtub he had to use a carefully designed experimental set-up, because contrary to an often

Ref. 48

heard assertion, no such effect can be seen in real bathtubs. He succeeded only by carefully eliminating all disturbances from the system, among others by waiting 24 hours after the filling of the reservoir (and never actually stepping in or out of it!) in order to avoid any left over spurious motion of water which would disturb the effect, and by building a carefully designed, completely rotationally symmetric opening mechanism. Others have repeated the experiment, confirming its result, also on the southern hemisphere. But let us go on with the story.

Also the aberration of light, discovered in 1728 by James Bradley, the astronomer royal, shows the rotation of the earth. At the equator, it adds an angular deviation of $0.32''$, changing sign every 12 hours, to the aberration due to the motion of the earth around the sun, about $20.5''$. We will discuss the aberration in more detail shortly. See page 147

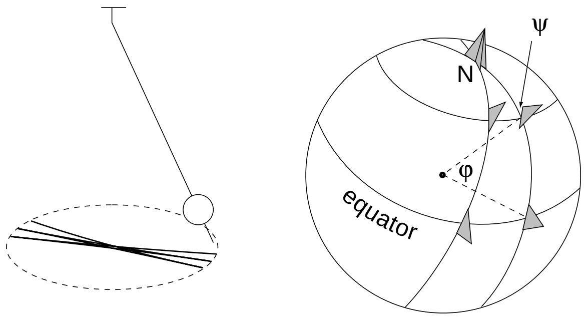


Figure 21 The motion of a pendulum on the rotating earth

In 1851, the french physician turned physicist Jean Bernard Léon Foucault (1819, Paris–1868, Paris) performed an experiment which rendered him world famous practically overnight. He suspended a 67 mlong pendulum in the Panthéon in Paris and showed the astonished public that the direction of its swing changes over time, rotating slowly. To everybody with a few minutes of patience to watch the change of direction, the experiment proved that the earth is rotating. More precisely, the rotation period T_F of the oscillation plane seen on earth is given by Challenge

$$T_F = \frac{24\,\mathrm{h}}{\sin\varphi} \qquad (26)$$

where φ is the latitude of the location of the pendulum, e.g. $0°$ at the equator and $90°$ at the north pole. This is perhaps the most beautiful result of galilean kinematics. Ref. 49

Why was such a long pendulum necessary? Understanding the reasons allows one to repeat the experience at home, using a pendulum only 70 cm long. Challenge Ref. 50

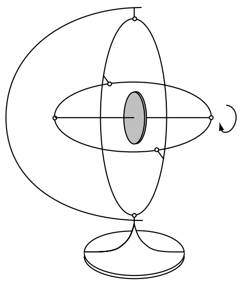


Figure 22 A gyroscope

Foucault is also the inventor and namer of the *gyroscope*, a device shown in figure 22, which he built in 1852, one year after his pendulum. With it, he again demonstrated the rotation of the earth. Can you imagine how? Such devices are now routinely used in ships and in airplanes to give the direction of north, because they are more precise and more reliable than magnetic compasses. Challenge

In 1910, E. Hagen published the results of an even simpler experiment also proving the rotation of the earth. If two masses on a horizontal bar are moved towards the supporting axis as shown in figure 23, and if the friction is kept low enough, the bar rotates. That would not happen if the earth was at rest. Can you explain why? This not so well-known effect is useful for winning bets among physicists. Challenge

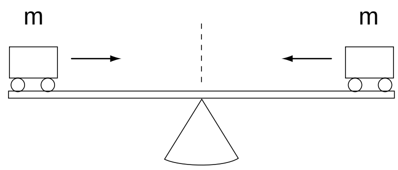


Figure 23 Showing the rotation of the earth through the rotation of an axis

In 1925, the Albert Michelson* and his collaborators in Illinois constructed an vacuum interferometer with a perimeter of 1.9 km – this is no typo – and found a fringe shift due to the rotation of the earth. The result uses an effect first measured in 1913 by the french physicist Georges Sagnac: the rotation of a complete ring interferometer with angular frequency Ω produces a fringe shift $\Delta\varphi$ given by

$$\Delta\varphi = \frac{8\pi \mathbf{\Omega} \, \mathbf{A}}{c\,\lambda} \qquad (27)$$

where **A** is the area enclosed by the two interfering light rays, λ the wavelength, and c the speed of light. The effect is now called the *Sagnac effect*; it had been predicted already 20 years earlier by Oliver Lodge (1851–1940), a british physicist who studied electromagnetic waves and also tried to communicate with the dead. Modern high precision versions use ring lasers with areas of only a few square meters, but are able to measure variations of the rotation rates of the earth of less than one part per million. They are used for research into the motion of the soil due to lunar tides, to earth quakes, and for checks of the theory of relativity. Anyway, over the course of a year, the length of a day varies irregularly by a few milliseconds, mostly due to the sun, the moon, the currents inside the earth and the weather. Ref. 51 Ref. 52

In summary, observations show that the earth surface turns with 463 m/s at the equator, a larger value than that of the speed of sound in air, about 340 m/s in usual conditions, and that we are in fact whirling through the universe.

Is the rotation of the earth constant over geological time scales? Can you find a way to check this? That is a really hard question. If you find any possible answer, publish it! (The same is valid for the question whether the length of the year is constant.) Challenge Ref. 53

But the earth does not simply rotate around an axis. This was known already long ago.

In 128 BCE, the greek astronomer Hipparchos discovered what is today called the *(equinoctial) precession*, by comparing a measurement he made himself with another made 169 years before. Hipparchos found that the earth axis was not pointing to the same position over time. He concluded that the sky was moving; today, using the above results, we prefer to say that the axis of the earth is moving. It turns out that in 26000 years it draws a cone with an opening angle of 23°.

But the axis of the earth is not even fixed compared to the earth's surface. In 1885, by measuring the exact angle above the horizon of the point around which stars turn at night, Küstner found that the axis of the earth moves with respect to the earth's crust. In practice, the north pole moves around an average position by about 15 m every 1.2 years. The reasons are probably the different concentrations of land masses and seasonal variations of air and water masses on both hemispheres. Ref. 54

In 1912, the german meteorologist and geophysicist Alfred Wegener (1880–1930), after studying the shapes of the continental shelves and the geological layers on both sides of the atlantic, conjectured that the continents *move*. Following the modern version of the model,

* Albert Albert Michelson (1852, Strelno–1931, Pasadena) prussian-polish-american physicist, Nobel prize in physics in 1907.

plate tectonics, the continents float on the fluid mantle of the earth like pieces of cork on water, and the convection inside the mantle provides the driving force for the motion. Ref. 55

Satellite measurements confirm that e.g. the american continent moves away from the european continent by about 10 mm/a. There are also speculations that this velocity may have been much higher for certain periods in the past. The way to check this is to look at magnetization of sedimental rocks. At present, this is still a hot topic of research.

Why does the earth rotate at all? The rotation is a result of the rotating gas cloud from which the solar system formed. This connection explains that the sun and all planets, except one, turn around themselves in the same direction, and that they also all turn around the sun in that same direction. But the complete story is outside the scope of this text. Nevertheless, it brings us to the next question: is the center of the earth at rest in the universe? Ref. 56

Already Aristarchos of Samos in the third century BCE had maintained that the earth turns around the sun. But a fundamental difficulty of the heliocentric system is that the stars look the same all year long. How can this be, if the earth goes around the sun? Only in 1837, the westphalian astronomer Friedrich Wilhelm Bessel (1784–1846), who left his successful business career to dedicate his life to the stars, was the first to measure the *parallax* of stars, i.e. the change of the direction of their image with time, proving in this way that the earth is not directly fixed compared to the stars and actually does move around the sun. This was a result of extremely careful measurements and complex calculation: he discovered the *Bessel functions* in order to realize it. He was able to find a star, 61 Cygni, whose apparent position changed with the month of the year. Seen over the whole year, the star describes a small ellipse on the sky, with an opening of 0.588″ (this is the modern value). After carefully eliminating all other possible explanations, he deduced that the change of position was due to the motion of the earth around the sun, and from the size of the ellipse he determined the distance of the star to be 105 Pm or 11.1 light years. Challenge

Also the abovementioned aberration of light, discovered already in 1728 by James Bradley and to be discussed shortly, shows that compared to the fixed stars, the earth moves around the sun. See page 147

With the improvement of telescopes, also other motions of the earth were discovered. In 1748, James Bradley announced that there is a small regular change of precession, which he called *nutation*, with a period of 18.6 years and an amplitude of 19.2 seconds of arc, due to the fact that the plane of the moon's orbit is not exactly the same as the plane of the earth's orbit around the sun. Are you able to confirm that this situation can produce nutation? Challenge

The tilt of the earth's axis, i.e. the angle between its intrinsic and its orbital angular momentum, which today is about 23.5 degrees, changes from 0 to 90 degrees with a period of 41000 years. This motion is due to the attraction of the sun and the deviations of the earth from spherical shape. In 1914, the serbian astronomer Milutin Milankovitch (1879–1958) understood that the orbital precession of 23000 years (or 26000, depending on the source) and the 41000 year period of the tilt, the *obliquity*, of the axis give rise to the *ice ages*. This happens through stronger or weaker irradiation of the poles, which then melt and lead to changes in average temperature. This has been confirmed spectacularly by measurements of the average temperature in the past million years, deduced from oxygen isotope ratios in sea sediments. The last ice age had is peak about 20000 years ago. Ref. 57

In addition, the earth's orbit changes its *eccentricity* with time, from completely circular to slightly oval and back. However, this happens in very complex ways, not with periodic

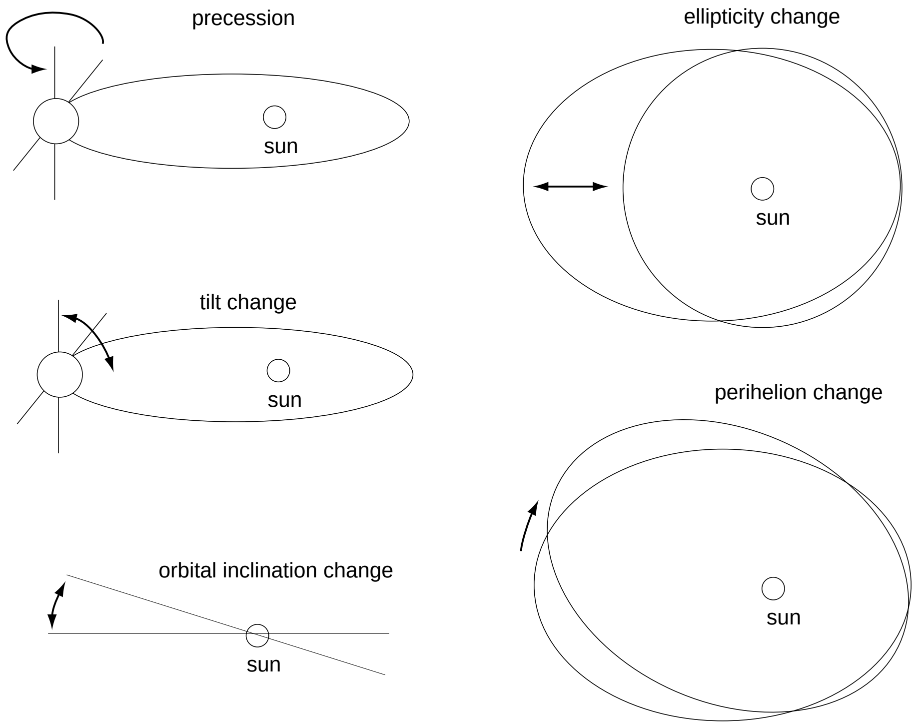


Figure 24 Changes in the earth's motion around the sun

regularity. The typical time scale is 100000 to 125000 years.

The earth's orbit changes also in *tilt* with respect to the orbits of the other planets; this seems to happen regularly every 100000 years. In this period the orbit changes from an inclination of 2.5 degrees back to zero and to minus 2.5 degrees, and so forth.

Even the direction in which the ellipse points changes with time. This so-called *perihelion shift* will not be studied at this point; but we return to this topic in the chapter of general relativity.

The next step is to ask whether the sun moves. Well, it does. Locally, it moves with a speed of 19.4 km/s towards the constellation of Hercules. But globally, the motion is more interesting. The diameter of the galaxy is 100000 light years, and we are located 25000 light years form the center. At our distance, the galaxy is 1300 light years thick; we are 68 light years from the centre plane. The sun, and with it the solar system, takes about 225 million years to turn once around the galactic centre, its orbital velocity being around 220 km/s. The formation of galaxies, like that of solar systems, always happens in a whirl. By the way, are you able to confirm by personal observation that our galaxy rotates? Ref. 58 Challenge

Finally, one can ask whether the galaxy moves. This can be measured because it is possible to give a value for the motion of the sun through the universe, defining it as the motion against the background radiation. This value has been measured to be 370 km/s. (The velocity of the *earth* through the background radiation of course depends on the season.) This value is a combination of the motion of the sun around the galaxy centre and of the motion of the galaxy itself. This latter motion is due to the attraction of other nearby galaxies. Ref. 59

Why don't we feel all these motion of the earth? Again, this question was answered by the master. Galileo explained in his lectures and books that only *relative* velocities produce effect, not absolute velocities. For the senses, there is no difference between constant motion and rest. We do not feel the motion of the earth because we move with it, and because at everyday scale, it is essentially constant. Nevertheless, many of these motions also induce measurable effects in atomic clocks.

See page 219

In summary, the earth really moves, and does so in rather complex ways. As Henri Poincaré would say, if we are on a given spot today, say the place of the Panthéon in Paris, and come back to the same spot tomorrow at the the same time, we are in fact 31 million kilometers further. But we stop this discussion at this point, and we return to motion in everyday life.

By the way, every physicist knows that the statement, "Daß die Sonne morgen aufgehen wird, ist eine Hypothese; und das heißt: wir wissen nicht, ob sie aufgehen wird." is *wrong*.* Can you show it?

Challenge

Curiosities of everyday motion

It is a mathematical fact that the casting of this pebble from my hand alters the centre of gravity of the universe.
Thomas Carlyle (1797–1881), *Sartor Resartus III*.

Here are a few facts to ponder.**

Challenge
• How often in 24 hours do the hour and minute hands of a clock lie on top of each other? How often does this happen for clocks having also a hand for seconds?

Ref. 5
Challenge
• How many times in twelve hours can the two hands of a clock be *exchanged* with the result that the new situation shows a *valid* time? What happens for clocks having also a third hand?

Ref. 60
Challenge
• A common fly on the stern of a 30,000 ton ship of 100 m length tilts it by less than the diameter of an atom. Today, distances that small are easily measured. Can you think of at least two methods, of which one should not cost more than 2000 Euro?

Challenge
• Does a wall get a stronger push when it is hit by a ball rebounding from it or when it is hit by a ball which remains stuck to it?

• The level of acceleration a human can survive depends on the time it takes. For a tenth of a second, $30\,g = 300\,m/s^2$, as generated in ejector seats in airplanes, are acceptable. (It seems that the record survived acceleration is about $80\,g = 800\,m/s^2$.) But in general it is said that accelerations of $15\,g = 150\,m/s^2$ or more are fatal.

Challenge
• The sliding ladder problem, shown schematically in the figure, asks for the the detailed motion of the ladder over time. The problem is more difficult than it looks, even if friction is not taken into account. Can you say whether the lower end always touches the floor?

* "It is an hypothesis that the sun will rise tomorrow; and this means that we do not *know* whether it will rise." The statement is by Ludwig Wittgenstein, *Tractatus*, 6.36311.

Challenge
** Sections entitled 'curiosities' are collections of different topics. – By the way, do yo agree with the quotation?

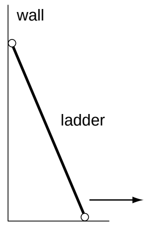


Figure 25 How does the ladder fall?

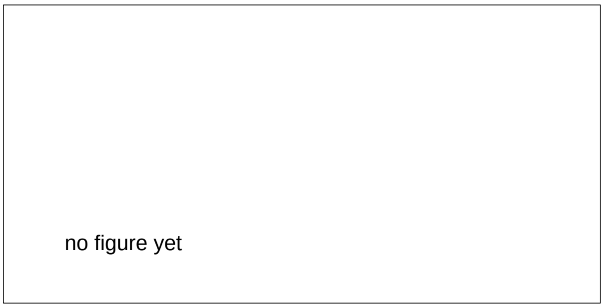


Figure 26 Observation of sonoluminescence with a diagram of the experimental set-up

• The highest microscopic accelerations are observed in particle collisions, where one gets values up to 10^{35} m/s^2. The highest macroscopic accelerations are probably found in the collapsing interiors of *supernovae*, the exploding stars which can be so bright to be visible in the sky even during day time, and on earth in the interior of collapsing bubbles, in what is called *sonoluminescence*. This latter effect appears when air bubbles in water are expanded and contracted by underwater loudspeakers at around 30 kHz. At a certain threshold intensity, the bubble radius changes with 1500 m/s in as little as a few μm, giving an acceleration of several 10^{11} m/s^2. Ref. 61

• If a canon located at the equator shoots a bullet in vertical direction, where does the bullet fall back? Challenge

• Is traveling through interplanetary space healthy? People often fantasize about long trips through the cosmos. Experiments have shown that on trips of long duration, cosmic radiation, bone weakening, and muscle degeneration are the biggest dangers. Many medical experts question the viability of space travel lasting longer than a couple of years. Other dangers are rapid sunburn, at least near the sun, and the vacuum. So far it happened to a man only once to find himself unprotected in vacuum. He lost consciousness after 14 seconds, but survived. Ref. 62

• How does the kinetic energy of a rifle bullet compare with that of a running man? Challenge

• How does a flame inside a jar change inclination from the vertical on a rotating turntable? Challenge

• A ping-pong ball is attached with a string to a stone, and the whole is put under water in a jar. The jar is moved. In which direction does the ball move? Challenge

• What happens to the size of an egg when one places it into a jar of vinegar for a few days? Challenge

• The *smallest experimentally probed distance*, in 1996, was 10^{-19} m, for quarks at Fermilab. What does this mean for the continuity of space? Ref. 63 Challenge

• Does centrifugal acceleration exist? Most students at university go through the shock of meeting a teacher saying that it doesn't because it is a 'fictitious' quantity, in the face of what one experiences every day in the car when driving around a bend. Simply ask the teacher who denies it to define 'existence'. (The definition physicists usually use is given in the intermezzo following this chapter.) Then check whether the definition applies to the term and make up your own mind. Challenge

• What is the best way to transport full coffee or tea cups while at the same time avoiding to spill any precious liquid? Challenge

• The moon recedes from the earth by 3.8 cm a year, due to friction. Can you find the responsible mechanism? Challenge

• What is the amplitude of a pendulum oscillating in such a way that the absolute value of its acceleration in the lowest point and in the return point is equal? Challenge

• Is it correct that the value of the acceleration of a drop of water falling through vapour is $g/7$? Challenge

Ref. 64 • In athletics, ong jumpers take off at angles of about 20°, as they are not able to achieve a higher angle at the speed of 9.5 m/s. How much would they gain if they could achieve 45°? Challenge

Legs or wheels? – again

The acceleration and deceleration of standard cars is never much higher than about $1g = 9.8\,\mathrm{m/s^2}$, the acceleration due to gravity on our planet. That is not a big feat. Higher accelerations are achieved by motor bikes and racing cars by the use of suspensions diverting weight to the axes and by the use of spoilers, so that the car is pushed downwards with more than its own weight. Modern spoilers are so efficient in producing a force pushing towards the track that racing cars could race on the roof of a tunnel without falling down. Through the use of special tires these downwards forces are transformed into grip; modern racing tires allow forward, backwards and sideways accelerations (necessary for acceleration, for braking and for turning corners) of about 1.1 to 1.3 times the load. Engineers once believed that a factor 1 was a theoretical limit and this limit is still sometimes found in textbooks; but advances in tire technology, mostly by making clever use of interlocking between the tire and the road surface as in a gear mechanism, have allowed to achieve these higher values. The highest accelerations, around $4g$, are achieved when part of the tire melts and glues to the surface. Special tires designed to make this happen are used for dragsters or for high performance radio controlled model cars.

How do all these efforts compare to legs? High jump athletes can achieve peak accelerations of about 2-4 times g, bushbabies up to $13g$, locusts about $18g$, and fleas have been measured to accelerate about $135g$. The maximum acceleration known for animals is that of click beetles, a small insect able to accelerate at over $2000\,\mathrm{m/s^2} = 200g$, about the same as an airgun pellet when fired. Legs are thus definitively more efficient accelerating devices than wheels, and evolution also developed them, instead of wheels, to improve the chances of an animal in danger to get to safety. In short, legs *outperform* wheels. Ref. 65

Challenge There other reasons to use legs instead of wheels. (Can you name some?) For example, legs, in contrast to wheels, allow to walk on water. Most famous for this ability is the *basilisk*,* a lizard living in Central America, about 50 cm long and with a mass of about 90 g. This reptile, looking like a miniature running Tyrannosaurus Rex, can actually run over water surfaces on his back legs. The motion has been studied in detail with high speed film cameras and by measurements using aluminium models of the animal's feet. The lizards run by pushing down the legs onto the water very rapidly. But the experiments shows that the slapping on the water provides only 25% of the needed force to run above water; the other 75% are provided by a pocket of compressed air that the basilisks create below their feet when these are in the water. In fact, basilisks mainly walk on air. It was calculated that a human can also walk on water, provided his feet hit the water with a speed of 100 km/h using the simultaneous physical power of 15 sprinters. Quite a feat for all those who ever did it. Ref. 66

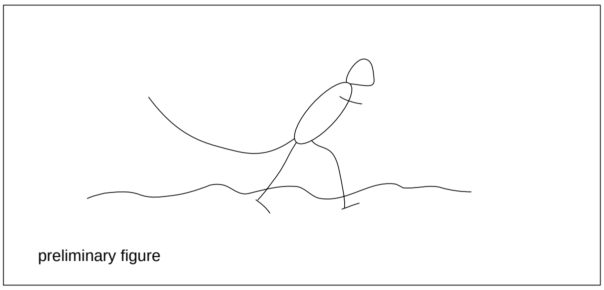


Figure 27 A basilisk lizard (Basiliscus basiliscus) running over water

After this short overview into motion based on contact, let us continue with the study of motion transmitted over distance, without any contact at all. It is easier and simpler to study.

Gravitation

Caddi come corpo morto cade.
Dante, *Inferno*, c. V, v. 142.**

The first contact-free method to generate motion one discovers is *height*. Waterfalls, snow, rain, and falling apples all rely on it. It was one of the fundamental discoveries of physics that height has this property because there is an interaction between every body and the earth, called *gravitation*, producing an acceleration along the line connecting the centers of mass of the two. Note that in order to make this statement, one has to realize that the earth is a body, in the same way as a stone or the moon, that this body is finite, and that therefore it has a center and a mass. Today, these statements are common knowledge, but they are by no means evident from personal everyday experience.***

Many years of observations of the movement of the moon and the planets were collected by numerous astronomers, but in particular by the dane Tycho Brahe (1546–1601), who organized an industrial search for astronomical facts sponsored by his king. He consumed almost 10% of the danish gross national product at doing so. His measurements were the basis for the research of the swabian astronomer Johannes Kepler (1571–1630), which in turn lead to an astonishingly simple result the english physicist Robert Hooke (1635–1703)

* In the middle ages, a basilisk was a mythical monster connected with the end of the world. Today, it is a small reptile in the Americas.

** 'I fell like dead bodies fall.' Dante Alighieri (1265, Firenze–1321, Ravenna), the powerful italian poet.

*** In several myths about the creation or the organization of the world, such as the biblical one, the earth is not an object, but an imprecisely defined entity, such as an island floating or surrounded by water with unclear boundaries or suspension method. Are you able to convince a friend that the earth is round and not flat? Can you find another argument apart from the fact that its shadow is round when it is visible on the moon? If the earth is round, the top of two buildings is further apart than their base. Can this effect be measured?

Challenge
Challenge

formulated in 1684: every body of mass M attracts any other body towards its centre with an acceleration whose magnitude a is given by

$$a = G\frac{M}{r^2} \tag{28}$$

where r is the center-to-center distance of the two bodies. This is called the "*universal law of gravitation*" for reasons to be explained shortly. If bodies are small compared to the distance r, or if they are spherical, the expression is correct as it stands, as shown by Newton; otherwise, i.e. for non-spherical shapes, the acceleration has to be calculated separately for each part of the bodies and then added together.

This inverse square property is often called Newton's "law", because the english physicist Isaac Newton (1642–1727) proved more elegantly than Hooke that it agreed with all astronomical observations. Above all however, he organized a better public relation campaign, in which he claimed to be the originator of the idea. Ref. 67

Most importantly, Newton showed that this description of astronomical motion also gives the correct description for the stones one throws through the air, down here on father earth. How did he do this? He compared the acceleration a_m of the moon with that of stones g. For the ratio between these two accelerations, the inverse square relation predicts a value $a_m/g = R^2/d_m^2$, where R is the radius of the earth. The moon's distance d_m can be measured by triangulation, i.e. by comparing the position of the moon against the starry background from two different points on earth.* The result is $d_m/R = 60 \pm 3$, depending on the orbital position of the moon, so that one predicts the average ratio $a_m/g = 3.6 \cdot 10^3$.

But both accelerations can also be measured directly. On the surface of the earth, stones feel an acceleration due to gravitation with magnitude $g = 9.8\,\mathrm{m/s^2}$, as determined by measuring the time stones need to fall a given distance. For the moon, the definition of acceleration, $a = dv/dt$, in the case of circular motion – roughly correct here – gives $a_m = d_m(2\pi/T)^2$, where $T = 2.4\,\mathrm{Ms}$ is the time the moon takes for an orbit around the earth.** The measurement of the radius of the earth*** yields $R = 6.4\,\mathrm{Mm}$, so that the average moon-earth distance is $d_m = 0.38\,\mathrm{Gm}$. One thus has $a_m/g = 3.6 \cdot 10^3$, in agreement

* This was realized for the first time in 1752 by the french astronomers Lalande and La Caille, by simultaneously measuring the position of the moon in Berlin and in Le Cap.

** This is deduced easily by noting that for an object in circular motion, the magnitude v of the velocity $\mathbf{v} = d\mathbf{x}/dt$ is given as $v = 2\pi r/T$. From a drawing of the vector $\mathbf{v}$ over time, the so-called *hodograph*, one sees that it behaves exactly like the position of the object. Therefore the magnitude a of the acceleration $\mathbf{a} = d\mathbf{v}/dt$ is given by the same expression, viz. $a = 2\pi v/T$. Challenge

*** This is the hardest quantity to measure oneself. The most surprising way to determine the earth's size is the following: watch a sunset, near a house, with a stopwatch in hand. When the last ray of the sun disappears, start the stopwatch and run upstairs. There, the sun is still visible; stop the stopwatch when the sun disappears again and note the time t. Measure the height distance h of the two eye positions where the sun was observed. The earth radius R is then given by $R = kl/t^2$, with $k = 378\,\mathrm{s}^2$. Ref. 68 Challenge

Ref. 69 There is also a simple way to measure the distance to the moon, once the size of the earth is known. Take a photograph of the moon when it is high in the sky, and call θ its zenith angle, i.e. its angle from the vertical. Make another photograph of the moon a few hours later, when it is just above the horizon. On this picture, contrary to a common optical illusion, the moon is smaller, because it is further away. (With a drawing the reason for this becomes clear immediately.) If q is the ratio of the two moon diameters, the earth-moon distance d_m is given by the relation $d_m^2 = R^2 + [2Rq\cos\theta/(1-q^2)]^2$. (Enjoy its derivation from the drawing.) Challenge

Another possibility is to determine the size of the moon by comparing it to the size of the shadow of the earth during an eclipse. The distance to the moon is then computed from its angular size, about 0.5°.

with the above prediction. With this famous 'moon calculation' one has thus shown that the inverse square property of gravitation indeed describes both the motion of the moon and that of stones.

From the observation that on earth all motion eventually comes to rest, whereas in sky all motion is eternal, Aristoteles and many others had followed that motion in the sublunar world follows different "laws" than motion in the translunar world. Several thinkers had criticized this distinction, notably the french philosopher and rector of the University of Paris, Jean Buridan. * The moon calculation was the most important result showing this Ref. 70 distinction to be wrong. This is the reason for calling the expression (28) the *universal* "law" of gravitation.

This result allows to answer another old question. Why does the moon not fall from the sky? Well, the preceding discussion showed that *fall* is motion due to gravitation. Therefore the moon actually *is* falling, with the particularity that instead of falling *towards* the earth, it is continuously falling *around* it. The moon is continuously missing the earth.**

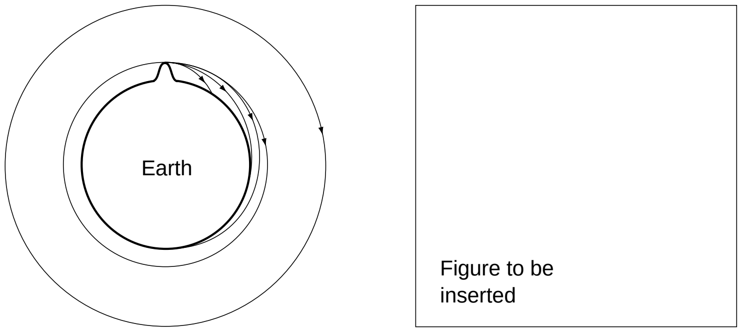


Figure 28 A physicists and an artists view of the fall of the moon: a graph by Christiaan Huygens a marble by Auguste Rodin

Properties of gravitation

If all objects attract each other, that should also be the case for everyday life objects. Gravity must also work *sideways*. This is indeed the case, even though the effects are so small that they were measured only long after universal gravity had predicted them. For example, two average people at the close distance of 0.1 m feel an acceleration towards each other which is smaller than that exerted by the leg of a common fly on the skin. Therefore one usually Challenge does not notice the attraction to other people. When one notices it, it is much stronger than that. This simple calculation thus proves that gravitation cannot be at the origin of people falling in love, and that sexual attraction is not of gravitational, but of different origin. This other interaction will be studied later in our walk; it is called *electromagnetism.*

Another consequence of gravitation is that the path of a stone is not a parabola, as stated earlier, but actually an ellipse around the center of the earth. This happens for exactly the same reason that makes the planets move in ellipses around the sun. Are you able to confirm Challenge

∗ Jean Buridan (ca. 1295– ca. 1366) was also one of the first modern thinkers to speculate about a rotation of the earth around an axis.

∗∗ Another way to put it is to use the answer of the dutch physicist Christiaan Huygens (1629–1695): the moon does not fall from the sky because of the centrifugal acceleration. As explained on page 72, this explanation is nowadays out of favor at most universities.

Ref. 71 Challenge There is a beautiful problem connected to the left part of the figure: Which points of the surface of the earth can be reached by shooting off from a mountain? And which points can be reached by shooting off horizontally?

this?*

Universal gravitation allows to lift a mystery. The puzzling acceleration value $g = 9.8\,\mathrm{m/s^2}$ we encountered in equation (3) is thus due to the relation

$$g = GM_{\text{earth}}/R^2_{\text{earth}} \quad . \qquad (29)$$

It can be deduced from equation (28) by taking the earth to be spherical. Obviously, the value for g is almost constant on the surface of the earth, because the earth is almost a sphere. The expression also explains why g is smaller if one rises in the air, and the deviations of the shape of the earth from sphericity explain why g is different at the poles and larger if one rises on a mountain.

Note that 9.8 is roughly π^2. This is *not* a coincidence: the meter has been chosen in such a way to make this correct. It is true that the original definition of the meter in the 18th century was as $1/40,000,000$ of the circumference of the earth. But this definition was chosen because it was almost identical to – but much more precise than – the definition that a meter should be that length for which a pendulum would make one swing in one second.

Challenge The period of a swinging pendulum, i.e. a back and forward swing, is given by**

$$T = 2\pi\sqrt{\frac{l}{g}} \qquad (30)$$

If one had defined the meter such that $T/2 = 1\,\mathrm{s}$, the value of the normal acceleration g would have been exactly $\pi^2\,\mathrm{m/s^2}$.*** This first proposal in 1790 by Talleyrand was rejected by the conference which defined the meter because variations in the value of g with geographical position and in the length of a pendulum with varying temperature induce errors which are too large to give a useful definition, and much larger than those using the circumference of the earth. Thus the meridian definition of the meter was adopted by the french national assembly on the 26 March 1791, with the statement that “a meridian passes under the feet of every human being, and all meridians are equal.”

By the way, it is possible to devise a simple machine, other than a yo-yo, which slows down the acceleration of gravity by a known amount, so that one can measure its value more easily. Can you imagine it? Challenge

But one still can go on asking: Why does the earth have the mass and size it has? And why does G have the value it has? The first question asks for a history of the solar system; this is still a topic of research and lies outside the scope of this book. The second question is addressed in appendix B.

The gravitational constant G is important also for more mundane reasons. Without it, it is impossible to determine the mass of the *earth*. In 1798, the english physicist Henry

* By the way, the hodograph of an elliptic orbit is a circle. In fact, in velocity space, elliptic, parabolic and hyperbolic motions are all described by circles. Ref. 72

** The formula (30) is noteworthy because it expresses that the period does *not* depend on the amplitude. (This is true as long as the oscillation angle is smaller than about 30 degrees.) Galileo discovered this as a student, when observing a chandelier hanging on a long rope in the dome of Pisa. Using his heartbeat as a clock he found that even though the amplitude of the swing got smaller and smaller, the time for the swing stayed the same.

*** A leg also moves like a pendulum, when one walks normally. Why then do taller people tend to walk faster? Challenge

Cavendish (1731–1810) was the first to measure G; therefore he called the result of his experiments "weighing the earth". Are you able to imagine how he did it? As a hint, his experiments were also the first to confirm that gravity works sideways as well. Challenge

But gravity has more interesting properties to offer. The effects of gravitation can also be described by another observable, namely the *(gravitational) potential* φ. One the has the simple relation that the acceleration is given by the *gradient* of the potential

$$\mathbf{a} = -\nabla\varphi \qquad \text{or} \qquad \mathbf{a} = -\operatorname{grad}\varphi \tag{31}$$

where the abbreviation ∇ for the gradient, pronounced 'nabla', is defined as the vector $\nabla\varphi = (\partial\varphi/\partial x, \partial\varphi/\partial y, \partial\varphi/\partial z) = \operatorname{grad}\varphi$. The minus sign in the above definitions is introduced by convention, in order to have a higher potential at larger height.* For a pointlike or a spherical body of mass M, the potential is

$$\varphi = -G\frac{M}{r} \tag{32}$$

For a general, extended body, the potential is found by requiring that the *divergence* of its gradient is given by the mass (or charge) distribution times some proportionality constant. More precisely, one has

$$\Delta\varphi = 4\pi G\rho \tag{33}$$

where $\rho = \rho(\mathbf{x}, t)$ is the mass volume density of the body and the operator Δ, pronounced 'delta', is defined through $\Delta f = \nabla\nabla f = \partial^2 f/\partial x^2 + \partial^2 f/\partial y^2 + \partial^2 f/\partial z^2$. Equation (33) is called the *Poisson equation* for the potential φ,** and the positions where ρ is not zero are called the *sources* of the potential. A potential simplifies considerably the description of motion, since a potential is additive: given the potential of a point particle, one can calculate the potential and then the motion around any other, irregularly shaped object. It is then easy to calculate the motion of a comet near the sun and all the planets. Note that not all accelerations can be derived from a potential; this is a special property of gravitation and some other interactions. In this case one speaks of *conservative* systems. Note also that the number of dimensions of space d is coded into the potential: its dependence on the radius r is in fact r^{d-2}. The exponent $d-2$ has been checked experimentally to high precision and no deviation of d from 3 has ever been found. Challenge Ref. 73

Note that the source $\Delta\varphi$ of a function is a measure for how much the function $\varphi(x)$ at a point x differs from the average value in a region around that point. (Can you show this, by showing that $\Delta\varphi \sim \bar{\varphi} - \varphi(x)$?) In other words, the Poisson equation (33) implies that the actual value of the potential at a point is the same as the average value around that point minus the mass density multiplied by $4\pi G$. In particular, in the case of empty space the potential at a point is equal to the average of the potential around that point. Challenge

Why is the potential φ such an interesting quantity? First of all, with a single quantity on can describe the vector aspects of gravitational acceleration. With a single function it

* In two or more dimensions slopes are written $\partial\varphi/\partial z$ – where ∂ is still pronounced 'd' – because in those cases the expression $d\varphi/dz$ has a slightly different meaning; but the details lie outside the scope of this walk.

** Siméon-Denis Poisson (1781–1840), eminent french mathematician and physicist.

describes that gravity in New Zealand acts in the opposite direction to gravity in Paris. Secondly, it allows to introduce the so-called *potential energy*

$$U = m\varphi \tag{34}$$

which allows to determine the change of *kinetic* energy of a falling body from a point 1 to a point 2 via

$$T_1 - T_2 = U_2 - U_1 \quad \text{or} \quad \tfrac{1}{2}m_1\mathbf{v}_1{}^2 - \tfrac{1}{2}m_2\mathbf{v}_2{}^2 = m\varphi_2 - m\varphi_1 \quad . \tag{35}$$

In other words, the *total energy*, defined as the sum of kinetic and potential energy, is *conserved* in motion due to gravity. Later we will see that the corresponding sum is conserved for all other interactions as well. We'll come back to this topic in more detail shortly.

Often, the concept of *gravitational field* is also introduced. We avoid this in our walk, because later we will discover that the theory of relativity shows that gravity is not due to a field at all; in fact even the concept of gravitational potential turns out to be only an approximation.

The concept of potential helps to understand a few important aspects of the earth. Since most of the earth is still liquid when seen on cosmic scale, the *shape* of the earth is such that
Ref. 74 its surface is always horizontal with respect to the direction determined by the combination of the accelerations of gravity and and of rotation. In short, the earth is *not* a sphere. The
Ref. 75 mathematical shape following from this requirement is called a *geoid*. (That shape is given approximately, with an error of less than 50 m, by an ellipsoid.) The geoid is an excellent approximation to the actual shape of the earth. Sea level is different from it only by less than twenty meters. These differences can be measured with satellite radar and are of great interest to geologists and geographers. For example, it turns out that the south pole is nearer to the equator plane than the north pole. The difference, about 30 m, is probably due to the large land masses in the northern hemisphere.

The inertia of matter (the so-called "centrifugal force") increases the radius of the earth at the equator; in other words, the earth is flattened at the poles. The equator has a radius a of 6.38 Mm, whereas the distance b from the poles to the centre of the earth is 6.36 Mm.
See appendix B The flattening $(a-b)/a$ has the value $1/298.3 = 0.0034$. The flattening of the earth implies that the top of mount Chimborazo in Ecuador, even though its height is only 6267 m above sea level, is about 20 km farther away from the centre of the earth than the top of mount Qomolangma (sometimes also called mount Everest) in Nepal, whose height above sea level is about 8850 m.

If the earth would stop to rotate, the water of the oceans would flow north; all of Europe would be under water, except for the few mountains of the Alps higher than about 4000 m. The northern parts of Europe would be between 6 km and 10 km under water; and mount Qomolangma would be over 11 km above sea level. By the way, if the earth would stop to rotate, the deflattening of the earth would also produce extremely strong earthquakes and storms. As long as these effects are lacking, we are *sure* that the sun will indeed rise
See page 71 tomorrow, despite what philosophers might pretend.

Gravitation in the sky

Taking the sun as gravitating mass, the expression $a = Gm/r^2$ also describes the motion of all the planets around the sun. Everybody can check for himself that the planets always stay within the *zodiac*, a narrow stripe across the night sky. The center line of the zodiac, on which the sun moves, is called the *ecliptic* since the moon has to be located on it to produce an eclipse. But the detailed motion of the planets is not easy to describe.* A few generations before Hooke, the swabian astronomer Johannes Kepler** had deduced several "laws" in his painstaking research about the the movements of the planets in the zodiac. The three main ones are:

(1604) Planets move on ellipses with the sun located at a focus;

(1604) Planets sweep equal areas in equal times;

(1618) The duration of a planet's orbit is related to its semimajor axis d by $T^2 \sim d^3$.

Can you show that all three laws follow from the expression of universal gravity? Challenge

Kepler and Newton became famous because they brought order in the description of the observations of the planets. This achievement of such small practical significance was so publicized because of the age-old prejudices linked to astrology.

But there is also another reason to value the result: the sheer work required to deduce it was enormous. Kepler had no calculation machine available at his time, not even a slide ruler. The calculation technology he used were the recently discovered logarithms. Whoever has ever used them and the tables that come with them to actually perform calculations can get a feeling for the large amount of work behind these three discoveries.

One should also remember that Newton was not able to write down, let even to handle differential equations at the time he published his results on gravitation. In fact his notation and calculation methods were poor; the english mathematician G.H. Hardy (1877–1947) used to say that insistence on Newton's integral and differential notation – instead of using the one of his rival Leibniz, common today – threw back english mathematics by 100 years. Ref. 15

Moreover, gravitation explains the motion and shape of the milky way and of the other galaxies, the motion of many weather phenomena, and explains why the earth has an atmosphere but the moon does not. (Can you?) Challenge

The moon is actually worth a few remarks on its own.

One often hears that the moon always shows the same side to the earth. But this is wrong. As one can check with the naked eye, a given feature in the centre of the face of the moon at full moon is not at the centre one week later. The various motions leading to this change are called *librations*; they result from the fact that the moon does not describe a circular, but an elliptic orbit around the earth, due to the fact that the axis of the moon is slightly inclined compare to that of its rotation around the earth and due to some smaller effects.

The moon seems to originate from earth itself: in the old times, an object hit the earth almost tangentially, and threw much material up in the sky. This is the only mechanism able

* The apparent height of the ecliptic changes with the time of the year and is the reason for the difference between seasons. Therefore seasons are gravitational effects as well.

** Johannes Kepler (1571, Weil der Stadt–1630); after helping his mother to defend herself in a trial where she is accused to be a witch, he studies protestant theology, and becomes teacher of mathematics, astronomy and rhetorics. His first book on astronomy makes him famous, he becomes pupil of Tycho Brahe and then, at his teacher's death, the imperial mathematician. He is the first to use mathematics in the description of astronomical observations, and introduces the concept of 'celestial physics'.

to explain the large size of the moon, and its anemia, i.e. its low iron content, as well as its general material composition.
Ref. 76

The moon is receding from the earth at 3.8 cm a year. This result confirms the old deduction that the tides slow down earth's rotation. Can you imagine how this measurement was performed?* Since the moon slows down the earth, the earth also changes shape due to this effect. (Remember that the shape of the earth depends on its rotation speed.) These changes in shape have influence on the tectonic activity of the earth, and maybe also in the drift of the continents.
Challenge

The moon has many effects on animal life. In 1995, D. Neumann published a study about a type of midge, of the genus Clunio, which lives on sea coasts with pronounced tides. Clunio lives between 6 and 12 weeks as a larva, then hatches and lives only one or two hours as adult insect, during which it reproduces. The reproduction is only successful if the adult hatches during the low tide phase of a spring tide. Spring tides are the especially strong tides during full and new moon, when the solar and lunar effects add, and occur only every 14.8 days. Neumann showed that the larvae have two built-in clocks, a circadian and a circalunar one, which control the hatching to precisely those few hours when the insect can reproduce. He also showed, among many other details, that the circalunar clocks is synchronized by the brightness of the moon at night. In other words, the larvae watch the moon at night and then decide when to hatch.
Ref. 77

If insects can have circalunar clocks and cycles, it should not be surprising that women also have such a cycle. The question is not yet settled.

The gravitation of the moon also helps to deflect asteroids from the earth, as its numerous craters show. The gravitational attraction of the moon saved the life of humans already many times over.

The moon also helps to stabilize the tilt of the earth's axis. If the axis were not more or less parallel to that of the motion around the sun, we would not have a regular day and night rhythm, and the evolution of life would have been made impossible.
Ref. 78

How often do planets have moon-sized moons? This question is important for the estimation of the probability that life exists on other planets. We have seen that the moon is necessary to create a large magnetic field on earth shielding us from cosmic radiation and to capture comets saving us from impacts. If the moon did not exist, the earth would rotate much faster and we would have a much more unfriendly weather.
Ref. 79 Ref. 80

Gravitation also implies that comets return regularly. The english astronomer Edmund Halley (1656–1742) was the first to take this conclusion, and to predict the return of a comet. It arrived at the predicted date in 1756, and is now called after him. This result finally settled a long dispute on whether comets were heavenly bodies or only images on the horizon. The period of Halley's comet is between 74 and 80 years; the first recorded sighting was 22 centuries ago, and it has been seen at every one of its thirty passages since, the last time in 1986.

See page 58
We have seen that mass is always positive and that gravitation is thus always attractive; there is no *antigravity*. Can gravity be used for *levitation* nevertheless? Yes; there are two

∗ If you want to read about the motion of the moon in all its fascinating details, have a look at Martin C. Gutzwiller, *Moon-earth-sun: the oldest three body problem*, Reviews of Modern Physics **70**, pp. 589–639, 1998.

examples.** The first one, of great practical importance, are the geostationary satellites used for easy transmission of television and other signals from and towards earth.

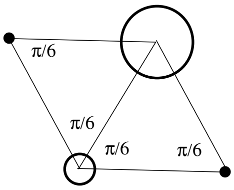


Figure 29 The lagrangian points

The second example are the *lagrangian libration points*; named after their discoverer, these are points in space near a two body system, such as moon-earth or earth-sun, in which small objects have a stable equilibrium position. Can you find them? There are also three additional unstable lagrangian points. There are thousands of asteroids, called *Trojan asteroids*, at and around the lagrangian points of the Sun-Jupiter system. Challenge

In 1990, a Trojan asteroid for the Mars-Sun system was found, and in 1997, an asteroid was found which follows the earth in its way around the sun. It is an asteroid of 5 km in diameter. Moreover, on the main lagrangian points of the earth-moon system a high concentration of dust has been observed. Ref. 81

In summary, the single equation $\mathbf{a} = -GM\mathbf{r}/r^3$ describes correctly a large number of phenomena. The first person to make clear that the expression was also sufficient to describe *everything* happening in the sky was Pierre Simon Laplace (1749, Beaumont–en-Auge-1827, Paris), in his famous *Mécanique céleste*, which appeared in 5 volumes between 1798 and 1825.

These results are quite a feat for such a simple expression. How precise is it? Astronomy allowing the most precise measurements, also it provides the most stringent tests. In 1846, the observed deviations of the motion of the planet Uranus from this law led to the prediction of another planet, Neptune, which was discovered shortly thereafter. In 1882, Simon Newcomb (1835–1909) concluded after intensive study that there was only one known example of discrepancy from universal gravity, namely for the planet Mercury. The direction of the long axis of the orbit of planet Mercury, the perihelion, changes with a rate slightly smaller than the predicted one: the difference is around 43 arc seconds per century. The study of motion had to wait for Albert Einstein to find the solution of this discrepancy. Ref. 82

Tides

Why do physics text always talk about something as boring as tides? Because, as general relativity shows, tides prove that space is curved! It is thus useful to study them a bit.

Indeed, gravitation also describes the sea *tides* as results of the attraction of the ocean water by the moon and the sun. Tides are fun; even though the amplitude of the tides is only about 0.5 m on the open sea, it can be up to 20 m at special places near the coast. Can you imagine why? Also the *soil* is lifted and lowered by the tides, by an amount of the order of 0.3 m, as satellite measurements show. Even the *atmosphere* is subject to tides, and the corresponding pressure variations can be filtered out from the weather pressure measurements. Challenge

In fact, all these tidal flows, deformations and stresses appear for any *extended* body moving in a gravitational field of a second one. To understand the origin, one can take a body in orbit, like the earth, and imagine its components, such as the five segments of figure 30, as kept together by springs. The segment on the outside of the orbit is *pulled* by the

** Levitation is discussed in detail on page 346.

rest of the body, as it would like to orbit more slowly, being more distant from the sun. The inside segment is *retained* by the rest, as it wants to orbit more rapidly. As a result, the inside segment wants to get near the sun, and the outside segment wants to get away from it: both segments feel a pull away from the centre of the body, until the inner springs say stop. In short, *extended bodies are deformed in direction of the field inhomogeneity.*

For example, as a result of tidal forces, the moon points always with (almost) the same face to the earth; and indeed, its radius towards the earth is larger by about 30 km than the radius perpendicular to it.

If on the other hand the inner springs are too weak, the body is torn into many small ones. In such a way a ring can appear, such as the asteroid ring or the rings around Saturn.

Of course, the earth orbits – no misprint – two bodies: the sun and the moon. Using the numbers from appendix B, the gravitational acceleration from the sun and the moon are

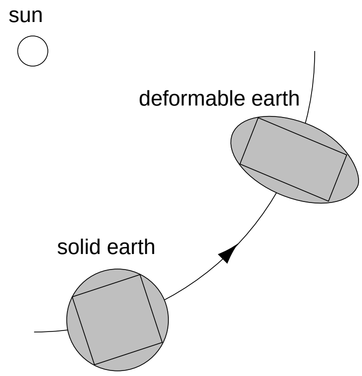


Figure 30 Tidal deformations due to gravity

$$a_{\text{sun}} = \frac{GM_{\text{sun}}}{d_{\text{sun}}^2} = 5.9\,\text{mm/s}^2$$
$$a_{\text{moon}} = \frac{GM_{\text{moon}}}{d_{\text{moon}}^2} = 0.033\,\text{mm/s}^2 \qquad (36)$$

which shows that the attraction of the moon is about 178 times weaker than that from the sun.

Now the *relative* acceleration $b = \nabla a$ of two nearby bodies falling together near a large spherical mass M is given by

$$b_{\text{rel}} = \frac{-2GM}{r^3} \qquad (37)$$

which gives the values

$$b_{\text{sun,rel}} = \frac{-2GM}{d_{\text{sun}}^3} = -0.8 \times 10^{-13}\,/\text{s}^2$$
$$b_{\text{moon,rel}} = \frac{-2GM}{d_{\text{moon}}^3} = -1.7 \times 10^{-13}\,/\text{s}^2 \qquad (38)$$

In other words, despite the much weaker pull of the moon, its tides are predicted to be over twice as *strong* as the tides from the sun, as indeed is observed.

Of course, tides also produce friction. The friction leads to a slowdown of earth's rotation. Nowadays, the slowdown can be measured by precise clocks (even though short time variations are often larger). The result fits well with fossil results showing than 400 million years ago, in the devon, a year had 400 days, and a day 22 hours.

As we saw, tides are due to relative accelerations of nearby objects. In the chapter on general relativity we will find that time and length divided by the speed of light play a similar role. Using this similarity, relative acceleration measures, when translating from

See page 205

time to space, something like relative distance increase with length. Two nearby particles falling together then correspond to two parallel lines. If the relative distance increase with length is zero, parallel lines stay parallel. If the relative distance increase is larger or smaller, the lines diverge or approach each other, as nearby falling particles do. But parallel lines can diverge or approach each other *only* if space (and thus space-time) is *curved*. Even though this simple reasoning could have been performed already in the 18th century, it took until the early 20th century and Albert Einstein's work to realize it. See page 234

Can light fall?

Once people found out that light has a finite velocity – a story which we will tell in detail later on – they of course started to ask how light *falls*. The value of the speed of light was actually known from the mid 17th century. See page 145

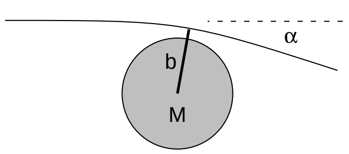


Figure 31 Masses bend light

Does the speed increase when light reaches the surface of the earth? For almost three centuries people had no measurement machines to detect any effect, so that the question was not investigated. Then, in 1801, the prussian astronomer Johann Soldner was the first to put the question in a different way. Being an astronomer, he was used to measure stars and their observation angles. He calculated that due to gravity, light passing *near* near a massive body would be *deflected*. Ref. 83

For a body on a hyperbolic path, moving with velocity c past a body of mass M at distance b, one gets a deflection angle Challenge

$$\alpha_{\text{univ.grav.}} = \frac{2}{b}\frac{GM}{c^2} \qquad (39)$$

In his time, this angle was far too small to be measured even for the mass of the sun, where it turns out to be a tiny $0.88'' = 4.3\,\mu$rad, and the issue was forgotten. Had it be pursued, general relativity would have started as an experimental science, and not as a theoretical effort by Albert Einstein! Why? The value so calculated is *different* from the measured one. And the reason is not easy to find. By the way, how would you measure the deflection of light near the bright sun? See page 228 Challenge

What is mass? – again

Mass describes how an object interacts with others. In our walk, we have encountered two of its aspects. *Inertial mass* is the property which keeps objects moving and which offers resistance to change of their motion. *Gravitational mass* is the property responsible for the acceleration of bodies nearby (the active aspect) or of being accelerated by objects nearby (the passive aspect). For example, the active aspect of the mass of the earth determines the surface acceleration of bodies; the passive aspect of the bodies allows one to weigh them in order to measure their mass using distances only, e.g. on a scale or a balance. The gravitational mass is the basis of *weight*, the difficulty to sustain something.*

Challenge * What is the value shown by balance for a person of 85 kg juggling with three balls of 0.5 kg each?

One then can ask whether the gravitational mass of a body is equal to the inertial mass. A rough idea is given by the experience that an object which is difficult to move is also difficult to lift. The simplest precision experiment is to take two bodies of different mass and let them fall. If the acceleration is the same for all bodies, inertial mass is equal to (passive) gravitational mass, because in the relation $ma = \nabla(GMm/r)$ the left m is actually the inertial mass, and the right m is actually the gravitational mass.

But already in the 17th century Galileo showed without a single experiment that the acceleration indeed is the same for all bodies. If e.g. larger masses would fall more rapidly than smaller masses, as one could imagine, one would get the following situation. Any body can be seen as composed from a large fragment attached to a small fragment. If small bodies really would fall less rapidly, the small fragment would slow the large fragment down, so that the complete body would have to fall *less* rapidly than the larger fragment (or even break into pieces). But at the same time, the body being larger than its fragment, it should fall *more* rapidly than that fragment. This is obviously impossible: all masses must fall with the same acceleration.

Many accurate experiments have been performed after Galileo's original discussion. In all of them the independence of the acceleration of free fall on mass and material composition Ref. 84 has been confirmed with the precision expected. In other words, as far as we can tell, the gravitational mass and the inertial mass are identical. What is the origin of this mysterious equality?

This so-called mystery is a typical example of student disinformation, now common See page 55 across the whole world of physics education. Let us go back to the definition of mass as negative inverse acceleration ratio. It was mentioned there that the physical origin of the accelerations do not play a role in the definition. In other words, the value of the mass is by definition independent of the interaction. That means in particular that inertial and gravitational mass are identical *by definition*.

Ref. 85 Another beautiful proof of this statement was given by A.E. Chubykalo, and S.J. Vlaev. They show that the total kinetic energy T of two bodies, like the earth and the moon, circling around their common centre of mass, is given by $T = GmM/2R$, where the two quantities m and M are the *gravitational* masses of the two bodies and R their distance. From this expression, in which the inertial masses do *not* appear, they prove that the inertial and gravitational Challenge mass must be proportional to each other. Can you see how?

No wonder that all measurements confirm this result. The argument for the equality of inertial and gravitational mass reappears in general relativity, as explained page 211. But what is the *origin* of mass? Why does it exist? This simple but deep question cannot be answered by classical physics. We will discover the answer later during our walk.

Curiosities about gravitation

Fallen ist weder gefährlich noch eine Schande;
Liegen bleiben ist beides.*
Konrad Adenauer

* 'Falling is neither dangerous nor a shame; keep lying is both.' Konrad Adenauer (1876, Köln–1967, Rhöndorf), german chancellor.

• It is easy to put a book on a table. Surprisingly, it is equally easy to put the whole earth on a table. Are you able to do it? Challenge

• Do all objects fall with the same acceleration of 9.8 m/s^2, assuming that air resistance can be neglected? No; every housekeeper knows that. You can check this by yourself. A broom angled at around 35 degrees hits the floor earlier than a stone, as the impact noises tell. Are you able to explain why? Challenge

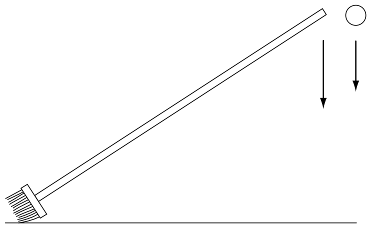


Figure 32 Brooms fall more rapidly than stones

• Does every spherical body fall with the same acceleration? No. If the weight of the object is comparable to that of the earth, the distance decreases in a different way. Can you confirm the statement? What then is wrong about Galileo's argument about the constancy of acceleration of free fall? Challenge

• When one runs to the east, one *loses weight*. There are even two different reasons for this: the "centrifugal" acceleration increases and thus the force with which we are pushed down diminishes, and the Coriolis force appears, with a similar result. Can you estimate the size of the two effects? Challenge

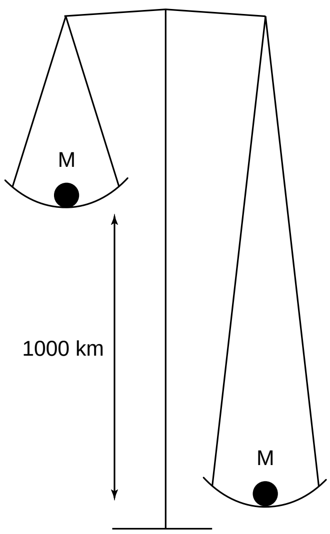


Figure 33 A honest balance?

• What is the time ratio between a stone falling through l and a pendulum swinging though half a circle of radius l? How many digits of the number π can one expect to determine in this way? Why? Challenge

• Why can a spacecraft accelerate through the *slingshot effect* when going round a planet, given that momentum is conserved? Challenge

• The acceleration g due to gravity is 10.05 m/s^2 at a depth of 3000 km, higher than at the surface of the earth. How is this possible? On the other hand, on the tibetan plateau, g is also higher than the standard seal level value of 9.8 m/s^2, even though it is farther away form the center of the earth. Why? Ref. 86 Challenge

• One can prove that objects attract each other (and that they are not attracted by the earth alone) with a simple experiment which everybody can perform at home, as described on the http://www.fourmilab.ch/gravitation/foobar/ web site.

• It is instructive to calculate the *escape velocity* of the earth, i.e. that velocity with which a body must be thrown so that it never falls back. It turns out to be 11 km/s. What is the escape velocity for the solar system? By the way, the escape velocity of the galaxy is 129 km/s. What would happen if a planet or a system would be so heavy that its escape velocity would be larger than the speed of light? Challenge

• For bodies of irregular shape, the center of gravity of a body is *not* the same as the center of mass. Are you able to confirm this? (Hint: find and use the simplest example possible.) Challenge

• The constellation in which the sun stands at 12h00 sharp (on the center of the time zone) is supposedly called the "zodiacal sign" of that day. Astrologers say there are twelve of them, namely Aries, Taurus, Gemini, Cancer, Leo, Virgo, Libra, Scorpius, Sagittarius, Capricornus, Acquarius, and Pisces and that each is a month long. Any check with a calendar shows that at present, the midday sun is never in the zodiacal sign during the days usually connected to it (the relation has shifted by about a month since it was defined, due to the precession of the earth), and any check with a map of the star sky shows that the twelve signs do not have the same length, and that there are fourteen of them, and not twelve (there is *Ophiuchus* between Scorpius and Sagittarius, and *Cetus* between Acquarius and Pisces). Ref. 87 In fact, not a single astronomical statement about zodiacal signs is correct. To put it clearly, astrology, in contrast to its name, is *not* about stars.

• The gravitational acceleration for a particle inside a spherical shell is zero. The vanishing of gravity in this case is independent of the particle shape and its position, and independent of the thickness of the shell.* Can you find the Challenge argument using the picture? This works only because of the $1/r^2$ dependence of gravity. Can you show that the result Challenge does not hold for non-spherical shells? Note that the vanishing of gravity inside a spherical shell usually does not hold if other matter is found outside the shell. Except in one case. How could one realize an experiment eliminating the effects Challenge of outside matter?

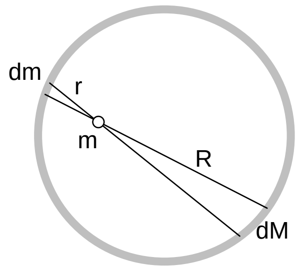


Figure 34 The vanishing of gravitational force inside a spherical shell of matter

• There is no planet X, i.e. no tenth planet in our solar sys- Ref. 88 tem outside Neptune and Pluto. But there are many small objects over there, in the so-called Kuiper belt and in the Oort cloud. Sometimes they change trajectory due to the attraction of a planet: a new comet is born.

• In astronomy many examples of motion are still studied and even regular discoveries are made. One example is the – still controversial – fall on the earth of mini comets (a few Ref. 89 dozens of kilograms of ice) every few seconds. Discovering objects hitting the earth is not at all easy, contrary to what one may believe. It is often claimed that an asteroid as large as that which led to the extinction of the dinosaurs could hit the earth without any astronomer noticing beforehand, if the direction is slightly unusual.

• Universal gravity allows only elliptical, parabolic, or hyperbolic orbits. It is impossible that a small object approaches a large one and gets captured. At least, that is what we learned so far. Nevertheless, all astronomy books tell stories of capture in our solar system, Challenge e.g. about several outer satellites of Saturn. How is this possible?

Challenge • Is it true that the centre of mass of the solar system is always inside the sun?

• Do all points on the earth have the same number of daylight hours in a year? The answer Challenge is no. Can you see why?

• There is an important difference between the heliocentric system and the old idea that all planets turn around the earth. The heliocentric system allows that certain planets, such as Mars or Venus, can be at times *between* the earth and the sun, and *behind* the sun at other times, in contrast to the geocentric system. Why did such an important difference not kill the geocentric system right away?

Ref. 90 • The strangest reformulation of the description $m\mathbf{a} = \nabla U$ is the almost absurd looking

* This is a small example from the beautiful text by Mark P. SILVERMAN, *And yet it moves: strange systems and subtle questions in physics*, Cambridge University Press, 1993. It is a treasure chest for anybody interested in the details of physics.

$$\nabla v = d\mathbf{v}/ds \qquad (40)$$

where s is the motion path length. It is called the *ray form* of Newton's equation of motion. Can you find an example of its application? Challenge

• Seen from Neptune, the size of the sun is the same as that of Jupiter seen from the earth at the time of closest approach.

• What happens to the solar system in future? This question is surprisingly hard to answer. The main expert of this topic, US physicist Gerald Sussman, simulated a few hundred million years of evolution, on specially built computers, following only the planets, without taking into account the smaller objects. He finds that the planetary orbits are stable, but that (Ref. 91) there is clear evidence of chaos in the evolution of the solar system, albeit at a small level. See page 109 The various planets influence each other in subtle and still not well understood manners. There is still a lot of research to be done in this field.

• What is gravity? This simple question is not easy. After Newton, already in 1747 Georges-Louis Lesage gave an explanation for the $1/r^2$ dependence. He argued that the (Ref. 92) world is full of small particles flying around randomly and hitting all objects. Single objects do not feel the hits, since they are hit continuously and randomly from all directions. But when two objects are near each other, they produce shadows for part of the flux to the other body, resulting in an attraction. It is easy to show that such an attraction has a $1/r^2$ dependence. Challenge This famous argument has resurfaced in physics regularly ever since, even though such particles have never been found. Only in the third part of our escalation we will settle the issue.

• How often does the earth rise and fall when seen from the moon? Challenge

• What is the weight of the moon? Is it smaller of larger than the weight of the Alps? Challenge

• One of the mysteries of the solar system is the description about planet distances discovered in 1766 by Johann Daniel Titius (1729–1796) and publicized by Johann Elert Bode (1747–1826). Titius discovered that planet distances from the sun can be approximated by

$$d = a + b\,2^n \quad \text{with} \quad a = 0.4\ ,\ b = 0.3 \qquad (41)$$

when distances are measured in astronomical units. The resulting approximation is shown in table 75.

Interestingly, the last three planets, as well as the planetoids, were discovered *after* Titius; he had successfully predicted Uranus´ distance from this rule, as well as the planetoids. Despite these successes – and the failure for the last two planets – nobody has yet found a model for the formation of the planets which explains this rule.

What is classical mechanics?

Given the mass of a body as its only permanent property, what types of motion can one describe? Quite a few, as it turns out. The study of these motion types is called *mechanics*, as are its experts. One can think of mechanics as the athletic part of physics;* like in athletics,

∗ This is in contrast to the actual origin of the term 'mechanics', which means 'machine science'. It derives from the greek μηχανή, which means 'machine' and even is the origin of the english word 'machine' itself.

Sometimes the term 'mechanics' is used for the study of motion of solid bodies only, excluding e.g. hydrodynamics. This use has fallen out of favor in physics since about a century.

Planet	n	predicted	measured
		distance in AU	
Mercury	$-\infty$	0.4	0.4
Venus	0	0.7	0.7
Earth	1	1.0	1.0
Mars	2	1.6	1.5
Planetoids	3	2.8	ca. 2.2 to 3.2
Jupiter	4	5.2	5.2
Saturn	5	10.0	9.5
Uranus	6	19.6	19.2
Neptune	7	38.8	30.1
Pluto	8	77.2	39.5

Table 13 Planet distances and the values resulting from Titius' rule

also in mechanics one only measures lengths, times, and masses. In general, that part of physics in which observables are described by real numbers is called *classical*, in contrast to *quantum* physics. Taken in this sense, classical physics is the description of interactions between bodies with *commuting continuous* quantities. All observables depending on space and time, such as field strengths, densities, currents, etc., are described with help of continuous functions of space and time. This is true even in the case of motion change due to contact. In the domain of motion, a classical description is possible only in the following domains: mechanics, thermal physics, relativity, gravitation, and electromagnetism. We encounter the results from each of these sections of classical physics in the present, first part of our walk.

The systematic name of the topic of this chapter is thus *classical mechanics*, whereas the historical name is *galilean physics* or *newtonian physics*. Like all other parts of *classical physics*, it is composed of kinematics and dynamics.

Classical mechanics consists of the description of motion only using space and time, such as $z(t) = z_o + v_o(t - t_o) - \frac{1}{2}g(t - t_o)^2$, called *kinematics*, and of the description of motion as consequence of interactions between bodies; *dynamics*, as this domain is called, forms the main part of mechanics. Like mechanics, also all other parts of *classical physics*, namely relativity, thermodynamics, and electrodynamics, can be seen as composed of kinematics and dynamics.

In galilean physics, whenever an object changes its momentum, one says there is an *interaction* responsible for it. That means that there is either a potential or a form of friction involved. Can friction be due to gravity? No, since in the skies friction is not observed;** and
Challenge on earth, it is independent of it. There must be another interaction responsible for friction. We study it shortly.

** This is not completely correct: in the 1980s, the first case of gravitational friction was discovered: the emission of gravity waves. We discuss it in detail later on.

Should one use force?

We all have to take a stand on this question. It has been debated in many discussions, also in the field of physics. Many types of forces are used to describe daily life. One speaks of muscular, gravitational, characterial, sexual, demoniac, supernatural, social, political, economic, and many other types of forces. We call *interactions* the different types of forces we observe between objects. The study of the details of all these forces shows that they all follow from only four *fundamental* types of interactions: the gravitational, the electromagnetic and the two nuclear interactions.

But what is force? *(Physical) force* is defined as the *flow of momentum*, i.e. as

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} \quad . \tag{42}$$

Using the fact that in galilean physics the linear momentum $\mathbf{p}$ is defined as $\mathbf{p} = m\mathbf{v}$, one can rewrite the definition of force as

$$\mathbf{F} = m\mathbf{a} \quad , \tag{43}$$

where $\mathbf{F} = \mathbf{F}(t,\mathbf{x})$ is the force acting on an object of mass m and $\mathbf{a} = \mathbf{a}(t,\mathbf{x}) = d\mathbf{v}/dt = d^2\mathbf{x}/dt^2$ is the acceleration of the same object, i.e. its change of velocity.* The expression states in precise terms that force is what changes the *velocity* of masses.The quantity is called 'force' because it corresponds in many, but *not* in all aspects to muscular force. For example, the more force one uses, the further one can throw a stone, because one can then impart it a higher acceleration.

However, whenever one uses the concept of force, one should remember that *physical force is different from everyday force or from everyday effort*. Effort is probably best approximated by the concept of *(physical) power*, usually abbreviated P, and defined as Ref. 93

$$P = \frac{dW}{dt} = \mathbf{F} \cdot \mathbf{v} \tag{44}$$

in which *(physical) work* W is defined as $W = \mathbf{F} \cdot \mathbf{s}$.**

When students in exams say that the force acting on a thrown stone is smallest at the highest point of the trajectory, it is customary to say that they are using the so-called *Aristotelian view*, in which force is proportional to velocity. It is then added with a tone of superiority that this view is all wrong. This is a typical example of intellectual disinformation. Every child knows from cycling, or from throwing a stone, or from pulling objects that increased effort results in increased speed. The child is right; wrong are those theoreticians which translate this by saying that the child has a mistaken concept of *force*. In fact, the child or Ref. 94

∗ This equation was first written down by the swiss mathematician and physicist Leonhard Euler (1707–1783) in 1747, over 70 years after Newton's first law and 20 years after Newton's death, to whom it is usually and falsely ascribed; it was Euler, not Newton, who first understood that this definition of force is useful in *every* case of motion, whatever the appearance, be it for point particles or extended objects, and be it rigid, deformable
Ref. 15 or fluid bodies. Surprisingly and in contrast to what is often said, equation (43) is even correct in relativity, as shown on page 174.

Challenge ∗∗ One has mastered the definition of work and of energy once one can answer the following puzzle: what happens to the electricity consumption of an escalator if one walks on it, instead of staying still?

student is just using, its *everyday* concept of force, namely effort, instead of the *physical* one. (One can also argue that Aristoteles or the student use a different concept of *state* of motion.) When one has understood the difference, including e.g. the strange consequence that a man walking carrying a heavy rucksack is not doing any work, one has taken the main hurdle in the learning of mechanics.***

Often equation (42) is not recognized as the definition of force. This is mainly due to the fact that there seem to be forces without any associated acceleration or momentum change, such as mechanical tension or water pressure. Pushing against a tree, there is no motion, yet there is force. If force is momentum flow, where does the momentum go? In fact, the tree is slightly deformed, and the associated momentum change, of the molecules, the atoms or the electrons of the bodies *is* observed, but only at the microscopic level. For this reason one never needs the concept of force in the microscopic description of nature. For the same reason, the concept of *weight*, also a force, will not be (much) used in our walk.

On the other hand, at the macroscopic level the concept of force is thus useful, especially when no apparent motion can be associated to it, such as in the case of pressure.

Through its definition the concept of force is distinguished clearly from 'mass', from 'momentum', from 'energy' and from 'power'. But where do forces originate from? In other words, which effects in nature have the capacity to accelerate bodies, to pump momentum into objects? The following table gives an overview.

All motion, such as that which lets us choose the direction of our look, that which carries a butterfly through the landscape, as well as all other types, can be divided into one of the two leftmost columns of table 14.

Physically, the two columns are separated by the following criterion: in the first class, acceleration of a body can be in a different direction than its velocity. The second class of examples only produces accelerations exactly *opposed* to the velocity of the moving body, as seen from the frame of reference of the braking medium. Such a resisting force is called *friction*, *drag*, or a *damping*. All examples in the second class are types of friction. Note that in the case of friction, macroscopic energy is not conserved.

Friction can be so large that all motion of a body is made impossible. In everyday life, this type of friction, called *static friction* or *sticking friction*, is common and important: without it, the turning wheels in bicycles, trains, cars could not make them advance. Without static friction, not a single screw would stay tightened. Without static friction, we could not walk in this forest, nor run. In fact not only our own motion, but all *voluntary motion* of living beings is *based* on friction. The same is the case for self-moving machines. In contrast, the motion of non-living objects, passive motion, is never based on friction, but only hindered by it. This so-called *dynamic friction* between bodies in motion is even more important; without it, the propellers in ships, airplanes and helicopters would turn but be ineffective, the wings of airplanes would produce no lift to keep them in the air, falling bodies would always rebound to the same height without ever stopping on the floor, neither parachutes nor brakes would not work, nor would we have memory, as we will see later on.*

*** This stepping stone is so high that many professional physicists do not really take it themselves; this is witnessed by the innumerable comments in papers which state that physical force is defined using mass, and that mass is defined using force (the latter part of the sentence being a fundamental mistake).

* For a general overview of the topic, from physics to economics, architecture and organizational theory, see N. Akerman, editor, *The necessity of friction – nineteen essays on a vital force*, Springer Verlag, 1993.

Accelerating situations	Decelerating situations	Motors and actuators
piezoelectricity		
quartz under applied voltage	thermoluminescence	walking piezo tripod
gravitation		
falling	emission of gravity waves	pulley
collisions		
satellite acc. by planet encounter	car crash	rocket motor
magnetic effects		
compass needle near magnet	electromagnetic braking	electromagnetic gun
magnetostriction	transformer losses	linear motor
current in wire near magnet	electric heating	galvanometer
electric effects		
rubbed comb near hair	friction between solids	electrostatic motor
bombs	fire	muscles, sperm flagella
television tube	electron microscope	brownian motor
light		
levitating objects by light	light bath stopping atoms	(true) light mill
solar sail for satellites	light pressure inside stars	solar cell
elasticity		
arrow and arch	trouser suspenders	ultrasound motor
bent trees standing up again	pillow	bimorphs
osmosis		
water rising in trees	conservation of food with salt	osmotic pendulum
electro-osmosis		variable X-ray screen
heat & pressure		
freezing champagne bottle	surfboard water resistance	hydraulic engines
tea kettle	quicksand	steam engine
barometer	parachute	air gun, sail
earthquakes	sliding resistance	seismometer
attraction of nearby trains	shock absorbers	water turbine
nuclei		
radioactivity	plunging into sun	supernova explosion

Table 14 A selection of processes and devices changing the motion of bodies

The same distinction between the first two columns is found when using a more abstract, mathematical criterion: on the left are forces which can be derived from a potential, on the right, forces which cannot. Like in the case of gravitation, the description of any kind of motion is much simplified by the use of a potential: at every position in space, one needs only the single value of the potential, instead of the three values of the acceleration or the force, to calculate the trajectory of an object; moreover, the magnitude of the velocity of an object at any point can be calculated directly from energy conservation.

In the second class of processes this is impossible. These are the cases where one necessarily has to use force, if one wants to describe the details of the motion of the system. For example, the wind resistance of a body is roughly given by

$$F = 1/2 c_{\mathrm{w}} \rho A v^2 \qquad (45)$$

where A is the area of its cross section, v its velocity relative to the air, ρ is the density of air, and c_{w} is a pure number, the *drag coefficient* which depends on the shape of the moving object. You may check that Challenge aerodynamic resistance cannot be derived from a potential.**

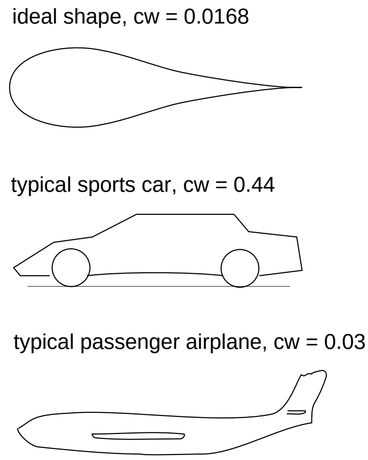


Figure 35 Shapes and air resistance

The drag coefficient is found experimentally to be always larger than 0.0168, which corresponds to the optimally streamlined tear shape; an aerodynamic car has a value of 0.25 to 0.3; but many sports cars share with trucks values such as 0.44 and higher.*

Ref. 97 Wind resistance is also of importance to humans, e.g. in athletics. It is estimated that 100 m sprinters spend between 3% and 6% of their power to overcome drag. This leads to Challenge varying sprint times t_{w} when wind of speed w is involved, as described by the expression

$$t_{\mathrm{o}} = t_{\mathrm{w}}(1.03 - 0.03(1 - \frac{w t_{\mathrm{w}}}{100})^2) \quad , \qquad (46)$$

when the more conservative estimate of 3% is used. An opposing wind speed of $-2\,\mathrm{m/s}$ gives a time increase of 0.13 s, enough to change an a potential world record into an “only” excellent result. (Are you able to deduce the c_{w} value for running humans from Challenge the formula?)

Friction between solids who do not glide against each other has different properties. It is proportional to the force pressing the two bodies against each other. For a man walking,

Recent research suggest that maybe in certain crystalline systems, such as tungsten bodies on silicon, under ideal conditions gliding friction can be extremely small and possibly even vanish (‘superlubrification’) in certain Ref. 95 directions of motion.

** Such a statement about friction is correct only in three dimensions, as is the case in nature; in the case of a single dimension, a potential can always be found.

* The issue is even more interesting for fluid bodies. They are kept together by a the so-called *surface tension*. For example, it keeps the hair of a wet brush together. Surface tension also determines the shape of rain drops. Experiments show that it is spherical for drops smaller than two millimeters, and that larger rain drops are *lens* shaped, with the flat part towards the bottom. The usual tear shape is *not* encountered in nature; something Ref. 96 vaguely similar to it appears only during drop detachment.

this is the lower surface of the shoe. Without this so-called *sticking friction* we would not be able to advance when walking, as ice surfaces remind us during winter. Studying the situation in more detail, one finds that sticking friction is proportional to the actual contact surface. It turns out that putting two solids together is rather like turning Switzerland upside down and standing it on Austria; the area of contact is much smaller than the one estimated macroscopically. One also finds that the actual contact area is proportional to the *normal* force. The study of what happens in the small percentage of contact area is still a topic of research, studied with atomic force microscopes, lateral force microscopes, and triboscopes. Ref. 98

These and all other examples of friction are accompanied by an increase of the temperature of the moving body. After the discovery of atoms, the reason became clear. Friction is not observed in few – e.g. 2, 3, or 4 – particle systems. Friction only appears in systems with many particles, usually millions or more. Such systems are called *dissipative*. Both the temperature changes as well as friction itself are due to motion of microscopic particles against each other. This motion is not included in the galilean description. When one does include it, friction and energy loss disappear, and potentials can then be used throughout, positive accelerations – of microscopic magnitude – also appear, and motion is found to be conserved. Therefore, on the microscopic scale it is possible to describe *all* motion without the concept of force.* In conclusion, one should use force only in one situation: in the case of friction, and only when one does not want to go into the details.**

Complete states: initial conditions

Quid sit futurum cras, fuge quaerere ...***
Horatius, *Odi*, lib. I, ode 9, v. 13. Ref. 41

When the motion of a body is given by an expression such as

$$\mathbf{x}(\mathbf{t}) = \mathbf{x}_o + \mathbf{v}_o(t-t_o) + \frac{1}{2}\mathbf{a}_o(t-t_o)^2 + \frac{1}{6}\mathbf{j}_o(t-t_o)^3 + \ldots \qquad (47)$$

the quantities with an index o, such as $\mathbf{x}_o, \mathbf{v}_o$, etc., are called *initial conditions*. They are necessary to describe the motion one is investigating. We see directly from the equation that different systems have different initial conditions. Initial conditions specify the individuality of any particular system. We also see that that initial conditions allow to distinguish the present situation of a system from that at previous times: initial conditions specify the changing aspects of a system.

But this is exactly the set of properties we required for a description of the *state* of a See page 34

* The first scientist who eliminated force from the description of nature was Heinrich Rudolf Hertz (1857, Hamburg–1894, Bonn), the famous discoverer of electromagnetic waves, in his textbook on mechanics, *Die Prinzipien der Mechanik*, Barth, 1894, republished by Wissenschaftliche Buchgesellschaft, Darmstadt, 1963. His idea was dismissed at that time; only a generation later, when quantum mechanics quietly got rid of the concept for good, the idea became commonly accepted. (Many have speculated about the role Hertz would have played in the development of quantum mechanics and general relativity, had he not died so young.) In his book, Hertz also formulated the principle of the straightest path: particles follow geodesics. This same description is used in general relativity, as we will see later on.

** Read: *microscopic* details. In the case of human relations the evaluation should be somewhat more discerning. A powerful book on human violence is James GILLIGAN, *Violence – our deadly epidemic and its causes*, Grosset/Putnam, 1992.

*** 'What future be tomorrow, never ask ...' Quintus Horatius Flaccus (65–8 BCE).

system. To find a complete description of states, we thus only need a complete description of the initial conditions. Now it turns out that for gravitation, as well as all microscopic and several macroscopic interactions there is usually no need for $\mathbf{a}_o$, $\mathbf{j}_o$, or higher order quantities. The reason for this is that one can find expressions for the acceleration depending on the properties of objects alone, and not depending on their past. Expressions like the expression $a = GM/r^2$, do not depend on the past at all, but only on the environment of the system. For the other interactions, we will deduce similar expressions below.

The *complete* state of a moving mass point – we defined it on page 51 – is thus described by specifying its position and its momentum for all instants of time. After a very short study, we have therefore achieved a rather complete description of point objects, namely by their mass, and of their states of motion, namely by their momentum, energy, position and time. For extended rigid objects, one also needs orientation and angular velocity. Can you guess what is necessary in the case of fluids?

Challenge

The set of all possible states of a system is given a special name, because it is often useful to treat it as a single concept and to discuss its properties: it is called the *phase space*. We will use the term repeatedly.

However, there are situations in nature where motion of an object depends on other characteristics than its mass; motion can depend on its colour (can you find an example?), on its temperature, and on a few other properties which we will soon discover. We must therefore conclude that we do not have a complete description of extended objects yet. That will happen during the coming parts of our walk.

Challenge

As a note, it is interesting to recall the previous challenge and ask again: does the universe have initial conditions? Does it have a phase space? As a hint, recall that when a stone is thrown, the initial conditions summarize the effects of the thrower, his history, the way he got there etc.; in other words, initial conditions summarize the effects the environment had during on the history of a system.

Challenge

An optimist is somebody who thinks that the future is uncertain.

Do surprises exist? Is the future determined?

Die Ereignisse der Zukunft *können* wir nicht aus den gegenwärtigen erschließen. Der Glaube an den Kausalnexus ist ein Aberglaube.*
Ludwig Wittgenstein, *Tractatus*, 5.1361

Freedom is the recognition of necessity.
Friedrich Engels (1820–1895)

If, after climbing a tree, one jumps down, one cannot stop the jump in the middle of the trajectory; once the jump is engaged, it is unavoidable and determined, like all passive motion. However, when one starts moving an arm, one can stop or change its motion from a hit to a caress. Voluntary motion does not seem unavoidable or predetermined. Which of these two cases is the general one?

∗ We *cannot* infer the events of the future from those of the present. Superstition is nothing but belief in the causal nexus.

Let us start with the examples we can describe most precisely so far: the fall of a body. Once the potential acting on a particle is given, for example, in the case of gravitation, by

$$\mathbf{a}(x) = -\nabla\varphi/m = -GM\mathbf{r}/r^3 \tag{48}$$

and the state at a given time is given by initial conditions such as

$$\mathbf{x}(t_o) = x_o \quad \text{and} \quad \mathbf{v}(t_o) = v_o \quad , \tag{49}$$

we can determine its motion, i.e. its complete trajectory $\mathbf{x}(t)$. Due to this possibility, such an equation is called an *evolution equation* for the motion of the object. (Note that the term 'evolution' is used in different meanings in physics and in biology.) An evolution equation always expresses the fact that not all types of change are observed in nature, but only some specific cases. An evolution equation expresses the fact that not all imaginable sequences of events are observed, but only a limited number of them. In particular, equation (48) expresses that from an instant to the next, objects change their motion in the way given by the potential acting on them. Thus, given an evolution equation and initial state, the whole motion of a system is *uniquely fixed*; this property of motion is often called *determinism*. Since this term is often used with various meanings in the description of motion, one has to distinguish it carefully from several similar concepts, to avoid misunderstandings.

Motion can be deterministic and at the same time *unpredictable*. The latter property can be the case due to impractically large numbers of particles involved, due to the complexity of the evolution equations, due to insufficient information on initial conditions, or, as we will see, due to strange shapes of space-time. The weather is a case where the first three conditions are fulfilled together.* Nevertheless, its motion still is deterministic. In the case of black holes all four cases apply. We will discuss them in the section on general relativity; nevertheless, their motion is still deterministic.

For the same reason motion can be deterministic and at the same time *random*, i.e. with different outcomes in similar experiments. A roulette ball's motion is deterministic, but it is also random.** As we will see later, quantum mechanical situations fall into this category, as do all examples of irreversible motion. In all these cases the randomness and the irreproducibility is only apparent, and disappears when one includes the microscopic domain.

On the other hand, it seems impossible to have deterministic motion (of matter and energy) which is also *acausal*, i.e. faster than light. This topic will be deepened in the section on special relativity.

Saying that motion is 'deterministic' also means that it is fixed in the future *and in the past*. It is sometimes stated that predictions of *future* observations are the crucial test for a successful description of nature. Due to our often impressive ability to influence the future, this is not necessarily a good test. Any theory must, first of all, describe the *past* observations correctly. It is our lack of freedom to change the past which gives origin to the lack of choice in the description of nature that is so central to physics. In this sense, the term 'initial

∗ For a beautiful view of clouds, see the http://www.goes.noass.gov web site.

∗∗ Mathematicians have developed a large number of tests do determine when a collection of numbers may be called random; roulette results pass these tests – in honest casinos only, however. Such tests typically check the equal distribution of numbers, of pairs of numbers, of triples of numbers, etc. Other tests are the χ^2 test and the
Ref. 99 Monte Carlo test.

condition' is an unfortunate choice, because it automatically leads to search for the initial condition of the universe and to look there for answers to questions which can be answered without that knowledge. The central ingredient of a deterministic description is that all motion can be reduced to an evolution equation plus one chosen state, which can be either initial, intermediate, or final.

To get a clear concept of determinism, it is useful to remind oneself why the concept of 'time' is introduced in our description of the world. We introduce time because we observe that we are able to define sequences among observations, and in particular because unrestricted change is impossible, in contrast to movies, where one person can walk through a door and exit into another continent or century. Neither do we observe metamorphoses, e.g. people changing into toasters or dogs into toothbrushes. We are able to introduce 'time' only because the sequential changes we observe are extremely restricted. If nature were not reproducible, time could not be used. In short, determinism expresses the experience that *sequential changes are restricted to a single possibility*.

Challenge

Since determinism is connected to the use of the concept of time, new questions arise whenever the content of the concept of time changes, as happens in special relativity, in general relativity, and in theoretical high energy physics. There is a lot of fun ahead.

In summary, every description of nature which uses the concept of time, such as that of everyday life, that of classical physics, and also that of quantum mechanics, is intrinsically and inescapably deterministic, since it connects observations of the past and the future *eliminating* alternatives. In short, *the use of time implies determinism, and vice versa*. When drawing metaphysical conclusions, as is so popular nowadays when discussing quantum theory, one should never forget this connection. Whoever uses clocks but denies determinism is nurturing a split personality!

See page 499

The idea that motion is determined often produces fear, because one associates lack of freedom with it. We do experience freedom in our actions and call it *free will*. We know that it is necessary for our creativity and for our happiness. Therefore it seems that determinism is in contrast with happiness.

But what is free will precisely? Much ink has been consumed on trying to find a precise definition. One can try to define free will as the arbitrariness of the choice of initial conditions. But this is impossible since initial conditions must themselves result from the evolution equations, so that there is in fact no freedom in their choice. One can try to define free will from the idea of unpredictability, or from similar properties, such as uncomputability. But all these definitions face the same problem: whatever the definition, there is *no way* to experimentally prove that an action was performed freely. Free will cannot be observed. (Psychologists also have a lot of their own data to underline this, but that is another topic.)

It is also clear from above that no process which is *gradual* – in contrast to *sudden* – has the chance to be due to free will; gradual processes are described by time and are deterministic. In this sense, the question about free will becomes one about the existence of sudden changes in nature. This will be a recurring topic in the rest of this walk. Does nature have the ability to surprise? In everyday life, nature does not. Sudden changes are not observed. We still have to investigate this question in other domains, in the very small and in the very large. But it will turn out at the end of our walk that the result still holds. Note that even curiosity is based on the idea that everything discovered is useful afterwards. If nature would continually surprise us, curiosity would make no sense.

Given that there are no sudden changes, there is only one consistent definition of free will: it is a *feeling*, usually of independence of others, or of independence from fear, or of accepting the consequences of one actions. It is a feeling of satisfaction.* This solves the apparent paradox: free will, being a feeling, exists as a human experience, even though all objects move without any possibility of choice. There is no contradiction.**

Even if an action is determined, it still is authentic. So why is determinism so frightening? That is a question everybody has to ask himself. What difference does determinism imply for one's life, for the actions, the choices, the responsibilities, and for the pleasures one encounters?* If one concludes that being determined is or would be different from being free, one should change one's life! The fear of determinism usually stems from the refusal to take the world the way it is. Paradoxically, it is precisely he who insists on the existence of free will who is running away from responsibility. Ref. 100 Challenge

You do have the ability to surprise yourself.
Richard Bandler and John Grinder

Darum kann es in der Logik auch *nie* Überraschungen geben.**
Ludwig Wittgenstein, *Tractatus*, 6.1251

A strange summary about motion

Classical mechanics describes nature in a rather simple way. *Objects* are permanent entities with mass localized in space-time. *States* are changing properties of objects, described by energy and momentum, position and instants. *Time* is the relation between events measured by a clock. *Space* and position is the relation between objects measured by a meter bar. *Clocks* are devices in undisturbed motion whose position can be observed. *Meter bars* are mobile devices at rest, whose shape is subdivided in an invariant and observable manner. *Motion* is change of position with time (times mass); it is determined without surprises, is conserved (even in death), and is due to gravitation and other interactions.

Even though this description works rather well, it contains a circular definition. Can you spot it? Physicists had worked for about two hundred years on classical mechanics without noticing it. Undoing this logical loop is one of the aims of the rest of our walk. To achieve this, we need to increase dramatically the level of precision in the description of motion. Challenge

* See for example the book by Bert Hellinger, *Zweierlei Glück*, Carl Auer Systeme Verlag, 1997. The author explains how to achieve a serene life, living with the highest possible responsibility of one's actions, by reducing entanglements with the destiny of others. He gives a simple and successful technique to realize this goal.

** That free will is a feeling can also be confirmed by careful introspection. The idea of free will appears always *after* an action has been started or decided. It is a beautiful experiment to go back to the actual moment when a decision was taken – e.g. to close a hand – and to try to investigate in detail what happened around that very Challenge moment of decision. One finds either a mechanism which lead to the decision, or a diffuse, unclear mist. One never finds free will. Nevertheless, such an experiment is a way to experience deeply the wonders of the self. Experiences of this kind might also be one of the origins of human spirituality.

Challenge * If nature's "laws" are deterministic, are they in contrast with moral "laws"?

** Hence there can *never* be surprises in logic.

4. Global descriptions of classical motion – the simplicity of complexity

Πλεῖν ἀνάγκε, ζῆν οὐκ ἀνάγκη.
Navigare necesse, vivere non necesse.*
Pompeius (106–48), as cited by Plutarchus.

Ref. 41

Typeset in April 2001

The story of the discovery of the universal law of gravity has an important moral. Since people on all places of the earth, even in Australia, say that stones fall “down”, one had to look for a description valid *globally*. It only took a small additional step to deduce $a = GM/r^2$.

In short, thinking globally, i.e. taking a helicopter view, provides an efficient way to increase the precision of motion description. If one pursues this helicopter approach more systematically, it turns out there are at least six general methods to follow. Each way helps in the way to the top of motion mountain.

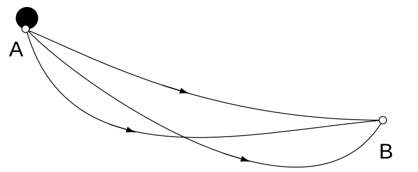


Figure 36 What shape of rail allows the black stone to glide most rapidly from point A to the lower point B?

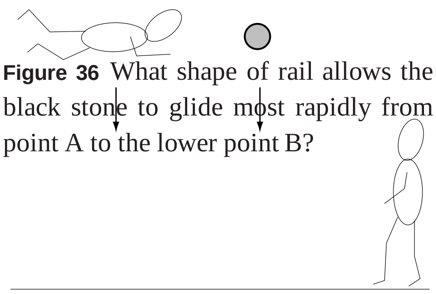


Figure 37 Can motion be described in a manner common to all observers?

• Whenever one calculates the evolution of a particle by calculating the acceleration it is subjected to, one is using the most *local* description of motion possible, with an extremely limited horizon. In fact one is using the idea that for an acceleration given at a certain place and instant of time, the position of the particle just after that moment and the motion just near that place can be calculated. Evolution equations have a mental horizon of radius zero. A famous example of the opposite approach is shown in figure 36. The problem asks for the properties of a complete movement, for *all* times and positions. Challenge

• The second global approach is to compare the various descriptions produced by *different* observers of the *same* example of motion. For example, the observations by a passenger in a roller coaster and those by an observer on the ground usually are different. Finding a connection will directly lead to the theory of relativity.

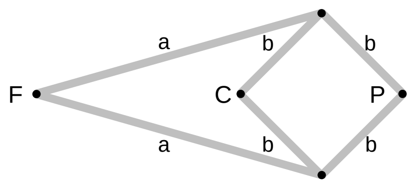


Figure 38 How to draw a straight line with a compass: fix point F, put a pencil into joint P, and move C with a compass along a circle

• A third way to look at motion globally is to turn from the study of point objects to that of *extended and rigid* bodies. Studying the interactions among several of them, such as the example of figure 38, is essential for the design of mechanisms, machines, and robots, and many other applications. For example, the mechanism of figure 38 connects the motion of points C and P; it implicitly defines a unit circle such that for the distances r from its center one always has the relation $r_C = 1/r_P$. Are you able to find that circle? Or do you know how the various Rubik’s cubes are built? Challenge Challenge

* To navigate is necessary, to live is not.

This is a section of the freely downloadable e-textbook

Motion Mountain

Hiking beyond space and time along the concepts of modern physics

available at
www.motionmountain.org

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Did you find any mistakes? Challenge

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

Ref. 101 Challenge Another famous puzzle is to devise a wooden carriage, with gearwheels connecting the wheels to an arrow, so that whatever path the carriage takes, the arrow always points south. Such a device is useful to understand general relativity, as we will see.

Ref. 102 In this class also belongs the research in the way that human movement, e.g. the motion of an arm, is built from a small number of basic motor primitives. However, we won't deepen these topics in our escalation.

• The motion of *non-rigid extended bodies* is another generalization of particle motion. The first part, *fluid mechanics*, studies the flow of honey, water, or air around solid bodies such as spoons, supports, ships, sails, and wings. For example, it investigates how insects, birds and airplanes fly,* why sailing boats can sail against the wind, what happens when a hard-boiled egg is made to spin on a thin layer of water, or how to empty a bottle full of wine in the fastest way possible. (Ref. 103, Challenge)

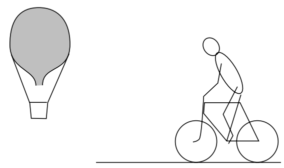


Figure 39 Why do balloons stay inflated? How do you measure the weight of a bicycle rider with a ruler only?

The other part of the study of extended bodies, the behaviour of deformable *solid* bodies, is called *continuum mechanics*. It explains for example why bells are made in the shape they always have, or where materials break when under load. We will mention a few issues in special relativity and in quantum theory.

• A fifth general viewpoint, related to the preceding, arises when one asks for the motion of *large number* of particles. In these cases the study of motion is called *statistical mechanics*. We will only briefly visit it, and introduce only those concepts we need for our escalation.

• A sixth and final set of moving systems require for their description most of the mentioned viewpoints so far at the same time. Such systems form an important part of everyday life: the formation of a specific number of petals in flowers, the differentiation of the embryo in the womb, the origin of the heart beat in the human body, and the formation of mountains are examples of such situations. Another much studied example is the formation of sea waves by the wind.

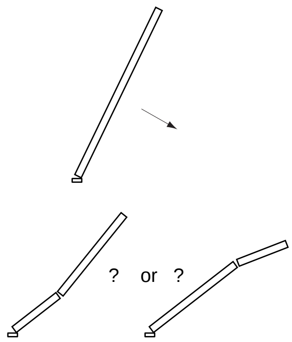


Figure 40 How and where does a falling brick chimney break?

All these are examples of *growth processes* or, as is usual to say today, of *selforganisation*. The study of these situations is a common research theme across many sciences, such as biology, chemistry, medicine, and engineering.

In the following, we give a short introduction into most of these global viewpoints. We start with the main achievement of classical mechanics, namely the two ways to describe the motion of an object globally: the principle of least action and Hamilton's equations of motion. One can say that these two methods, after several centuries of collective effort, form

* The mechanisms of insect flight are still a research subject. Traditionally, fluid dynamics concentrated on large systems like boats, ships and airplanes. Indeed, the smallest human made object which can fly in a controlled way, say a radio controlled plane or helicopter, is much larger and heavier than what evolution was able to engineer. It turns out that there are many more tricks and much more knowledge involved in letting small things fly than is the case for large things. More on page 570.

the highlight of mechanics. The methods also form the basis for all further refinements of the description of motion which we will encounter in the rest of our walk.

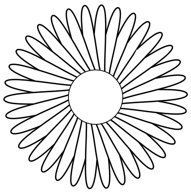


Figure 41 What determines the number of petals in a daisy?

The principle of least action

Observations only make sense if everybody agrees on what is seen. But that is rare. On the other hand, everybody agrees on the value of a number. Thus an ideal description of motion should use a *single* number. One of the most beautiful results of mechanics is the discovery that this is indeed possible.

It turns out that describing the trajectory $x(t)$ of a body does not require an evolution equation; there is another, equivalent description, called the *principle of least action*.* The principle states that the *actual* trajectory $x(t)$ of a particle between two instants of time t_0 and t_1 is precisely that specific trajectory for which the so-called *action* S, a pure number defined as

$$S = \int_{t_0}^{t_1} L\, dt \quad , \qquad \text{where} \quad L = T - U = \frac{1}{2} m v^2 - U \quad , \tag{50}$$

is *minimal*.** The quantity $L = L(\mathbf{x}, \dot{\mathbf{x}}, t)$ appearing under the integration sign is called the *lagrangian (function)* of the particle, after the italian-french mathematician Giuseppe Lagrangia.*** The quantity T is the well-known *kinetic energy* of the system; the quantity $U(\mathbf{x}, t) = m\varphi(\mathbf{x}, t)$ is called the *potential energy* and is defined with help of the *potential* $\varphi(\mathbf{x}, t)$ introduced earlier on. The change of the action for different trajectories, written δS, is called its *variation*. See page 78

The principle of least action therefore states:

$$\triangleright \textit{ The trajectory is given by } \delta S = \delta \int_{t_0}^{t_1} L dt = 0 \textit{ given that } \delta x(t_0) = \delta x(t_1) = 0. \tag{51}$$

It is a *variational principle*, since it states that the variation of the action vanishes. An equivalent way to read it is to say that the actual trajectory is the one for which the *average* of the lagrangian over the whole trajectory is minimal. Can you deduce this? Challenge

The description of motion with the principle of least action distinguishes the actual trajectory from all other imaginable ones. This fact lead to Leibniz' famous interpretation that

* The naming of the principle varies. Originally, the term 'action' was defined in a slightly different way, and the principle of least action was different from what is often called *Hamilton's principle* in anglo-saxon usage, even though the latter is (mostly) due to others, first among all to *Leibniz*. Therefore the naming is not used here. In fact, for about two centuries there was an intense search to describe motion with so-called *extremal* or *variational principles*; these are only of historical value today, because they are all special cases of the one
Ref. 104 described here, which forms the key to all the others.

** The meaning of the expression $\int f(t)\, dt$, the *integral* over t of the function $f(t)$, is explained in appendix D. Summarizing in a few words, it quantifies the effect of f does over t; it is an abbreviation of $\lim_{\Delta t \to 0} \sum_i f(t_i) \Delta t$. In other words, the integral measures the area between the curve $f(t)$ and the t-axis.

In fact, in some rare situations the action is maximal, so that the snobbish form of the principle states that the action is 'stationary,' or an 'extremum,' i.e. either minimal *or* maximal.

*** Giuseppe Luigi Lagrangia (Torino 1736–Paris 1813), better known as Joseph Louis Lagrange.

the world is the "best of all possible worlds."* But apart from the metaphysical speculations, expression (51) was the first example of a description of nature which singled out observations from all other imaginable possibilities. For the first time the search for reasons why things are not different from what they are became a part of physical investigation. The deep underlying question is: could the world be different from what it is? The answer will become clear only in the last part of our escalation.

The lagrangian description of motion is often equivalent to that using an evolution equation, as can be seen by the standard procedure for finding the minimum of the action S. (This procedure is part of the so-called "calculus of variations.") The condition $\delta S = 0$ means** Challenge

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_i}\right) = \frac{\partial L}{\partial x_i} \quad . \tag{52}$$

When the above definition of the lagrangian is introduced, these so-called *Lagrange's equations of motion* reduce to Challenge

$$ma = \nabla U \tag{53}$$

i.e. to the original evolution equation, which is what we wanted to show.

For example, the lagrangian of a mass attached to a spring is given by $L = mv^2/2 - kx$. Can you confirm it? Challenge

Compared to the description with evolution equations, the description of motion with a lagrangian has several advantages. First of all, writing a evolution equation with a lagrangian is usually more *compact* than writing the corresponding evolution equations. For example, only *one* lagrangian is needed for a system, *independently* of the number of particles. (One also makes fewer mistakes, especially sign mistakes, as one rapidly learns when performing calculations.)

Moreover, the description with a lagrangian is valid with *any* type of coordinates describing the objects of investigation. Coordinates do not have to be cartesian, they can be chosen as one prefers, cylindrical, spherical, hyperbolical etc. The advantage to use these so-called *generalized coordinates* will not be studied in our walk; they allow to calculate rapidly the behaviour of many mechanical systems which are complicated to describe in cartesian coordinates. For example for the programming of the motion of robot arms they simplify the calculation of simplest ways to move the hand from one point to the other.

More importantly, the lagrangian allows to read off quickly a number of key properties of the evolution of a system, namely its symmetries and its conserved quantities. We will develop this ability shortly, and then use it regularly throughout our walk.

Furthermore, the lagrangian formulation can be generalized to encompass all types of interactions; the terms kinetic and potential energy are interaction independent. We will see that the formulation can also be used in electricity, in magnetism, and in optics. It is also important in general relativity and in quantum theory, and allows one to easily relate both fields to classical mechanics.

* This idea was drawn into ridicule by the french philosopher Voltaire (1694–1778) in his lucid writings, notably in the brilliant book *Candide ou l'optimisme*, 1759, available e.g. as paperback from Folio-Gallimard, 1992.

** One needs the relation $\delta\dot{q} = d/dt(\delta q)$ which is valid only for *holonomic* coordinates; they are discussed below.

When the principle of least action became well-known, people applied it to an incredible amount of problems; today, using lagrangians in calculations is a standard technique in theoretical physics and in engineering, where they are used from the study of elementary particle collisions to the programming of robot motion. However, one should not forget that despite its simplicity and usefulness, the lagrangian formulation is equivalent to the original evolution equations. It is neither more general nor more specific. In particular, it is *not* an explanation for any evolution equation, but only a rephrasing of it. In fact, the correspondence is so close that one can say that the search of a new physical "law" of motion is 'simply' or 'only' the search for a new lagrangian. Ref. 104

Limits of lagrangians

There are a few comments to be made, for those interested in the topic, on the equivalence of lagrangians and evolution equations. Otherwise just skip this section.

First of all, lagrangians do not exist for *dissipative*, i.e. non-conservative systems, such as for any motions involving friction, which means that it cannot be derived from a potential. However, there is a generalized formulation of the principle of least action in that case. Whenever there is no potential, one can express the work differences δW between different trajectories as

$$\delta W = \sum_i m_i \ddot{x}_i \delta x_i \quad . \tag{54}$$

The actual motion is then given as that trajectory for which

$$\int_{t_0}^{t_1} (\delta T + \delta W) dt = 0 \quad , \quad \textit{given that} \quad \delta x(t_0) = \delta x(t_1) = 0 \quad . \tag{55}$$

In short, proper lagrangian descriptions exist only for *conservative* systems. But there are also additional limitations of the lagrangian formulation. They were discovered in times when computers where not available, when this topic was fashionable.

The generalized coordinates used in lagrangians are not necessarily the usual ones. Generalized coordinates are especially useful when there are *constraints*, such as in the case of a pendulum, where the weight has always to be at the same distance from the suspension, or in the case of an ice skater, where the skate has to move in the direction it is pointing to. Generalized coordinates may even be mixtures of positions and momenta. They can be divided into a few general cases. Ref. 105

Generalized coordinates are called *holonomic-scleronomic* when they are related to cartesian coordinates in a fixed way, i.e. independently of time; the pendulum is an example, as is a particle in a potential. The coordinates are called *holonomic-rheonomic* when the dependence involves the situation itself, such as the case of an ice skater, who can move only along the skates, not perpendicular to them.* The more general situation is called *anholonomic*, a term due, like the term holonomic, to Heinrich Hertz). Lagrangians work well only See page 94

* The two terms rheonomic and scleronomic are due to the important austrian physicist Ludwig Boltzmann (1844, Wien–1906). He is mostly famous for his work on thermodynamics, in which he explained all thermodynamic phenomena, inclusive entropy, as results of the behaviour of atoms. The naming of the Boltzmann constant resulted from these investigations. He was one of the most important physicists of the ending 19th century, and stimulated many developments which lead to quantum theory.

for holonomic systems.

It should be noted that the lagrangian for a moving system is not unique; the study of how the various lagrangians for a given moving system are related among each other is not part of this walk, however. Ref. 106

To sum up, even though the use of lagrangians has its limits, they do not bother us, since microscopic systems are always conservative, holonomic and scleronomic. We therefore can continue our walk with the result that for fundamental examples of motion, evolution equations and lagrangians are indeed equivalent.

Curiosities about lagrangians

Lagrangians form a fascinating topic.

• When in 1788 Lagrange published his book *Mécanique analytique*, it formed one of the high points in the history of mechanics. He was very proud of the fact that he had written a systematic exposure of mechanics without a single figure. Obviously the book was difficult to read and was not a sales success. Therefore his methods took another generation to get into general use.

• In galilean physics, the lagrangian is given by the difference between potential and kinetic energy. Later on, this will be generalized in a way that sharpens the understanding of this distinction: the lagrangian becomes the difference between a term for free particles a term due to its interactions with other particles. In other words, particle motion is a continuous compromise between what the particle would do if it were free and what the other particles want it to do. In this they behave a lot like humans beings.

• There is a principle of *least effort* describing the growth of trees. When a tree grows, all the mass it consists of has to be lifted upwards from the ground. A tree does this in such a way that it gets the best possible result, which means as many branches as high up in the air as possible using the smallest amount of energy.

• Another minimum principle can be used to understand the construction of animal bodies, especially their size and the proportions of their inner structures. For example, the heart pulse and breathing frequency both vary with animal mass as $m^{-1/4}$, and the dissipated power as $m^{3/4}$. It turns out that such exponents result from three properties of living beings. First, they transport energy and material through the organism via a branched network of few large and more and more smaller and smaller vessels. Second, the vessels all have a universal minimum size. And third, the networks are optimized in order to minimize the energy needed for transport. These relations might explain many more known connections, e.g. why animal life span scales as $m^{-1/4}$. A competing explanation states that 1/4 power laws arises in any network built so that the flow arrives to the destination by the most direct path. Ref. 107 Ref. 108

• The minimum principle for the motion of light will be mentioned later on.

Are you able to confirm that all these minimum principles are special cases of the principle of least action? In fact this is true for all existing minimum principles. Challenge

Why is motion so often bound?

Looking around oneself on earth and in the sky, one finds that matter is not evenly distributed. Matter tends to be near other matter; it is lumped together in *aggregates*, of which the main ones are shown in figure 42 and listed in table 15. Obviously, the stronger the interaction, the smaller the aggregate. But why is nature mainly made of lumps of matter at all? Ref. 109

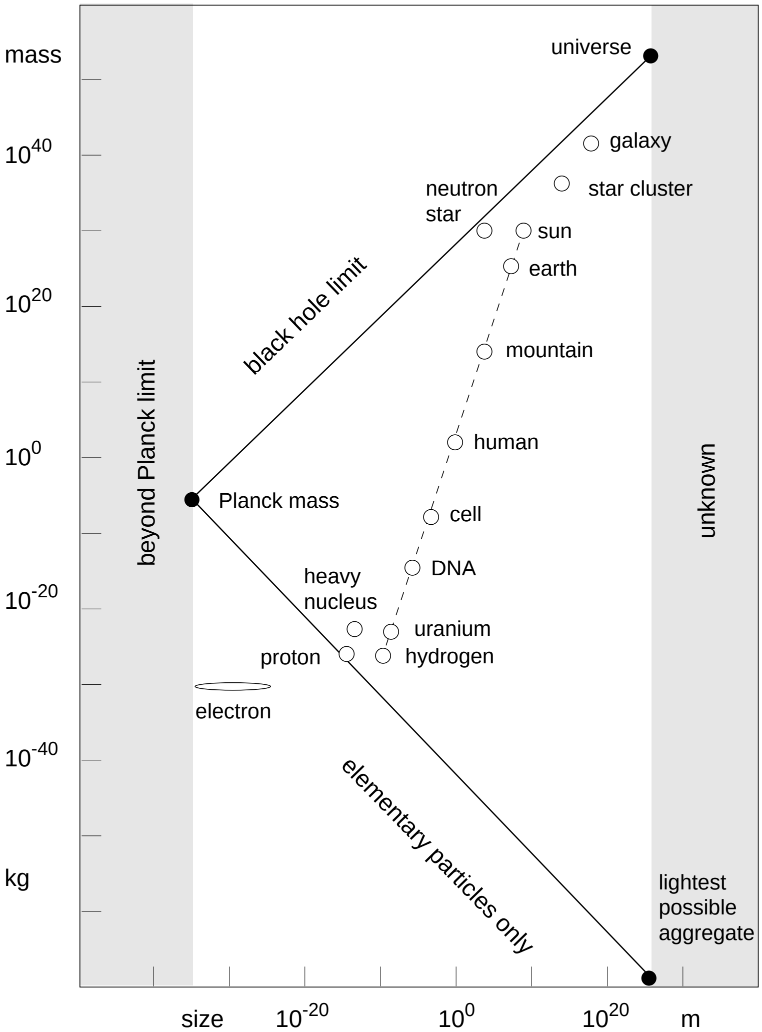


Figure 42 Aggregates in nature

First of all, aggregates form because of the existence of *attractive* interactions between objects. Secondly, they stay together because of *friction*, i.e. because the energy released when the objects approach can be changed into heat, which prevents the objects to leave again. Third, aggregates exist because of *repulsive* effects, which avoid that the components collapse completely. Together, these three characteristics ensure that bound motion is much more frequent than unbound, 'free' motion.*

By the way, the same arguments also explain why so many aggregates rotate. Can you explain the connection? Challenge

But why does friction exist at all? And why do attractive interactions exist? Or repulsive ones? And anyway, the above answers assume that in some distant past matter was *not* found in lumps. Is this correct? We start with the first investigations straight away.

Table 15 The main aggregates observed in nature

Interaction aggregate	size (diameter)	observed number	constituents
gravitationally bound aggregates			
matter across universe	ca. 100 Ym	1	photons, hydrogen & helium atoms, galaxy groups
galaxy group	100 Zm	10^8	10 to 1000 galaxies
our galaxy group	50 Zm	1	ca. 20 galaxies
general galaxy	0.5 to 2 Zm	10^{10}	10^{10} stars, dust
our galaxy	1.0(0.1) Zm	1	10^{10} stars, solar systems, clouds
interstellar clouds	ca. 1 PM	$\gg 10^5$	hydrogen, ice and dust
solar system[a]	30 TM	ca. 1	sun, planets, moons, comets, asteroids, gas, dust
Oort cloud	6 to 30 Pm	1	comets, dust
Kuiper belt	60 Tm	1	comets, dust
Pluto's orbit	11.8 Tm		
star[b]	10 km to 100 Gm	$10^{22\pm1}$	ionized gas: protons, neutrons, electrons, neutrinos, photons
neutron stars (with gravity)	10 km	ca. 1000	neutrons
our star	1.39 Mm		
planet[a] (Jupiter, Earth)	143 Mm 12.8 Mm	ca. 9	solids, liquids, gases; high Z atoms
moons	10 - 1000km	ca. 50	solids
electromagnetically bound aggregates[c]			
asteroids, mountains[d]	10 km	10^9	solids
comets	...	...	solids, e.g. ice, dust
planetoids, solids, liquids, gases, cheese	1 nm to > 100 km	n.a.	molecules, atoms
animals, plants, kefir			organs, cells
brain	0.15 m	10^{10}	neurons and other cell types
cell		$10^{28\pm2}$	organelles and membranes, all made of molecules
smallest: ca. 5 μm		molecules	
amoeba	600 μm		molecules
largest: (whale nerve,	ca. 30 m		molecules

* The are only three types of attractions which lead to aggregates: gravity, the electric interaction, and the strong nuclear interaction. There are only three types of repulsive effects: rotation, pressure, and the Pauli exclusion principle. Of the nine combinations, only some appear in figure 42; can you see which ones are missing, and why? Challenge

interaction aggregate	size (diameter)	observed number	constituents
single cell plants)			
molecules		ca. $10^{78\pm2}$	atoms
H_2	ca. 50 pm		atoms
DNA	2 m (total)		atoms
atoms, ions	100 pm to 300 pm	$10^{79\pm2}$	electrons, nuclei
weak interaction bound aggregates[c]			
none			
strong interaction bound aggregates[c]			
nucleus	$> 10^{-15}$ m	$10^{79\pm2}$	nucleons
nucleon (proton, neutron)	ca. 10^{-15} m	$10^{80\pm2}$	quarks
mesons	ca. 10^{-15} m	n.a.	quarks

a. Only in the year 1994 the first evidence was found for objects circling other stars than our sun; most of them were found around neutron stars. In particular, three objects circle the pulsar PSR1257+12; moreover, a matter ring circles the star β Pictoris. The objects seem to be dark stars, i.e. brown dwarfs; none of the systems found so far forms solar systems of the type we live in. Ref. 110
b. The sun is among the top 5% stars, when ranked in brightness (most fainter stars, namely 70%, are red M dwarfs, 15% are orange K dwarfs, and 10% are white dwarfs). However, almost all stars in the night sky belong to the bright 5%; they consist of the rare blue O class and the blue B class such as Spica, Regulus, and Riga; 1% consist of the bright, white A class such as Sirius, Vega, and Altair, and of the yellow-white F class such as Canopus, Procyon and Polaris; 4% form the yellow G class, like Alpha Centauri, Capella, or the sun. Exceptions are the few K giants, such as Arcturus and Aldebaran, and the M supergiants, such as Betelgeuse and Antares. For more details, see page 264.
c. For more details on microscopic aggregates, see the table of composites in appendix C.
d. There are about 10^9 asteroids (or planetoids) larger than 1 km and about 10^{20} heavier than 100 kg. Ref. 111

The phase space and Hamilton's equations of motion

How can one distinguish bound and free motion most clearly? The best tool for visualization is the *phase space*, i.e. the set of all possible states of a system. For a single particle, the phase space is therefore the set of all its possible velocities and positions. At every instant in time, the particle has a certain position and velocity, and during motion, it traces a certain curve in phase space, as shown in figure 43.

Rest corresponds to a point in phase space, constant motion to a line, *periodic* motion to closed curves, and other curves correspond to irregular motion. The latter ones are often divided into *quasiperiodic*, *chaotic* and *unbound*, as shown in figure 43. For motion in more than one spatial dimension or for more than one particle, the phase space is more than two dimensional; its dimension is twice the number of degrees of freedom of each particle. The number of *(mechanical) degrees of freedom* is the number of observables necessary to define the coordinates of a system.* For example, for two particles in three dimensions, phase space is twelve dimensional.

* Note that in thermodynamics a different definition of the number of degrees of freedom is used.

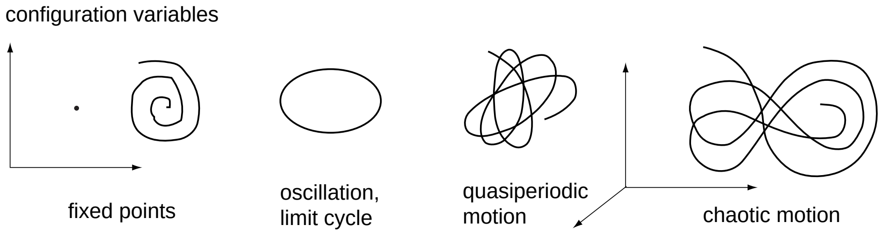


Figure 43 Examples of different types of motion in configuration space

For physicists, chaotic motion is the most irregular type of motion.[**] Chaos can be defined as the motion of systems for which small changes in initial conditions evolve into large changes of the motion. The weather is such a system, but it turns out that also dropping faucet, many flows of liquids, the fall of dice, and many others are chaotic. Research on the mechanisms by which the heart beat is generated showed that the heart is not an oscillator, but a chaotic system with irregular cycles. This allows the heart to be continuously ready for demands of changes in beat rate which appear once the body has to increase or decrease its efforts. Ref. 112

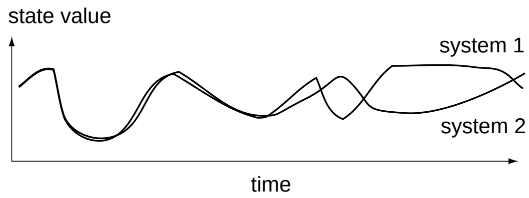


Figure 44 Sensitivity to initial conditions

And of course there is chaotic motion also in machines: chaos in the motion of trains on the rail, chaos in gear mechanisms, chaos in firemen hoses. The author predicts that the precise study of the motion in a zippo will also yield an example of chaos. The mathematical description of chaos, simple in the textbook examples, but extremely involved in other cases, remains an important topic of research. Despite their fascination we will not study the quasiperiodic and the chaotic cases because they do not lead towards the top of motion mountain.

It is possible to describe motion in such a way that the trajectories in phase space are calculated directly. This was discovered by the irish mathematician and physicist William Rowan Hamilton (1805–1865), one of the most important mathematicians of his time. He used Lagrange's ideas as a starting point. For every generalized coordinate q_i he

[**] On the topic of chaos, see the beautiful book by H.-O. Peitgen, H. Jürgens & D. Saupe, *Chaos and fractals*, Springer Verlag, 1992. It includes stunning pictures, the mathematical background, and the computer programs allowing personal exploration. 'Chaos' is an old word; according to greek mythology, the first goddess, Gaia or the earth, emerged from the chaos existing at the beginning. She then gave birth then the other gods, the animals, and the first humans.

defined the *conjugate momentum* p_i as

$$p_i = \frac{\partial L}{\partial q_i} \quad . \tag{56}$$

He discovered that the whole system is then described by what is nowadays called the *hamiltonian (function)*, given as

$$H = \sum_i \dot{q}_i p_i - L \quad . \tag{57}$$

Using it, the equations of motion, called *Hamilton's equations of motion*, of the system become

$$\frac{\partial H}{\partial q_i} = -\frac{\partial L}{\partial q_i} = -\dot{p}_i \tag{58}$$

$$\frac{\partial H}{\partial p_i} = \dot{q}_i \quad . \tag{59}$$

These equations describe the motion of systems in phase space. The hamiltonian description is complete, is equivalent to the local equations of motion, contains all permanent properties of a system under investigation, allows a compact description of numerous interacting particles, and can be generalized for continuum systems. In short, it is worth knowing it.

Additionally, one has the relation

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t} = \frac{dH}{dt} \quad . \tag{60}$$

For conservative systems, i.e. when there is no friction – and this is always the case on microscopic scale – one finds that

$$\frac{dH}{dt} = 0 \quad \text{or} \quad H = \text{const} \quad . \tag{61}$$

In the rest of our walk the coordinate systems are usually constant in time, so that H is the *total energy* of the system. This means that the relation $H(p,q)$ between the total energy and coordinates and momenta completely determines the motion of a conservative system. This is the reason that the concept of energy plays such an important role in the description of motion.

One can also write Hamilton's equations of motion in a more compact form, using a neat abbreviation called the Poisson bracket. The *Poisson bracket* (g,h) between two functions $g(q_i, p_i)$ and $h(q_i, p_i)$ is defined as

$$(g,h) = \sum_i \left(\frac{\partial g}{\partial p_i}\frac{\partial h}{\partial q_i} - \frac{\partial g}{\partial q_i}\frac{\partial h}{\partial p_i} \right) \quad . \tag{62}$$

One then has, for every observable g, the equation of motion

$$\dot{g} = (H,g) \quad . \tag{63}$$

Note that in this equation, space, in contrast to time, does not appear explicitly any more. This will be helpful in some cases and limiting in others, as we will see. By the way, using Poisson brackets, Hamilton's equations of motion become

$$\dot{q}_i = (H, q_i) \quad \text{and} \quad \dot{p}_i = (H, p_i) \tag{64}$$

and one also finds the useful relations

$$\begin{aligned} &(q_i, q_k) = 0 \quad , \quad (p_i, p_k) = 0 \quad , \quad (p_i, q_k) = 0 \\ &(p_k, q_k) = 1 \quad , \quad (q_k, p_k) = -1 \quad , \quad (H, t) = 1 \quad . \end{aligned} \tag{65}$$

We will encounter these expressions again in quantum mechanics, where they play a central role.

When does motion exist?

For the curious reader, a digression. It may seem that conceptually, a phase space and usual space are not very different, apart from the number of dimensions and the interpretation of the axes. One may be led to ask what determines the difference between the two types of spaces.

To see this, rewrite Hamilton's equations of motion in phase space as

$$\dot{\xi} = J \cdot \nabla H(\xi) \quad \text{where} \quad \xi = \begin{pmatrix} q \\ p \end{pmatrix} \quad \text{and} \quad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{66}$$

with 1 being the identity matrix of half the size and ξ the coordinates in phase space. Note that for the matrix J one has $J^{-1} = -J$, $J^T = -J$, and $J^2 = -1$, of full size, this time.* In general, what we write q and p will be strings of f coordinates q_{i} and p_{i}. The quantity $\nabla H = (\partial H/\partial q^i, \partial H/\partial p^i)$ is called the *hamiltonian vector field*** with energy H.

A trajectory $\xi(t)$ satisfies Hamilton's equations if and only if $\xi(t)$ is an integral curve of X_H, that is when $\dot{\xi}(t) = X_H(\xi(t))$. Defining a skew-symmetric bilinear form ω, the *symplectic form*, by

$$\omega(\xi_1, \xi_2) = \xi_1 \cdot J \cdot \xi_2 \tag{67}$$

* A matrix M for which $M^T JM = J$ is called *symplectic*. The set of all such matrices forms the (real) symplectic group, abbreviated SP(2f).

** We remark in passing that the term 'field' is used with various meanings in the exact sciences:

- Originally, a *(physical) field* is the space where a force is experienced, as in the expressions 'electric field', 'magnetic field' or 'gravitational field'. More on the topic of physical fields will be said on page 303. By extension, a field is also any part of space which carries an additional physical property, i.e. any quantity written as $\psi(t, \mathbf{x})$, such as 'velocity field', 'temperature field', etc. Obviously, most observables can be seen as examples of a physical fields. (Can you find an exception?) Challenge Therefore the concept will be used frequently, throughout our escalation, as synonym for observable. In mathematical physics the use of this term is generalized, and *any* function depending on space and time, such as the example above, is called a field.

- Additionally, a *(mathematical) field* is a specific mathematical structure, a special type of set, connected with the properties of numbers, as explained in the intermezzo following this chapter. It is has no relation to the previous meaning.

one has for all ξ and χ in the phase space, the relation

$$\omega(X_H(\xi),\chi) = dH(\xi)\cdot\chi \tag{68}$$

where $dH(\xi) = (\partial H/\partial q^i, \partial H/\partial p^i)$. Such a space is called a *symplectic space*.

Mathematical physicists have fun to study such spaces like any other space. They study their symmetries, the effect of coordinate transformations, their curvature, their topology, etc.

For example, a change of coordinates in phase space $h = f(x)$ implies, if $x(t)$ satisfies Hamilton's equations, that $h(t)$ satisfies $\dot{h} = A\dot{x} = AJ\nabla_x H(x) = AJA^T\nabla_h H(x(h))$. The equations for $h(t)$ will be hamiltonian – with energy $K(h) = H(x(h))$ – if and only if $AJA^T = J$. Any transformation A satisfying $AJA^T = J$ is called *canonical* or *symplectic* or a *symplectomorphism*.

The essence remains that a general space can be called a phase space if it is symplectic, i.e. if it contains a symplectic form. Actually, one can argue that the space only needs a symplectic form *locally*; one then speaks of a *presymplectic* space. The existence of a symplectic form is thus the condition necessary to be able to interpret a curve in an abstract space as the trajectory of an object. For a mathematical physicist, the existence of motion, and thus the possibility to define time, is equivalent the existence of a presymplectic form. However, this requirement does not help us much in our escalation. We therefore end this discussion here, and we continue with another global aspect of motion.

Motion and symmetry

The second way to describe motion globally is to describe it in such a way that all observers *agree*. It is well known that an object of observation is called *symmetric* if it looks the same when seen from some different points of view, i.e. when it looks the same for various observers. For example, the forget-me-not of figure 45 is symmetrical because it looks the same when watched after turning around it by 72 degrees; many flowers of fruit trees have the same symmetry. One also says that under change of viewpoint the flower has an *invariant property*, namely its shape. If there are many such viewpoints one talks about a *high* symmetry, otherwise about a *low* symmetry. For example, a four-leaved clover has a higher symmetry than a usual, three-leaved one. In physics, the viewpoints are often called *frames of reference* and are described mathematically by coordinate systems.

Figure 45 Forget-me-not, also called *myosotis* (Barraginaceae)

At first sight, not many objects or observations in nature seem to be symmetrical. But this is a mistake due to a too narrow interpretation of the term. On the contrary, one can deduce that nature as a whole is symmetric from the simple fact that people have the ability to talk about it! Moreover, contrary to what one might think, the symmetry of nature is considerably higher than that of a forget-me-not and among others is at the basis of $E = mc^2$. Challenge

Why can we think and talk?

The hidden harmony is stronger than the apparent.
Ref. 14 Heraclitos of Ephesos, about 500 B.C.

Why can we understand somebody when he is talking about the world, even though we are not in his shoes? We can for two reasons: because most things look *similar* from different viewpoints, and because most of us have made already similar experiences *before*.

'Similar' means that what *we* observe and what *others* observe somehow correspond. Somehow our observations do not depend much on our viewpoint. For example, the number of petals of a flower has the same value for all observers. We can therefore say that this quantity has the highest possible symmetry. We will see below that mass is another such example. Observables with this highest possible symmetry are called *scalars* in physics. Other aspects change from observer to observer, such as apparent size changes with distance. However, the actual size is observer independent. In general terms, any type of *viewpoint independence* is a form of symmetry, and the fact that two people looking at the same thing from different viewpoints can understand each other proves that nature is symmetric. The details of this symmetry are explored in this section, as well as during most of the rest of our escalation.

In the world around us, we note another general property: not only does the same phenomenon look similar to different observers, also *different* phenomena look similar to the *same* observer. For example, we know that if the fire in the kitchen burns the finger, it will do so outside the house as well, and also in other places and at other times. Our memory and our thinking are only possible because of this basic property of nature. (Can you confirm
Challenge this?) As we will see, this property leads to additional strong restrictions on the description of nature

Without viewpoint independence and without reproducibility, talking to others or to oneself would be impossible. Even more importantly, we will discover that viewpoint independence and reproducibility not only determine the possibility of talking to each other; they also fix the *content* of what we can say to each other. In other words, we will see in the following that the description of nature follows logically, almost without choice, from the simple fact that we can talk about nature to our friends!

Viewpoints

Tolerance ... is the suspicion that the other might be right.
Kurt Tucholski (1890–1935), german writer.

When in early childhood one starts to meet other people, one quickly finds out that certain experiences are shared, while others, such as dreams, are not. Learning to make this distinction is one of the adventures of human life. On this escalation, we concentrate on a section of the first type of experiences, the *physical* observations. However, even among these, distinctions are to be made. For example, in daily life one is used to imply that weights, volumes, lengths, and time intervals are independent of the viewpoint of the observer. One can talk about these observed quantities to anybody, and there are no disagreements on their values, provided they have been measured correctly.

However, other quantities do depend on the observer. Imagine talking to a friend after he jumped from one of the trees along our path, while he is still falling downwards. He will say that the forest floor is approaching with high speed, whereas the observer below will maintain that the floor is stationary. Obviously, the difference between the statements is due to their different viewpoints. The velocity of an object, in this example that of the forest floor or of the friend itself, is thus a less symmetric property than weight or size. Not all observers agree on the value of velocity, nor even on its direction.

In such a case, understanding is still possible with help of little effort: each observer can *imagine* to observe from the point of view of the other, and *check* whether the imagined result agrees with the statement of the other.* If the thus imagined statement and the actual statement of the other observer agree, the observations are consistent, and the difference in statements is only due the different viewpoints; in the other case the difference is fundamental, and one cannot talk to each other. One can even argue whether human feelings, judgments or taste lead to fundamental differences. Challenge

The distinction between viewpoint invariant and viewpoint dependent quantities is essential. Invariant quantities such as mass or shape describe *intrinsic* properties, and quantities depending on the observer make up the *state* of the system. Therefore, in order to find a *complete* description of the state of physical systems, we have to answer the following questions:

- What type of viewpoints are possible?
- What types of symmetry properties do these viewpoints allow?
- What types of observable quantities are possible?
- How does one translate descriptions from one to the other?
- What follows about motion and from these results?

In the discussion so far, we studied viewpoints differing in location, in orientation, in time and, most importantly, in their motion. With respect to each other, observers can be at rest, move with constant speed, or accelerate. These "concrete" changes of viewpoints are the obvious ones, which we will study first. In this case the requirement of consistency of observations made by different observers is called the *principle of relativity*. The symmetries associated to this type of invariance are also called *external* symmetries, as listed in table 17. See page 122

A second class of fundamental changes of viewpoint concerns "abstract" changes. Viewpoints can differ by the mathematical description used, and then are generally called *changes of gauge*. They will be introduced first in the section of electrodynamics. Again, one is lead to require that all statements must be consistent across different mathematical descriptions. This second requirement of consistency is called the *principle of gauge invariance*. The symmetries associated to this type of invariance are also called *internal* symmetries. The consistency requirements are called 'principles' because these basic statements are so strong that they almost completely determine the "laws" of physics, as will be seen shortly.

* Only at the age of about four years, humans develop the ability to imagine that others can be in situations different from their own. Therefore one deduces that before the age of four, humans are unable to understand special relativity; afterwards, they can. Ref. 113

The third principle, whose importance is not evident from everyday life either, is the behaviour of a system under exchange of its parts; it is called *permutation symmetry*. It is a *discrete* symmetry, and we will encounter it in the second part.

We will also discover in our escalation that looking for a complete description of the state will also yield a complete description of intrinsic properties. But enough of introduction; let's come to the heart of the topic.

Symmetries and groups

A system which appears identical when observed from different viewpoints is said to be symmetric or to possess a *symmetry*. One also says that the system possesses an *invariance* under the specified changes from one viewpoint to the other, which are called *symmetry operations* or *transformations*. A symmetry is thus a set of transformations. But it is more than a set: the concatenation of two elements, i.e. of two symmetry operations is another symmetry operation. To be more precise, a symmetry is a set $G = \{a, b, c, ...\}$ of elements, the transformations, together with a binary operation $\circ$, called *concatenation* and pronounced 'after' in which the following properties hold for all elements a, b, c :

$$\begin{aligned} \textit{associativity}\text{, i.e. } \quad & (a \circ b) \circ c = a \circ (b \circ c) \\ \text{a } \textit{neutral element } e \text{ exists such that} \quad & e \circ a = a \circ e = a \\ \text{an } \textit{inverse element } a^{-1} \text{ exists such that} \quad & a^{-1} \circ a = a \circ a^{-1} = e \quad . \end{aligned} \tag{69}$$

Any set which fulfills these defining properties or axioms is called a *(mathematical) group*. Historically, the notion of group was the first example of a mathematical structure which was
Challenge defined in a completely abstract manner. Can you give an example of a group in daily life? Groups are rather common entities which appear frequently in physics and mathematics, because symmetries are almost everywhere, as we will see.* Can you extract the symmetry
Challenge operations for figure 46?**

Representations

In a symmetric and composed system such as the one shown in figure 46, one quickly
Challenge notices that each of its parts belongs to a set of similar objects, usually called a *multiplet*, which all together have the full object symmetry. For certain coloured patches one needs four objects to make up a full multiplet, whereas for others one needs two or only one, such as for the central star. In fact, in any symmetric system each part can be classified by saying

* In principle, mathematical groups need not be symmetry groups; but one can prove that all groups can be seen as transformation groups on some suitably defined mathematical space, so that in mathematics one can use the terms 'symmetry group' and 'group' interchangeably.

A group is called *abelian* if the multiplication is commutative, i.e. if one has $a \circ b = b \circ a$ for all couples of elements. In this case the multiplication is sometimes called *addition*. A subset $G_1 \in G$ of a group G can itself be a group; one then calls it a *subgroup* and often says sloppily that G is *larger* than G_1 or that G is a *higher* symmetry group than G_1.

** The most beautiful book on this topic is the text by Branko Grünbaum & G.C. Shephard, *Tilings and Patterns*, W.H. Freeman and Company, New York, 1987. It has been translated into several languages and republished several times.

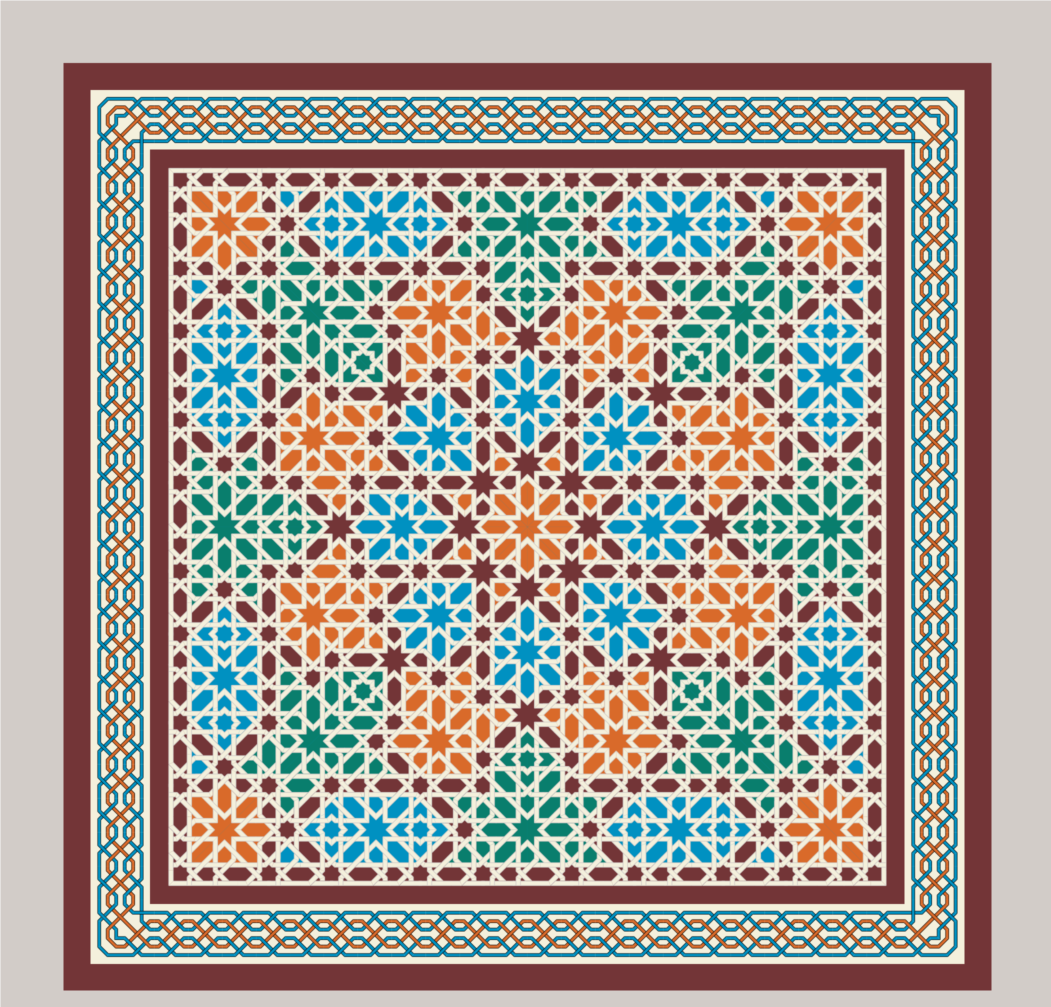


Figure 46 A hispano-arabic ornament from the governor's palace in Sevilla

to what type of multiplet it belongs. Throughout our escalation we will perform the same classification with the parts of nature, with ever-increasing precision.

A multiplet is a set of parts which transform into each other under all the symmetry transformations. Mathematicians often call abstract multiplets *representations*. Multiplets or representations describe in which way this object is part of the whole system. In mathematical language, symmetry transformations are usually described by matrices. For example, in the plane, a reflection along the first diagonal is represented by the matrix Challenge

$$D(\text{refl}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \qquad (70)$$

since every point (x, y) becomes transformed to (y, x) when multiplied by the matrix $D(\text{refl})$. Therefore, for a mathematician a *representation* of a symmetry group G is a mapping D into the group of (non-singular) matrices, where to each group element $a \in G$ is assigned a nonsingular matrix $D(a)$ such that the representation of the product of two elements a and b is nothing else than the product of the representations of the elements

$$D(a \circ b) = D(a)D(b) \quad , \tag{71}$$

together with the requirement that the identity operation of G is mapped to the unit matrix. In even more compact language, a representation is a *homomorphism* from G into the group of non-singular (i.e. invertible) matrices.*

For every symmetry group, the construction and classification of all possible representations is an important task. It corresponds to the classification of all possible part types a symmetric system can be made of. In this way, the classification of all multiplets and parts which can appear in figure 46 will teach us how to classify all possible parts an object or an observation can be made of.

A representation is called *faithful* or *true* or *proper* if it is also an isomorphism. A representation is called *unitary* if all matrices D are unitary.* All representations appearing in physics, with only a handful of exceptions, are unitary: this term is the most restrictive, since it specifies among others that the corresponding transformations are one-to-one and invertible, which means that one observer never sees less or more than another. Obviously, if an observer can talk to a second one, the second one can also talk to the first.

In addition, there is a third important property for a multiplet or representation, namely whether it can be seen as composed of smaller multiplets or not. Mathematicians therefore distinguish *reducible* and *irreducible* representations. The irreducible representations are those which cannot be decomposed any further. For example, the symmetry group of figure 46, commonly called D_4, has the general, faithful, unitary and irreducible matrix representation:

$$\begin{pmatrix} \cos n\pi/2 & -\sin n\pi/2 \\ \sin n\pi/2 & \cos n\pi/2 \end{pmatrix} \text{ for } n = 0..3, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}. \tag{73}$$

The complete list of irreducible representations of this group is given by *singlets*, *doublets*, and *quartets*. Are you able to find them all? These representations allow to classify all

Challenge

* If there exists a mapping f from a group G to another G' such that

$$f(a \circ_G b) = f(a) \circ_{G'} f(b) \quad , \tag{72}$$

the two groups are called *homomorphic*, and the mapping f an *homomorphism*. A mapping which is also one-to-one is called a *isomorphism*.

In the same way as groups, also more complex mathematical structures such as rings, fields and associative algebras may be represented by suitable classes of matrices. A representation of the field of complex numbers is given in appendix D.

* The *transpose* A^T of a matrix A is defined element by element by $(A^T)_{ik} = A_{ki}$. The *complex conjugate* A^* of a matrix A is defined by $A^*_{ik} = A^*_{ki}$. The *adjoint* $A^\dagger$ of a matrix A is defined by $A^\dagger = (A^T)^*$. A matrix is called *symmetric* if $A^T = A$, *orthogonal* if $A^T = A^{-1}$, *hermitian* or *self-adjoint* (the two are synonymous in all physical applications) if $A^\dagger = A$ (hermitian matrices have real eigenvalues), and *unitary* if $A^\dagger = A^{-1}$. Unitary matrices have eigenvalues of norm one; multiplication by a unitary matrix is a one-to-one mapping; therefore the time evolution of physical systems is always described by a unitary matrix. A *real* matrix obeys $A^* = A$, an *antisymmetric* or *skew-symmetric* matrix is defined by $A^T = -A$, an *anti-hermitian* one by $A^\dagger = -A$ and an *anti-unitary* matrix by $A^\dagger = -A^{-1}$. All the mappings described by these special types of matrices are one-to-one. A matrix is *singular*, i.e. not one-to-one, if $\det A = 0$.

white and black types of ribbons that appear in the figure, as well as all colored patches. The most symmetric elements are singlets, the least symmetric ones are the quartets. And the complete system is always a singlet as well. Challenge

With these concepts we are now ready to talk about motion with improved precision.

Table 16 The correspondences between the symmetries of an ornament, of a flower, and those of nature as a whole

Concept/system	Hispano-arabic pattern	Flower	Motion
Structure and components	setof ribbons and patches	set of petals, stem	motion path and observables
System symmetry	pattern symmetry	flower symmetry	symmetry of lagrangian
Mathematical description	D4	C5	in galilean relativity: position, orientation, instant, and velocity changes
Invariants	number of multiplet elements	petal number	number of coordinates, norm of vectors and tensors
Representations	multiplet types	multiplet types	tensors, including scalars and vectors
Most symmetric representation	singlet	part with circular symmetry	scalar
Simplestfaithful representation	quartet	quintet	vector
Leastsymmetric representation	quartet	quintet	no limit (tensor of infinite rank)

Symmetries, motion and galilean physics

We experience every day that we are able to talk to each other about motion. Following the arguments just presented, it must therefore be possible first of all to find an *invariant* quantity which describes it. But we already know the invariant quantity: it is the *action*. Indeed the (galilean) action is a number whose value is the same for each observer at rest. If from the symmetry of figure 46 it was possible to deduce a list of possible multiplets (representations) for its possible building blocks, this must be possible for motion as well. In the case of figure 46, the classifications of the ribbons in singlets, doublets, etc. was deduced from the various possible viewpoints. The building blocks of a moving system, corresponding to the ribbons, are the *observables*. Since we observe that nature is symmetric under many different changes of viewpoints, we can therefore classify all possible observables. To do so, we first need to know all the representations of these transformations.

Experiments show that the transformations under which the world of galilean physics, i.e. our everyday life, does not change are changes in position, in orientation, and in observation time. But all these transformations are different from those of figure 46 in two respects: they are *continuous*, and they are *unbounded*. As a result, the representations will also be concepts which can vary continuously and without bounds: they will be *quanti-*

ties or *magnitudes*, i.e. *numbers*. Note that in this way we have deduced why numbers are necessary in the description of motion.*

In addition, since observers can differ in orientation, the representations will be objects which possess a direction. To make a long story short, the change of observation position, orientation or instant leads to the results that all observables are either 'scalars', 'vectors' or higher order 'tensors.'**

A *scalar* corresponds to a singlet; it is an observable quantity which is the same for all observers. Examples are the mass or the charge of an object, the distance between two points, the distance of the horizon, and many others. Their possible values are (usually) continuous and unbounded. On the other hand, velocity is obviously not a scalar, nor is the coordinate of a point. Can you find more examples and counterexamples?

Challenge

Any quantity which has a length, a direction, and which "stays the same" when one changes viewpoint is a *vector*. For example, the arrow between two fixed points on the floor is a vector. Its length and direction are the same for all observers. On the other hand, the arrow between a tree and the place where a rainbow touches the earth is *not* a vector, since that point does not stay the same for all observers. Mathematicians say that vectors – as well as all other tensors – are objects staying invariant under coordinate transformations. Vectors are characterized by their magnitude, their direction, and their coordinate transformation properties. Velocities of objects and forces are examples of vectors. (Can you show that?) For all vectors their magnitude is a scalar. By the way, a famous and baffling result of 19th century experiments is that the velocity of light is *not* a vector. This mystery will be solved shortly.

Challenge

A (second order) *tensor* is the most general proportionality factor possible between two vectors. If two vector quantities are proportional in the sense that doubling the magnitude of one vector doubles the magnitude of the other, given however that the two vectors are not parallel to each other, then the proportionality factor is a tensor. The proportionality factor between angular momentum and circular frequency for a rotating body, the already introduced *moment of inertia*, is an example of a tensor. Can you name another one? Tensors have magnitudes, directions, and connections between directions.

A *n*th order tensor is the proportionality factor between a first order tensor, i.e. between a vector, and a $n-2$nd order tensor. Vectors and scalars are thus 1st and 0th order tensors. The rank, by the way, also gives the number of indices an observable has. Can you show this?

Challenge

Let us get back to the description of motion. Table 16 shows that in physical systems one always has to distinguish between the symmetry of the observables – corresponding to the symmetry of the ribbons –, the symmetry properties of the path of the motion – corresponding to a particular set of ribbons –, and the symmetry of the lagrangian – corresponding to the symmetry of the complete pattern.

Since the action and the lagrangian must be scalars, and since all observables must be tensors, lagrangians contain sums and products of tensors only in combinations forming scalars, i.e. as scalar products or its generalizations. In short, lagrangians always look like

$$L = \alpha a_i b^i + \beta c_{jk} d^{jk} + \gamma e_{lmn} f^{lmn} + \ldots \qquad (74)$$

* Only scalars, in contrast to vectors and higher order tensors, may also be *discrete* observables.

** Later on, *spinors* will be added to this list, which then is complete.

where the indices always come in matching pairs to be summed over (this is so obvious that they are often simply left out) and the greek letters are some constants. For example, the action of a free point particle in galilean physics was given as

$$S = \int L\, dt = \frac{m}{2} \int v^2\, dt \qquad (75)$$

which is indeed of the form just mentioned. We will encounter several other important cases during our escalation.

By the way, is the usual list of possible observation viewpoints, namely different positions, different observation instants, different orientations, and different velocities, also *complete* for the action (75)? Surprisingly, the answer is no. One of the first who seems to have noted the fact was U. Niederer in 1972. Studying the quantum theory of point particles, he found that even the action of a galilean free point particle is invariant under some additional transformations. If the two observers are described by coordinates $(t, \mathbf{x})$ and (τ, ξ), the action is invariant under the transformations

Ref. 114

Challenge

$$\xi = \frac{\mathbf{R}\mathbf{x} + \mathbf{x}_o + \mathbf{v}t}{\gamma t + \delta} \text{and} \tau = \frac{\alpha t + \beta}{\gamma t + \delta} \quad \text{with} \quad \mathbf{R}^T\mathbf{R} = 1 \text{and} \alpha\delta - \beta\gamma = 1 \quad . \qquad (76)$$

where $\mathbf{R}$ describes the rotation from the orientation of one observer to the other, $\mathbf{v}$ the velocity between the two observers, and $\mathbf{x}_o$ the vector between the two origins at time zero.

The important special cases of these transformations are

$$\begin{array}{llll} \text{The connected, static galilei group} & \xi = \mathbf{R}\mathbf{x} + \mathbf{x}_o + \mathbf{v}t & \text{and} & \tau = t \\ \text{The transformation group SL(2,R)} & \xi = \dfrac{\mathbf{x}}{\gamma t + \delta} & \text{and} & \tau = \dfrac{\alpha t + \beta}{\gamma t + \delta} \end{array} \qquad (77)$$

The latter, three-parameter group includes *spatial inversion*, *dilations*, *time translation*, and a set of time-dependent transformations such as $\xi = \mathbf{x}/t$, $\tau = 1/t$ called *expansions*. Dilations and expansions are rarely mentioned, as they are symmetries of point particles only, and do not apply to everyday life objects and systems. However, this will change later on in our escalation, when more careful considerations will be applied.

Reproducibility, conservation, and Noether's theorem

I will leave my mass, charge and momentum to science.
Graffito

It is now obvious that the reproducibility of observations, i.e. the symmetry under change of instant of time, is also a case of viewpoint independence. (Can you find its irreducible representations?) Therefore it is not necessary to describe it again in much detail. However, there is an additional topic which is of general interest. Challenge

We have seen that symmetry implies invariance. It turns out that for continuous symmetries, this sentence can be made more precise: to any continuous symmetry of the lagrangian there is an associated conserved constant of motion and vice versa. The exact formulation

of this connection is the theorem of Emmy Noether,* who found the result in 1918 when helping Albert Einstein and David Hilbert, who were both struggling and competing at constructing general relativity; however, the result applies to any type of lagrangian.
Ref. 115

Noether investigated continuous symmetries $h(s)$ depending on a continuous parameter s. An example could be the change of position along a curve, where s could label the distance from the starting point. If such a symmetry transforms the generalized coordinates q of the observed system following

$$q \mapsto h_s(q) \quad \text{with} \quad h_o(q) = q \tag{78}$$

without changing the lagrangian $L(q_k, \dot{q}_k, t)$, then there exists an invariant quantity I given by

$$I = \sum_{\mathrm{i}} \frac{\partial L}{\partial \dot{q}_{\mathrm{i}}} \frac{d}{ds} h_s(q_{\mathrm{i}})|_{s=0} \quad . \tag{79}$$

For example, if a lagrangian is symmetrical under space translations, this expression shows that the total momentum is conserved. If the lagrangian is invariant under temporal translations, one finds that total energy is conserved. (Check this!) Since both deductions are also valid in the opposite direction, one also says that energy and momentum are the *generators* of time and space translations.
Challenge
Challenge

The conserved quantities for a continuous symmetry are sometimes called the *Noether charges*, because the term *charge* is used in theoretical physics to designate conserved extensive observables; 'electric charge', 'gravitational charge' (i.e. mass), and 'topological charge' are other common example. What is the Noether charge for rotation invariance?
Challenge

We note that the expression "energy is conserved" has several meanings. First of all, it means that the energy of a *single* free particle is constant in time. Secondly, it means that the energy of a *set* of particles is constant. Finally, it means that the energy of a *system* of particles, i.e. including their interactions, is constant in time, as in the case of a collision. Noether's theorem makes all of these points, as one can verify using the corresponding lagrangians.
Challenge

But Noether's theorem makes, or better repeats, an even stronger statement: if energy were not conserved, time could not be defined. The whole description of nature requires the existence of conserved quantities, as we noticed when we introduced the concepts of object, state, and environment. We also saw that the introduction of time is possible only because in nature there are no surprises. Noether's theorem describes exactly what such a surprise would have to be: the non-conservation of energy. It has never been observed yet.
See page 34
See page 95

Since symmetries are so important for the description of nature, in table 17 we give an overview of all the symmetries of nature we will encounter in the first two parts of our walk, together with their main properties. Except for those marked as 'approximate' or 'speculative', an experimental proof of incorrectness of any of them would be a big surprise indeed.

* Emmy Noether (Erlangen, 1882–Bryn Mayr, 1935), german mathematician. By the way, the theorem is only a sideline in her career which she dedicated mostly to number theory. The theorem also applies to gauge symmetries, where it states that to every gauge symmetry corresponds an identity of the equation of motion, and vice versa.

Summarizing, from the fact that we can talk about nature we can deduce several of its symmetries, in particular its symmetry under time and space translations. From nature's symmetries, using Noether's theorem, we can deduce the conserved charges, such as conservation of energy, of linear and of angular momentum, etc. In other words, the definition of mass, space, and time, together with their symmetry properties, is *equivalent* to the conservation of energy and momentum. Conservation and symmetry are two ways to express the same property of nature.

In other words, our ability to talk about nature means that energy and momentum is conserved. Therefore, to uncover the "laws" of nature, the most elegant way is to search for nature's symmetries. Historically, once this connection had been understood, physics made rapid progress. For example, Albert Einstein discovered the theory of relativity in this way. We will use the same method throughout our walk; in the third part of our escalation we will uncover some symmetries which are much more mind-boggling than those of relativity.

Table 17 The symmetries of relativity and quantum theory and their properties; at the same time, the complete list of *inductions* used in the two fields

Symmetry	type [param. number]	space of action	group topology	possible representations	conserved quantity	vacuum/ matter is symmetric	main effect
Geometric or space-time, external, symmetries							
Time and space translation	$R \times R^3$ [4 par.]	space, time	not compact	scalars, vectors,	momentum and energy	yes/yes	allow everyday
Rotation	SO(3) [3 par.]	space	S^2	tensors	angular momentum	yes/yes	communication
Galilei boost	R^3[3 *par.*]	space, time	not compact	same	center of mass velocity	approximately; at low speeds	
Lorentz	homog. Lie SO(3,1) [6 par.]	space-time	not compact	tensors, spinors	energy-momentum $T^{\mu\nu}$	yes/yes	constant light speed
Poincaré ISL(2,C)	inhomog. Lie [10 par.]	space-time	not compact	tensors, spinors		yes/yes	
Dilation invariance	R^+ [1 par.]	space-time			none	yes/no	massless particles
Special conformal invariance	R^4 [4 par.]	space-time			none	yes/no	
Conformal invariance	[15 par.]	space-time				yes/no	
Dynamical, interaction dependent symmetries: gravity							
$1/r^2$ gravity	SO(4) [6 par.]	config. space			perihelion direction	yes/yes	

Symmetry	type [param. number]	space of action	group topology	possible representations	conserved quantity	vacuum/ matter is symmetric	main effect
Diffeomorphism invariance	[∞ par.]	space-time			locally vanishing energy-momentum divergence	yes/no	perihelion shift
Dynamical, classical and quantum mechanical motion symmetries							
Motion (“time”) reversal T		Hilbert or phase space	discrete	even, odd	T-parity	yes/no	
Parity reversal P		Hilbert or phase space	discrete	even, odd	P-parity	yes/no	
Charge conjugation C	global, antilinear, anti-hermitian	Hilbert or phase space	discrete	even, odd	C-parity	yes/no	
CPT		Hilbert or phase space	discrete	even	CPT-parity	yes/yes	makes field theory possible
Chiral symmetry		Hilbert space	discrete			approximately	“massless” fermions[a]
Dynamical, interaction dependent, gauge symmetries							
Electromagnetic classical gauge inv.	[∞ par]						
Electromagnetic q.m. gauge inv.	abelian Lie U(1) [1 par.]	Hilbert space	circle S^1		electric charge	yes/yes	massless photon
Electromagnetic duality	abelian Lie U(1) [1 par.]		circle S^1			yes/no	
Weak gauge	non-abelian Lie SU(2) [3 par.]	Hilbert space			weak charge	no/ approx.	
Colour gauge	non-abelian Lie SU(3) [8 par.]	Hilbert space			colour	yes/yes	massless gluons
Permutation symmetries							
Particle exchange	discrete	Fock space and simil.	discrete	fermions and bosons	none	n.a./yes	Gibbs’ paradox
Selected speculative symmetries of nature							

Symmetry	type [param. number]	space of action	group topology	possible represen-tations	conserved quantity	vacuum/ matter is symmetric	main effect
GUT	E8, SO(10)	Hilbert				yes/no	coupling constant convergence
N-super-symmetry[b]	global	Hilbert		particles, sparticles	T_{mn} and N spinors[c] Q_{imn}	no/no	"massless"[a] particles
R-parity	discrete	Hilbert				yes/yes	
Braid symmetry						yes/?	
Space-time duality	discrete	all				yes/?	fixes mass?
Event symmetry		space-time				yes/no	

The explanation of the terms in the table will be completed in the rest of the walk. For details about the connection between symmetry and induction, see page 406.
[a] Only approximate; "massless" means that $m \ll m_{\text{Pl}} \approx 22\,\mu\text{g}$.
[b] $N = 1$ supersymmetry, but not $N = 1$ supergravity, is probably a good approximation for nature at everyday energies.
[c] $i = 1..N$.

We now continue with the next method allowing a global description of motion.

Do extended bodies exist?

Strangely enough, this extension of the description of motion is one of the most intensely discussed questions in physics. It appeared again and again, at each improvement of the description, and the answer alternated between the affirmative and the negative. In classical physics the issue appears in several settings.

Mountains, manifolds and fractals

Whenever we climb a mountain, we automatically follow its outline. We usually describe this outline by a curved two-dimensional surface. (Mathematicians speak of a 'differentiable manifold' or of a 'smooth' manifold, a term explained in appendix D.) In everyday life we are used to think that this is a good approximation. But there are alternatives.

The most popular one is the idea that mountains are fractal surfaces. A *fractal* was defined by Benoit Mandelbrodt as a set which is selfsimilar under a countable infinite number of magnification values. An example of an algorithm building a (stochastic) fractal surface is See page 26 shown on the right side of figure 47. It produces shapes which look incredibly similar to real mountains. The results are so realistic that they are used in Hollywood movies. Ref. 116

But worse, even a flat surface could be a fractal, since it could have an infinity of small holes, as shown in the example on the left side of the same figure. In short, fractals are very different from manifolds. Can you devise an experiment to decide whether fractals or manifolds provide the correct description for matter? Or could a mountain be a fractal, even Challenge

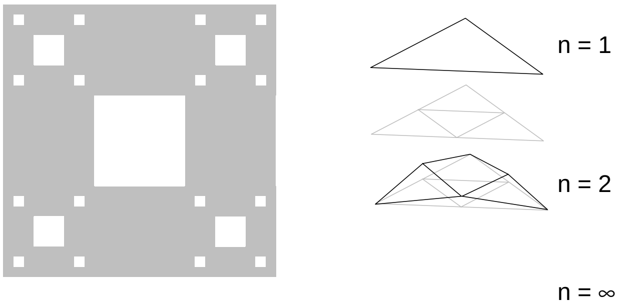


Figure 47 Mountains as fractals

a three-dimensional version, full of holes, thus solving the just mentioned box paradox? Chocolate bars can tell.

Can a chocolate bar last forever?

> From a drop of water a logician could predict
> an Atlantic or a Niagara.
> Arthur Conan Doyle (1859–1930), *A study in Scarlet*.

Any child knows how to make a chocolate bar last for ever: by eating every day only one half of what is left over from the previous day. But this method works only if matter is scale invariant. Fractal matter would be scale invariant for a discrete set of zoom factors, and continuous matter would be scale invariant for any zoom factor. Is this really the case?

Challenge

See page 44

We already encountered a fact making continuity a questionable assumption: continuity would allow, as Banach and Tarski showed, to multiply food and any other matter by smart cutting and reassembling. It would allow children to eat the *same* amount of chocolate every day, without ever buying a new bar. There is only one way to settle the question: by experiment. Let us take fluid chocolate; or even simpler, let us take simply some oil – which is the main ingredient of chocolate anyway – and spread it out over an ever increasing surface. How far can this be done? This question was answered first by the Benjamin Franklin. He spread droplets of oil onto a pond on a day without rain or wind and found that a small droplet of oil cannot cover a surface larger than about – can you guess the value? Trying to spread the film further inevitably rips it apart.

Challenge

In short, experiments show that there is a minimum thickness of oil films, with a value of about 5 nm. This was the first measurement of the smallest size. But its existence – not its value – was already deduced much earlier from observation, namely by Galileo. These arguments were the real reason for his trial.

See page 416

How high can animals jump?

Fleas can jump to heights hundred times their size, humans only to heights about their own size. In fact, biological studies yield a simple observation: all animals, independently of their size, achieve jumping heights of 1.5 ± 0.7 m, whether they are humans, cats, grasshoppers, apes, horses, leopards, etc. Explaining this observation takes only two lines. Are you able to do it?

Ref. 117

Challenge

At first sight, this seems an example of scale invariance. But there are some interesting exceptions. It turns out that they exist at both ends of the mass range. On the small side, mites and other small insects do not achieve such heights because, like all small objects, they encounter the problem of air resistance. At the large end, elephants do not jump that high, because doing so would break their bones. But this connection is interesting: why do bones break at all?

Why are all humans of about the same size? Why are there no giant adults with a height of ten metres? Why aren't there any land animals larger than elephants? The answer yields the key for the understanding of the structure of matter. In fact, the materials of which we are made would not allow such rescaling; for too high people, the bones would collapse. Can you explain why? The answer is obvious but important, because the argument does *not* apply for continuous matter. Challenge

Felling trees

We are still on the first gentle slopes of motion mountain, surrounded by trees. Trees are fascinating structures. Take their size. Why do trees have limited size? Already Galileo knew that increasing the tree height is not possible without limits: such a tree would not have the strength to withstand its own weight. He estimated the maximum height to be 90 m; the actual record, unknown to him at the time, is 152 m. Why does a limit exist at all?

The fact that trees must not break under strong winds also provide restriction. It limits the height to thickness ratio h/d to about 50 for standard sized trees (for $0.2\,\mathrm{m} < d < 2\,\mathrm{m}$). Are you able to deduce this limit? Thinner trees are limited in height to less than 10 m by the fact that they should be able to return to the vertical after being bent by the wind. Challenge Ref. 118

By the way, such studies of natural constraints also answer the question why trees are made from wood and not for example, from steel. Wood is actually the best material for making a light and stiff column. Only recently a few selected engineering composites managed to achieve slightly better performance. Ref. 119

photograph to be included

Figure 48 Atomic steps in broken gallium arsenide crystals seen under a conventional light microscope

But why do materials break at all? All collected data yield the same answer: because there is a smallest size in materials. Continuous matter could not break. Are you able to confirm the deduction?* Challenge

Experimental confirmation is not difficult. If one breaks a thin single crystal in two, such as a gallium arsenide wafer, the breaking surface is either completely flat or shows extremely small steps, as shown in figure 48. These steps are visible in a normal light microscope. It turns out that all observed step heights are multiples of a smallest height; it value is about 1 nm.

* For example, bodies under stress are torn apart at the position at which their strength is minimal. If a body were completely homogeneous, it could not be torn apart. True? Challenge

Listening to silence

Climbing the slopes of our mountain, we arrived in a region of the forest covered with deep snow. Stop one minute and look around this place. It is dark, all the animals are asleep, there is no wind, and there is no source of sound around. We stand still, without breathing, and listen to the silence. (You can also have this experience in a sound studio such as those used for musical recordings, or in your sleeping room at night.) In situations of complete silence, the ear automatically increases its sensitivity;[**] one then makes a strange discovery. One hears two noises, a lower and a higher pitched one, which obviously are generated inside the ear. Experiments show that the lower note is due to pulsating blood streaming through the head, and the higher note is due to the activity of the nerve cells in the inner ear.

This and many other, similar experiments show that whatever one does, one can never eliminate noise from measurements. This unavoidable type of noise is called *shot noise* in physics. When the properties of this type of noise are measured, one can show that they arise because what looks like a continuous flow is in fact a transport of a large number of equal, small, and discrete entities. By listening to noise, one can show that electric current is madded of electrons, that liquids are made of molecules, an that light is made of photons. In a sense, the sound of silence is the sound of atoms.

Little hard balls

> I prefer knowing the cause of a single thing to being king of Persia.
>
> Democritos

All these and many other observations show that matter is neither continuous nor a fractal; matter is made of smallest particles. Galileo, strengthened by the mathematical arguments on giants and trees, called them 'smallest quanta.' Today they are called 'atoms', in honour of a famous speculation of the ancient greeks.

Already 2500 years ago, a group of people started asking the following question. Motion and matter are conserved; however, in nature we observe change and transformation. How does this fit with conservation? The philosophical school of Leucippos of Elea, and Democritos of Elea[*] deduced that there is only one possible solution for this contradiction: nature is made of void and of small, hard, indivisible, and conserved particles.[**] In this way any example of observed motion, change or transformation is due to rearrangements of these particles; change and conservation are reconciled.

In short, matter, being hard, divisible, and having a shape, was imagined as being made of atoms, i.e. of particles which are hard, have a shape, but are indivisible. The greek imagined nature as a big lego set. lego is made of interacting elements which are *impenetrable*, i.e. repulsive at very small distances, *attractive*, i.e. attracting at small distances, and *without interaction* at large distances. Atoms are similar. (Actually, what the greek called 'atoms'

∗∗ It seems that the ear can measure pressure variations of at least as small as20 μPa.

∗ Leucippos of Elea, an island, (ca. 490–ca. 430 BCE); Democritos of Elea, greek philosopher (ca. 460–ca. 356 or 370 BCE), together with his teacher Leucippos founder of the atomic theory; much admired thinker in antiquity, contemporary of Socrates, wrote many books which have been lost, as they have not been copied during the middle ages due to his scientific and thus atheistic world view.

∗∗ A typically male idea, isn't it?

partly corresponds to what today we call ‘molecules’, a term invented and introduced by Amadeo Avogadro in 1811. But let us forget this detailed nitty-gritty for the moment.)

In the nineteenth century, the idea of atoms was first verified by the discovery of the “laws” of chemistry. Since atoms are so small, the experiments showing their existence took many years to convince everybody. We’ll meet another such experiment shortly. But nowadays, with the advances of technology, things are easier. Single atoms can be seen, photographed, hologrammed, counted, touched, moved, lifted, levitated, and thrown around. Ref. 120, 121 And indeed, like matter, atoms have mass, size, shape, and colour. Several sections of modern physics research have fun to play with atoms in the same way that children play with lego.

no figure yet

Figure 49 The principle and the realization of an atomic force microscope

Maybe the most beautiful example for this is provided by the many applications of the atomic force microscope.*** It is a simple device which follows the surface of an object with an atomically sharp needle; such needles, usually of tungsten, are easily fabricated with a simple etching method. The height changes are recorded Ref. 122 with help of a deflected light ray. With a little care, the atoms of the object can be felt, and made visible on a computer screen. With special types of such microscopes, the needle can be used to move atoms one by one to specified places on the surface. Ref. 123 People have also been able to scan a surface, pick up an atom, throw it towards a mass spectrometer to determine what sort of atom it is. Ref. 124

no image yet

Figure 50 The surface of a silicon crystal mapped with an atomic force microscope

As a note, the construction of an atomic force microscope is only a small improvement on what nature is building already by the millions; when we use our ears to listen, we are actually detecting changes in eardrum position of about 1 nm. In other words, we all have two atomic force “microscopes” built into our heads.

In summary, matter is not smooth, nor a fractal, nor scale invariant; it is made of atoms. Different types of atoms, as well as their various combinations, produce different types of substances. Pictures from atomic force microscopes show that size and arrangement of the atoms produce the *shape* and the *extension* of objects.

Material objects are thus somehow organized like lego constructions, but using very small bricks. Lego constructions are made of bricks, i.e. interacting elements which are impenetrable, i.e. repulsive at very small distances, attractive, i.e. attracting at small distances, and

∗∗∗ This is the answer to the question on page 71.

without interaction at large distances, i.e. isolable. The different colours of the bricks correspond to the different types of atoms. lego bricks are thus a correct analogy to show that Ref. 125 shape and substance are inextricably connected in nature. Lego provides a mental image for all properties of *matter* we mentioned so far in our walk:

- its impenetrability,
- its isolability,
- its *interactions*,
- its divisibility, and
- its shape.

These properties are all connected to motion; that will be uncovered in the second part of our escalation. For the moment we only conclude that describing the motion of extended objects can be reduced to the description of the motion of their atoms.* That will be a major theme in the following. One particular consequence of this atomic motion is especially important.

Why are objects warm?

We continue our short strife through the field of global descriptions of motion with an overview of heat and its main concepts. In our escalation, we do not need to know much Ref. 126 about heat. The main points you learned in high school are almost sufficient:

- Macroscopic bodies, i.e. bodies made of many atoms, are described by temperature. Temperature is an aspect of the state of each body. Bodies in contact tend to the same temperature. In other words, temperature describes an equilibrium situation. This is often called the *zeroth law of thermodynamics*.

Challenge • Why does water boil at 99.975 °C instead of 100 °C?

- Heat flows from one body to another, and accumulates. It has no measurable mass.* The content of heat inside a body increases with increasing temperature. We'll see below how this happens precisely.
- Heat is a form of energy. For example, friction creates heat. To heat 1 kg of water by one degree, one needs 4.2 kJ of mechanical energy transformed through friction.*
- The sum of mechanical energy and of thermal energy is constant. This is usually called the *first law of thermodynamics*. Equivalently, it is impossible to produce mechanical energy without paying with thermal energy. Therefore engines need fires and muscles get warm. This is also called the impossibility of building a perpetuum mobile of the first kind. This is an important statement, because among others it means that humanity will stop living one

∗ Studying matter in even more detail yields the now well-known picture that matter, when going to higher and higher magnification, is made of molecules, atoms, nuclei, nucleons, i.e. protons and neutrons, and finally, quarks. In atoms one also finds electrons, and a final type of matter, neutrinos, is observed coming from the sun and from certain types of radioactivity. Even though the fundamental bricks have become smaller with time, the basic idea remained. The second part of our escalation explains this connection in detail; appendix C gives numerical details on the properties of the most important elementary particles.

∗ This might change in near future, when mass measurements improve in precision, allowing to detect relativistic effects.

∗ The first to measure this with precision was the english physicist James Prescott Joule (1818–1889) (pronounced such that it rhymes with 'cool', as his descendants like to stress). His experiments were made public by William Thomson (1824–1907) (later Lord Kelvin), and eventually lead to the naming of the unit of work after him.

Observation	Temperature
Unachievable temperature	0 K
Temperature a perfect vacuum would have at earth's surface	40 zK
Lithium gas in certain laboratories - lowest value achieved by man, and possibly coldest matter system in the universe	1 nK
Temperature of neutrino background in the universe	ca. 2 K
Temperature of photon gas background (or background radiaton) in the universe	2.7 K
Liquid helium	4 K
Oxygen triple point	54.3584 K
Liquid nitrogen	77 K
Average temperature of the earth surface	287.2 K
Human body interior	305.3 K
Water boiling	373.13 K or 99.975 °C
Liquid iron	1 kK
Gold freezing point	1337.33 K
Sun's surface	5.8 kK
Space between earth and moon (no typo)	ca. 1 MK
Sun's centre	20 MK
Inside the JET fusion tokamak	100 MK
Heavy ion collisions - highest man made value	up to 3 TK or 300 MeV
Planck temperature - nature's limit temperature	10^{32} K

Table 18 Some temperature measurements

day. Indeed, we live mostly on energy form the sun; since the sun is finite, it means that its energy content will be consumed one day. Can you estimate when this will happen? Challenge

The study of the details of these topics is called *thermostatics* if it concerns situations at equilibrium, and *thermodynamics* if the situations envisaged are away from equilibrium. In the latter case on distinguishes situations *near* equilibrium, when equilibrium concepts such as temperature can still be used, and situations *far* from equilibrium, such as selforganisation, where such concepts usually cannot be applied.

Brownian motion

In 1827, the english botanist Robert Brown (1773–1858) observed with his microscope that small pollen particles in a liquid never come to rest, but keep executing a random zigzag movement. In 1905 and 1906, Albert Einstein argued that this effect is due to the hits that Ref. 127 the pollen particles undergo from the atoms or molecules of the liquid and proposed an experiment to check this, even though at that time nobody was able to see atoms.

It had already been clear from a long time that if smallest matter particles existed, heat had to be disordered motion of these constituents, and temperature had to be the average energy per degree of freedom of the constituents. The model of the next section shows that Challenge

$$T_{\text{kin}} = \frac{3}{2}kT \qquad (80)$$

where the left T is the kinetic energy per particle, and the right T is temperature.** At a room temperature of 298 K, the kinetic energy is thus 6 zJ.

Expression (80) is valid for gases, liquids, and solids. But calculating the speed of air molecules for room temperature gives values of several hundred meters per second. Why then does smoke from a candle take so long to diffuse through a room? Rudolph Clausius (1822–1888) answered this question in the mid 19th century: diffusion is slowed down by the hits of the air molecules, in the same way that pollen particles are accelerated in liquids.

Obviously, the average distance the pollen particle has moved after n hits will be zero, because all hits are random, even though between two hits the particle may move on average by a distance l. However, one *can* measure the average *square* displacement $< d^2 >$, which, as you should

Challenge

be able to show yourself, is given by

$$< d^2 >= nl^2 \qquad (81)$$

Now, for molecules with an average velocity v one has

figure to be added

Figure 51 A typical path for a particle undergoing brownian motion, its displacement distribution, and its average square displacement

$$nl^2 = vlt \quad , \qquad (82)$$

in other words, the average square displacement increases proportionally with time. If one repeatedly measures the position of a particle, one should find the distribution shown in figure 51 for the probability of the particle of being displaced by a given distance from the starting point. It is called the *(Gaussian) normal distribution*. In 1908, the french physicist

Ref. 128

Jean Perrin (1870–1942) performed extensive experiments in order to test this prediction. He found complete correspondence of equation (82) with observations, thus convincing everybody that brownian motion is indeed due to hits by the molecules of the surrounding liquid, as Einstein had predicted.*

Einstein also showed that the same experiment could be used to determine the amount of

Challenge

molecules in a liter of water. Can you find out how?

Why do balloons take up space?

Only with the idea that matter is made of constituents people were able to understand gases.** In particular, it became clear that the *pressure* of a gas in a container is given

** A *thermodynamic degree of freedom* is, for each particle in a system, the number of dimensions in which it can move plus the number of dimensions in which it is kept in a potential. Particles in a solid have 6, whereas particles in monoatomic gases have 3. An excellent introduction into the physics of heat is the book by Linda Reichl, *A modern course in statistical physics*, Wiley, 2nd edition, 1998.

* In a delightful piece of research, in 1998, Pierre Gaspard & al., Nature *Experimental evidence for microscopic chaos*, **394**, p. 865, 27 August 1998, showed that brownian motion is also chaotic.

** By the way, the word '*gas*' is invented. It was coined by the Brussels alchemist and physician Johan Baptista von Helmont (1579–1644), to make it sound similar to 'chaos'. It is one of the few words which have been invented by a particular person and which then were adopted all over the world.

by the steady flow of constituents hitting the wall. It is then not difficult to show that if the constituents are assumed to behave as small hard balls, i.e. for the so-called *ideal gas*, the quantities pressure p, volume V, and temperature T must be related by Challenge

$$pV = \frac{3}{2} NkT \qquad (83)$$

where N is the number of particles contained in the gas and the so-called *Boltzmann constant* $k = 1.4 \cdot 10^{-23}\,\mathrm{J/K}$ is the standard conversion factor between temperature and energy.

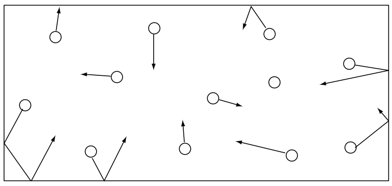


Figure 52 The basic idea of statistical mechanics

Relation (83) is indeed confirmed by experiment, and thus provides another argument for the existence of atoms and their behaviour as normal, though small objects. (Can you imagine how N may be determined experimentally?) Challenge

This model for gases also helps to decide questions such as the following. Take two *identical* rubber balloons, one filled up to a larger size than the other, and connect them via a pipe and a valve. The valve is opened. Which one deflates? Challenge

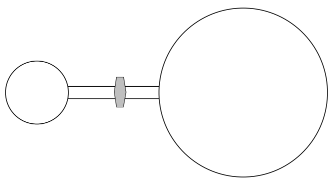


Figure 53 Which one wins?

Expression (83) allows an easy measurement of temperature itself, and that is indeed the way temperature has been defined and measured for about a century. However, the approximation of hard constituents without any long distance interactions is not always valid at low temperatures; but working with improvements of the ideal gas relation, taking into account the deviations due to interactions between the atoms or molecules, is now standard practice in this domain.* Ref. 129 Ref. 130

Now you are able to take up the following challenge: how can you measure the weight of a car with a ruler only? Challenge

Relation (83) also implies that the temperature scale has an absolute minimum value when, sloppily speaking, all the balls are at rest. That happens, as is well known, at $T = 0$, i.e. at $-273.15\,°\mathrm{C}$. The effects observed near that domain, such as the solidification of air, frictionless electrical current transport, or frictionless liquid flow, form a fascinating world of their own; however, the beautiful domain of low temperature physics will not be explored during this walk. Ref. 131, 132

Curiosities around heat

Heat is disordered motion. Nevertheless, heat follows simple rules.

• The highest recorded air temperature in which a man survived is 127 °C. This was tested in 1775 in London, by the secretary of the Royal Society, Blagden, together with a few friends, who remained in a room of that temperature for 45 minutes. Interestingly, the steak Ref. 133

* Liquid pressure depends on height; for example, if the average human blood pressure at the height of the heart is 13.3 kPa, can you guess what it is inside the feet when standing? Challenge

which he had taken with him was cooked "well done" when he left the room. What conditions had to be strictly followed in order to avoid cooking the people in the same way as the
Challenge steak?

Challenge • Can one fill a bottle with 1 ± 10^{30} kg of water?

• If you do not like this text, you can use it for other aims. For example, you can use the paper to make a cup, as shown in figure 54, and boil water in it over an open flame.
Challenge However, to succeed, you have to be a little careful. Can you find out how?

• One gram of fat contains 38 kJ of chemical energy (or, in old units more familiar to nutritionists, 9 kcal). That is the same value as that of car fuel.

• A famous exam question: How does one measure the height of a building with a barometer, a rope, and a ruler? Find at least
Challenge six different ways.

• What is the probability that out of one million throws of a
Challenge coin you get exactly 500 000 heads and as many tails? (You may want to use Stirling's formula: $n! \approx \sqrt{2\pi n}\,(n/e)^n$.)

• All friction processes, such as osmosis, diffusion, evaporation, or decay, are *slow*. They have a characteristic time. It turns out that any (macroscopic) process with a time-scale is irreversible. In short, at everyday scale, everything which takes time is irreversible. Well, this is not real news: we know that intuitively; in this world, undoing things always more time that doing them. That is the second "law" of thermodynamics.

• Mixing 1 kg of water at 0 °C and 1 kg of water at 100 °C gives 2 kg of water at 50 °C. What is the result of mixing 1 kg of
Challenge *ice* at 0 °C and 1 kg of water at 100 °C?

• It turns out that storing information into a memory can be done with negligible entropy generation. However, *erasing information* requires entropy. This is one of the prettier results of
Ref. 134 the discussions into irreversibility of macroscopic motion.

• Why aren't there any small humans, e.g. 10 mm in size? For such small people, the body temperature would fall too low through cooling. (In fact, there are no warm blooded animals of that size of any kind at all.)

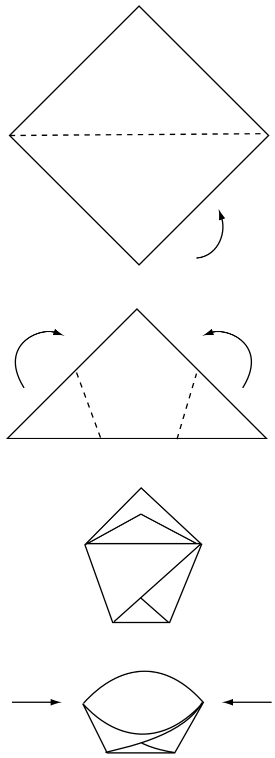


Figure 54 Are you able to boil water in this paper cup?

• Shining rapidly interrupted light on a body of matter produces sound. This is called the *photoacoustic effect*, and is due to the thermal expansion of the material. By changing the frequency of the light, and measuring the intensity of the noise, one gets a characteristic photoelectric spectrum for the material. This method allows to detect gas concentrations in air of 10^{-9}, and is used among others to study the gases emitted by plants. Plants emit methane, alcohol, and acetaldehyde in small quantities; this technique can detect them and help understanding the processes behind them.

• What is the rough probability that all oxygen molecules in the air move away from a
Challenge given city for a few minutes, killing all people?

• If one pours a liter of water into the sea, stirs thoroughly and takes a liter out of the
Challenge mixture, how many of the original atoms will one find?

Domain	extensive quantity i.e. energy carrier	current i.e. flow intensity	intensive quantity i.e. driving strength	energy flow i.e. power	resistance to transport i.e. intensity of entropy generation
Rivers	mass m	mass flow m/t	height difference gh	$P = gh\,m/t$	$R_{\mathrm{m}} = ght/m$ $[\mathrm{m^2/s\,kg}]$
Gases	volume V	volume flow V/t	pressure p	$P = pV/t$	$R_{\mathrm{f}} = pt/V$ $[\mathrm{kg/s\,m^5}]$
Mechanics	momentum $\mathbf{p}$	force $\mathbf{F} = d\mathbf{p}/dt$	velocity $\mathbf{v}$	$P = \mathbf{v}\mathbf{F}$	$R_{\mathrm{p}} = t/m$ $[\mathrm{s/kg}]$
	angular momentum $\mathbf{L}$	torque $\mathbf{M} = d\mathbf{L}/dt$	angular velocity $\boldsymbol{\omega}$	$P = \boldsymbol{\omega}\mathbf{M}$	$R_{\mathrm{L}} = t/mr^2$ $[\mathrm{s/kg\,m^2}]$
Electricity	charge q	electrical current $I = dq/dt$	electrical potential U	$P = U\,I$	$R = U/I$ $[\Omega]$
Thermodynamics	entropy S	entropy flow $I_S = dS/dt$	temperature T	$P = T\,I_S$	$R_S = Tt/S$ $[\mathrm{K^2/W}]$
Chemistry	amount of substance n	substance flow $I_n = dn/dt$	chemical potential μ	$P = \mu\,I_n$	$R_n = \mu t/n$ $[\mathrm{Js/mol^2}]$

Table 19 Extensive quantities in nature, i.e. quantities which *flow* and *accumulate*

• How long could you breathe in the room you are in, assuming it would be air tight? Challenge

• If heat is disordered motion of atoms, there is a big problem. When two atoms collide head-on, in the instant of smallest distance, none has velocity. Where did the kinetic energy go? The standard answer is: it is transformed into potential energy. But that implies that atoms can be deformed, that they have internal structure, and thus that they can be split. In short, if heat is disordered atomic motion, *atoms are not indivisible*! In the 19th century this argument was brought forward as a supposedly absurd consequence in order to show that heat must be something else; instead it turned out that heat is kinetic energy, and that atoms were indeed divisible. The issue will be closed in the second part of our escalation.

Entropy

- It's irreversible.
- Like my raincoat!
Mel Brooks, *Spaceballs*, 1987.

Every domain of physics describes change with two quantities: with energy and with an extensive quantity characteristic of the domain. An overview is given in table 19. It turns out that even though heat is a form of energy, the quantity physicists call heat is *not* an extensive quantity. For example, when physicists' heat is added to transform liquid water into steam, the heat is *not* contained in the steam. The extensive quantity corresponding to what is called "heat" in everyday language is called *entropy*,* in the same way as momentum is the Ref. 135

* The term 'entropy' was invented by the german physicist Rudolph Clausius (1822–1888) in 1865. He formed it from the greek ἐν 'in' and τρόπος 'direction', to make it sound similar to energy. It always had the meaning given here.

extensive quantity describing motion. When two objects differ in temperature, an entropy flow takes place between them, like the flow of water between two lakes of different height, or like the flow of momentum taking place when two objects of different speed collide. Let us define the concept more precisely and explore its properties in more detail.

First of all, the entropy is proportional to the volume of the system under consideration. Like any other extensive quantity, entropy can be accumulated in a body; it can flow into or out of bodies. When the water is transformed into steam, the entropy added is still contained in the steam. In short, entropy should be called ‘heat’. The next question is how entropy behaves.

– CS – Some parts to be filled in. – CS –

When a piece of rock detaches from a mountain, it falls, tumbles into the valley, heats up a bit, and eventually stops. The opposite process, that a rock cools and tumbles upwards, is never observed. Why? The opposite motion does not contradict any rule or pattern about
Challenge motion that we have deduced so far.

The main reason lies hidden in the fact that mountains, valleys and rocks are made of many particles. Motion of many particle systems, especially in the domain of thermostatics, are called *processes*. Central to thermostatics is the distinction between *reversible* processes, such as the flight of a stone, and *irreversible* processes, such as the mentioned example. Irreversible processes are all those processes in which friction and its generalisations play a role. By the way, they are important: if there were friction, shirt buttons and of shoelaces would not work, one could not walk or run, coffee machines would not work, and maybe most importantly of all, our memory would not work, as we will see.

Irreversible processes transform macroscopic motion into the disorganized motion of all the small microscopic components involved. It is therefore not impossible to reverse irreversible motion; it is only highly improbable. Can one measure the degree of irreversibility? Yes, entropy does precisely this; in a sense it measures the degree of decay.

The second “law” of thermodynamics states that “entropy isn’t what it used to be.” More precisely, *the entropy in a closed system tends towards its maximum*. Here, by a *closed system* one means a system which does not exchange energy or matter with its environment.
Challenge Can you mention an example?

A simple consequence of the second “law” of thermodynamics is that *white colour does not last*. Whenever disorder increases, the colour white becomes “dirty”, usually grey or brown. Perhaps for this reason white objects, such as white clothes, white houses, white underwear, are so valued in our society; all white objects defy decay.

Entropy allows to define the concept of equilibrium more precisely as the state of maximum entropy. Note that for any system whose entropy S increases, the increase is due to contact with some bath. (More precisely, this is correct for systems which are not driven by
Ref. 136 outside influences; in this latter case, entropy can be produced inside the system.)

It is a everyday experience that in a closed system, the disorder increases with time, until it reaches some maximum. To reduce disorder, one needs effort, i.e. work and energy. In other words, in order to reduce the disorder in a system, one needs to connect the system to an energy source in some smart way.

– CS – Some parts to be added. – CS –

Transport of an extensive quantity always includes friction. That means that the flow of entropy itself produces additional entropy. For example, when a house is heated, entropy is produced in the wall. Heating means to keep a temperature difference between the interior and the exterior of the house. The heat flow J traversing a square meter of wall is given by

$$J = \kappa(T_\mathrm{i} - T_\mathrm{e}) \tag{84}$$

where κ is a constant characterizing the insulating power of the wall. It turns out that the wall produces an entropy σ proportional to the difference of entropy flows between the interior and the exterior, in other words,

$$\sigma = \frac{J}{T_\mathrm{e}} - \frac{J}{T_\mathrm{i}} = \kappa \frac{(T_\mathrm{i} - T_\mathrm{e})^2}{T_\mathrm{i} T_\mathrm{e}} \quad . \tag{85}$$

Note that we assumed that we are near equilibrium (in each slice) in this calculation, a reasonable assumption in everyday life. A typical case is a good wall with $\kappa = 1\,\mathrm{W/m^2K}$ insulating between 273 K and 293 K. One gets the value

$$\sigma = 5 \cdot 10^{-3}\,\mathrm{W/m^2 \cdot K} \quad . \tag{86}$$

How does the amount of entropy produced in the flow compare with the amount transported? In comparison, clothes typically have $\kappa = 0.1\,\mathrm{W/m^2K}$, whereas warm duvets have $\kappa = 1.5\,\mathrm{W/m^2K}$, which in shops is also called 15 tog.* Challenge

As a note, the insulation of materials is usually measured by the constant $\lambda = \kappa d$ which is independent of the thickness d of the insulating layer. Values in nature range from about $2000\,\mathrm{W/m \cdot K}$ for diamond, which is the best conductor of all, down to $0.1\,\mathrm{W/m \cdot K}$ to $0.2\,\mathrm{W/m \cdot K}$ for wood, between $0.015\,\mathrm{W/m \cdot K}$ and $0.05\,\mathrm{W/m \cdot K}$ for wools, cork and foams, and $5 \cdot 10^{-3}\,\mathrm{W/m \cdot K}$ for Krypton.

There are two other ways, apart from heat conduction, to transport entropy: *convection*, i.e. matter transport, and *radiation*, which is possible also through empty space. For example, the earth radiates about $1.2\,\mathrm{W/m^2K}$ into the cosmos, in total thus about $0.51\,\mathrm{PW/K}$. If more entropy had to be radiated away, the temperature of the surface of the earth would have to increase. Let's hope that the effect remains small in the near future.

Once it became clear that heat and temperature are due to the motion of microscopic constituents, people directly asked what entropy was microscopically. The answer can be formulated in various ways; the two most extreme answers are:

- Entropy is the expected number of yes-no questions, multiplied by $k \ln 2$, needed to be answered for knowing everything about the system, i.e. for knowing its microscopic state.
- Entropy measures the (logarithm of the) probability of the microscopic state.

In short, entropy is higher the more microstates are possible. Through either of these definitions, entropy measures the quantity of randomness in a system; in other words, it measures the transformability of energy; higher entropy means lower transformability. For example,

* That unit is still better than the official (not a joke) $\mathrm{BthU \cdot h/sqft/cm/^\circ F}$ used in some remote provinces of this planet.

when a molecule of glucose (a type of sugar) is produced by photosynthesis, about 40 bits of entropy are released. This means that after the glucose is formed, 40 additional yes/no questions must be asked to know again the full microscopic state of the system. Physicists often use a macroscopic unit; most systems are large, and thus 10^{23} bit are abbreviated as $1\,\mathrm{J/K}$.**

Entropy is thus a specific measure for the characterization of disorder. In this context, Ref. 137 there are a few points worth mentioning. First of all, entropy is not *the* measure for disorder, but *one* measure for disorder. It is therefore *not* correct to use entropy as a *synonym* for the concept of disorder, as is often done in the popular literature. Entropy is only defined for systems which have a temperature, in other words, only for systems which are in or near equilibrium. (For systems far from equilibrium, no measure for disorder has been found yet; probably none is possible.) In fact, the use of the term entropy has degenerated so much that sometimes one has to call it *thermodynamical* entropy for clarity.

Secondly, entropy is related to information *only if* information is defined also as $-k \ln W$. To make this point clear, take a book of about one kilogram of mass. At room temperature, its entropy content is about $4\,\mathrm{kJ/K}$. The printed information inside a book, say 500 pages of 40 lines with each 80 characters out of 64 possibilities, corresponds to an entropy of $4 \cdot 10^{-17}\,\mathrm{J/K}$. In short, what in everyday life is usually called 'information' is a negligible fraction of what a physicist calls information. Entropy is defined using the *physical* concept of information.

Finally, entropy is thus *not* a measure for what in normal life is called the *complexity* of a Ref. 138 situation. In fact, nobody has yet found a quantity describing this everyday experience. The Challenge task is surprisingly difficult. Have a try!

Challenge By the way, does it make sense to say that the universe have an entropy?

In summary, if you hear sentences where the term entropy is used with a different meaning than $S = k \ln W$, beware. Somebody is trying to get you, probably with some ideology.

Do isolated systems exist?

In all the discussions so far, we assumed that we could distinguish the system under investigation and keep it apart from the environment. In fact we assumed that at least in principle such *isolated systems*, i.e. systems not interacting with their environment, actually exist. Probably our own human condition was the original model: we do experience having the possibility to act independently of the environment. Following this idea, an isolated system is a system not exchanging energy or matter with its environment. Experiments for many centuries had shown no reason to question this definition. By the way, are you able to decide Challenge whether the universe is an isolated system or not?*

The concept of isolated system had to be refined somewhat with the advent of quantum mechanics. Nevertheless, the concept still remains valid and useful also in quantum theory. However, in the third part of our walk, the situation will change drastically. The investigation of whether the universe is an isolated system will lead to surprising results. We'll take the first steps towards the answer shortly.

Challenge ** This is only approximate; can you find the correct number?

* Hint: your answer is most probably wrong.

Why can't we remember the future?

It's a poor sort of memory which only works backwards.
Lewis Carroll (1832–1898), *Alice in Wonderland*

In the section where time was introduced, right from the start we ignored the difference between past and future. Obviously, there is a difference: we do not have the ability to remember the future. This is not simply a limitation of our brain alone. Also the devices around us, such as tape recorders, photographic cameras, newspapers, and books only tell us about the past. Is there a way to build a *video recorder* with a 'future' button? To find out, one only has to think about the following issue: how would such a device distinguish between the near and the far future? It does not take much to find out that any way to do this Challenge conflicts with the second law of thermodynamics. Bad luck. By the way, one would need precisely the same device device to show that there is faster than light motion. Can you see the connection? Challenge

In summary, the future cannot be remembered because entropy in closed systems tends towards a maximum. Put even more simply, memory exists because the brain is made of many particles, and for the very same reason the brain is limited to the past. However, for the most simple types of motion, when only a few particles are involved, the difference between past and future we note in everyday life disappears. For few particle systems, there is no difference between times gone by by and times approaching. Even more sloppily, the future differs from the past only in our brain, or equivalently, only because of friction. Therefore the difference between the past and the future is not mentioned frequently in this walk, even though it is an essential part of our human experience. But the fun of the present adventure is precisely to overcome our own limits.

To live, change the past –
and cut and paste it into the present.

Is everything made of particles?

A physicist is an atom's way of knowing about atoms.
George Wald

Historically, the study of statistical mechanics has been of fundamental importance for physics. It was the first time one could demonstrate that physical objects are made of interacting particles. The story of the topic is in fact a long chain of arguments showing that every single property we ascribe to objects, such as size, stiffness, colour, mass density, magnetism, heat conductivity, electrical conductivity, etc., results from the interaction of the many particles they consist of. (Some of these properties need the quantum version of statistical mechanics.) The discovery that *all objects are made of interacting particles* has often been called the main result of modern science.

This conceptual success has also lead many people to generalize it to the statement: "Everything we observe is made of parts." This approach has been applied to chemistry with molecules,* to material science and geology with crystals, to electricity with electrons, to atoms with elementary particles, to space with points, to time with instants, to light with photons, to biology with cells, to genetics with genes, to neurology with neurons, to mathematics with sets and relations, to logic with elementary propositions, and even to linguistics with morphemes and phonemes. All these sciences have flourished on the idea that everything is made of related parts. The basic idea seems so self-evident that we have difficulties even in naming an alternative. Just try.

Challenge

However, as we will see later on, in the case of the whole of nature this idea is incorrect. It turns out to be a prejudice. A prejudice so entrenched that for at least thirty years it has retarded further developments in physics. In particular, it does *not* at all apply to elementary particles and to space-time. Finding the correct description makes our escalation a challenging adventure, especially in the third part.

> Jede Aussage über Komplexe läßt sich in eine Aussage über deren Bestandteile und in diejenigen Sätze zerlegen, welche die Komplexe vollständig beschreiben.*
> Ludwig Wittgenstein, *Tractatus*, 2.0201

Selforganisation

> The study of non-linear physics is like the study of non-elephant biology.

Selforganisation is the most general generalization in the description of motion. It studies the appearance of order. *Order* is taken here as a collective term for *shape*, such as the complex symmetries of snow flakes, for *pattern*, such as the stripes of zebras, and for *cycle*, such as the creation of sound when singing. Many studies into the conditions under which these types of order appear or disappear have shown that the same concepts are necessary, independently of the actual system. You might check that every example of what we call *beauty* is a combination of shapes, patterns, and cycles. Selforganisation can thus simply be called the study of the origin of beauty.

Challenge

The topic is vast, since it touches all aspects of growth, such as the cell differentiation in an embryo inside a woman's body, the formation of colour patterns on tigers, tropical fish and butterflies, the formation of the symmetrical arrangements of flower petals, or the formation of biological rhythms.

The behaviour of fluids, such as the flickering of a burning candle, the flapping of a flag in the wind, the regular stream of bubbles emerging from small irregularities in the surface of a champagne glass, or the regular or irregular dripping of a water tap, provides a large amount of phenomena where the conditions for appearance and disappearance of order can be studied.

All growth processes are selforganisation phenomena. Have you ever pondered about how teeth grow, a practically inorganic material, forming shapes in the the upper and the lower

* An fascinating introduction into Chemistry is the text by John Emsley, *Molecules at an Exhibition*, Oxford University Press, 1998.

* Every statement about complexes can be resolved into a statement about their constituents and into the propositions that describe the complexes completely.

rows fitting exactly into each other? Also the formation, before and after birth, of neural networks in the brain is a selforganisation process. Even the physical processes at the basis of thinking, with all its changing electrical signals, are at least partly to be described along these lines.

A special case of growth is biological evolution. Even though its mechanisms are rarely described quantitatively, the topic can be fascinating nevertheless. It is often concerned with shapes and explains for example that snakes tongues are forked because that is the most efficient shape for following chemical trails left by prey and conspecifics. Ref. 139

Figure(s) to be added

Figure 55 Examples of selforganisation for sand

Many problems of selforganisation are mechanical problems, such as the formation of mountains when continents move, the creation of earth quakes, or the creation of regular cloud arrangements in the sky. Pondering the mechanisms behind the formation of clouds one sees from an airplane can transform a boring flight into a fascinating intellectual adventure. Challenge

A simple system showing all the richness of the topic is the study of plain sand. Why do sand dunes have ripples, why does the sand floor at the bottom of the sea? People are also interested to know how avalanches form on steep sand heaps, how sand behaves in hourglasses, in mixers, in vibrating containers, etc. For example, only recently Umbanhowar and Swinney have found that when Ref. 140 one shakes a flat container with tiny bronze balls (less than a millimeter in diameter) up and down in vacuum, the surface of this bronze 'sand' shows stable heaps at certain frequencies. These heaps, "oscillons", also bob up and down. The heaps can move and interact with other heaps. Sand is proving to be such a beautiful and fascinating topic that the prospect of each human, namely of once becoming dust again, does not look so grim any more.

Figure(s) to be added

Figure 56 An oscillon formed by shaken bronze balls

A second simple but beautiful example is the effect discovered in 1999 by Klaus Kötter and his group. They discovered that the behaviour of a set of spheres swirled in a dish depends on the number of spheres used. For certain 'magical' numbers, such as 21, stable ring patterns emerge, but not for others. The rings are best visualized by coloring the spheres. Ref. 141

As a whole, the study of selforganisation has changed the way nature is to be described in a number of ways. First of all, the field has shown that patterns and shapes are similar to cycles: all are due to motion. Without motion, there are no patterns nor shapes. Every Ref. 142

pattern has a history; every pattern is an example of motion.

Secondly, in all complex systems the motion is an organized motion of – usually small – constituents; systems which selforganize are always *cooperative structures*.

Thirdly, all these systems show evolution equations which are *nonlinear* in the configuration variables. Linear systems do not selforganize. Many selforganizing systems therefore also show *chaotic* motion.

Figure to be included

Figure 57 Magic numbers of spheres swirled in a dish

Moreover, all order and all structure appears when two general types of motion compete with each other, namely a "driving", energy adding process, and a "dissipating", braking mechanism. There is no selforganisation without thermodynamics playing a role. Selforganizing systems are always *dissipative systems* and far from equilibrium. When both the driving and the dissipation are of the same order of magnitude, and when the key behaviour of the system is not a linear function of the driving action, order may appear.*

For example, all selforganizing systems at the onset of order appearance, can be described by equations of the general form

$$\frac{\partial A(t,x)}{\partial t} = \lambda A - \mu |A|^2 A + \kappa \Delta A + \text{higher orders} \tag{87}$$

Here, the possibly complex observable A is the one which appears when order appears, such as an oscillation amplitude or a pattern amplitude. One notes the driving term λA in which λ describes the strength of the driving, the nonlinearity in A, and the dissipative term. In cases that the dissipative term plays no role ($\kappa = 0$), when λ increases above zero, a *temporal*
Challenge oscillation, i.e. a stable cycle with nonvanishing amplitude appears.

In case that the diffusive term does play a role, equation (87) describes how an amplitude for a *spatial* oscillation appears when the the driving parameter becomes positive, as the
Challenge solution $A = 0$ becomes unstable.

The onset of order is called a *bifurcation*, because at these critical values the situation with amplitude zero, i.e. the disordered state, becomes unstable, and the ordered state becomes stable. *In nonlinear systems, order is stable.* This is the main conceptual result of the field. But the equation (87) and its numerous variations allow to describe many additional phenomena, such as spirals, waves, hexagonal patterns, topological defects, some forms of
Ref. 143 turbulence, etc. The main point is to distill the observable A and the parameters λ, μ and κ from the physical system under consideration.

* When one wishes to describe the "mystery" of human life, often terms like 'fire', 'river', or 'tree' are used as analogies. They all are examples of selforganized systems; they have many degrees of freedom, have competing driving and breaking forces, depend critically on the initial conditions, show chaos and irregular behaviour, and sometimes show cycles and regular behaviour. Humans and their life resemble them in all these aspects; thus there is a solid basis to their use as metaphors. One could even go further and speculate that pure beauty *is* pure selforganisation. The lack of beauty indeed often results from a disturbed equilibrium between external breaking and external driving.

Finally, selforganisation is of interest also for a more general reason. Sometimes one hears that the ability to formulate the patterns or rules of nature from observation does not include the ability to predict *all* observations from these rules. In this view, so-called "emergent" properties exist, i.e. properties appearing in complex systems as something *new* that cannot be deduced from the properties of their parts and their interactions. (The ideological background of these views is obvious; it was the last try to fight the deterministic description of the world.) The study of selforganisation has definitely settled this debate. The properties of water molecules do allow to predict Niagara falls[**] and the diffusion of signal molecules determines the developments of a single cell into a full human being. In particular, cooperative phenomena determine the place where arms are formed, they ensure the (approximate) right-left symmetry of human bodies, prevent mix-ups of connections when the cells in the retina are wired to the brain, and explain the patterns on zebras, to cite only a few examples. Similarly, the mechanisms behind the heart beat and many other cycles have been deciphered.

Selforganisation provides the general principles which allow to predict the behaviour of complex systems of any kind. They are presently being applied to the most complex system in the universe: the human brain. The details of how it learns to coordinate the motion of the body, and how it extracts information from the images in the eye are being studied. The work ongoing in this domain is fascinating.

In any case, the lack of understanding of complex systems in the past was due mainly to the restrictive teaching of the subject, which usually concentrated – as does this walk – on examples of motion in simple systems. Even though selforganisation is and will remain fascinating for many years to come, we now leave it and continue with our own adventure on the fundaments of motion.[*]

5. Limits of galilean physics

I only know that I know nothing.
Socrates (470–399 BCE), as cited by Plato.

We give a few proofs for the truth of this quote in the case of galilean physics, despite its general success in engineering and everyday life.

Research topics in classical dynamics

Even though mechanics is now several hundred years old, research into its details is still not concluded.

• Above already we mentioned the study of the stability of the solar system. Its future is unknown. In general, the behaviour of few body systems is still a research topic of mathematical physics. Answering the simple question on how long a given set of bodies gravitat- Ref. 145

** Already small versions of Niagara fall, namely dripping water taps, show a large range of cooperative phe-
Ref. 144 nomena, including the chaotic, i.e. non-periodic, fall of water drops. This happens when the water flow has the
Challenge correct value, as you might want to test in your own kitchen.

* An important case of selforganisation is *humour*. An overview of science humour can be found in the famous anthology compiled by R.L. Weber, edited by E. Mendoza, *A random walk in science*, Institute of Physics, London 19▷▷. It exists in several expanded translations.

ing around each other will stay together is a formidable challenge. This so-called *many-body problem* is one of the seemingly neverending stories of theoretical physics.

• The spinning top, and in general the description of rotating bodies, including bicycles, motorcycles and other such dangerous contraptions is still ongoing. The mathematical difficulties of this topic fascinate many.

• The challenges of selforganisation, of nonlinear evolution equations, and of chaotic motion are still plenty and motivate numerous researchers.

• Perhaps the toughest of all problems in modern physics is the description of turbulence. When the young Werner Heisenberg was asked to do a Ph.D. on turbulence, he refused – rightly so – saying it was too difficult; he turned to something easier and discovered quantum mechanics instead. Turbulence is such a vast topic, with many of the concepts still not settled, that only now, towards the end of the twentieth century, its secrets start to be unraveled, despite the fact that its applications are numerous and important. And the math behind it is mind boggling. The topic is still in flux. Ref. 146

What is contact?

Near the start of our escalation we defined mass through the measurement of velocity changes during collisions. But why do objects change their motion in such instances? Why are collisions among two balls made of chewing gum different from those between two stainless steel balls? What happens during those moments of contact? See page 55

Obviously, contact is related to material properties, and these influence motion in a complex way. The complexity is so large that the sciences of material properties developed independently from physics for a long time; for example, the techniques of metallurgy – often called the oldest science of all – of cooking and of chemistry were related to the properties of motion only in the twentieth century, after having been independent for thousands of years. But if material properties determine the essence of contact, we *need* knowledge about matter and about materials to understand mass, and thus motion. The second part of the escalation will uncover the connections.

Precision and accuracy

When we started climbing motion mountain, we stated that to gain height means to increase the *precision* of our description of nature. Well, this statement is itself not precise. It turns out that one has to distinguish two often confused terms: *precision* is the degree of reproducibility; *accuracy* is the degree of correspondence to the actual situation. Both concepts apply to measurements, to statements, and to physical concepts.* Therefore, in our walk concepts have mainly to be *precise*, and descriptions have to be *accurate*. Inaccuracy is a proof of lack of understanding. Instead of 'inaccurate' we simply say *wrong*. In other words, in our walk, more than to an increase in the precision of our description of nature, we actually aim at increasing its accuracy.

What should one think of a car company claiming that the air friction coefficient c_w is 0.375? Or of the fact that the world record in fuel consumption for cars is given as Challenge

* For measurements, both precision and accuracy are best described by their *standard deviation*, as explained in appendix B, on page 731.

2315.473 km/l? Or of a census bureau giving the population of a country with a precision of one person? One lesson one gains from the investigations into measurement errors is that one should never provide more digits for results than one can put one's hand into fire for.

Taking an overview of the most precise and accurate measurements at present, one finds that the record number of digits is 14. Why so few? Classical physics doesn't cover the issue. What is the maximum number of digits one can expect in measurements, what is its origin, and how can one achieve it? These questions are still open at this point of our escalation; they will be covered in the second part of our escalation. See appendix B

Why is measurement possible?

In the description of gravity given so far, the one that everybody learns – or should learn – at high school, acceleration is connected to mass and distance via $a = GM/r^2$. That's all. But this simplicity is deceiving. In order to check whether this description is correct, one has to measure lengths and times. However, it is *impossible* to measure lengths and time intervals with any clock or any meter bar based on the gravitational interaction alone! Try to conceive such an apparatus and you will be inevitably lead to disappointment. One always needs a non-gravitational method to start and stop the stopwatch. Similarly, when one measures length, e.g. of a table, one has to put a meter bar or some other device along it. The interaction necessary to line up the meter and the table cannot be gravitational. Challenge

A similar limitation even applies to mass measurements. Try to measure mass using gravitation alone. Any scale or balance, like any clock or meter bar, needs other, usually mechanical, electromagnetic or optical interactions to achieve their function. Whatever method one uses, *in order to define length, time, and mass, one needs interactions other than gravity.* In short, the fact that we are able to measure at all shows that gravity is not all there is. Challenge

But since we need the concepts of space, time and mass to *talk* about motion, we cannot give the term 'motion' any meaning if we neglect the non-gravitational interactions in nature. We are thus forced to investigate these other interactions as well. To get as far as possible, we start with an example of motion which we mentioned at the beginning but which we excluded of our investigations so far, even though it is used for the definition both of the meter and the second: the motion of light.

6. Special relativity – rest at any speed

Typeset in April 2001

Light is important to describe motion precisely. We need it to read these lines; but there are more important reasons. How do we check whether a line or a path of motion is straight? We look along it; in other words, we use light. How do we decide whether a plane is flat? We look across it, i.e. we use light. How do we measure length to high precision? With light. How do we measure time to high precision? With light; once that from the sun, now that from Cesium atoms. In other words, light is important because it is the official standard for *undisturbed motion*. Ref. 147

But is light a moving phenomenon at all? It was known already in ancient Greece that this is proven by 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 obvious conclusion that light takes a certain amount of time to travel from the source to the surface showing the shadow was already taken by the greek thinker Empedocles (ca. 490–ca. 430 BCE).

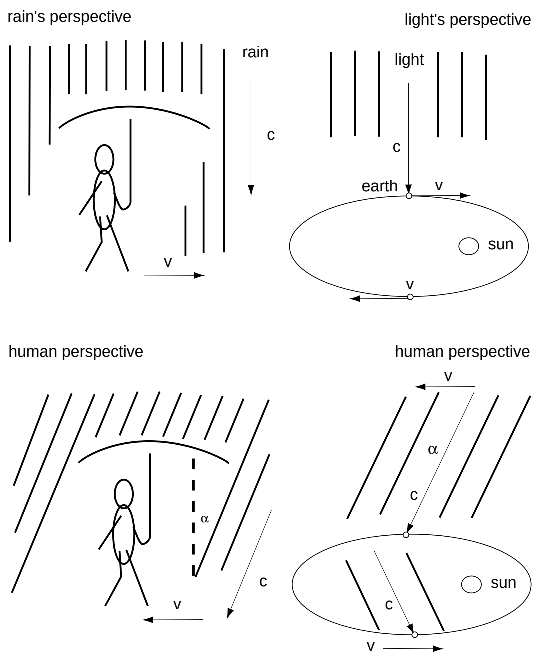


Figure 58 The rain method to measure the speed of light

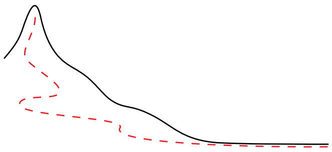
We can confirm his result with a different, but equally simple argument. Speed can be measured. Therefore the *perfect* speed, which is used as the implicit measurement standard, must have a finite value. An infinite velocity standard would not allow measurements at all, as we will confirm later on. Of course, we are implying here that the speed of light actually *is* the perfect speed. We will show this in a minute. The story of physics would have been much more rapid if light propagation had been recognized as the ideal example of motion at some earlier time. Challenge

The obviously rather high speed of light was measured for the first time only in 1676, even though many, including Galileo, had tried to do so earlier. The first measurement was performed by the danish astronomer Olaf Rømer (1644–1710) when he studied the moons of Jupiter. He got an incorrect value because he used the wrong value for the distance to Jupiter. However, this was quickly corrected by his peers, such as Newton himself. His

∗ Whenever one a source produces shadows, one calls the emitted entity *radiation* or *rays*. 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 – high frequency – version of light, *channel rays*, which turned out to be traveling ionized atoms, and the three types of radioactivity, namely α-*rays* (Helium nuclei), β-*rays* (again electrons), and γ-*rays* (high energy X-rays) which will be discussed later on. All these discoveries were made between 1890 and 1910; those were the “ray days” of physics. Ref. 148

This is a section of the freely downloadable e-textbook

Motion Mountain



Hiking beyond space and time along the concepts of modern physics

available at
www.motionmountain.org

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Did you find any mistakes?

Challenge

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

result was confirmed most beautifully by the next measurement, which was only performed fifty years later, in 1726, by the astronomer James Bradley (1693–1762). Being english, he
Ref. 149 thought of the "rain method" to measure the speed of light.

How can one measure the speed of falling rain? One walks rapidly with an umbrella, measures the angle α under which the rain appears to fall, and then measures one's own velocity v. As shown in figure 58, the velocity c of the rain is then given by

$$c = v/\tan\alpha \quad . \tag{88}$$

The same can be done for light by measuring the angle under which the light from a star above the ecliptic flies towards the earth. This effect is called the *aberration* of light; the angle is best found comparing measurements distant by six months. The angle has a value of 20.5" which nowadays can be measured with a precision of five digits. Given that the velocity of the earth around the sun is $v = 2\pi R/T = 29.7\,\mathrm{km/s}$, one finds for the speed of light the value $c = 3.00 \cdot 10^8\,\mathrm{m/s}$.* This is quite an astonishing value, especially when compared with the fastest velocity ever achieved by a man made object, namely the Voyager satellites, which travel at $52\,\mathrm{Mm/h} = 14\,\mathrm{km/s}$, the growth speed of children, about $3\,\mathrm{nm/s}$, or the growth speed of stalactites in caves, about $10\,\mathrm{nm/s}$. One gets a feeling why it took so long to measure the speed of light with some precision.**

Perhaps the most beautiful way to prove hat light is a moving phenomenon with a *finite* velocity is to photograph a light pulse flying across one's field of view, in the same way as one takes the picture of a car driving by or of a bullet flying along. Figure 59 shows

Ref. 151 the first such photograph, produced with a standard off the shelf reflex camera, a very fast shutter invented by the photographers, and most noteworthy, not a single piece of electronic equipment. (How fast does such a shutter have to be? How would you build such a shutter? Challenge And how would you open it at the right instant?)

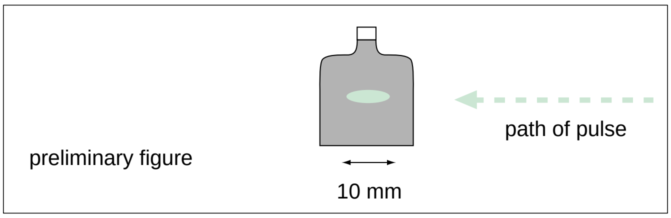


Figure 59 A photograph of a light pulse moving from right to left through milky water, taken by Dugay and Mattick

In short, light is thus much faster than lightning, as you might like to check yourself. Challenge But once the velocity of light could be measured routinely, two surprising properties were discovered in the late nineteenth century, which form the basis of special relativity.*

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

Everyone knows that in order to throw a stone as far as possible, one throws it running; one knows instinctively that in that case the stone's speed with respect to the ground is larger. However, to the initial astonishment of everybody, experiments show that light emitted from

∗ Umbrellas were not common at that time yet in Britain; they became fashionable later, after being introduced from China. The umbrella part of the story is made up. In reality, Bradley got the understanding of his unexpected results while sailing on the Themse, when he noted that on a moving ship the apparent wind has a different direction than on land. He had observed 50 stars for many years, and during that time he had been puzzled by the *sign* of the aberration, which was *opposite* to the effect he hoped to find, namely the star parallax.

Challenge By the way, it follows from special relativity that the correct formula is $c = v/\sin\alpha$; can you see why?

To determine the velocity of the earth, its distance to the sun has to be determined. This is done most simply by a method published already by the greek thinker Aristarchos of Samos (ca. 310–ca. 230 BCE). One measures the angle between the moon and the sun at the moment that the moon is precisely half full. The cosine of that angle gives the ratio between the distance to the moon (determined e.g. via the methods of page 75) and the Challenge distance to the sun. The explanation is a puzzle left to the reader.

The angle in question is almost a right angle (which would yield an infinite distance), and one needs good Ref. 150 instruments to measure it precisely, as Hipparchos noted in an extensive discussion of the problem around 130 BCE. The measurement became possible only in the late 17th century, showing that its value is 89.86 See page 734 degrees, and the distance ratio about 400. Today, through radar measurements of planets, the distance to the sun is known with the incredible precision of 30 meters. Moon distance variations can even be measured down to Challenge the 1 centimeter range; can you guess how this is achieved?

Ref. 46 Aristarchos also determined the radius of the sun and of the moon as multiples of those of the earth. Aristarchos was a remarkable thinker: he was the first to propose the heliocentric system, and perhaps the first to propose that stars were other, far away suns. For these ideas, several contemporaries of Aristarchos proposed that he should be condemned to death for impiety. When the polish monk and astronomer Nicolaus Copernicus (1473–1543) reproposed the heliocentric system two thousand years later, he kept this reference unmentioned, even though he got the idea from him.

∗∗ The first precise measurement of the speed of light was in 1849 by the french physicist Hippolyte L. Fizeau (1819–1896), whose value was only 5% above that measured today. He sent a beam of light to a mirror and back. How far away does the mirror have to be? How do you think did he measure the time, without using any Challenge electric device?

∗ One can learn the basics of special relativity also with help of the web, without any book, by using as a starting point the http://physics.syr.edu/research/relativity/RELATIVITY.html web page. It mentions most (english language) relativity pages available on the web. Some links in other languages are mentioned on the motion mountain web site.

a moving lamp has the same speed as light emitted from a resting one. Many carefully and
Ref. 152 specially designed experiments 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
Challenge than 290000 km/s no differences have been found. (Can you guess what lamps were used?) In short, experiments show that the velocity of light has the *same value* for all observers, even if they are moving with respect to each other or if they are moving with respect to the light source. Light velocity is indeed the ideal, perfect measurement standard.*

Ref. 153 There is also a second group of experimental evidence for the constancy of light speed: every electromagnetic device, such as every electric toothbrush, shows that the speed of light
See page 314 is constant. We will discover that magnetic fields would not result from electric currents, as they do every day in every motor and in every loudspeaker, if the speed of light were not constant. This was actually the historical way the constancy was first deduced; only after realizing this connection, the german-swiss physicist Albert Einstein** showed that the constancy is also in agreement with the motion of bodies, as we will do in this section. The connection between electric toothbrushes and relativity will be elucidated in the chapter on electrodynamics.***

The constancy of light is in complete contrast with galilean mechanics, and proves that the latter is *wrong* at large velocities. At small velocities the description remains good, because the error is small. But if one looks for a description valid at *all* velocities, galilean mechanics has to be discarded. For example, when one plays tennis one uses the fact that hitting the ball in the right way, one can increase or decrease its speed. But with light this is impossible. Even if one takes an airplane and flies after a light beam, it still moves away with the same speed. Riding an accelerating car, the cars on the other side of the road pass by with higher and higher speed, as we are driving faster. For light, this is *not* so; light always passes by with the *same* speed.****

Why is this result almost unbelievable, even though the measurements show it unambiguously? Take two observers O and Ω moving with relative velocity v; imagine that at the moment they pass each other, a light flash is emitted by a lamp in the hand of O. The light flash moves along positions $x(t)$ for O, respectively $\xi(\tau)$ (pronounced "xi of tau") for

* An equivalent alternative for light speed is 'radar speed' or 'radio speed'; we will understand below why this is the case.

Experiments also show that the speed of light is the same in all directions of space to at least 21 digits of
Ref. 155 precision. Other data, taken from gamma ray bursts, show that the speed of light is independent of frequency
Ref. 156 for at least its first 20 digits.

** Albert Einstein (1879, Ulm–1955, Princeton); one of the greatest physicists of all time. He published three important papers in 1905, namely about the idea of light quanta, about Brownian motion, and about special relativity. Each paper was worth a Nobel prize, but he got it only for the first one. In 1905, he discovered the famous formula $E_o = mc^2$ (published early 1906). Although one of the founders of quantum theory, he later tuned against it; his famous discussions with his friend Niels Bohr though helped a lot to clarify the field in its most counterintuitive aspects. In 1915 and 1916, he published the general theory of relativity, one of the most beautiful and remarkable works of science ever. Being jewish and famous, he was a favorite target of antisemitic
Ref. 154 attacks; in 1933 he emigrated to the USA. He was not only a great physicist, but also a great man; reading his collection of thoughts about topics outside physics is time well spent.

*** About the influences of relativity on machine design see the textbook by J. VAN BLADEL, *Relativity and engineering*, Springer 1984.

**** Indeed, the presently possible measurement precision of $2 \cdot 10^{-13}$ does not allow to discern any changes
Ref. 155 of the speed of light with the speed of the observer.

Ω ("Omega"). Since the light speed is the same for both, one has

$$\frac{x}{t} = c = \frac{\xi}{\tau} \quad . \tag{89}$$

In the situation described, one obviously has $x \neq \xi$. In other words, the constancy of light speed implies that $t \neq \tau$, i.e. that *time is different for observers moving against each other*. Time is thus not unique. This surprising result, which in the meantime has been confirmed by many experiments, was first stated in detail in 1905 by Albert Einstein. The discussion of this issue, especially in connection with view point invariance, had been named already in 1895 the *theory of relativity* by the important french mathematician and physicist Henri Poincaré (1854–1912).* Einstein called the description of motion without gravity the theory of *special relativity*, and the description with them the theory of *general relativity*. Both fields are full of fascinating and counterintuitive results. In particular, they show that galilean physics is wrong at large speeds.

Challenge

Ref. 157

Ref. 153

Of course, many people tried to find all sorts of arguments to avoid the strange conclusion that time differs from observer to observer. But all had to bow to the experimental results.

Acceleration of light and the Doppler effect

Light *can* be accelerated. Every mirror does so! We will see in the chapter on electromagnetism that matter also has the power to *bend* light, and thus to accelerate it. However, it will turn out that all these means only change the propagation direction; none has the power to change the vacuum light speed.

See page 328

What would happen if one could accelerate light to higher speeds? It would mean that light is 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 then would be lower and depend on the kinetic energy of those massive particles. Light speed would not be constant, but the massless energy speed would still be so. Massive light particles could be captured, stopped, and stored in a box. Such boxes would render electric illumination superfluous; it would be sufficient to store in them some daylight and release it, slowly, the following night.**

People have therefore checked this issue in quite some detail. Observations now put any possible mass of light (particles) at less than $1.3 \cdot 10^{-52}$ kg from terrestrial arguments, and at less than $4 \cdot 10^{-62}$ kg from astrophysical arguments. In other words, light is not heavy, light is light.

Ref. 158

But what happens when light hits a *moving* mirror? If the speed of light does not change, something else must. In this case, as in the situation when the light source moves with

* The most beautiful and simple introduction into relativity is still given by Albert Einstein himself, such as in *Über die spezielle und allgemeine Relativitätstheorie*, Vieweg, 1997, or in *The meaning of relativity*, Methuen, London, 1951. See also his text *Relativity, the special and general theory*,, which can also be found at http://ourworld.compuserve.com/homepages/eric_baird/rel_main.htm. Only a century later there are books almost as beautiful as that, such as Edwin F. Taylor & John A. Wheeler, *Spacetime Physics – Introduction to special relativity*, second edition, Freeman, 1992.

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

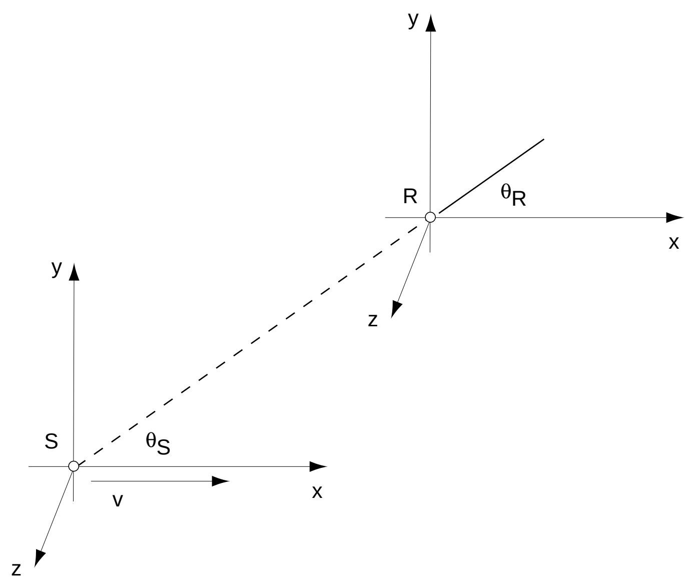


Figure 60 The set-up for the observation of the Doppler effect

respect to the receiver, the receiver will observe a *different colour* than the sender. This result is called the *Doppler effect*. He had studied the frequency shift in the case of sound waves - the well-known change in whistle tone between approaching and departing trains.* As we will see later on, light is (also) a wave, and its colour is determined by its frequency. Like the tone change for moving trains, also a moving light source produces a colour for the receiver which is different from the sent colour. Simple geometry, starting from the fact that all maxima and minima emitted must also be received, leads to the result

Challenge

$$\frac{\lambda_R}{\lambda_S} = \gamma(1 - \frac{v}{c}\cos\theta_R) \tag{90}$$

Light from an approaching source becomes more bluish, and light from a departing source more reddish. One speaks of a *red shift* and a *blue shift*. The first observation of the Doppler effect for light was by Stark, who studied the light emitted by moving atoms. All subsequent experiments confirmed the colour shift within measurement errors; the latest checks found agreement to within two parts per million. Note that in contrast to sound waves, a colour effect is also found when the motion is *transverse* to the line of sight. (How does the colour change in this case?)

Ref. 159

Challenge

The colour shift is used in many applications. Almost all bodies are mirrors for radio waves. When one enters a building, often the doors open automatically. A little sensor above the door detects the approaching person. Usually (but not always) this is done 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 two sides of the same phenomenon.) Police radar also works in this way.**

* Christian Doppler (1803, Salzburg–1853, Venezia); austrian physicist.

Challenge

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

In summary, whenever one tries to change the *speed* of light, one only manages to change its *colour*. That is the *Doppler effect*. But the Doppler effect for light is much more important than the Doppler effect for sound. Even if the speed of light were not yet known to be constant, the colour change alone already *proves* that time is different for observers moving against 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 back and forward between the two. In short, we need to use light signals. And a colour change appearing when light moves from one observer to another implies that the watches run differently, and thus means that time is *different* at the two places. Are you able to confirm this conclusion in more detail? And why Challenge is the conclusion about time differences not possible when the Doppler effect for sound is used? Challenge

The Doppler effect also allows to measure the speed of light sources. Indeed, it is used to measure the speed of far away stars. Can you imagine how this is done in detail? Challenge

Can one shoot faster than one's shadow?

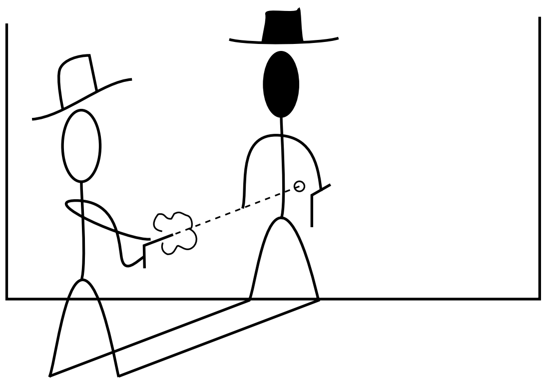


Figure 61 Lucky Luke

To achieve such a feat, the bullets and the hand have to move faster than the speed of light. To try this, certain people use the largest practical amounts of energy possible, taken directly from an electrical power station, accelerate the lightest known objects, namely electrons, and measure how fast they can get them. This is carried out daily in particle accelerators such as the Large Electron Positron ring, the LEP, of 27 km circumference located partly in France and partly in Switzerland, near Genève. In that place, 40 MW of electrical power, the same amount used by a small city, accelerate electrons and positrons to energies of over 16 nJ (100 GeV) each. The result is shown in figure 62: even with these impressive means it is impossible to make Ref. 160 electrons move more rapidly than light. These and many similar observations thus show that there is a *limit* to the velocity of objects. Velocities of bodies (or of radiation) higher than the speed of light do not exist.*

The people most unhappy with this limit are the computer engineers; if the limit would be higher, it would be possible to make faster microprocessors and thus faster computers; this would allow e.g. a more rapid progress towards the construction of computers which understand and use language.

∗ There are still people refusing to accept these results, as well as the ensuing theory of relativity. Every physicist should enjoy the experience, at least once in his life, of discussing with at least one of these men. (Strangely, no woman has yet been reported as member of this group of people.) This can be done e.g. via the internet, in the
Ref. 161 sci.physics.relativity news group. See also the http://www.crank.net web site. Crackpots are a fascinating lot, especially since they teach the importance of *precision* in language and in reasoning, which they all, without exception, neglect. Encounters with several of them provided the inspiration for this section.

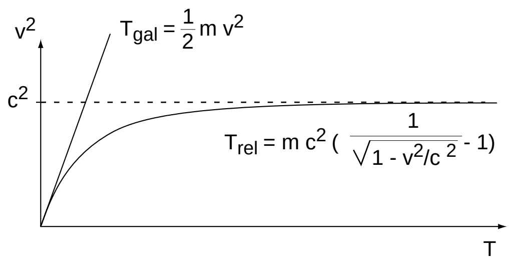


Figure 62 Experimental values for the electron velocity v as function of kinetic energy T

The observation of a limit speed is in complete contrast to galilean mechanics. In fact, it means that for velocities near that of light, say about 15000 km/s or more, the expression $mv^2/2$ is *not* equal to the kinetic energy E of the particle. Such high speeds are rather common: many families have an example in their home. Just determine the speed of electrons inside a television, given that the transformer inside produces 30 kV. Challenge Historically, the accuracy of galilean mechanics was taken for granted for more than three centuries, so that nobody ever thought of checking it; but when this was finally done, it was found wrong.

The observation of light speed as a *limit* speed for objects is easily seen to be a consequence its *constancy*. Bodies which can be at rest in one frame of reference obviously move more slowly than the maximum velocity (light) in that frame. Now, if something moves slower than something else for *one* observer, it does so for all other observers as well. (Trying to imagine a world in which this would not be so is very interesting: funny things would happen, such as things interpenetrate each other.) Challenge Therefore no object which can be at rest can move faster than the limit speed. But any body which can be at rest does have different speeds for different observers. Conversely, if a phenomenon exists whose speed is the same for all observers, then this speed must necessarily be the limit speed. Challenge Among others, one deduces that the maximum speed is the speed of *massless* entities. Light, all the other types of electromagnetic waves, and (probably) neutrinos are the only known examples. Ref. 162

A consequence of the existence of a limit velocity is important: velocities cannot simply be added or subtracted, as one is used to in everyday life. If a train is riding at velocity v_{te} compared to the earth, and somebody throws a stone inside it with velocity v_{st} in the same direction, 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, measurements show a different result. The combined velocity is given by

$$v_{se} = \frac{v_{st} + v_{te}}{1 + v_{st} v_{te}/c^2} \qquad (91)$$

Challenge Note that the result is never larger than c. We will deduce the expression in a moment, from reasoning alone.* In addition, the result has been confirmed by literally all millions of cases in which it has been checked so far.

The principle of special relativity

The next question to ask is *how* the different time intervals and lengths measured by two observers are related to each other. We start with a situation where neither gravitation nor any other interaction plays a role; in other words, we start with *relativistic kinematics*.

If an undisturbed body travels along a straight line with a constant velocity, or if it stays at rest, one calls the observer making this observation *inertial*, and the coordinates he uses an

Ref. 164 * One can even deduce the Lorentz transformation from this expression.

inertial frame of reference. Examples of inertial observers (or frames) are the *two* horizontal dimensions on a frictionless ice surface or on the floor inside a smoothly running train or ship; another example are all *three* spatial dimensions for the cosmonauts in the Apollo 13 capsule while they were traveling between the moon and the earth, as long as they kept their engine switched off. Inertial observers in three dimensions might also be called *free floating* observers. They are thus not so common.

Special relativity is built on a simple principle:

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

Or:

▷ *The speed of light is the same for all inertial observers.*

Or, as we will show below, the equivalent:

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

$$v \leqslant c \tag{92}$$

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

This experimental statement was checked with high precision by Michelson and Morely* in the years from 1887 onwards. It has been confirmed in all cases so far. Therefore the following conclusions can be drawn from it, using various implicit assumptions which will become clear during the rest of the escalation: Ref. 165

- In a closed free-floating room, there is no way to tell the speed of the room.
- There is no absolute rest; rest is an observer-dependent concept.
- All inertial observers are equivalent: they describe the world with the same equations. (This statement was called the *principle of relativity* by Henri Poincaré.)
- Any two inertial observers move with constant velocity against each other. (Are you able to show this?) Challenge

Historically, it was the equivalence of all inertial observers which used to be called the principle of special relativity. Nowadays this habit is changing, though slowly, mainly because they are connected to Einstein himself. The essence of relativity is the existence of a limit speed.

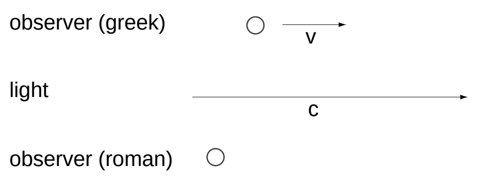


Figure 63 Two inertial observers, using coordinates (t, x) and (τ, ξ), and a beam of light

But let us continue with the original topic of this section. To see how length and space intervals change from one observer to the other, assume that two observers, a roman one using coordinates x, y, z and t, and and a greek one using coordinates ξ, υ, ζ and τ,** move with velocity $\mathbf{v}$ against each other. The axes are chosen in such a way that the velocity points in the x-direction. We start by noting that the constancy

* Albert Abraham Michelson (1852, Strelno–1931, Pasadena) prussian-polish-US-american physicist, Nobel prize in physics in 1907. Michelson called the set-up he devised an *interferometer*, a term still in use today. Edward William Morely (1838–1923), US-american chemist.

** The names of the greek letters are explained in appendix A.

of the speed of light in any direction for any two observers means that

$$(cdt)^2-(dx)^2-(dy)^2-(dz)^2=(cd\tau)^2-(d\xi)^2-(d\upsilon)^2-(d\zeta)^2 \quad . \qquad (93)$$

Assume also that a flashlamp at rest for the greek observer, thus with $d\xi=0$, flashes twice, with flash interval $d\tau$. For the roman observer, the flashlamp moves, so that $dx=vdt$. Inserting this into the previous expression, and assuming linearity and speed direction independence for the general case, one finds that intervals are related by

Challenge

$$\begin{aligned} dt &= \gamma(d\tau+vd\xi/c^2)=\frac{d\tau+vd\xi/c^2}{\sqrt{1-v^2/c^2}} \qquad \text{with} \quad v=dx/dt \\ dx &= \gamma(d\xi+vd\tau)=\frac{d\xi+vd\tau}{\sqrt{1-v^2/c^2}} \\ dy &= d\upsilon \\ dz &= d\zeta \quad . \end{aligned} \qquad (94)$$

These expressions describe how length and time intervals measured by different observers are related. At relative speeds v which are small compared to the velocity of light, such as in everyday life, the time intervals are essentially equal; the factor γ is then equal to one for all practical purposes. However, for velocities *near* that of light the measurements of the two observers give different values. In these cases, space and time *mix*, as shown in figure 64.

The expressions are strange also in another respect. When two observers look at each other, each of them says that he measures shorter intervals than the other. In other words, special relativity shows that the grass on the other side of the fence is always *shorter* – when one rides along the fence in a bicycle. We explore this bizarre result in more detail shortly.

Challenge

The factor γ is equal to one in everyday life and for most practical purposes. The largest value humans ever achieved is about $2\cdot10^5$; the largest observed value in nature is about 10^{12}. Can you imagine their occurrences?

Challenge

Once one knows how space and time *intervals* change, one can easily deduce how *coordinates* change. Figures 63 and 64 show that the x coordinate of an event L is the sum of two intervals: the ξ coordinate and the length of the distance between the two origins. In other words, one has

$$\xi=\gamma(x-vt) \quad \text{and} \quad v=\frac{dx}{dt} \quad . \qquad (95)$$

Using the invariance of the space-time interval, one gets

$$\tau=\gamma(t-xv/c^2) \qquad (96)$$

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 beautiful discoveries of physics, in 1892 and 1904, Lorentz deduced them from the equa-

Ref. 166

* Hendrik Antoon Lorentz (Arnhem, 1853– Haarlem, 1928) was, together with Boltzmann and Kelvin, the most important physicist of his time. He was the first to understand, long before quantum theory confirmed the idea, that almost all material properties are due to interacting electrons. He showed this in particular for the dispersion of light, for the Zeeman effect, for the Hall effect, for the Faraday effect, and many others. He understood that Maxwell's equations of the vacuum describe matter as well, if only charged particles are included. He also gave the correct description of the Lorentz force. Outside physics, he was active in the internationalization of scientific collaborations.

tions of electrodynamics, which contained them, waiting to be discovered, since 1865.* In See page 316 that year James Clerk Maxwell has published them in order to describe everything electric and magnetic.

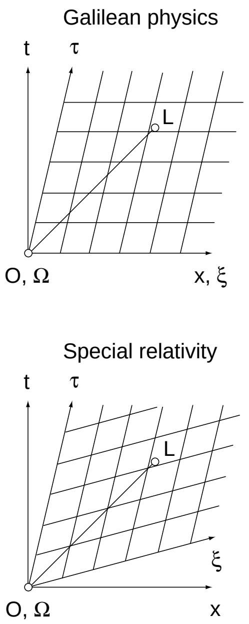


Figure 64 Space-time diagrams for light seen from two different observers

The formulas in this section form the basis of the theory of relativity, both the special and the general one. Many alternative formulas have also been explored, such as expressions in which the relative acceleration of the two observers enters as well, instead of only the relative velocity. However, all had to be discarded when Ref. 163 compared to experimental results. But before we have a look at such experiments, we continue with a few logical deductions from the above relations.

What is space-time?

The Lorentz transformations tell something important: space and time are two aspects of the same "stuff", they are two aspects of the same basic entity. They mix in different ways for different observers. This fact is commonly expressed by stating that time is the *fourth dimension*. This makes sense because the common entity, called *space-time*, can be defined as the set of all possible events, because events are described by 4 coordinates in time and space, and because the set of all events behaves like a manifold. (Can you confirm this?) In the theory of special relativity, the space-time Challenge manifold is characterized by property that the *space-time interval* between two nearby events, defined as Ref. 167

$$di^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 = c^2dt^2(1 - \frac{v^2}{c^2}) \quad , \qquad (97)$$

is *independent* of the (inertial) observer. Such a space-time is also called Minkowski space-time, after the german physicist Hermann Minkowski (1864–1909), the prematurely passed away teacher of Albert Einstein; he was the first physicist, in 1904, to define the concept and to understand its usefulness and importance.

The space-time interval of equation (97) has a simple interpretation. It is the time measured by an observer moving between from event (t,x) to event $(t+dt, x+dx)$, the so-called *proper time*, multiplied by c^2. One could simply call it wristwatch time.

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 are infinite, i.e. open, with the topology of the punctured sphere. (Can you confirm this?) Both manifolds are flat, Challenge i.e. free of curvature. In both cases, space is what one measures with a meter bar or with a light ray, and time is what one reads from a clock. In both cases, space-time is fundamental; it is and remains the *background* and the *container* of things and events. We *live* in a Minkowski space-time, so to say. Minkowski space-time exists independently of things.

* The irishman George F. Fitzgerald had discovered the Lorentz transformations already in 1889, but had, in contrast to Lorentz, not continued his research in the field.

And even though coordinate systems can be different from observer to observer, the underlying entity, space-time, is still *unique*, even though space and time by themselves are not.

The central difference, in fact the only one, is that Minkowski space-time, in contrast to the galilean case, *mixes* space and time, and does so differently for different observers, as shown by figure 64.

Time and causality

Given that time is different between different observers, does time stillorder events in sequences? The answer of relativity is a clear yes and no. Certain sets of events are not in any given sequence; others sets are. This is best seen in a space-time diagram.

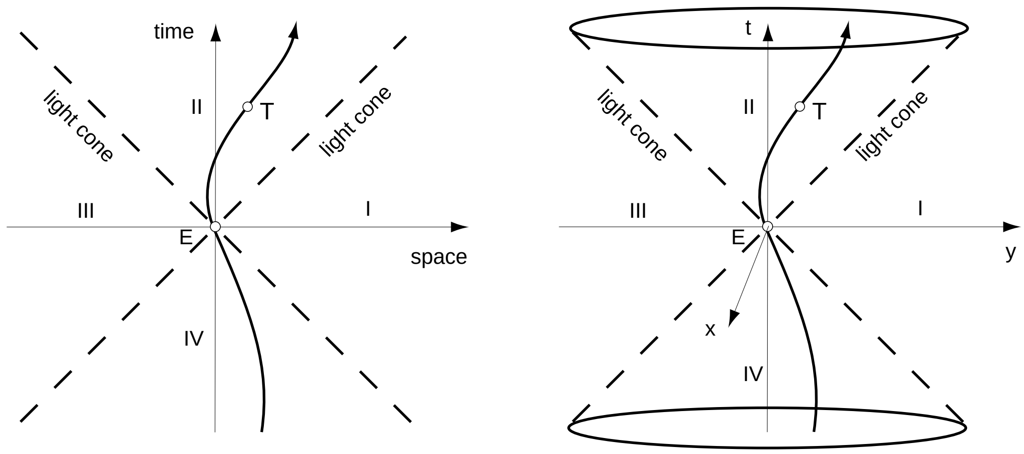


Figure 65 A space-time diagram of an object T seen from an inertial observer O in case of one and of two spatial dimensions

Sequences of events can surely be defined if an event is the cause of another. But this can only be the case if energy or signals travel from one event to another at most with the speed of light. Figure 65 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 itself can influence only events in the quadrant II (the *future light cone*). Events in the quadrants I and III do not influence, nor are they influenced by event E. In other words, the light cone defines the boundary between events which *can* be ordered with respect to their origin – namely those inside the cones – and those which *cannot* – those outside the cones. In short, time orders events only *partially*. For example, for two events that are not causally connected, their simultaneity and their temporal order depends on the observer!

Challenge

In particular, the past light cone gives the complete set of events which can influence what happens at the origin. One says that the origin is *causally connected* only to the past light cone. This is a consequence of the fact that any influence is transport of energy, and thus travels at most with the speed of light. Note that causal connection is an invariant concept: all observers agree on whether it applies to two given events or not. Are you able to confirm this?

Challenge

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

If the speed of light could be surpassed in some way, the future could influence the past. Are you able to confirm this? In such situations one would speak of *acausal* effects. However, there is an everyday observation telling us 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 "law" of thermodynamics would not be valid, and then our memory would not work.* Also all other data from everyday life and from experiments provide no evidence that the future can influence the past. How the situation changes in other domains will be discovered later on. Challenge Ref. 168

Curiosities of special relativity

Faster than light: how far can one travel?

How far away from earth can one travel, given that the trip should not last more than a lifetime, say 80 years, and given that one can use a rocket whose speed can approach the speed of light as much as desired? Given the time t one is 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 one can move away is given by Challenge

$$d = \frac{vt}{\sqrt{1 - v^2/c^2}} \qquad (98)$$

The distance d is larger than ct already for $v > 0.71c$, and, by choosing v large enough, it increases beyond all bounds! In other words, relativity itself thus does *not* limit the distance one can travel, not even that covered in a single second. One could, in principle, roam the entire universe in less than a second. In situations such as these it makes sense to introduce the concept of *proper velocity* w , defined as

$$w = d/t = \frac{v}{\sqrt{1 - v^2/c^2}} = \gamma v \quad . \qquad (99)$$

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

∗ Another related result is slowly becoming common knowledge. Even if spacetime had a non-trivial shape, such as a cylindrical topology, one still would not be able to travel into the past, in contrast to what many science fiction novels suggest. This point is made clear by Stephen Blau in a recent pedagogical paper. Ref. 169

∗∗ Using proper velocity, the relation given in relation (91) for the superposition of two velocities $\mathbf{w}_a = \gamma_a \mathbf{v}_a$ and $\mathbf{w}_b = \gamma_b \mathbf{v}_b$ simplifies to Challenge

$$w_{s\parallel} = \gamma_a \gamma_b (v_a + v_{b\parallel}) \quad \text{and} \quad w_{s\perp} = w_{b\perp} \quad , \qquad (100)$$

where the signs $\parallel$ and $\perp$ mean the components in direction of motion and that perpendicular to it, respectively. One can in fact write all of special relativity using 'proper' quantities, even though this is not done in this text. Ref. 170

Synchronization and aging: can a mother stay younger than her own daughter?

In the theory of special relativity time is different for different observers moving against each other. This implies that one has to be careful on how to synchronize clocks which are far apart, even if they are at rest with respect to each other in an inertial reference frame. For example, if one has two identical watches showing the same time, and if one carries one of the two 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 a plane going around the earth once at around 900 km/h is of the order of 100 ns – not very noticeable in everyday life. In fact, the delay is easily calculated from the expression

Ref. 171, 172

Challenge

$$\frac{t}{t'} = \gamma \quad . \tag{101}$$

Also human bodies are clocks; they show the time elapsed, usually called the *age*, by various changes in their shape, weight, hair colour, etc. If a person goes onto a long and fast trip, on her return she will have aged *less* than a second person who stayed at her (inertial) home. Special relativity thus confirms, in a surprising fashion, the well-known result that those who travel a lot remain younger. In short, the title question receives a positive answer. This result is usually called the *clock paradox* or the *twin paradox*. It has been confirmed in many experiments. We give a simple example below.

Challenge

Ref. 173

One concludes that one cannot synchronize clocks simply by walking, clock in hand, from one place to the next. The correct way to do it is to exchange light signals. Can you describe how?

Challenge

In summary, only with a clear definition of synchronization can one call two distant events simultaneous. Special relativity shows that simultaneity depends on the observer, as is confirmed by all experiments so far.

Length contraction

Can a rapid skier fall in a hole a bit shorter than his skis? Imagine a skier so fast that the length contraction factor $\gamma = d/d'$ is 4.* For an observer on the ground, the ski is 4 times shorter, and when it passes over the hole, it will fall into it. However, for the skier, it is the hole which is 4 times shorter; it seems that he cannot fall into it.

Ref. 174

More careful analysis shows that in contrast to the observation of hole digger, the skier does not experience the ski shape as fixed; while passing over the hole, he observes that the ski gets a parabola shape and falls into it. Can you confirm this? In other words, shape is *not* an observer invariant concept. (However, rigidity is one, if defined properly; can you confirm this?)

Challenge

Challenge

Challenge

* Even the earth is contracted in its direction of motion around the sun. How much is it? Is this measurable?

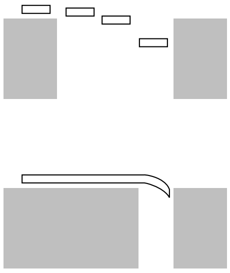


Figure 66 The observations of the trap digger and of the skier

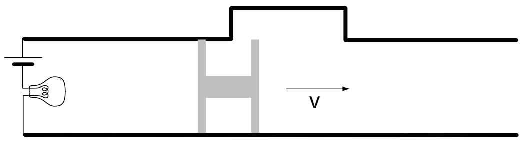


Figure 67 Does the conducting glider keep the lamp lit at large speeds?

The situation becomes more interesting in the case that the ski is replaced by a conductive bar and makes electrical contact between the two sides of the hole, with gravity switched off. This is achieved by putting the whole arrangement on the side, as in figure 67. Are you able to find out whether a lamp connected in series stays lit when the glider moves along the contacts? Do you get the same result for all observers? And what happens when the glider is longer than the detour? (Warning: this problem gives rise to *heated* debates!) Ref. 175 Challenge

Figure 68 What happens to the rope?

Another example of the phenomenon of *length contraction* appears when two objects, say two cars, are connected over a distance d by a straight rope. 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 both cars keep the same distance. On the other hand, the rope needs to span a distance $d' = d/\sqrt{1 - v^2/c^2}$, which has to expand when the two cars are moving. In other words, the rope will break. Is this prediction confirmed by observers on each of the two cars? Ref. 176 Challenge

Which is the best seat in a bus?

The last example provides another surprise. Imagine two twins inside the two identically accelerated cars, starting from standstill at the same time as described by an observer at rest with respect to them at time $t = 0$. Both cars contain the same amount of fuel. (Let's forget the rope now.) One easily finds that the acceleration of the two twins stops at the same time in the frame of the outside observer, that the distance between the cars has remained the same all along for the outside observer, and that the two cars continue rolling with an identical constant velocity, 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 are related simply by $t_f = t_b$. Inserting the Lorentz transformations you can deduce for the frame of the freely rolling twins the relation Ref. 176 Challenge

$$t_b = \gamma \Delta x v / c^2 + t_f \quad , \tag{102}$$

which means that the front twin has aged *more* than the back twin. Therefore one is inclined to conclude that if one wants to avoid grey hair as much as possible, one should always sit in the back of a bus or train when traveling. Is the conclusion correct? Or is it correct to deduce that people on high mountains age faster than people in valleys?

Challenge
Challenge

How fast can one walk?

To walk means to move the feet in such a way that at least one of the two feet is on the ground at any time. This is one of the rules athletes have to follow in olympic walking competitions, and they are disqualified if they break it. A certain student athlete was thinking about the theoretical maximum speed one could achieve in the olympics. He assumed that each foot could accelerate instantly to (almost) the speed of light. Then he thought about increasing his walking speed by taking the second foot off the ground in exactly the same instant at which he puts down the first. By 'same instant' he meant 'as seen by a competition judge at rest with respect to earth.' The motion of the feet is shown in the left of figure 69; it gives a limit speed for walking of half the speed of light. But then he noticed that a *moving* judge will see both feet off the ground and thus disqualify the athlete for running. To avoid disqualification from *any* judge, the second foot has to wait for a light signal from the first. The limit speed for olympic walking thus is one third of the speed of light.

Ref. 177

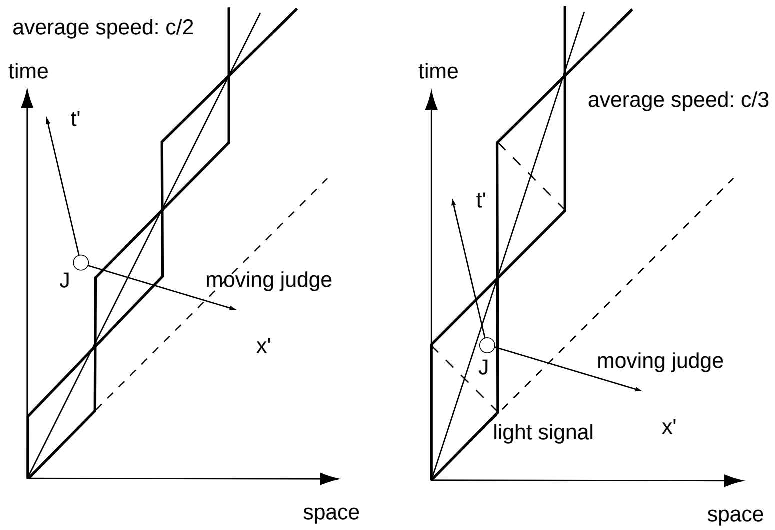


Figure 69 For the athlete on the left, the judge moving in opposite direction sees both feet off the ground at certain times, but not for the athlete on the right

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

Contrary to what is often implied, motion faster than light does exist and is even rather common. Special relativity only constraints the motion of mass and energy. However, non-material points, non-energy transporting features, or images *can* move faster than light. We give a few simple examples. Note that we are not talking about *proper* velocity, which in these cases cannot be defined anyway. (Why?) See page 158 Challenge

Neither are we talking of the situation where a particle moves faster than the velocity of light in matter, but slower than the velocity of light in vacuum. This situation gives rise to the so-called *Cerenkov radiation*, if the particle is charged. It corresponds to the v-shaped wave created by a motor boat on the sea or by the cone-shaped shock wave around an airplane moving beyond the speed of sound. Cerenkov radiation is often seen; for example it is the cause of the blue glow of the water in nuclear reactors. By the way, it is not always difficult to let objects move faster than the speed of light in matter; in the center of the sun, the speed of light is estimated to be only around 10 km/year, and in the laboratory, for some materials it has been reduced down to 0.3 m/s. The following examples show genuine faster than light velocity, as measured externally, and in vacuum. Ref. 178, 18

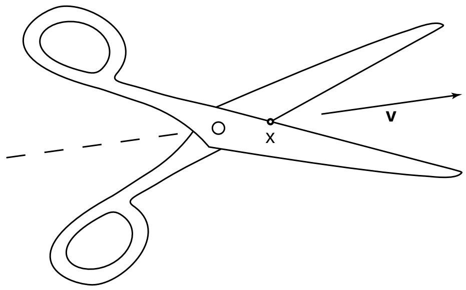


Figure 70 A simple example of faster than light motion

In contrast, an example of faster than light motion is the point marked ‘x’ in figure (70), the point at which scissors cut paper. If the scissors are closed rapidly enough, that point moves faster than light. Similar geometries can also be found in every window frame, and in fact in any device which has twisting parts.

Another example is the speed with which a music record – remember LP’s? – disappears in its sleeve, as shown in figure (71).

Finally, a standard example is the motion of a spot of light produced by shining a laser beam onto the moon. If one moves the laser, the spot can easily move faster than light. The same happens for the light spot on the screen of an oscilloscope when a high frequency signal is put on the input.

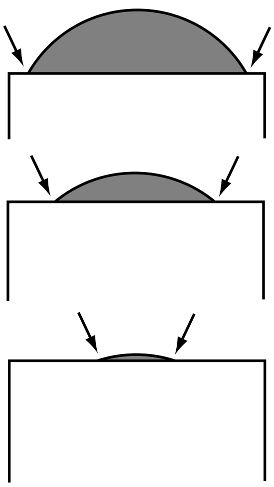


Figure 71 Another example of faster than light motion

All these are typical examples for 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 for their speed. Can you find another example? Challenge

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 larger than c. They regularly make headlines in the newspapers, usually of the type “light moves faster than light.” This surprising result is discussed in more detail later on. See page 333

For a different example, imagine to stand at the exit of a tunnel of length l. We see a car entering the other end of the tunnel whose speed we know to be v, driving towards us. We know that it entered the tunnel because the car is not in the sun any more or because its headlights switched on at that moment. At what time t does it drive past us? Simple reasoning shows that t is given by

$$t = l/v - l/c \tag{103}$$

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

$$v_{appr} = \frac{l}{t} = \frac{vc}{(c-v)} \qquad (104)$$

which is larger than c for any car velocity v larger than $c/2$. For cars this does not happen too often, but astronomers know a type of bright object called a *quasar* (a contraction of 'quasi-stellar'), which sometimes emit high speed gas jets. If the emission is in or near the direction to the earth it is possible that the apparent speed is larger than c; such situations are now regularly observed with telescopes. Ref. 179

Note that to a second observer at the *entrance* of the tunnel, the apparent speed of the car *leaving* from him is given by

$$v_{leav} = \frac{vc}{(c+v)} \qquad (105)$$

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

The story has a final twist. We have just seen that motion faster than light can be observed in several ways. But could an *object* moving faster than light be observed at all? Surprisingly, the answer is no, at least not in the common sense of the expression. First of all, since such an imaginary object, usually called a *tachyon*, moves faster than light, one could never see it approaching. If at all, tachyons can only bee seen departing.

Seeing a tachyon is very similar to hearing a supersonic jet. Only *after* a tachyon has passed nearby, assuming that it is visible in daylight, one could notice it. One would first see a flash of light, corresponding to the bang of a plane passing with supersonic speed. Then one would see *two* images of the tachyon, appearing somewhere in space and departing in opposite directions, as can be deduced from figure 72. Even if one of the two images would come nearer, it would be getting fainter and smaller. This is, to say the least, rather unusual behaviour. Moreover, if one wants to look at a tachyon at night, illuminating it with a torch, one has to turn the head in the direction opposite to the arm with the torch! This requirement also follows from the space-time diagram; are you able to deduce this? Nobody has ever seen such phenomena; tachyons do not exist. Tachyons would be strange objects: they have imaginary mass, they accelerate when they lose energy, and a zero energy tachyon would be infinitely fast. But no object

Challenge

Ref. 180

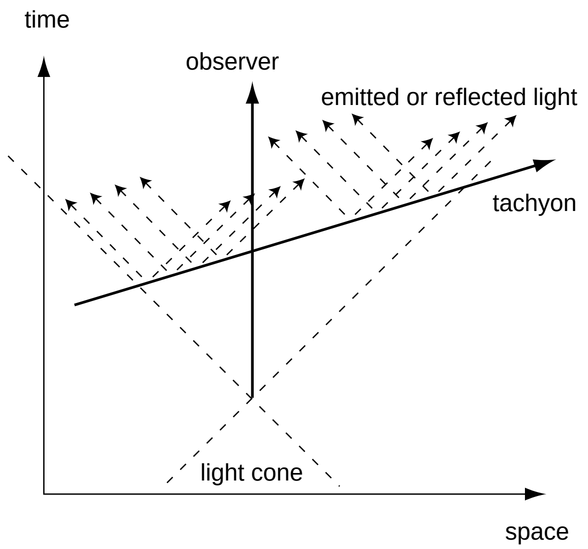


Figure 72 Hypothetical space-time diagram for tachyon observation

with these properties has ever been observed. Worse, as we just saw, tachyons would seem to appear from nothing, defying conservation laws; and note that since 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 escalation. Tachyons therefore cannot be objects in the usual sense. In the second part of our escalation we will show that quantum theory actually *rules out* the existence of (free) tachyons. However, it also requires the existence of virtual tachyons, as we will see.

But the best is still to come. Not only is it impossible to see approaching tachyons; it follows from equation (105) that departing ones seem to move with a velocity smaller than the speed of light. In other words, we just found that if one ever sees something move faster than light, it can be anything *but* a tachyon!

A neverending story: temperature and relativity

Not everything is settled in special relativity. Do you want to have trouble? Just deduce how temperature changes from one frame of reference to another, and publish the result. Just have a try. There are many opinions on the matter. True, Albert Einstein and Wolfgang Pauli agreed that the temperature T seen by a moving observer is related to the temperature T_o measured by the observer at rest with respect to the bath via Challenge

$$T = T_o \sqrt{1 - v^2/c^2} \quad , \tag{106}$$

thus always yielding lower values. But others maintain that T and T_o should be interchanged in this expression. Even different powers than the simple square root have been proposed. Ref. 181

The origin for these discrepancies is simple: temperature is only defined for equilibrium situations, i.e. for baths. But a bath for one observer usually is not a bath for the other; for small speeds, a moving observer sees *almost* a bath; but then the issue becomes tricky. The resulting temperature change might even depend on the energy range measured. So far, there do not seem to be any experimental observations which would allow to settle the issue. Realizing such a measurement is a challenge for experimental physics.

Relativistic mechanics

The speed of light being constant, and distances in space and time behaving so strangely, we need to rethink the definition of mass, as well as those of momentum and energy.

Mass in relativity

In galilean physics, the mass ratio between two bodies was defined using collisions and was given by the negative inverse of the velocity change ratio See page 55

$$\frac{m_2}{m_1} = -\frac{\Delta v_1}{\Delta v_2} \quad . \tag{107}$$

But experiments show that the expression is not fixed for speeds near that of light. In fact one can show that by thinking alone; are you able to do so? Challenge

The solution to this situation was found by Albert Einstein. He discovered that in collisions, the two galilean conservation theorems $\sum_i m_i \mathbf{v}_i =$ const and $\sum_i m_i =$ const had to be changed into

$$\sum_i \gamma_i m_i \mathbf{v}_i = \text{const} \qquad (108)$$

and

$$\sum_i \gamma_i m_i = \text{const} \quad . \qquad (109)$$

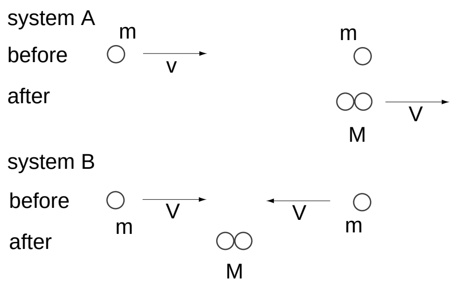


Figure 73 An inelastic collision of two identical particles seen from two different inertial frames of reference

These expressions, which are correct throughout the rest of our escalation, imply, among others, that teleportation is not possible in nature.

Challenge (Can you confirm this?) Obviously, in order to recover galilean physics, the relativistic correction factors γ_i have to be equal to 1 for everyday life velocities, and have to differ noticeably from that value only for velocities near the speed of light. Even if one did not know the value of the relativistic correction, one could find it by a simple deduction from the collision shown in figure 73.

In the first frame of reference one has $\gamma_1 mv = \gamma_2 MV$ and $\gamma_1 m + m = \gamma_2 M$. From the

Challenge observations of the second frame of reference one deduces that V composed with V gives v, i.e. that

$$v = \frac{2V}{1 + V^2/c^2} \quad . \qquad (110)$$

All put together, one finds that the relativistic correction γ depends on the magnitude of the velocity v through*

$$\gamma_v = \frac{1}{\sqrt{1 - v^2/c^2}} \quad . \qquad (111)$$

Using this expression, and generalizing the situation of galilean physics, one defines the *mass* ratio between two colliding particles as the ratio

$$\frac{m_1}{m_2} = -\frac{\Delta(\gamma_2 v_2)}{\Delta(\gamma_1 v_1)} \quad . \qquad (112)$$

(We do not give the generalized mass definition mentioned in galilean mechanics and based on acceleration ratios, because it contains some subtleties which we will discover shortly.) 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 might have. In this way, the concept of mass remains a number characterizing the difficulty to accelerate a body, and it can still be used for systems of bodies as well.

Following the example of galilean physics, one calls the quantity

$$\mathbf{p} = \gamma m \mathbf{v} \qquad (113)$$

the *(linear) relativistic (three-) momentum* of a particle. Again, the 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.

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

Why relativistic pool 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 pool and similar games, such as snooker or billiard.

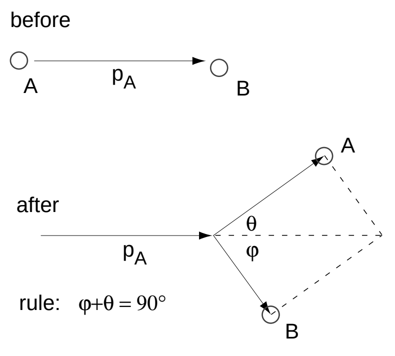


Figure 74 A useful rule for playing non-relativistic pool

After the collision, the two spheres will depart at a *right angle* from each other. However, experiments show that this rule is *not* realized for relativistic collisions. Indeed, using the conservation of momentum one finds with a bit of dexterity that Challenge

$$\tan\theta \tan\varphi = \frac{2}{\gamma + 1} \quad , \tag{114}$$

in other words, the sum $\theta + \varphi$ is *smaller* than a right angle in the relativistic case. Relativistic speeds completely change the game of pool, as every accelerator physicist knows. For electrons or protons, such angles can be easily deduced from photographs taken with cloud chambers, which show the tracks of particles when they fly through them. They all confirm the above expression. If relativity was wrong, most of these detectors would not work.

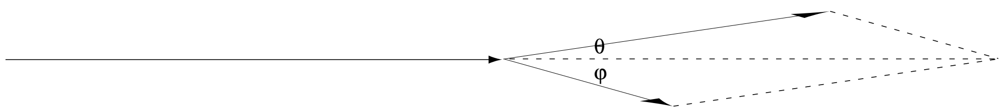


figure to be included

Figure 75 The construction of detectors is based on the relativistic angle rule

Mass is concentrated energy

Let us go back to the simple collinear collision of figure 73. What is the mass M of the final system? A somewhat boring calculation shows that one has Challenge

$$M/m = \sqrt{2(1+\gamma_V)} > 2 \quad . \tag{115}$$

In other words, the mass of the final system is *larger* than the sum 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 (109) of the corrected masses is conserved.

The solution of this puzzle was also given by Einstein. In one of the magic moments of physics history he saw that everything fell into place if for the *energy* of an object of mass m and velocity v he used the expression

$$E = \gamma m c^2 \qquad (116)$$

applying it both to the total system and to each component. The conservation of the corrected mass can then been 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 particles 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 particles. In particular, it follows that the energy E_o and the mass m of a body *at rest* are related by

$$E_o = m c^2 \qquad (117)$$

which is perhaps the most beautiful and famous discovery of modern physics. Since the value for c^2 is so large, we can say that *mass is concentrated energy*. The usual galilean kinetic energy T is then given by

$$T = \gamma m c^2 - m c^2 = \tfrac{1}{2} m v^2 + \tfrac{1 \cdot 3}{2 \cdot 4} m \frac{v^4}{c^2} + \tfrac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \frac{v^6}{c^4} + \ldots \quad . \qquad (118)$$

In other words, special relativity tells that every mass has energy, and that every form of energy, when stored into a system, has mass. Increasing the energy of a system increases its mass, and decreasing the energy content decreases the mass. In short, if a bomb explodes inside a closed box, the mass, weight, and momentum of the box is the same before and after the explosion, but the combined masses of the debris inside the box will be *smaller* than before. All bombs – not only nuclear ones – thus take their energy from a reduction in mass.

By the way, one should be careful and distinguish the transformation of *mass* into energy from the transformation of *matter* into energy. the latter is much more rare. Can you give some examples?

Challenge

On the other hand, expression (116) is the death for many science fiction dreams. The mass energy relation 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 and still are looking for such effects, with negative result. Free energy is unavailable in nature.*

The mass-energy relation $m = E_o/c^2$ also implies that one needs about 90 thousand million kJ (or 21 thousand million kcal) to increase one's weight by one single gram – even though diet experts have slightly different opinions on this matter. In fact, humans do get their everyday energy from the material they eat, drink and breathe by reducing its combined weight before expelling it again. However, this *chemical mass defect* appearing when fuel is burned cannot yet be measured by weighing the materials before and after the reaction; the difference is too small, due to the large conversion factor involved. Indeed, for

* Actually, in the universe there might be still some undiscovered form of energy, called *dark matter*. The issue is not closed yet.

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 mass differences between food and excrements. Therefore, for chemical processes mass can be approximated to be constant, as is indeed done in galilean physics and in everyday life.

However, modern methods of mass measurement of *single molecules* have made it possible to measure the chemical mass defect through comparisons of the mass of a single molecule with that of its constituent atoms. The group of David Pritchard has developed Penning traps which allow to determine masses by measuring frequencies; the attainable precision of these cyclotron resonance experiments is sufficient to confirm $E_o = mc^2$ for chemical bonds. In future, increased precision will even allow to determine precise bond energies in this way. Since binding energy is often radiated away as light, one can say that these modern techniques allow to *weigh* light. Ref. 182

Thinking about light and its mass was also the basis for Einstein's first derivation of the mass-energy relation. When an object emits two equal light beams in opposite directions, its energy decreases by the emitted amount. Since the two light beams are equal in energy and momentum, the body does not move. If the same situation is described from the viewpoint of a moving observer, one gets again that the *rest energy* of the object is Challenge

$$E_o = mc^2 \tag{119}$$

In summary, collisions and any other physical processes need relativistic treatment whenever the energies involved are a sizable fraction of the rest energies.

As a note, human senses detect energies of quite different magnitudes. The eyes can detect light energies of about 1 aJ, whereas the sense of touch can detect only about 10^{μ} J. Which of the two systems is relativistic? Challenge

How are energy and momentum related? The definition of momentum (113) and of energy (116) lead to two basic relations. First of all, their magnitudes are related by

$$m^2c^4 = E^2 - p^2c^2 \tag{120}$$

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

$$\mathbf{p} = \frac{E}{c^2}\mathbf{v} \tag{121}$$

which is equally valid for *any* type of moving energy, be it an object or a beam or a pulse of radiation. We will use both relations regularly in the rest of the walk, including the following situation. Challenge

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. So let us have a look at collisions in more detail, using these new concepts. Obviously 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 a *hit*, i.e. a short distance interaction, is the event at which both objects change momentum. But the two hitting objects are located at *different* points when this happens. A collision is therefore described by a space-time diagram such as the one of figure 76, reminiscent of the Orion constellation. It is easy to check that the process described by such a diagram shows all the properties of a collision.

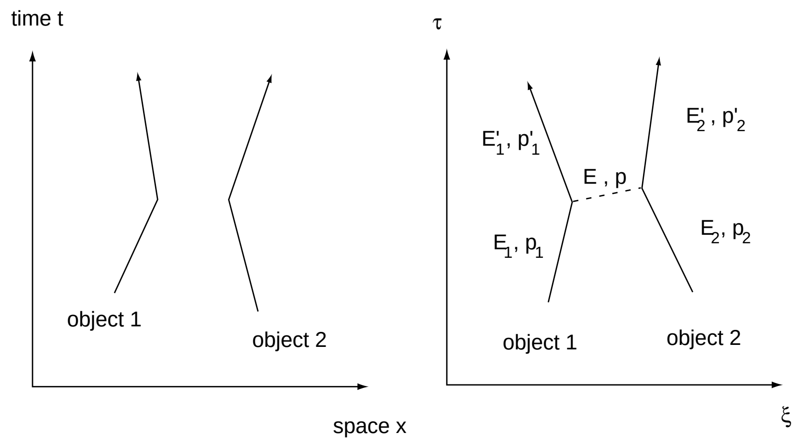


Figure 76 Space-time diagram of two interacting particles

Ref. 183

The right hand side of figure 76 shows the same process seen from another, greek frame of reference. The greek observer says that the first objects has changed its momentum *before* the second one. That would mean that there would be a short interval where momentum and energy are *not* conserved!

The only way to save 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. If we give numerical subscripts to the masses, energies and momenta of the two bodies, and give them an apostrophe after the collision, the unknown mass obeys

Challenge

$$m^2c^4 = E^2 - p^2c^2 = (E_1 - E_1')^2 - (p_1 - p_1')^2c^2 = 2E_1E_1'(\frac{v_1v_1'}{c^2} - 1) < 0 \quad . \qquad (122)$$

This is a strange result, because it means that the mass is an *imaginary* number, not a real and positive one. On top of that, we also see directly from the second graph that the exchanged object moves faster than light. It is a *tachyon*. In other words, collisions involve faster than light motion! 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 can move freely without restriction.* We will study their properties later on, in the part on quantum theory. Only virtual objects may be tachyons. Note that tachyons do not allow energy transport faster than light. Note also that imaginary masses do not violate causality if and only if they are emitted and absorbed with the same probability. Can you confirm all this?

Challenge

There is an additional secret hidden in collisions. In the right part of the figure, the tachyon is emitted by the first object and absorbed by the second one. However, it is easy to find an observer where the opposite happens. In short, the direction of travel of a tachyon depends on the observer! In fact, this is the first hint about *antimatter* we encounter in our walk. We will come back to the topic in detail in the part on quantum theory.

Challenge

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

There we will also discover that a general contact interaction between objects is not described by the exchange of a *single* virtual object, but by a continuous *stream* of virtual particles. In addition, 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 a hand touches another, when it pushes a stone, or while a mountain keeps the trees on it in place, streams of virtual photons are continuously exchanged. This is one of the strange ways we will need to describe nature.

Systems of particles: no centre of mass

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

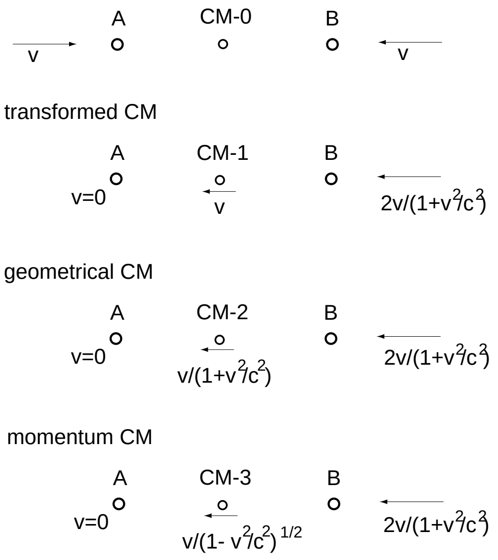


Figure 77 No way to define a center of mass

Figure 77 shows that from the viewpoint in which one of the two particles is at rest, there are at least three different ways to define the center of mass. In other words, the center of mass is not an observer invariant concept. One can also deduce from the figure that the concept does only make sense for systems whose components move against each other with *small* velocities. For other cases, it is not uniquely definable. Will this hinder us in our walk? No. We are more interested in the motion of single particles than that of composite objects or systems. Ref. 184

Why is most motion so slow?

For most everyday cases, the time intervals measured by two different observers are practically equal; only at large relative speeds, typically at more than a few percent of the speed of light, a difference is noted. Most such situations are microscopic. We mentioned already the electrons inside a television tube or inside accelerators. Another example are the particles making up cosmic radiation, which produced so many of the mutations at 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 rapid *macroscopic* bodies? Moving bodies with relativistic velocities, including observers, have a property not found in everyday life; when they get involved in a collision, part of their energy gets converted into new matter via $E = \gamma mc^2$. In the history of the universe this has happened so many times that practically all the bodies still in relativistic motion are microscopic particles only.

A second reason for the disappearance of rapid relative motion is radiation damping. Can you imagine what happens to charges during collisions or in a bath of light? Challenge

In short, almost all matter in the universe moves with small velocity against each other. The few known counterexamples are either very old, such as the quasar jets mentioned above, or stop after a short time. The huge energies necessary for macroscopic relativistic motion are found e.g. in supernova explosions, but stop after only a few weeks. Therefore the universe is filled mainly with slow motion because it is *old*. We universe will determine its age shortly.

Four-vectors

To describe motion consistently for *all* observers, we have to introduce some new quantities. Two ideas are used. First of all, motion of particles is seen as a sequence of events. Secondly, given the coordinates of a particle, we cannot define its 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 *proper time* τ, which is is defined as the time shown by a clock attached to the object. *Relativistic velocity* or *4-velocity* $\mathbf{u}$ of a body is thus defined as the change of the event coordinates or *4-coordinates* $\mathbf{r} = (ct, \mathbf{x})$ with proper time, i.e. as

$$\mathbf{u} = d\mathbf{r}/d\tau \quad . \tag{123}$$

Using $dt = \gamma d\tau$ and thus $dx/d\tau = dx/dt\ dt/d\tau = \gamma dx/dt$, one gets the relation with the 3-velocity $\mathbf{v} = d\mathbf{x}/dt$:

$$u^0 = \gamma c \quad , \quad u^{\mathrm{i}} = \gamma v_{\mathrm{i}} \quad . \tag{124}$$

For small velocities one has $\gamma \approx 1$, 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 $\mathbf{u}$ one finds $\mathbf{uu} = \eta_{ab} u^a u^b = c^2$, which is therefore independent of the magnitude of the 3-velocity $\mathbf{v}$ and makes it a timelike vector, i.e. a vector *inside* the light cone.*

Note that the magnitude of a 4-vector can be zero even though all components of such Challenge so-called *null* vectors are different from zero. Can you find an example for velocity?

* In general, a *4-vector* is defined as a quantity (h_0, h_1, h_2, h_3) which transforms as

$$\begin{aligned} h'_0 &= \gamma(h_0 - v h_1/c) \\ h'_1 &= \gamma(h_1 - v h_0/c) \\ h'_2 &= h_2 \\ h'_3 &= h_3 \end{aligned} \tag{125}$$

when changing form one inertial observer to another moving with a relative velocity v in x direction and the corresponding generalization of the other coordinates. In special relativity, the length of a vector is always defined using the metric η^{ab}, an abbreviation of the matrix

$$\eta^{ab} = \eta_{ab} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad . \tag{126}$$

so that the (square of the) length of a 4-vector $\mathbf{x}$ is defined through $\mathbf{xx} = \eta_{ab} x^a x^b = \eta^{ab} x_a x_b$.

Similarly, the *relativistic acceleration* or *4-acceleration* $\mathbf{b}$ of a body is defined as

$$\mathbf{b} = d\mathbf{u}/d\tau = d\mathbf{x}^2/d\tau^2 \quad . \tag{127}$$

Using $d\gamma/d\tau = \gamma d\gamma/dt = \gamma^4 \mathbf{va}/c^2$, one gets the following relations between the four components of $\mathbf{b}$ and the 3-acceleration $\mathbf{a} = d\mathbf{v}/dt$: Ref. 186

$$b^0 = \gamma^4 \frac{\mathbf{va}}{c} = \frac{\mathbf{va}}{c(1-\frac{v^2}{c^2})^2} \quad , \quad b^i = \gamma^2 a_i + \gamma^4 \frac{(\mathbf{va})v_i}{c^2} = \frac{a_i}{1-\frac{v^2}{c^2}} + \frac{(\mathbf{va})v_i}{c^2(1-\frac{v^2}{c^2})^2} \tag{128}$$

The magnitude of the 4-acceleration is rapidly found via $\mathbf{bb} = \eta_{cd} b^c b^d = -\gamma^6(a^2 - (\mathbf{v} \times \mathbf{a})^2/c^2)$ and thus does depend on the value of the 3-acceleration $\mathbf{a}$. (What is the connection between 4-acceleration and 3-acceleration for a comoving observer?) We note that 4-acceleration lies *outside* the light cone, i.e. that it is a space-like vector, and that $\mathbf{bu} = \eta_{cd} b^c u^d = 0$, which means that the 4-acceleration is always perpendicular to the 4-velocity.* We also note from the expression that accelerations, in contrast to velocities, cannot be called relativistic; the difference between b_i and a_i or between their two magnitudes does not depend on the value of a_i, but only on the value of the speed v. In other words, accelerations require relativistic treatment only when the involved velocities are relativistic. If the involved velocities are small, even the highest accelerations can be treated with galilean methods. Challenge

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

$$\mathbf{P} = m\mathbf{u} \tag{131}$$

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

$$\mathbf{P} = (\gamma mc, \gamma m\mathbf{v}) = (E/c, \mathbf{p}) \quad . \tag{132}$$

For this reason 4-momentum is also called the *energy-momentum 4-vector*. In short, *the 4-momentum of a body is given by mass times 4-displacement per proper time*. This is the simplest possible definition of momentum and energy. The energy-momentum 4-vector, like the 4-velocity, is *tangent* to the world line of a particle. This follows directly from the definition, since

$$(E/c, \mathbf{p}) = (\gamma mc, \gamma m\mathbf{v}) = m(\gamma c, \gamma\mathbf{v}) = m(dt/d\tau, d\mathbf{x}/d\tau) \quad . \tag{133}$$

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

$$\mathbf{J} = d\mathbf{b}/d\tau = d^2\mathbf{u}/d\tau^2 \quad . \tag{129}$$

Challenge For the relation with the 3-jerk $\mathbf{j} = d\mathbf{a}/dt$ one then gets

$$\mathbf{J} = (J^o, J^i) = (\frac{\gamma^5}{c}(\mathbf{jv} + a^2 + 4\gamma^2 \frac{(\mathbf{va})^2}{c^2}) \quad , \quad \gamma^3 j_i + \frac{\gamma^5}{c^2}((\mathbf{jv})v_i + a^2 v_i + 4\gamma^2 \frac{(\mathbf{va})^2 v_i}{c^2} + 3(\mathbf{va})a_i)\,) \tag{130}$$

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

The (square of the) length $\mathbf{PP} = \eta_{ab} P^a P^b$ of the energy-momentum vector is by definition the same for all observers and found to be

$$E^2/c^2 - p^2 = m^2 c^2 \quad , \tag{134}$$

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

Challenge (What is their energy-momentum relation?)

Note that in this escalation, by the term 'mass' m we always mean what is sometimes also called the *rest mass*. This name derives from the bad habit of many science fiction and high school books to call the product γm the *relativistic mass*. Workers in the field reject this concept, as did Einstein himself, and they also reject the often heard sentence that "(relativistic) mass increases with velocity." This last statement is more of the intellectual level of mass tabloids, and not worthy of any motion expert. Ref. 185

We note that 4-force $\mathbf{K}$ is defined as

$$\mathbf{K} = d\mathbf{P}/d\tau = m\mathbf{b} \tag{135}$$

and that therefore, contrary to an often heard statement, force remains mass times acceleration in relativity. From the definition of $\mathbf{K}$ one deduces the relation with 3-force $\mathbf{F} = d\mathbf{p}/dt = md(\gamma\mathbf{v})/dt$, namely* Ref. 186, 187

$$\mathbf{K} = (K^o, K^i) = (\gamma^4 m\mathbf{va}/c, \gamma^2 m a_i + \gamma^4 v_i \frac{m\mathbf{va}}{c^2}) = (\frac{\gamma}{c}\frac{dE}{dt}, \gamma\frac{d\mathbf{p}}{dt}) = (\gamma\frac{\mathbf{Fv}}{c}, \gamma\mathbf{F}) \quad . \tag{136}$$

Challenge Also the 4-force, like the 4-acceleration, is orthogonal to the 4-velocity. One easily recognizes the meaning of the zeroth component of the 4-force: it is the *power* required to accelerate the object. But since force is not an important concept in physics, we turn to a different topic. See page 89

Rotation in relativity

If at night one turns around one's own axis while looking at the sky, the stars move with a much higher velocity 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?

The 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 is not in contrast with the concept of a limit speed. Mathematically speaking, the speed of light limits relative velocities *only* between objects which come *near* to each other. To compare velocities of distant objects is only possible if all velocities involved are constant in time; this is not the case in the present example. Avoiding this limitation is one of the reasons to prefer the differential version of the Lorentz transformations. In many general cases relative velocities of *distant* objects can be larger than the speed of light. We encountered a first ex- See page 162

* Some authors define 3-force as $\mathbf{F} = d\mathbf{p}/d\tau$. It is important to note that in relativity, 3-force $\mathbf{F}$ is indeed proportional to 3-acceleration $\mathbf{a}$; however, force and acceleration are not parallel to each other. In fact one finds $\mathbf{F} = \gamma m\mathbf{a} + (\mathbf{Fv})\mathbf{v}/c^2$. In contrast, in relativity 3-momentum is not even proportional to 3-velocity. (Challenge)

ample above, when discussing the car in the tunnel, and we will encounter a few additional examples shortly. See page 184

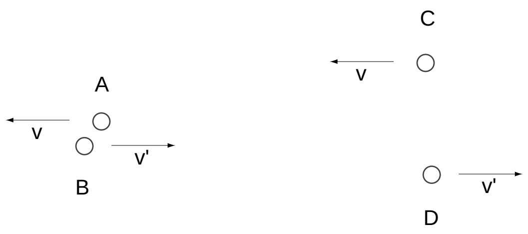


Figure 78 On the definition of relative velocity

With this clarification, we can have a short look at *rotation* in relativity. The first question is how lengths and times change in a rotating frame of reference. You might want to check that an observer in a rotating frame agrees with a non-rotating colleague on the radius of rotating bodies; however, both find that the rotating body has a *different circumference* than before it started rotating. Sloppily speaking, the value of *changes* for rotating observers.

We mention that 4-angular momentum is defined naturally as

$$l^{ab} = x^a p^b - x^b p^a \quad . \tag{137}$$

In other words, 4-angular momentum is a *tensor*, not a vector, as shown by its two indices. Angular momentum is obviously conserved also in special relativity, so that there are no surprises on this topic. Challenge

Is angular velocity limited? Yes; the tangential speed in an inertial frame of reference cannot exceed that of light. The limit thus depends on the *size* of the body in question. That leads to a neat puzzle: can one *see* objects rotating very rapidly? Challenge

The moment of inertia is defined as the proportionality factor between angular velocity and angular momentum.

– CS – Text to be filled in. – CS –

Obviously, for a rotating particle, the rotational energy is part of rest mass. You might want to calculate the fraction for the earth and the sun. By the way, how would you determine whether a small particle is rotating? Challenge Challenge

The action of a free particle

If we want to describe relativistic motion of a free particle with an extremal principle, we can easily guess that the lagrangian of a free particle is

$$S = -mc^2 \int_{\tau_1}^{\tau_2} d\tau \tag{138}$$

for the proper time τ. The action can also be written as Challenge

$$S = \int L\,dt = -mc \int_{\tau_1}^{\tau_2} \sqrt{u_a u^a}\,d\tau = -mc \int_{\tau_1}^{\tau_2} \sqrt{\eta^{ab} \frac{dx_a}{d\tau} \frac{dx_b}{d\tau}}\,d\tau \tag{139}$$

where the *metric* $\eta^{\alpha\beta}$ of special relativity is given as

$$\eta^{ab} = \eta_{ab} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad . \tag{140}$$

One can easily confirm the form of the lagrangian by deducing the equation of motion by
Challenge the usual procedure.

In short, nature is in not 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 principle is also valid for particles under the influence of gravity, as we will see in the section on general relativity. In nature, *proper time is always maximal*. Speculating on the deeper meaning of this result is left to your personal preferences; enjoy it!

Conformal transformations: Why is the speed of light constant?

The distinction between space and time in special relativity depends on the observer. On the other hand, all inertial observers at a point do agree on the position and the orientation of the light cone. The light cone is thus the basic physical "object" with which space-time is described in the theory of relativity. Given its importance, one might ask if the set of all those observers which observe the same light cones corresponds to the set of all inertial observers. Interestingly, it turns out that there are other such observers.

The first group of these additional observers are those using different units of measurement, namely units in which all time and length intervals are multiplied by the same *scale factor* λ. The transformations among these points of view are given by

$$x_a \mapsto \lambda x_a \tag{141}$$

and are called *dilations*.

There is a second type of additional observers, which one finds by applying the so-called *special conformal transformations*. They are combinations of an *inversion*

$$x_a \mapsto \frac{x_a}{x^2} \tag{142}$$

with a *translation* by a vector b_a, namely

$$x_a \mapsto x_a + b_a \tag{143}$$

and a second inversion. This gives for the expression for the special conformal transformations

$$x_a \mapsto \frac{x_a + b_a x^2}{1 + 2bx + b^2 x^2} \quad \text{or} \quad \frac{x_a}{x^2} \mapsto \frac{x_a}{x^2} + b_a \quad . \tag{144}$$

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

Challenge * For the massless neutrinos, the action does not work. Why? Can you find an alternative?

imally) small shapes; they thus leave the *form* (of infinitesimally small objects) unchanged. For example, they transform infinitesimal circles into infinitesimal circles. They are called *special* because the complete *conformal group* is the one including the dilations and the inhomogeneous Lorentz transformations as well.*

The way in which special conformal transformations leave light cones invariant is a bit subtle.

– CS – Text to be filled in. – CS –

Note that since dilations do not commute with time translations, there is no conserved quantity associated with this symmetry. (The same happens with Lorentz boosts; in contrast, rotations and spatial translations do commute with time translations and thus do lead to conserved quantities.)

In summary, the vacuum is conformally invariant – in the special way just mentioned – and thus also dilation invariant. This is another way to say that vacuum alone is not sufficient to define lengths; matter is necessary to do so. Indeed, (special) conformal transformations are not symmetries of situations containing matter. Only the 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. Obviously, conformal invariance is also *necessary* for velocity measurements, as you might want to check. Conformal invariance includes inversion symmetry. Challenge
Inversion symmetry means that the large and the small scales of the vacuum are related. It thus seems as if the constancy of the speed of light is related to the existence of inversion symmetry. This surprising and mysterious connection gives us a glimpse of the adventures we will encounter in the third part of our escalation. As we will see there, conformal invariance turns out to be an important property of the vacuum, even though it is hidden by the presence of matter. It will produce many incredible surprises.*

Challenge
* 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 conformal group. The conformal group is locally isomorph to SU(2,2) and to the simple group SO(4,2). Note that all this is true only *four* space-time dimensions; in *two* dimensions, the other important case, especially in string theory, the conformal group is isomorphic to the the group of arbitrary analytic coordinate transformations, and is (thus) infinite-dimensional.

* The conformal group does not appear only in the kinematics of special relativity; it is the symmetry group of all physical interactions, such as electromagnetism, if all the particles involved have zero mass, as is the case for the photon. Any 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 is not conformally invariant? Challenge

But since all particles observed up to now have masses which are many orders of magnitude smaller than the Planck mass, from a global viewpoint it can be said that they have almost vanishing mass; conformal symmetry then can be seen as an approximate symmetry of nature. In this view, all massive particles should 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.

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. But when one or both of the observers is accelerating, the story gets even more interesting.**

As an appetizer, let us see what an accelerating, greek observer says about the clock of
Ref. 188 an inertial, roman one, and vice versa. Assume that the greek observer moves along $\mathbf{x}(t)$, as observed by the inertial roman one. In general, the roman/greek clock rate ratio is given by $\Delta\tau/\Delta t = (\tau_2 - \tau_1)/(t_2 - t_1)$, where the greek coordinates given are those constructed with a simple procedure: take the set of events defined by $t = t_1$ and $t = t_2$, and determine where these sets intersect the time axis of the other, greek observer, and call them τ_1 and τ_2.*** One sees that the clock ratio of a greek observer, in the case that the greek observer is inertial and moving with velocity v as observed by the roman one, is given by

$$\frac{\Delta\tau}{\Delta t} = \frac{d\tau}{dt} = \sqrt{1 - v^2/c^2} = \frac{1}{\gamma_v} \qquad (145)$$

Challenge as we are now used to. We see again that moving clocks run slow.

For accelerated motions, the differential ver-
Ref. 188 sion of the reasoning is necessary. In other words, the roman/greek clock rate ratio is again $d\tau/dt$, and τ and $\tau + d\tau$ are calculated in the same way as just defined from the times t and $t + dt$. Assume again that the greek observer moves along $\mathbf{x}(t)$, as measured by the roman one. One finds directly that

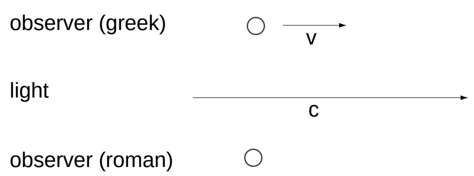


Figure 79 Simplified coordinate diagram for an inertial and an accelerated observer

$$\tau = t - \mathbf{x}(t)\mathbf{v}(t) \quad . \qquad (146)$$

Similarly, one gets that

$$\tau + d\tau = (t + dt) - [\mathbf{x}(t) - dt\,\mathbf{v}(t)][\mathbf{v}(t) + dt\,\mathbf{a}(t)] \quad ; \qquad (147)$$

together one finds

$$\text{“}d\tau/dt\text{”} = \gamma_v(1 - \mathbf{v}\mathbf{v}/c^2 - \mathbf{x}\mathbf{a}/c^2) \quad . \qquad (148)$$

This result shows that accelerated clocks can run *fast* instead of slow, depending on their position $\mathbf{x}$ and the sign of their acceleration $\mathbf{a}$. There are quotes in the expression because one sees directly that the other, greek observer notes that

$$\text{“}dt/d\tau\text{”} = \gamma_v \qquad (149)$$

** One sometimes hears that special relativity cannot be used to describe accelerating observers. The reasoning is wrong: it would imply that galilean physics could not be used for accelerating observers either, in contrast to everyday experience. Special relativity's only limitation is that it cannot be used in non-flat space-time. Accelerating bodies do exist in flat space-times, and therefore can be discussed in special relativity.

*** These sets form what mathematicians call *hypersurfaces*, as we will see later on.

which is *not* the inverse of equation (148). This difference becomes most apparent in the simple case of two clocks with the same velocity, of which one is accelerated constantly towards the origin with magnitude g, whereas the other moves inertially. One then has

$$\text{“}d\tau/dt\text{”} = 1 + gx \tag{150}$$

and

$$\text{“}dt/d\tau\text{”} = 1 \quad . \tag{151}$$

We will encounter the situation shortly.

Another difference with the case for velocities is the way accelerations change under change of viewpoints. Let us only take the simple case in which everything moves along the x-axis: the object and the two inertial observers. If the roman observer measures an acceleration $a = dv/dt = d^2x/dt^2$, the greek an acceleration $\alpha = d^2\xi/dt^2$, and if their relative velocity is V, one gets Challenge

$$\gamma_V^3 \alpha(t) = \frac{a(t)}{(1 + v(t)V/c^2)^3} \quad . \tag{152}$$

The relation shows that accelerations are *not* Lorentz invariant; they are so only if velocities are small compared to the speed of light. Note that expression (152) simplifies in the case that the accelerations are measured at a time t in which v vanishes.

In summary, for accelerated observers results get more complicated than for inertial ones. This is such an interesting feature that it merits a deeper investigation. To keep matters simple, we only study *constant* accelerations. Interestingly, this situation is a good introduction to black holes and, as we will see shortly, to the whole universe. See page 278

Accelerating frames of reference

How does one check whether one lives in an inertial frame of reference? An *inertial frame (of reference)* has two properties: firstly, for any two observers of that frame the ratio c between twice the rod distance and the time taken by light to travel from one to the other and back is always the same: it is independent of time and of the choice of the pair of observers. Secondly, rod lengths are described by Euclidean geometry. 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 remain always at rest with respect to each other. Interestingly, there are other, non-inertial situations where this is the case.

In special relativity it indeed makes sense to talk about non-inertial frames, i.e. about *accelerating frames*. In these frames, many strange things happen, of which we will mention the most important ones. Ref. 189

A general *frame of reference* is a continuous set of observers who remain at rest with respect to each other. Here, ‘at rest to each other’ means that the time for a light signal to go from one observer to another and back again is constant in time, or, if one prefers, that the rod distance between the two observers is constant in time. Any frame of reference can therefore also be called a *rigid* collection of observers. One has to note that therefore a general frame of reference is *not* the same as a set of coordinates; the latter usually is *not* rigid. In the special case that one has chosen the coordinate system in such a way that all the rigidly connected observers have constant coordinate values, one speaks of a *rigid coordinate system*. Obviously, these are the most useful to describe accelerating frames of references.*

Ref. 189 Note that if two observers both move with a velocity **v**, as measured in some *inertial* frame, they observe to be at rest to each other *only* if this velocity is *constant*. Again we (Challenge) find, as above, that two persons 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 together they accelerate to (or decelerate from) relativistic speeds in precisely the same way. Relativistic acceleration requires careful thinking.

An observer which feels always the same force on his body is called *uniformly* accelerating. More precisely, a uniformly accelerating observer Ω is thus an observer whose acceleration at every moment, measured by the inertial frame with respect to which he is at rest *at that moment*, has always the same value **b**. It is important to note that uniform acceleration is *not* uniformly accelerating when observed always from the *same* inertial frame K. This is an important difference with respect to the galilean case.

For uniformly accelerated motion in the sense just defined, one needs

$$\mathbf{b} \cdot \mathbf{b} = -g^2 \qquad (153)$$

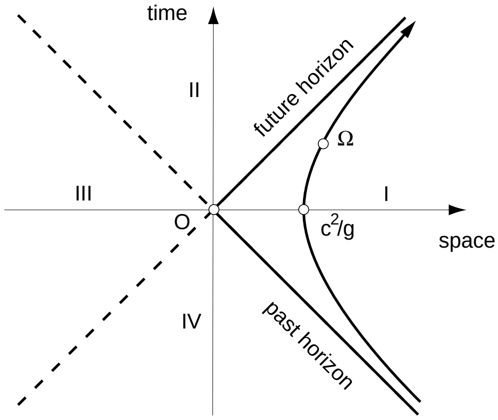


Figure 80 The hyperbolic motion of an rectilinearly, uniformly accelerating observer and its event horizons

where g is a constant independent of t. The simplest case is uniformly accelerating motion which is also *rectilinear*, i.e. for which the acceleration **a** is parallel to **v** at one instant of time Ref. 191 and (therefore) for all other times as well. In this case one can write, using three-vectors,

$$\gamma^3 \mathbf{a} = \mathbf{g} \quad \text{or} \quad \frac{d\gamma\mathbf{v}}{dt} = \mathbf{g} \quad . \qquad (154)$$

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

$$v = \frac{gt}{\sqrt{1 + \frac{g^2 t^2}{c^2}}} \qquad (155)$$

where it was assumed that $v_o = 0$. One notes that for small times, one gets $v = gt$, as expected, and for large times one gets $v = c$, also as expected. The momentum of the greek (Challenge) observer increases linearly with time, gain as expected. Integrating, one finds that the accel-

Ref. 195 * There are essentially only two other types of rigid coordinate frames, apart from the inertial frames:

- the frame $ds^2 = dx^2 + dy^2 + dz^2 - c^2 dt(1 + g_k x_k / c^2)^2$ with arbitrary, but constant acceleration of the origin. The acceleration is $\mathbf{a} = -\mathbf{g}(1 + \mathbf{g}x/c^2)$;

- the uniformly rotating frame $ds^2 = dx^2 + dy^2 + dz^2 + 2\omega(-y\,dx + x\,dy)dt - (1 - r^2\omega^2/c^2)dt$. Here the z-axis is the rotation axis, and $r^2 = x^2 + y^2$.

erated observer Ω moves along the path

$$x(t) = \frac{c^2}{g}\sqrt{1+\frac{g^2t^2}{c^2}} \qquad (156)$$

where it was assumed that $x_o = c^2/g$, in order to keep the expression simple. Due to this result, visualized in figure 80, a rectilinearly and uniformly accelerating observer is said to undergo *hyperbolic* motion. For small times, the world-line reduces to the usual $x = gt^2/2 + x_o$, whereas for large times the result is $x = ct$, as expected. The motion is thus uniformly accelerated only for the moving body itself, *not* for an outside observer.

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 (155) one gets* Ref. 191, 192

$$t = \frac{c}{g}\sinh\frac{g\tau}{c} \quad \text{and} \quad x = \frac{c^2}{g}\cosh\frac{g\tau}{c} \quad . \qquad (157)$$

as relation between proper time τ and the time t and the position x measured by the external, inertial roman observer. We will encounter this relation again during the study of black holes. All this sounds boring? Just imagine to accelerate with the motor bike value $g = 10\,\mathrm{m/s^2}$ 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 motor bikes nor missiles accelerating in this way do exist, as their fuel tank would be enormous. Can you confirm this even for the most optimistic case? Challenge

The coordinates transform as

$$\begin{aligned} t &= \left(\frac{c}{g}+\frac{\xi}{c}\right)\sinh\frac{g\tau}{c} \\ x &= \left(\frac{c^2}{g}+\xi\right)\cosh\frac{g\tau}{c} \\ y &= \upsilon \\ z &= \zeta \end{aligned} \qquad (158)$$

where τ now is the time coordinate in the greek frame. One notes also that

$$d\sigma^2 = (1+g\xi/c^2)^2c^2d\tau^2 - d\xi^2 - d\upsilon^2 - d\zeta^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 \quad ; \qquad (159)$$

and since for $d\tau = 0$ distances are given by the pythagorean theorem, the greek reference frame is indeed rigid. Ref. 190

After this forest of formulas, let's tackle a simple question. The roman observer O sees the greek observer Ω departing with acceleration g, moving further and further away, following equation (156). What does the greek observer say about his roman colleague? With all the experience we have now, that is easy. At each point of his trajectory the greek observer sees that O has the coordinate $\tau = 0$, (can you confirm this?) which means that the distance to the roman observer, as seen by greek one, is the same as the space-time interval OΩ. Using expression (156) that turns out to be Challenge Ref. 193

* Use your favorite mathematical formula collection to deduce this. The best and cheapest remains the one by K. Rottmann, *Mathematische Formelsammlung*, BI Hochschultaschenbücher, 1960. The abbreviations $\sinh y = (e^y - e^{-y})/2$ and $\cosh y = (e^y + e^{-y})/2$ imply that $\int dy/\sqrt{y^2+a^2} = \mathrm{arsinh}\, y/a = \mathrm{Arsh}\, y/a = \ln(y+\sqrt{y^2+a^2})$.

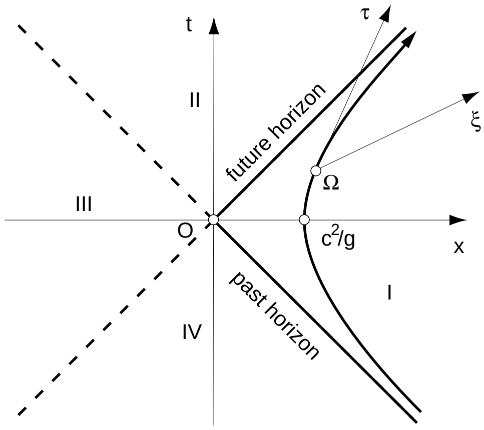


Figure 81 Do accelerated objects depart form inertial ones?

$$d_{O\Omega} = \sqrt{\xi^2} = \sqrt{x^2 - c^2 t^2} = c^2/g \tag{160}$$

which, surprisingly enough, is constant in time! In other words, the greek observer will say that he stays at 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. (Are you able to guess the relation to this issue?) Challenge

As a note, a nice and challenging problem is to deduce the addition theorem for accelerations. Challenge

Event horizons

The surprises of accelerated motion are not finished yet. Of special interest is the trajectory, in the rigidly accelerated frame coordinates ξ and τ, of an object located at the departure point $x = x_o = c^2/g$ at all times t. One gets* Challenge

$$\begin{aligned} \xi &= -\frac{c^2}{g}(1 - \operatorname{sech}\frac{g\tau}{c}) \quad \text{and} \\ d\xi/d\tau &= -c \operatorname{sech}\frac{g\tau}{c} \tanh\frac{g\tau}{c} \quad . \end{aligned} \tag{162}$$

These equations are strange. One notes that for large times τ the coordinate ξ approaches the limit value $-c^2/g$ and that $d\xi/d\tau$ approaches zero. The situation is similar to a car accelerating away from a man standing on a long road. Seen from the car, the man does move away; however, after a while, the only thing one notices is that he slowly approaches the horizon. In galilean physics, both the car driver and the man of the road see the other

* The functions appearing above are defined using the expressions from the footnote on page 182:

$$\operatorname{sech} y = \frac{1}{\cosh y} \quad \text{and} \quad \tanh y = \frac{\sinh y}{\cosh y} \tag{161}$$

approaching each others horizon; in special relativity, only the accelerated observer makes this observation.

Studying a graph of the situation confirms the result. In figure 81 one notes that light emitted from any event in regions II and III cannot reach the greek observer. Those events are hidden from him; he can never see them. Strangely enough, light from the observer *can* reach region II. The boundary between the part of space-time that can be observed and that which cannot is called the *event horizon*. In relativity, event horizons act like one-way gates for light and for any other signal. For completeness, the graph also shows the past event horizon.

In summary, not all events observed in an inertial frame of reference can be observed in an uniformly accelerating frame of reference. Uniformly accelerating frames of references produce event horizons at a distance $-c^2/g$. A person can never see further than this distance below his feet.

By the way, is it true that a light beam cannot catch up with an observer in hyperbolic motion, if the observer has a sufficient distance advantage at the start? Challenge

Acceleration changes colours

We saw above that a moving receiver sees different colours than the sender. This colour shift or Doppler effect was discussed above for inertial motion only. For accelerating frames the situation is even stranger: sender S and receiver R do not agree on colours even if they are at *rest* with respect to each other. Indeed, if light is emitted in the direction of the acceleration, Ref. 191, 194 the expression for the space-time interval gives

$$c^2 d\tau^2 = (1 + \frac{g_o x}{c^2})^2 dt^2 \tag{163}$$

in which g_o is the proper acceleration of an observer located at $x = 0$. One can deduce in a straightforward way that Challenge

$$f_R/f_S = 1 - \frac{g_R h}{c^2} = 1/(1 + \frac{g_S h}{c^2}) \tag{164}$$

where h is the rod distance between the source and the receiver, and where $g_S = g_o/(1 + g_o x_S/c^2)$ and $g_R = g_o/(1 + g_o x_R/c^2)$ are the proper accelerations measured at the x-coordinates of the source and at the detector. In short, the frequency of light decreases when light moves in the direction of acceleration. By the way, does this have an effect on the colour of trees along their vertical extension? Challenge

The formula usually given, namely

$$f_R/f_S = 1 - \frac{gh}{c^2} \quad , \tag{165}$$

is only correct to first approximation, and not exactly what was just found. In accelerated frames of reference, one has to be careful with the meaning of every quantity used. For everyday accelerations however, the differences between the two formulas are negligible. Are you able to confirm this? Challenge

Can light move faster than c?

What light speed is measured by an accelerating observer? Using expression (165) above, an accelerated observer deduces that

Challenge

$$v_{\text{light}} = c(1 + \frac{gh}{c}) \qquad (166)$$

which is larger than c in case that light moves in front or "above" him, and lower than c for light moving behind or "below" him. This strange result on the speed of light follows from the fact that in an accelerating frame of reference, even though all observers are at rest with respect to each other, clocks do *not* remain synchronized. This effect has been also confirmed by experiment. In other words, the speed of light is only constant when it is defined as $c = dx/dt$, *and* if dx and dt are measured with a rod located at a point *inside* the interval dx and a clock read off *during* an instant inside the interval dt. If the speed of light is defined as $\Delta x/\Delta t$, or if the rod defining distances or the clock measuring times are located away from the propagating light, the speed of light comes out to be different from c for accelerating observers. For the same reason, turning around your axis at night leads to star velocities much higher than the speed of light.

Ref. 196

Note that this result does not imply that signals or energy can be moved faster than c, as you may want to check for yourself. In fact, these difficulties are only noticeable for distances l which do not obey the relation $l \ll c^2/a$. This means that for an acceleration of $9.5\,\mathrm{m/s^2}$, about that of free fall, distances would have to be of the order of one light year, i.e. larger than $9.5 \cdot 10^{12}$ km, in order to observe any sizable effects. In short, *c is the speed of light relative to* nearby *matter only.*

Challenge

By the way, everyday gravity is equivalent to a constant acceleration. Why do distant objects, such as stars, still not move faster than light nevertheless?

Challenge

What is the speed of light?

We have thus seen that the speed of light, as usually defined, is given by c only if either the observer is inertial or if the observer measures the speed of light passing him nearby, instead of light passing at a distance. In short, the speed of light has to be measured locally. But this request does not eliminate all subtleties.

Ref. 198

An additional point is often forgotten. Usually, length is measured by light travel time. In that case the speed of light will be obviously constant. How does one check the constancy in this case? One needs to eliminate length measurements. The simplest way is to reflect light from a mirror. The constancy 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

$$t_3 - t_1 = 2\,(t_2 - t_1) \quad . \qquad (167)$$

Here it was assumed that the clocks were synchronized according to the prescription of page 159. If the factor would

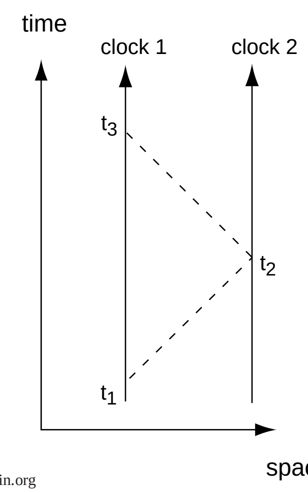


Figure 82 Clocks and the measurement of the speed of light as two-way velocity

not be exactly two, the speed of light would not be constant. Of course, experiments always yield a factor of two within measurement errors.

This result is often expressed by saying that it is impossible to measure the *one-way velocity of light*; only the *two way* velocity of light is measurable. Can you confirm this? Challenge

Limits on length of rigid bodies

An

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

$$\frac{v^2}{c^2} \geqslant \frac{2d}{l} \quad . \tag{168}$$

One sees that one can avoid the breaking of fragile objects by packing them into foam rubber – which increases the stopping distance – of roughly the same thickness than the object's size. This may explain why boxes with presents usually are so much larger than their contents...

Similarly, as proposed by Kewin Warnick, the speed of light limits the size of solid bodies. Ref. 199 Imagine to accelerate the front of a *rigid* body with some proper acceleration α. The back end cannot move with an acceleration equal or larger than infinity; therefore the length l of a rigid body must follow

$$l < \frac{c^2}{\alpha} \quad . \tag{169}$$

Are you able to confirm this? Challenge There is a simple argument leading to the result, by looking at the speeds of the front and the back ends. In any case, for $9.8\,\mathrm{m/s^2}$, the acceleration of a decent motor bike, this gives a length limit of 9.2 Pm, about a light year. Not a big restriction; most motor bikes are shorter.

But there are other, more interesting situations. The highest accelerations achievable today are produced in particle accelerators. Atomic nuclei have a size of about 1 fm. Are you able to deduce at which energies they break when smashed together in an accelerator?** Challenge

But an additional point deserves mention. No physical body can actually be rigid, however short it may be, because the speed of light is finite, not infinite. When one pushes a body on one end, the other end always moves a little bit later. What does this mean for the size of elementary particles? When the acceleration is electrostatic, one finds that the particle size must be smaller than the classical electron radius given by e^2/mc^2. Ref. 199 We will come back to the size issue of elementary particles later on.

* For glass and metals the (longitudinal) *speed of sound* is about 5.9 km/s for glass, iron or steel, and 4.5 km/s for gold; for lead about 2 km/s, for beryllium it is about 12.8 km/s. In comparison, it is 1.5 km/s for rubber and water, about 1.1 km/s for most other liquids, and about 0.3 km/s for air and almost all gases, except helium, where it is 1.1 km/s. (All these values are at room temperature and standard pressure.)

** However, inside a nucleus, the nucleons move with accelerations of the order of $v^2/r \approx \hbar^2/m^2 r^3 \approx 10^{31}\,\mathrm{m/s^2}$; this is one of the highest values found in nature.

Special relativity in four sentences

This section of our walk is rapidly summarized.

• All free floating observers find that there is a perfect velocity in nature, namely a common maximum energy velocity, which is realized by massless radiation such as light or radio signals, or by neutrinos (provided they are indeed massless), but cannot be achieved by ordinary material systems.

• Therefore, even though space-time is the same for every observer, times and lengths vary from one observer to the other, as described by the Lorentz transformations (95) and (96), and as confirmed by experiment.

• Collisions show that this implies that mass is concentrated energy, and that the total energy of a body is given by $E = \gamma m c^2$, as again confirmed by experiment.

• Applied to accelerated objects, these results lead to numerous counterintuitive consequences, such as the twin paradox, the appearance of event horizons, and the appearance of short-lived tachyons in collisions.

In summary, special relativity shows that motion is relative, defined using propagation of light, conserved, reversible, and deterministic.

During our earlier exploration of galilean physics, once we had defined the basic concepts of velocity, space, and time, we turned our attention to gravitation. Since experiments forced us to change these everyday concepts, we now return to study gravitation in light of these changes.

Aiunt enim multum legendum esse, non multa.
Plinius*

References

1 A beautiful book explaining physics and its many applications in nature and technology vividly and throughly is Paul G. HEWITT, John SUCHOCKI & Leslie A. HEWITT, *Conceptual physical science*, Second Edition, 1999. Cited on page 26.

2 A well-known principle in the social sciences states that given a question, for every possible answer, however weird it may seem, there is somebody – and often a whole group – who holds it as his opinion. One just has to go through literature to find out. This sociological "law of nature" is discussed in ...

About group behaviour in general, see R. AXELROD, *The evolution of cooperation*, Harper Collins, 1984. The propagation and acceptance of ideas, such as those of physics, are also an example of human cooperation, with all its potential dangers and weaknesses. Cited on page 27.

3 All the known texts by Parmenides and Heraclitos can be found in J.P. DUMONT, *Les écoles présocratiques*, Folio – Gallimard, 1988. Views about the non-existence of motion have also been put forward by much more modern, and much more despicable authors, such as in 1710 by Berkeley. Cited on page 27.

4 An example of people worried by Zeno is given by *William McLaughlin*, *Resolving Zeno's paradoxes*, Scientific American pp. 66–71, November 1994. The actual argument was not about a hand slapping a face, but about an arrow hitting the target. See also reference 30. Cited on page 27.

5 For other interesting physical effects in everyday life, see also Erwein FLACHSEL, *Hundertfünfzig Physikrätsel*, Ernst Klett Verlag, 1985. The book also covers several clock puzzles, in puzzle numbers 126 to 128. Cited on page 29, 71.

6 A concise and informative introduction into the history of classical physics is given in the first chapter of the book by F.K. RICHTMEYER, E.H. KENNARD & J.N. COOPER, *Introduction to modern physics*, McGraw-Hill, New York, 1969. Cited on page 29.

7 A good overview over the arguments used to prove the existence of god from motion is given by Michael BUCKLEY, *Motion and motion's god*, Princeton University Press, 1971. The intensity of the battles waged around these failed attempts is one of the tragicomic chapters of history. Cited on page 29.

8 A good and fun book on behaviour change is the well-known text by R. BANDLER, *Using your brain for a change*, Real People Press, 1985. See also Richard BANDLER & John GRINDER, *Frogs into princes – Neuro Linguistic Programming*, Eden Grove Editions, 1990. Cited on page 30, 31.

9 A beautiful book about the mechanisms of human growth from the original cell to full size is Lewis WOLPERT, *The triumph of the embryo*, Oxford University Press, 1991. Cited on page 30.

10 On the topic of grace and poise, see e.g. the numerous books on the Alexander technique, such as M. GELB, *Body learning – an introduction to the Alexander technique*, Aurum Press, 1981, and Richard BRENNAN, *Introduction to the Alexander Technique*, Little Brown and Company, 1996. Among others, the idea of the Alexander technique is to return to the situation that the muscles groups for sustention and those for motion are used only for their respective function,

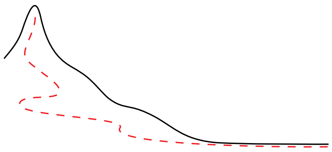
Typeset in April 2001

See page 411

∗ 'Read a lot, not anything.' Ep. 7, 9, 15

This is a section of the freely downloadable e-textbook

Motion Mountain



Hiking beyond space and time along the concepts of modern physics

available at
www.motionmountain.org

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?

Challenge
- Did you find any mistakes?

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

and not vice versa. Any unnecessary muscle tension, such as neck stiffness, is a waste of energy due to the use of sustention muscles for movement and of motion muscles for sustention. The technique teaches the way to return to the natural use of muscles.

Motion of animals was discussed extensively already in the 17th century by G. BORELLI, *De motu animalium*, 1680. An example of a more modern approach is J.J. COLLINS & I. STEWART, *Hexapodal gaits and coupled nonlinear oscillator models*, Biological Cybernetics **68**, pp. 287–298, 1993. See also I. STEWART & M. GOLUBITSKY, *Fearful Symmetry*, Blackwell, 1992. Cited on page 31, 65.

11 A description of the reptile brain in comparison to the mammalian and the human one can be found in ... Cited on page 32.

12 The lower left corner movie can be reproduced on a computer after typing the following lines in the Mathematica software package: Cited on page 33, 33.

```
<< Graphics'Animation'
Nxpixels=72; Nypixels=54; Nframes=Nxpixels 4/3;
Nxwind=Round[Nxpixels/4]; Nywind=Round[Nypixels/3];
front=Table[Round[Random[]],{y,1,Nypixels},{x,1,Nxpixels}];
back =Table[Round[Random[]],{y,1,Nypixels},{x,1,Nxpixels}];
frame=Table[front,{nf,1,Nframes}];
Do[ If[ x>n-Nxwind && x<n && y>Nywind && y<2Nywind,
   frame[[n,y,x]]=back[[y,x-n+1]] ],
  ]      {x,1,Nxpixels}, {y,1,Nypixels}, {n,1,Nframes}]
film=Table[ListDensityPlot[frame[[nf]], Mesh-> False,
   Frame-> False, AspectRatio-> N[Nypixels/Nxpixels],
   DisplayFunction-> Identity],    {nf,1,Nframes}]
ShowAnimation[film]
```

But our motion detection system is much more powerful than the example shown in the lower left corners. The following, different movie makes the point.

```
<< Graphics'Animation'
Nxpixels=72; Nypixels=54; Nframes=Nxpixels 4/3;
Nxwind=Round[Nxpixels/4]; Nywind=Round[Nypixels/3];
front=Table[Round[Random[]],{y,1,Nypixels},{x,1,Nxpixels}];
back =Table[Round[Random[]],{y,1,Nypixels},{x,1,Nxpixels}];
frame=Table[front,{nf,1,Nframes}];
Do[ If[ x>n-Nxwind && x<n && y>Nywind && y<2Nywind,
   frame[[n,y,x]]=back[[y,x]] ],
  ]      {x,1,Nxpixels}, {y,1,Nypixels}, {n,1,Nframes}]
film=Table[ListDensityPlot[frame[[nf]], Mesh-> False,
   Frame-> False, AspectRatio-> N[Nypixels/Nxpixels],
   DisplayFunction-> Identity],    {nf,1,Nframes}]
ShowAnimation[film]
```

Similar experiments, e.g. using randomly changing random patterns, show that the eye perceives motion even in cases where all Fourier components of the image are practically zero; such image motion is called *drift-balanced* or *non-Fourier* motion. Several examples are presented in J. ZANKER, *Modelling human motion perception I: classical stimuli*, Naturwissenschaften **81**, pp. 156–163, 1994, and J. ZANKER, *Modelling human motion perception II: beyond Fourier motion stimuli*, Naturwissenschaften **81**, pp. 200–209, 1994.

13 An introduction into perception research is E. Bruce GOLDSTEIN, *Perception*, Books/Cole, 5th edition, 1998. Cited on page 33.

14 All fragments from Heraclitos are from John Mansley ROBINSON, *An introduction to early greek philosophy*, Houghton Muffin 1968, chapter 5. Cited on page 33, 112.

15 An introduction to the story of classical mechanics which also destroys a few of the myths surrounding it, such as the idea that he could solve differential equations or that he introduced the expression $F = ma$, is given by Clifford A. TRUESDELL, *Essays in the history of mechanics*, Springer, 1968. Cited on page 35, 80, 90.

16 An introduction to Newton the alchemist are the two books by Betty Jo TEETER DOBBS, *The foundations of Newton's alchemy*, 1983, and *The Janus face of genius*, Cambridge University Press, 1992. Newton is found to be a sort of highly intellectual magician, desperately looking for examples of processes where god interacts with the material world. An intense but tragic tale. A good overview is provided by R.G. KEESING, *Essay Review: Newton's Alchemy*, Contemporary Physics **36**, pp. 117–119, 1995.

Newton's infantile theology, typical for god seekers who grew up without a father, can be found in the many books summarizing the exchanges between Clarke, his secretary, and Leibniz, Newton's rival to fame. Cited on page 35.

17 Almost all textbooks, both for schools and for university start with the definition of space and time. Even otherwise excellent relativity textbooks cannot avoid this habit, even those which introduce the now standard k-calculus (which is in fact the approach mentioned here). Cited on page 36.

18 [4] C. Liu, Z. Dutton, C.H. Behroozi, L.V. Han/ *Observation of coherent optical storage in an atomic medium using halted light pulses*, Nature **409**, pp. 490–493, 2001. There is also a comment of the paper by E.A: CORNELL, *Stopping light in its track*, **409**, pp. 461–462, 2001. However, despite the claim, the light pulses of course have *not* been halted. Can you give at least two reasons without even reading the paper, and maybe a third after reading it?

Challenge

The work was an improvement of the previous experiment where a group velocity of light of 17 m/s had been achieved, in an ultracold gas of sodium atoms, at nanokelvin temperatures. This was reported by Lene VESTERGAARD HAU, S.E. HARRIS, Zachary DUTTON & Cyrus H. BERTOZZI, *Light speed reduction to 17 meters per second in an ultracold atomic gas*, Nature **397**, pp. 594–598, 1999. Cited on page 37, 161.

19 Two jets with that speed have been observed by I.F. MIRABEL & L.F. RODRÍGUEZ, Nature **371**, pp. 46–48, and p. 18, 1994. Cited on page 37.

20 The results on the development of children mentioned here and in the following have been drawn mainly from the studies initiated by Jean Piaget; for more details on child development, see the intermezzo following this chapter, on page 372. At vanbc.wimsey.com/ chris/JPS/JPS.html one can find the web site maintained by the Jean Piaget Society. Cited on page 38, 39.

21 The clocks in our brain are described in ...The clocks in our body are described in ... Cited on page 39.

22 This has been shown among others by the work of Anna Wierzbicka mentioned in more detail in the intermezzo following this chapter, discussed on page 380. Also the passionate bestseller by the chomskian author Steven PINKER, *The language instinct – How the mind creates language*, Harper Perennial, 1994, discusses issues related to this matter, refuting amongst others on page 63 the often repeated false statement that the *Hopi* language is an exception. Cited on page 40.

23 Galileo used a water tube pointing in a bucket which he kept closed with his thumb. To start the stopwatch, he removed the thumb, to stop it, he put it back on. The volume of water in the bucket then gave him a measure of the time interval. Cited on page 40.

24 Aristoteles rejects the idea of the flow of time in ... Cited on page 41.

25 Perhaps the most informative of the books about the "arrow of time" is ..

A typical conference proceeding is J.J. HALLIWELL, J. PÉREZ-MERCADER & W.H.

ZUREK, *Physical origins of time asymmetry*, Cambridge University Press, 1994. Cited on page 41.

26 R. DOUGHERTY & M. FOREMAN, *Banach-Tarski decompositions using sets with the property of Baire*, Journal of the American Mathematical Society **7**, pp. 75–124, 1994. See also Alan L.T. PATERSON, *Amenability*, American Mathematical Society, 1998, and Robert M. FRENCH, *The Banach-Tarski theorem*, The Mathematical Intelligencer **10**, pp. 21–28, 1998. Finally, there are the books by B.R. GERLBAUM & J.M.H. OLMSTED, *Counterexamples in ananlysis*, Holden-Day, 1964, and *Theorems and counterexamples in mathematics*, Springer Verlag. Cited on page 46.

27 About the shapes of salt water bacteria, see the corresponding section in the interesting book by Bernard DIXON, *Power unseen – how microbes rule the world*, W.H. Freeman, New York, 1994. The book has about 80 sections, in which as many microorganisms are vividly presented. Cited on page 47.

28 There is a whole story behind the variations of *g*. It can be read in Chuji TSUBOI, *Gravity*, Allen & Unwin, 1979, or in Wolfgang TORGE, *Gravimetry*, de Gruyter, 1989, or in Milan BURˇSA & Karel PˇEˇC, *The gravity field and the dynamics of the earth*, Springer, 1993. The variation of the height of the soil by around 0.3 m due to the moon is one of the more interesting effects described by these investigations. Cited on page 48.

29 This was discussed in the Frankfurter Allgemeine Zeitung, 2.8.1997, at the time of the world athletics championship. The values are for the fastest part of a 100 m sprinter; the exact values cited were called the running speed world records in 1997, and were given as 12.048 m/s= 43.372 km/h by Ben Johnson for men, and 10.99 m/s= 39.56 km/h for women.

The 1998 world record for ball juggling is nine balls. Cited on page 48.

30 The arguments of Zeno can be found in ARISTOTELES, *The ...*, which can be found translated in almost any language. Cited on page 50, 188.

31 Professor to student: What is the derivative of velocity? Acceleration! What is the derivative of acceleration? I don't know. *Jerk*! The fourth, fifth and sixth derivatives of position are sometimes called *snap*, *crackle* and *pop*. Cited on page 50.

32 Etymology can be a fascinating topic, e.g. when it discovers the origin of the genus of the german word 'Weib' ('woman', related to english 'wife'). It was discovered, via a few Thocharian texts – an extinct indoeuropean language from a region inside modern China – to mean originally 'shame'. It was used for the female genital region in an expression meaning 'place of shame'. With time, this expression became to mean 'woman' in general, while being shortened to the second term only. This story was discovered by the german linguist Klaus T. Schmidt; it explains in particular why the word is not feminine but neutral, i.e. why it uses the article 'das' instead of 'die'. Julia Simon, private communication.

Etymology can also be simple and plain fun, for example when one discovers that 'testimony' and 'testicle' have the same origin; indeed in latin the same word 'testis' was used for both concepts. Cited on page 51, 56.

33 An overview of the latest developments is given by J.T. ARMSTRONG, D.J. HUNTER, K.J. JOHNSTON & D. MOZURKEWICH, *Stellar optical interferometry in the 1990s*, Physics Today pp. 42–49, May 1995. More than 100 stellar diameters have been measured in 1995, and several new powerful instruments are being planned. Cited on page 52.

34 A good biology textbook on growth is ... Cited on page 53.

35 This is discussed for example in C.L. STONG, *The amateur scientist – how to supply electric power to something which is turning*, Scientific American pp. 120–125, 19▷▷. It also discusses how to make a still picture of something rotating simply using a few prisms, the so called *Dove prisms*. Other examples of attaching something to a rotating body are given by E. RIEFLIN, *Some mechanisms related to Dirac's strings*, American Journal of Physics **47**, pp. 379–381, 1979. By the way, the real reason for the impossibility of wheels turns out to be that it is impossible to make an *axis* using a single piece of skin. Cited on page 53.

36 James A. YOUNG, *Tumbleweed*, Scientific American **264**, p. 82›87, March 1991. Cited on page 53.

37 The first experiments to prove the rotation of the flagella were by M. SILVERMAN & M.I. SIMON, *Flagellar rotation and the mechanism of bacterial motility*, Nature **249**, pp. 73–74, 1974. For some pretty pictures of the molecules involved, see K. NAMBA, *A biological molecular machine: bacterial flagellar motor and filament*, Wear **168**, pp. 189–193, 1993. The present record speed of rotation, 1700 rotations per second, is reported by Y. MAGARIYAMA, S. SUGIYAMA, K. MURAMOTO, Y. MAEKAWA, I. KAWAGISHI, Y. IMAE & S. KUDO, *Very fast flagellar rotation*, Nature **371**, p. 752, 1994.
More on bacteria can be learned from David DUSENBERY, *Life at a small scale*, Scientific American Library, 1996. Cited on page 54.

38 There is also the beautiful book by Penelope FARRANT, *Colour in Nature*, Blandford, 1997. Cited on page 54.

39 The laws of cartoon physics can be found in various places on the internet. Cited on page 54.

40 For the curious, an overview of the illusions used in the cinema and in television, which lead to some of the strange behaviour of images mentioned above, is given in Bernard WILKIE, *The technique of special effects in television*, Focal Press, 1993, and his other books, or in the *Cinefex* magazine. Cited on page 55.

41 Giuseppe FUMAGALLI, *Chi l'ha detto?*, Hoepli, 1983. The sentence is also the motto of the cities of the Hansa. Cited on page 56, 94, 99.

42 For the role and chemistry of adenosinetriphosphate (ATP) in cells and in living beings, see ... Cited on page 60.

43 A picture of this unique clock can be found in the article by A. GARRETT, *Perpetual motion – a delicious delirium*, Physics World pp. 23–26, December 1990. Cited on page 60.

44 Natthi L. SHARMA, *A new observation about rolling motion*, European Journal of Physics **17**, pp. 353–356, 1996. Cited on page 65.

45 Serge GRACOVETSKY, *The spinal engine*, Springer Verlag, 1990. Cited on page 65.

46 Thomas HEATH, *Aristarchus of Samos – the ancient Copernicus*, Dover, 1981, reprinted from the original 1913 edition. Aristarchos' treaty is given in greek and english. Aristarchos was the first proposer of the heliocentric system. Aristarchos had measured the length of the day (in fact, by determining the number of days per year) to the astonishing precision of less than one second. This excellent book also gives an overview of greek astronomy before Aristarchos, explained in detail for each greek thinker. Aristarchos' text is also reprinted in ARISTARCHOS, *On the sizes and the distances of the sun and the moon*, ca. 280 BCE in Michael J. CROWE, *Theories of the world from antiquity to the Copernican revolution*, Dover, 1990, especially on pp. 27–29. Cited on page 66, 148.

47 An overview of the effects of the Coriolis acceleration $\mathbf{a} = -2\omega \times \mathbf{v}$ in the rotating frame is given by Edward A. DESLOGE, *Classical mechanics*, Volume 1, John Wiley & Sons,1982. Even the *gulf stream* is influenced by it. Cited on page 66.

48 The original publication is by A.H. SHAPIRO, *Bath-tub vortex*, Nature **196**, pp. 1080–1081, 1962. He also produced two movies of the experiment. The experiment has been repeated many times in the northern and in the southern hemisphere, where the water drains clockwise; a complete literature list is found in the letters to the editor of the American Journal of Physics **62**, p. 1063, 1994. Cited on page 66.

49 The calculation of the period of Foucault's pendulum ... Cited on page 67.

50 The tricks are explained by H. Richard CRANE, *Short Foucault pendulum: a way to eliminate the precession due to ellipticity*, American Journal of Physics **49**, pp. 1004–1006, 1981, and particularly in H. Richard CRANE, *Foucault pendulum wall clock*, American Journal of Physics **63**, pp. 33–39, 1993. Cited on page 67.

51 R. ANDERSON, H.R. BILGER & G.E. STEDMAN, *The Sagnac-effect: a century of earth-rotated interferometers*, American Journal of Physics **62**, pp. 975–985, 1994.

See also the clear and extensive paper by G.E. STEDMAN, *Ring laser tests of fundamental physics and geophysics*, Reports on progress of physics **60**, pp. 615–688, 1997. Cited on page 68.

52 About the length of the day, see the maia.usno.navy.mil web site, or the books K. LAMBECK, *The earth's variable rotation: geophysical causes and consequences*, Cambridge University Press, 1980, and W.H. MUNK & G.J.F. MACDONALD, *The rotation of the earth*, Cambridge University Press, 1960. Cited on page 68.

53 One example of data is the one by C.P. SONETT, E.P. KVALE, A. ZAKHARIAN, M.A. CHAN & T.M. DEMKO, *Late proterozoic and paleozoic tides, retreat of the moon, and rotation of the earth*, Science **273**, pp. 100–104, 5 July 1996. They deduce from tidal sediment analysis that days were only 18 to 19 hours long in the Proterozoic, i.e. 900 million years ago; they assume that the year was 31 million seconds long all along. Cited on page 68.

54 The motion of the earth's axis in the last ten years is shown on the mentioned maia.usno.navy.mil web site. The latest idea is that this motion, which in the US is called Chandlers pole height motion after the first US scientist to have heard about the original discovery by Küstner, is due to the seasonal pressure changes in the atmosphere of the earth, as explained by R.S. GROSS, *The excitation of the Chandler wobble*, Geophysical Physics Letters **27**, p. 2329, 1st August 2000. Cited on page 68.

55 For more information about Alfred Wegener, read ...; about plate tectonics, see the www.scotese.com web site. On earthquakes, see the www.geo.ed.ac.uk/quakexe/quakes web site. On volcanoes, see the vulcan.wr.usgs.gov and www.dartmouth.edu/ volcano/ web sites. Cited on page 69.

56 The rotation and history of the solar system ... Cited on page 69.

57 J.D. HAYS, J. IMBRIE & N.J. SHACKLETON, *Variations in the earth's orbit: pacemaker of the ice ages*, Science **194**, pp. 1121–1132, 1976. They did this by literally digging in the mud which covers the ocean floor in certain places. Cited on page 69.

58 Roberta HUMPHREYS & Jeffrey LARSEN, *The sun's distance above the galactic plane*, Astronomical Journal **110**, pp. 2183–2188, November 1995. Cited on page 70.

59 C.L. BENNET, M.S. TURNER & M. WHITE, *The cosmic rosetta stone*, Physics Today **50**, pp. 32–38, November 1997. Cited on page 70.

60 The equilibrium of ships, so important in car ferries, is an interesting part of ship building; an introduction is given by ... Cited on page 71.

61 K.R. WENINGER, B.P. BARBER & S.J. PUTTERMAN, *Pulsed Mie scattering measurements of the collapse of a sonoluminescing bubble*, Physical Review Letters **78**, pp. 1799–1802, 1997. Cited on page 72.

62 On http://www.sff.net/people/geoffrey.landis/vacuum.html one can read a description of what happened. See also the http://www.sff.net/people/geoffrey.landis/ebullism.html and http://imagine.gsfc.nasa.gov/docs/ask_astro/answers/970603.html web sites. Cited on page 72.

63 The smallest distances are probed in particle accelerators; the distance can be determined from the energy of the particle beam. Cited on page 72. Challenge

64 Long jump data and literature can be found in three articles all entitled *Is a good long jumper a good high jumper?*, in the American Journal of Physics **69**, pp. 104–105, 2001. Cited on page 73.

65 R. McN. ALEXANDER, *Leg design and jumping technique for humans, other vertebrates and insects*, Philosophical Transactions of the Royal Society in London B **347**, pp. 235–249, 1995. Cited on page 73.

66 Jim W. GLASHEEN & Thomas A. MCMAHON, *A hydrodynamic model of locomotion in the basilisk lizard*, Nature **380**, pp. 340–342, For pictures, see also New Scientist p. 18, 30 March 1996, or Scientific American pp. 48–49, September 1997, or http://rjf2.biol.berkeley.edu/Full_Lab/FL_ Personnel/J_Glasheen/J_Glasheen.html

Several shore birds also have the ability to run over water, using the same mechanism. Cited on page 74.

67 The story is told in M. NAUENBERG, *Hooke, orbital motion, and Newton's Principia*, American Journal of Physics **62**, 1994, pp. 331–350. Cited on page 75.

68 More details are given by D. RAWLINS, in *Doubling your sunsets or how anyone can measure the earth's size with wristwatch and meter stick*, American Journal of Physics **47**, 1979, pp. 126–128. Another simple measurement of the earth radius, using only a sextant, is given by R. O'KEEFE & B. GHAVIMI-ALAGHA, in *The world trade center and the distance to the world's center*, American Journal of Physics **60**, pp. 183–185, 1992. Cited on page 75.

69 More details on astronomical distance measurements can be found in the beautiful little book by A. VAN HELDEN, *Measuring the universe*, University of Chicago Press, 1985, and in Nigel HENBEST & Heather COOPER, *The guide to the galaxy*, Cambridge University Press, 1994. Cited on page 75.

70 A lot of details can be found in M. JAMMER, *Concepts of mass in classical and modern physics*, reprinted by Dover, 1997, and in *Concepts of force, a study in the foundations of mechanics*, Harvard University Press, 1957. These eclectic and confusing texts are full of details and of various philosophical viewpoints, but lack clear statements and conclusions on their relation to accurate descriptions of nature, and thus cloud many fundamental issues.

Jean Buridan (ca. 1295– ca. 1366) criticizes the distinction of sublunar and translunar motion in his book *De Caelo*, one of his numerous works. Cited on page 76.

71 D. TOPPER & D.E. VINCENT, *An analysis of Newton's projectile diagram*, European Journal of Physics **20**, pp. 59–66, 1999. Cited on page 76.

72 This is explained for example by D.K. FIRPI´ C & I.V. ANIÇIN, *The planets, after all, may run only in perfect circles – but in the velocity space!*, European Journal of Physics **14**, pp. 255–258, 1993. Cited on page 77.

73 About the measurement of spatial dimensions via gravity, see ... Cited on page 78.

74 There are many books explaining this situation, such as the pocket book S. ANDERS, *Weil die Erde rotiert*, Verlag Harri Deutsch, Thun, Frankfurt am Main, 1985. Cited on page 79.

75 *a*. The shape of the earth is described most precisely with the World Geodetic System. For an extensive presentation of its background and its details, see the http://www.eurocontrol.be/projects/eatchip/wgs84/start.html web site.

IERS web site ...? Cited on page 79.

76 W.K. HARTMAN, R.J. PHILLIPS & G.J. TAYLOR, editors, *Origin of the Moon*, Lunar and Planetary Institute, Houston, 1986. Cited on page 80.

77 Dietrich NEUMANN, *Physiologische Uhren von Insekten – Zur Ökophysiologie lunarperiodisch kontrollierter Fortpflanzungszeiten*, Naturwissenschaften **82**, pp. 310–320, 1995. Cited on page 81.

78 J. Laskkar, F. Joutel & P. Robutel, *Stability of the earth's obliquity by the moon*, Nature **361**, pp. 615–617, 1993. However, the question is not completely settled, and other opinions exist. Cited on page 81.

79 See for example the discussion by J.J. Lissauer, *It is not easy to make the moon*, Nature **389**, pp. 327–328, 1997. Cited on page 81.

80 Neil F. Comins, *What if the moon did not exist? – Voyages to earths that might have been*, Harper Collins, 1993. Cited on page 81.

81 Paul A. Wiegert, Kimmo A. Innanen & Seppo Mikkola, *An asteroidal companion to the earth*, Nature **387**, pp. 685–686, 12 June 1997, together with the comment on pp. 651–652, of the same issue. Cited on page 82.

82 Simon Newcomb, Astronomical Papers of the American Ephemeris **1**, p. 472, 1882. Cited on page 82.

83 J. Soldner, *Berliner Astronomisches Jahrbuch auf das Jahr 1804*, 1801, p. 161. Cited on page 84.

84 The equality was first tested with precision by Roland von Eötvös, Annalen der Physik & Chemie **59**, p. 354, 1896, and by R. von Eötvös, V. Pekár & E. Fekete, *Beiträge zum Gesetz der Proportionalität von Trägheit und Gravität*, Annalen der Physik 4, Leipzig **68**, pp. 11–66, 1922. He found agreement to 5 parts in 10^9. More experiments were performed by P.G. Roll, R. Krotkow & R.H. Dicke, Annals of Physics (NY) **26**, pp. 442–517, 1964, and by V.B. Braginsky & V.I. Panov, Soviet Physics - JETP **34**, pp. 463–466, 1971. Modern results, with less than one part in 10^{12}, are by Y. Su & al., *New tests of the universality of free fall*, Physical Review **D50**, pp. 3614–3636, 1994. Several experiments have been proposed to test the equality in space to less than one part in 10^{16}. Cited on page 85.

85 The issue is explained in A.E. Chubykalo & S.J. Vlaev, *Theorem on the proportionality of inertial and gravitational masses in classical mechanics*, European Journal of Physics **19**, pp. 1–6, 1998. Cited on page 85.

86 See L. Hodges, *Gravitational field strength inside the earth*, American Journal of Physics, **59**, pages 954-956, 1991. Cited on page 86.

87 From Neil de Gasse Tyson, *The universe down to earth*, Columbia University Press, 1994. Cited on page 86.

88 Gerald Quinlan, Nature **363**, pp. 18–19, 1993. See also reference 111. Cited on page 87.

89 See Robert Matthews, *Not a snowball's chance ...*, these three dots are part of the title and correct! New Scientist 12 July 1997, pp. 24–27. The original claim is by Louis A. Frank, J.B. Sigwarth & J.D. Craven, *On the influx of small comets into the earth's upper atmosphere*, parts I and II Geophysical Research Letters **13**, pp. 303–306, pp. 307–310, 1986. The latest observations seem to disprove the claim. Cited on page 87.

90 The ray form is beautifully explained by J. Evans, *The ray form of Newton's law of motion*, American Journal of Physics **61**, pp. 347–350, 1993. Cited on page 87.

91 G.J. Sussman & J. Wisdom, *Chaotic Evolution of the Solar System*, Science **257**, pp. 56–62, 1992. Cited on page 87.

92 George-Louis Lesage, *Lucrece Newtonien*, Nouveaux mémoires de l'academie royale des sciences et belles letters pp. 404–431, 1747. Cited on page 88.

93 Knowledge is power. Time is money. Now, power is work/time. Inserting the previous equations and transforming, one gets:

$$\text{money} = \frac{\text{work}}{\text{knowledge}} \quad ; \qquad (170)$$

in other words, the less you know, the more, money you make. That is why scientists don't make a lot of money. Cited on page 90.

94 D. HESTENES, M. WELLS & G. SWACKHAMER, *Force Concept inventory*, Physics Teacher **30**, pp. 141–158, 1982. They have developed tests checking the understanding of the concept of physical force in students which have attracted a lot of attention in the didactics of physics circles. Cited on page 90.

95 See M. HIRANO & K. SHINJO, ..., Physical Review Letters **78**, pp. 1448–1451, 1997. See also the discussion of their results by Serge FAYEULLE, *Superlubricity: when friction stops*, Physics World pp. 29–30, May 1997. Cited on page 92.

96 C.D. AHRENS, *Meteorology today: an introduction to the weather, climate, and the environment*, West Publishing Company, St. Paul, 1991. Cited on page 93.

97 This topic is discussed with lucidity by J.R. MUREIKA, *What really are the best 100 m performances?*, Athletics: Canada's National Track and Field Running Magazine, July 1997. It can also be found under http://xxx.lanl.gov/physics/9705004 together with other papers on similar topics by the same author. Cited on page 93.

98 F.P. BOWDEN & D. TABOR, *The friction and lubrication of solids*, Oxford University Press, Part I, revised edition, 1954, and part II, 1964. Cited on page 93.

99 The test of randomness can be found in the text ...The topic has many philosophical implications, discussed e.g. by ... Cited on page 96.

100 Henrik WALTER, *Neurophilosophie der Willensfreiheit*, ... Cited on page 97.

101 The beautiful story of the south pointing carriage is told in the Appendix B of James FOSTER & J.D. NIGHTINGALE, *A short course in general relativity*, Springer Verlag, 2nd edition 1998. Such carriages have existed in China, as told (as usual) by Joseph Needham, but their construction is unknown. The carriage described by Foster and Nightingale is the one (re-)invented in 1947 by George Lancaster, a british engineer. Cited on page 101.

102 See for example Z. GHAHRAMANI, *Building blocks of movement*, Nature **407**, pp. 682–683, 2000. Researchers in robot control are also interested in such topics. Cited on page 101.

103 G. GUTIERREZ, C. FEHR, A. CALZADILLA & D. FIGUEROA, *Fluid flow up the wall of a spinning egg*, American Journal of Physics **66**, pp. 442–445, 1998. Cited on page 101.

104 A historical account is given in Wolfgang YOURGRAY & Stanley MANDELSTAM, *Variational principles in dynamics and quantum theory*, Dover, 1968. Cited on page 102, 103.

105 Max PÄSLER, *Prinzipe der Mechanik*, Walter de Gruyter & Co., 1968. Cited on page 104.

106 The relations between the possible lagrangians is studied by ... Cited on page 105.

107 G.B. WEST, J.H. BROWN & B.J. ENQUIST, *A general model for the origin of allometric scaling laws in biology*, Science **276**, pp. 122–126, 4 April 1997, with a comment on page 34 of the same issue. The rules governing branching properties of blood vessels, of lymph systems, and of vessel systems in plants are explained. About plants, see also the paper by the same group in the 1999, August 12th issue of Nature. Cited on page 105.

108 J.R. BANAVAR & A. RINALDO, Nature 13 May 1999. Cited on page 105.

109 J.A. MOORE, *An innovation in physics instruction for nonscience majors*, American Journal of Physics **46**, pp. 607–612, 1978. Cited on page 106.

110 See e.g. Alan P. BOSS, *Extrasolar planets*, Physics Today September 1996, pp. 32–38. The most recent information can be found at the "Extrasolar Planet Encyclopaedia"maintained at http://www.obspm.fr/planets by Jean Schneider at the Observatoire de Paris. Cited on page 108.

111 A good review article is the one by David W. HUGHES, *Comets and Asteroids*, Contemporary Physics **35**, pp. 75–93, 1994. Cited on page 108, 196.

112 A review on the fascinating mechanisms at the basis of the heart beat is given by the text being written by A. BABLOYANTZ, ..., Cited on page 109.

113 See e.g. B. BOWER, *A child's theory of mind*, Science News **144**, pp. 40–41. Cited on page 113.
The maximal kinematical invariance group of the free Schrödinger equation, ," Helvetica Physica Acta, 45 (1972) 802810.

114 U. Niederer/ *The maximal kinematical invariance group of the free Schrödinger equation*, Helvetica Physica Acta **45**, pp. 802–810, 1972. See also the introduction by O. JAHN & V.V. SREEDHAR, *The maximal invariance group of Newton's equations for a free point particle*, math-ph/0102011 Cited on page 119.

115 The story is told in the interesting biography of Einstein by A. PAIS, *'Subtle is the Lord...' – The science and the life of Albert Einstein*, Oxford University Press, 1982. Cited on page 120.

116 M. AUSLOOS, Proceedings of the Royal Society in London A **400**, pp. 331–350, 1985. Cited on page 124.

117 T.A. MCMAHOB & J. TYLER BONNER, *Form und Leben – Konstruktion vom Reißbrett der Natur*, Spektrum Verlag, 1985. Using the fact that for jumps of an animal of mass m the necessary energy E is given as $E = mgh$, and that roughly speaking the work available to a muscle is proportional to its mass $W \sim m$, one gets that the height h is independent of the mass of the animal. In other words, the specific mechanical energy of animals is around $1.5 \pm 0.7\,\mathrm{J/kg}$. Cited on page 125.

118 A simple article explaining the tallness of trees is A. MINEYEV, *Trees worthy of Paul Bunyan*, Quantum, pages 4-10, January-February 1994. (Paul Bunyan is a mythical giant lumberjack who is the hero of the frontier pioneers in the beginning of United States history.) Note that the transport of liquids in trees poses no limits on their height, since water is pumped up along tree stems (except in spring, when it is pumped up from the roots) by evaporation from the leaves. This works almost without limits because water columns, when nucleation is carefully avoided, can be put under tensile stresses of over 100 bar, corresponding to 1000 m. Cited on page 126.

119 Such information can be taken from the excellent overview article by M.F. ASHBY, *On the engineering properties of materials*, Acta Metallurgica **37**, pp. 1273–1293, 1989. The article explains the various general criteria which determine the selection of materials, and gives numerous tables to guide the selection. Cited on page 126.

120 For a photograph of a single Barium *atom* – named Astrid – see Hans DEHMELT, *Experiments with an isolated subatomic particle at rest*, Reviews of Modern Physics **62**, pp. 525–530, 1990, for an earlier photograph of a Barium *ion*, see W. NEUHAUSER, M. HOHENSTATT & P.E. TOSCHEK, *Localized visible Ba^{+} mono-ion oscillator*, Physical Review A **22**, pp. 1137–1140, 1980. Cited on page 127.

121 Holograms of atoms have been produced first by HANS-WERNER FINK & al., ..., Physical Review Letters **67**, pp. 1543–1546, 1991. Cited on page 127.

122 See for example C. SCHILLER, A.A. KOOMANS, T.L. van ROOY, C. SCHÖNENBERGER & H.B. ELSWIJK, *Decapitation of tungsten field emitter tips during sputter sharpening*, Surface Science Letters **339**, pp. $L925$–$L930$, 1996. Cited on page 127.

123 Good pictures of atoms taken by atomic force microscopes can be found in Herman B. ELSWIJK, ..., Cited on page 127.

124 J. SPENCE & U. WEIERSTRASS, ..., end 1997. Cited on page 127.

125 Some religions incredibly, still pretend that people believe the opposite. Cited on page 128.

126 Thermostatics is difficult to learn also because it was not discovered in a systematic way, as told by C. TRUESDELL, *The tragicomical history of thermodynamics 1822-1854*, Springer

Verlag, 1980.

An excellent advanced textbook on thermostatics and thermodynamics is Linda REICHL, *A modern course in statistical physics*, Wiley, 2nd edition, 1998. Cited on page 129.

127 The famous paper on brownian motion which so much contributed to Einstein's fame is *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, Annalen der Physik **13**, pp. 549–560, 1905. Cited on page 130.

128 The first tests of the prediction were performed by Jean Perrin ... Cited on page 131.

129 Gas expansion was the main method used for the definition of the official temperature scale. Only in 1990 other methods were introduced officially, such as total radiation thermometry (140 K-373 K), noise thermometry (2 K-4 K and (900 K-1235 K)), acoustical thermometry (303 K), magnetic thermometry (0.5 K-2.6 K), and optical radiation thermometry ($>$ 730 K). But radiation thermometry is still the central method in the domain from about 3 K to about 1000 K. This is explained in detail in R.L. RUSBY, R.P. HUDSON, M. DURIEUX, J.F. SCHOOLEY, P.P.M. STEUR & C.A. SWENSON, *The basis of the ITS-90*, Metrologia **28**, p. 9, 1991. On the water boiling point see also page 737. Cited on page 132.

130 See e.g. the article by H. PRESTON-THOMAS, *The international temperature scale of 1990 (ITS-90)*, Metrologia **27**, pp. 3–10, 1990, and the errata H. PRESTON-THOMAS, *The international temperature scale of 1990 (ITS-90)*, Metrologia **27**, p. 107, 1990, Cited on page 132.

131 The weight is given by the pressure of the tires times the contact area. Cited on page 132.

132 An overview of low temperature physics is given by ... Cited on page 132.

133 K. SCHMIDT-NIELSEN, *Desert animals: physiological problems of heat and water*, Oxford University Press, 1964. Cited on page 132.

134 The story that entropy is created when information is erased, not when it is acquired, is told in Charles H. BENNETT & Rolf LANDAUER, *Fundamental Limits of Computation*, Scientific American **253:1**, pp. 48–56, 1985. We should pay to throw the newspaper away, not to buy it. Cited on page 133.

135 F. HERRMANN, *Mengenartige Größen im Physikunterricht*, Physikalische Blätter **54**, pp. 830–832, September 1998. See also his lecture notes on general introductory physics on the http://www-tfp.physik.uni-karlsruhe.de/ didaktik web site. Cited on page 134.

136 I. PRIGOGINE, ... Cited on page 135.

137 These points are made clearly and forcibly, as is his style, by N.G. van KAMPEN, *Entropie*, nederlands tijdschrift voor natuurkunde **62**, pp. 395–396, 3 December 1996. Cited on page 136.

138 This is a disappointing result of all efforts so far, as Grégoire Nicolis always stresses in his university courses. Seth Lloyd has compiled a list of 31 different propositions to define complexity, containing among others, fractal dimension, grammatical complexity, computational complexity, thermodynamic depth, etc. See e.g. a short summary in Scientific American p. 77, June 1995. Cited on page 137.

139 K. SCHWENK, *Why snakes have forked tongues*, Science **263**, pp. 1573–1577, 1994. Cited on page 140.

140 P.B. UMBANHOWAR & H.L. SWINNEY, *Localized excitations in a vertically vibrated granular container*, Nature **382**, pp. 793–796, 29 August 1996. Cited on page 140.

141 K. KÖTTER, E. GOLES & M. MARKUS, *Shell structures with "magic numbers" of spheres in a swirleed disk*, Physical Review E **60**, p. 7182, 1999, See also the http://www.aip.org/physnews/graphics/html/magicball.html web site. Cited on page 140.

142 A good introduction is the text by Daniel WALGRAEF, *Spatiotemporal pattern formation, with examples in physics, chemistry and material science*, Springer 1996. Cited on page 140.

143 For an overview, see the Ph.D. thesis by Joceline LEGA, *Défauts topologiques associés à la brisure de l'invariance de translation dans le temps*, Université de Nice, 1989. Cited on page 141.

144 For a clear introduction, see T. SCHMIDT & M. MARHL, *A simple mathematical model of a dripping tap*, European Journal of Physics **18**, pp. 377–383, 1997. Cited on page 141.

145 J.J. LISSAUER, *Chaotic motion in the solar system*, Reviews of Modern Physics **71**, pp. 835–845, 1999. Cited on page 142.

146 The present status of the understanding of turbulence is given in ... Cited on page 143.

147 For more details, see the appendix on units on page 723. Cited on page 145.

148 ... ..., *The ray days* .., .. Cited on page 145.

149 The history of the measurement of the velocity of light can be found in chapter 19 of the text by F.A. JENKINS & H.E. WHITE, *Fundamentals of optics*, McGraw-Hill, New York, 1957. Cited on page 147.

150 On the way to perform such a measurement, see Sydney G. BREWER, *Do-it-yourself astronomy*, Edinborough University Press, 1988. Kepler himself never measured the distances of planets to the sun, but only ratios planetary distances. The parallax of the sun from two points of the earth is at most 8.79 seconds of arc, as has been measured already in the 18th century.

The measurement of moon distance is now performed with laser beams fired onto a reflector deposited there during the Apollo missions. Distance measurements are done by time of flight measurement of 35 ps laser pulses. See http://almagest.as.utexas.edu/~ rlr/mlrs.html and http://cddisa.gsfc.nasa.gov/cstg/slr_cstg.html Cited on page 148.

151 The art to make pictures of light pulses with a usual photographic camera, without any electronics, is told by M.A. DUGAY & A.T. MATTICK, *Ultrahigh speed photography of picosecond light pulses and echoes*, Applied Optics **10**, 1971, pp. 2162–2170. The picture on page 148 is also taken from there. Cited on page 148.

152 The constancy of light speed deduced from gamma ray bursts shows that the speed of light does not depend on the lamp speed to within one part in 10^{20}, as shown by K. BRECHER, ...2000. See also his older paper K. BRECHER, *Is the speed of light independent of the source?*, Physics Letters **39**, pp. 1051–1054, Errata 1236, 1977. Measuring the light speed from rapidly moving stars is another way. Some of these experiments are not convincing however, because there is a competing theory of electrodynamics, by Ritz, which maintains that the speed of light is c only when measured with respect to the source. The light from stars however passes through the atmosphere, and its speed thus might be reduced to c.

The famous experiment with light emitted from rapid pions at CERN does not suffer from this criticism. It is described in T. ALVÄGER, J.M. BAILEY, F.J.M. FARLEY, J. KJELLMAN & I. WALLIN, *Test of the second postulate of relativity in the GeV region*, Physics Letters **12**, pp. 260–262, 1964. See also Ark. Fys. **31**, p. 145, 1965.

Another precise experiment at extreme speeds is described by G.R. KALBFLEISCH, N. BAGGETT, E.C. FOWLER & J. ALSPECTOR, ..., Physical Review Letters **43**, pp. 1361–1364, 1979. Cited on page 149.

153 The start of the modern theory of relativity is the famous paper by Albert EINSTEIN, *Zur Elektrodynamik bewegter Körper*, Annalen der Physik **17**, pp. 891–921, 1905. Cited on page 149, 150.

154 Albert EINSTEIN, *Mein Weltbild*, ... Cited on page 149.

155 See e.g. C. WILL, *Theory and Experiment in Gravitational Physics*, Revised edition, Cambridge University Press, 1993. Cited on page 149, 149.

156 B.E. SCHAEFER, *Severe limits on variations of the speed of light with frequency*, Physical Review Letters **82**, pp. 4964–4966, 21 June 1999. Cited on page 149.

157 R.J. KENNEDY & E.M. THORNDIKE, *Experimental establishment of the relativity of time*, Physical Review **42**, pp. 400–418, 1932. See also H.E. IVES & G.R. STILWELL, *An experimental study of the rate of a moving atomic clock*, Journal of the Optical Society of America **28**, pp. 215–226, 1938, and **31**, pp. 369–374, 1941, Cited on page 150.

158 Rod S. LAKES, *Experimental limits on the photon mass and cosmic vector potential*, Physical Review Letters **80**, pp. 1826–1829, 1998. The speed of light is independent of frequency within a factor of $6 \cdot 10^{-21}$, as was shown from gamma ray studies by Bradley E. SCHAEFER, *Severe limits on variations of the speed of light with frequency*, Physical Review Letters **82**, pp. 4964–4966, 1999.

An overview of previous results is given in Yuan-Chung CHANG & Yuan Zhong ZHANG, *Special relativity and its experimental foundations*, World Scientific, 1998. Cited on page 150.

159 R.W. MCGOWAN & D.M. GILTNER, *New measurement of the relativistic Doppler shift in neon*, Physical Review Letters **70**, pp. 251–254, 1993. Cited on page 151.

160 The beautiful graph is taken from Z.G.J. GUIRAGOSSIAN & al., .., Physical Review Letters **34**, pp. 335–338, 1975. Cited on page 152.

161 To know more about the best known crackpots, their ideas, e-mail to `majordomo@zikzak.net` with the one line body `subscribe psychoceramics`. Cited on page 152.

162 The speed of neutrinos is the same as that of light within 9 decimal digits, as shown from the fact that neutrinos and light arrived together after a trip of 170 000 light years from a supernova explosion. This is explained by Leo STODOLSKY, *The speed of light and the speed of neutrinos*, Physics Letters B **201**, p. 353, 1983. An observation of a small mass for the neutrino has been published by the japanese Super-Kamiokande collaboration, as .... However, the paper has been strongly criticized on its deductions, and the result has not yet been confirmed. Cited on page 153.

163 A general refutation of such proposals is discussed by S.R. MAINWARING & G.E. STEDMAN, *Accelerated clock principles*, Physical Review A **47**, pp. 3611–3619, 1993. Already experiments on *myons* at CERN in 1968 showed that *accelerations* of up to 10^{20} m/s^2 have no effect, as explained by D.H. PERKINS, *Introduction to high energy physics*, Addison-Wesley, 1972, or by J. BAILEY & al., Il Nuovo Cimento **9A**, p. 369, 1972. Cited on page 156.

164 B. ROTHENSTEIN & G. ECKSTEIN, *Lorentz transformations directly from the speed of light*, American Journal of Physics **63**, p. 1150, 1995. See also the comment by E. KAPUSCIK, American Journal of Physics **65**, p. 1210, 1997. Cited on page 153.

165 A.A. MICHELSON & E.W. MORLEY, *On the relative motion of the earth and the luminiferous ether*, American Journal of Science (3rd series) **34**, pp. 333–345, 1887. Michelson published many other papers on the topic after this one. Cited on page 154.

166 H.A. LORENTZ, *De relative beweging van de aarde en dem aether*, Amst. Versl. **1**, p. 74, 1892, and also H.A. LORENTZ, *Electromagnetic phenomena in a system moving with any velocity smaller than that of light*, Amst. Proc. **6**, p. 809, 1904, or Amst. Versl. **12**, p. 986, 1904, Cited on page 155.

167 W. RINDLER, *General relativity before special relativity: an unconventional overview of relativity theory*, American Journal of Physics **62**, pp. 887–893, 1994. Cited on page 156.

168 If the future could influence the past ... Cited on page 158.

169 Steven K. BLAU, *Would topology change allow Ms. Bright to travel backward in time?*, American Journal of Physics **66**, pp. 179–185, 1998. Cited on page 158.

170 On the 'proper' formulation of relativity, see for example ... Cited on page 158.

171 The simple experiment to take a precise clock on a plane, fly it around the world, and then compare it with an identical one left in place was first performed by J.C. HAFELE & Richard E. KEATING, *Around-the-world atomic clocks: predicted relativistic time gains*, Science **177**, pp. 166–167, and *Around-the-world atomic clocks: observed relativistic time gains*, pp. 168–170, 14 July 1972.

There is also the text by Yuan-Chung CHANG & Yuan Zhong ZHANG, *Special relativity and its experimental foundations*, World Scientific, 1998. Cited on page 159.

172 A readable introduction into the change of time with observers and relativity in general is R. SEXL & H.K. SCHMIDT, *Raum-Zeit-Relativität*, 2. Auflage, Vieweg & Sohn, Braunschweig, 1982. Cited on page 159.

173 Most famous is the result that moving muons stay younger, as shown for example by D.H. FRISCH & J.B. SMITH, *Measurement of the relativistic time dilation using μ-mesons*, American Journal of Physics **31**, pp. 342–335, 1963. Cited on page 159.

174 W. RINDLER, *Length contraction paradox*, American Journal of Physics **29**, 1961, pp. 365–366, For a variation without gravity, see R. SHAW, *Length contraction paradox*, American Journal of Physics **30**, p. 72, 1962. Cited on page 159.

175 This situation is discussed by G.P. SASTRY, *Is length contraction paradoxical?*, American Journal of Physics **55**, 1987, pp. 943–946, It also contains an extensive literature list of all variants of length contraction paradoxes. By the way, the bar does not light the lamp at high speeds, even though in the frame of the bar there is always contact. Can you see why? Cited on page 159. Challenge

176 S.P. BOUGHN, *The case of the identically accelerated twins*, American Journal of Physics **57**, pp. 791–793, 1989. Cited on page 160, 160.

177 G.R. RYBICKI, *Speed limit on walking*, American Journal of Physics **59**, pp. 368–369, 1991. Cited on page 161.

178 The slowness of the speed of light inside stars is explained e.g. by ... Cited on page 161.

179 The first examples of such astronomical observations were provided by A.R. WHITNEY & al., Science **173**, pp. 225–230, 1971, and by M.H. COHEN & al., Astrophysical Journal **170**, pp. 207–217, 1971. See also T.J. PEARSON, S.C. UNWIN, M.H. COHEN, R.P. LINFIELD, A.C.S. READHEAD, G.A. SEIELSTAD, R.S. SIMON & R.C. WALKER, Nature **290**, pp. 365–368, 1981. An overview is given in J.A. ZENSUS & T.J. PEARSON, editors, *Superluminal radio sources*, Cambridge University Press, 1987. A more recent example was reported in Science News **152**, p. 357, 6 December 1997.

A pedagogical explanation is given by D.C. GABUZDA, *The use of quasars in teaching introductory special relativity*, American Journal of Physics **55**, pp. 214–215, 1987, and by Edwin F. TAYLOR & John A. WHEELER, *Spacetime Physics - Introduction to special relativity*, second edition, Freeman, 1992, pages 89-92. This excellent book was mentioned already in the text. Cited on page 162.

180 O.M. BILANIUK & E.C. SUDARSHAN, *Particles beyond the light barrier*, Physics Today **22**, pp. 43–51, 1969, O.M.P. BILANUK, V.K. DESHPANDE & E.C.G. SUDARSHAN, *"Meta" relativity*, American Journal of Physics **30**, pp. 718–723, 1962. See also E. RECAMI, editor, *Tachyons, monopoles, and related topics*, North-Holland, Amsterdam, 1978. Cited on page 163.

181 See S.S. COSTA, G.E.A. MATSAS, *Temperature and relativity*, http://arxiv.org/gr-qc/9505045. Cited on page 164.

182 This information is due to a private communication by Frank DiFilippo; part of the story is given in F. DIFILIPPO, V. NATARAJAN, K.R. BOYCE & D.E. PRITCHARD, *Accurate* *atomic masses for fundamental metrology*, Physical Review Letters **73**, pp. 1481–1484, 1994. These measurements were performed with Penning traps; a review of the possibilities they offer is given by R.C. THOMPSON, Measurement Science and Technology **1**, p. 93, 1990. The most important experimenters in the world of single particle levitation were awarded the Nobel prize in 1989. Their Nobel prize lectures can be found in W. PAUL, *Electromagnetic traps for neutral and charged particles*, Reviews of Modern Physics **62**, pp. 531–540, 1990. Cited on page 168.

183 J.L. SYNGE, *Relativity: the special theory*, North-Holland, 1956, pp. 208–213. More about antiparticles in special relativity can be found in J.P. COSTELLA, B.H.J. MCKELLAR & A.A. RAWLINSON, *Classical antiparticles*, American Journal of Physics **65**, pp. 835–841, 1997. See also reference 193. Cited on page 169.

184 A. PAPAPETROU, *Drehimpuls- und Schwerpunktsatz in der relativistischen Mechanik*, Praktika Acad. Athenes **14**, p. 540, 1939, A. PAPAPETROU, *Drehimpuls and Schwerpunktsatz in der Diracschen Theorie*, Praktika Acad. Athenes **15**, p. 404, 1940, M.H.L. PRYCE, *The mass-centre in the restricted theory of relativity and its connexion with the quantum theory of elementary particle*, Physical Review A **195**, pp. 62–81, 1948. Cited on page 170.

185 A readable article showing a photocopy of a letter by Einstein making this point is Lev B. OKUN, *The concept of mass*, Physics Today pp. 31–36, June 1989. The topic is not without controversy, as the letters following that article show; they are found in *Putting to rest mass misconceptions*, Physics Today pp. 13–14, as well as pp. 115–117, May 1990. The topic is still source of debates. Cited on page 174.

186 A jewel among the textbooks on special relativity is the booklet by Ulrich E. SCHRÖDER, *Spezielle Relativitätstheorie*, Verlag Harri Deutsch, Thun, 1981. Cited on page 173, 174.

187 C. MØLLER, *The theory of relativity*, Clarendon Press, Oxford, 1952, 1972. This standard text has also been translated in several other languages. Cited on page 174.

188 R.J. LOW, *When moving clocks run fast*, European Journal of Physics **16**, pp. 228–229, 1995. Cited on page 178, 179.

189 Edward A. DESLOGE & R.J. PHILPOTT, *Uniformly accelerated reference frames in special relativity*, American Journal of Physics **55**, pp. 252–261, 1987. Cited on page 180, 180.

190 Carl G. ADLER & Robert W. BREHME, *Relativistic solutions to a falling body in a uniform gravitation field*, American Journal of Physics **59**, pp. 209–213, 1991. Cited on page 182.

191 R.H. GOOD, *Uniformly accelerated reference frame and twin paradox*, American Journal of Physics **50**, pp. 232–238, 1982. Cited on page 181, 181, 184.

192 J. DWAYNE HAMILTON, *The uniformly accelerated reference frame*, American Journal of Physics **46**, pp. 83–89, 1978. Cited on page 181.

193 See e.g. D.J. RAYMOND, *Radical freshman physics*, on the http://www.physics.nmt.edu/raymond/ph131/book1/index.html web site. Cited on page 183, 203.

194 Edward A. DESLOGE, *The gravitational red shift in a uniform field*, American Journal of Physics **58**, pp. 856–858, 1990. Cited on page 184.

195 The impossibility to define rigid coordinate frames for non-uniformly accelerating observers is discussed by Charles MISNER, Kip THORNE & John A. WHEELER, *Gravitation*, Freeman, 1973, page 168. Cited on page 180.

196 The propagation delays to be discussed in the chapter on general relativity can be seen as confirmations of this effect. Cited on page 185.

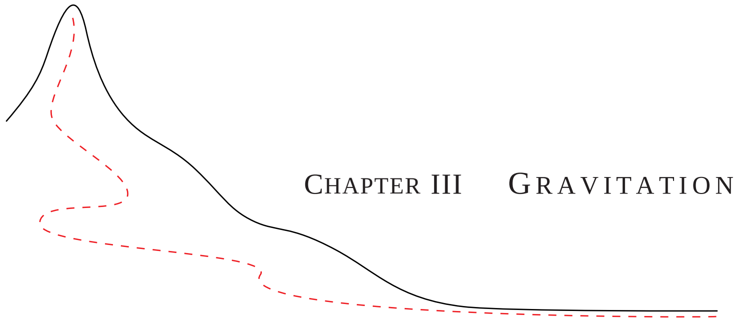
197 W. RINDLER & L. MISHRA, *The nonreciprocity of relative acceleration in relativity*, Physics Letters A **173**, 1993, pp. 105–108. Cited on page .

198 The subtleties of the one-way and two-way speed of light are explained in ... Cited on page 185.

199 Edwin F. Taylor & A.P. French, *Limitation on proper length in special relativity*, American Journal of Physics **51**, pp. 889–893, 1983. Cited on page 186, 186.

200 Richard P. Feynman, R.B. Leighton & M. Sands, *The Feynman lectures on physics*, Addison Wesley, 1977. A famous book. Cited on page .

Typeset in April 2001



Chapter III Gravitation and Relativity

Gravitational influences do transport energy.* In the description of motion, the next goal therefore must be to increase the precision in such a way that this transport happens at most with the speed of light, as Henri Poincaré stated already in 1905. The results will be fascinating; it will be found that empty space can move, that the universe has a finite age, that objects can be in permanent free fall, and that space can be bent – despite being much stiffer than steel.

Describing motion due to gravity using $a = GM/r^2$ not only allows speeds larger than light; it is also unclear how the values of a and r depend on the observer. It cannot be correct. In order to achieve a consistent description, called *general relativity* by Albert Einstein, we have to throw quite a few preconceptions overboard. Ref. 1

7. The new ideas on space, time, and gravity

Sapere aude.
Horatius**

What is the opposite of motion in daily life? A body at rest, such as a child sleeping. Or a man listening. Or a rock defying the waves. And when is a body at rest? When it is not disturbed by other bodies. In the galilean description of the world, rest thus is the *absence of velocity*. With special relativity, rest became *inertial motion*, since no inertially moving observer can distinguish its own motion from rest: nothing disturbs him. This is the case for the rock in the waves and for the rapid protons crossing the galaxy as cosmic rays. But the study of gravity leads to an even more general definition.

Rest and free fall

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 stepped out of a ballon 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. He was so surprised that he had to turn upwards in order to convince himself that he was really getting away from his balloon! In fact he was falling at up to 988 km/h with respect to the

Challenge

Ref. 3

* About the details of this statement, see page 207 and page 224.

** 'Venture to be wise.' Horatius, Quintus Flaccus, Ep. 1, 2, 40.

earth's surface. He started feeling something only from the moment that he encountered the first layers of air, thus when his free fall started to be disturbed. Later, after four and a half minutes of fall, his special parachute opened, and he landed safely in New Mexico.

He and all other observers in free fall, such as the cosmonauts circling the earth, 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: *rest is free fall*. The set of all free falling observers that meet at a point in space-time generalize the set of the inertial observers that can meet at a point in special relativity.

As a further consequence, true *motion is the opposite of free fall*. This conclusion directly produces a number of questions: Most trees or mountains are not in free fall, thus they are
Challenge not at rest. What motion are they undergoing? And if free fall is rest, what is weight? And what then is gravity anyway?

Is gravity a force?

One tends to answer affirmatively, as in galilean physics gravity was seen as an influence on the motion of bodies. In fact, it was described by a potential, meaning that gravity *produces* motion. But let us be careful. A force or an interaction is what changes the motion of objects. However, we just saw that when two bodies attract each other through gravitation, both always remain in free fall. For example, the moon circles the earth because it continuously falls around it. Since any freely falling observer continuously remains at rest, the statement that gravity changes the motion of bodies is not correct for all observers. Indeed, we will soon discover that in a sense to be discussed shortly, the moon and the earth both follow "straight" paths.

Is this correction of our idea of gravity only a question of words? Not at all. Since gravity is not a force, it is *not* due to a field, and there is *no* potential.

See page 408 Let us check this strange result in yet another way. The most fundamental definition of 'interaction' is the difference between the whole and the sum of its parts. In the case of gravity, an observer in free fall could claim that nothing special is going on, independently of whether the other body is present or not, and could claim that gravity is is not an interaction.

However, that is going too far. An interaction transports energy between systems. We will indeed find out that gravity can be said to transport energy only approximately. Gravitation is thus an interaction only approximately. But that is a sufficient reason for keep this characterization. In consistence with the strange conclusion, it turns out that the concept of energy is not useful for gravity outside of everyday life. For the general case, namely for a general observer, gravity is thus fundamentally different from electricity or magnetism. To find out what it is, let us start with the measurement data.

The speed of light and the constant of gravitation

Si morior, moror.*

* 'If I rest, I die.' Motto of the bird of paradise.

In order to find a more precise definition of gravity, we start with two basic experimental observations:

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

$$v \leqslant c \tag{171}$$

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

This description following from this first principle, *special* relativity, is extended to *general* relativity by a second principle:

▷ *For any physical system of extension L and mass M, their ratio is bound by the limit*

$$\frac{L}{M} \geqslant \frac{4G}{c^2} \tag{172}$$

for all observers, where G is the gravitational constant.

We will discuss and motivate this second, new limit all over this chapter, especially in the section on black holes, and show that no exceptions are known. The principles will tell us *what* gravity is and *how* exactly it behaves. Together, they imply all of general relativity.* The principles are valid for *all* observers. It makes no difference whether an observer feels gravity, is in free fall, is accelerated, or is in inertial motion. No exceptions are known.

The escalation so far has taught us that a precise description of motion requires the listing of all possible viewpoints, their characteristics and their differences, as well as the transformations from one viewpoint to the other. As this time, *all* viewpoints are allowed, without exception, anybody must be able to talk to anybody else. People who exchange left and right, people who exchange up and down, people who say that the sun turns around the earth; all these observers can talk to each other. This gives a much larger set of viewpoint transformations than in the case of special relativity, and makes general relativity both difficult and fascinating. And since all viewpoints are possible, the resulting description of motion is *complete*.**

What is gravity?

As William Unruh likes to explain, the constancy of the speed of light implies the following conclusion: *gravity is the uneven running of clocks at different places*.*** Of course, this seemingly absurd definition needs to be checked. Challenge

Note that the definition does not talk about a single situation seen by *different* observers, as we often did in special relativity. The definition states that neighboring, 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 defined to be 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.

* This somewhat unconventional didactic approach has been developed by the author.
** Were it not for a small deviation called quantum theory.
*** Gravity is also the uneven length of meter bars at different places, as we will see below. Both effects are needed to describe it completely; but for daily life on earth, the clock effect is sufficient, since it is much larger than the length effect, which can be usually be neglected. Can you see why? Challenge

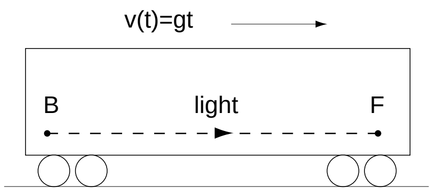


Figure 83 Colours inside an accelerating train or bus

An observer feels no difference between gravity and constant acceleration. Thus we can use a result See page 160 we encountered already in special relativity. If light is emitted at the back end of an accelerating train of length Δh, it arrives at the front end after a time $t = \Delta h/c$. However, during this time the accelerating train has picked up some additional velocity, namely $\Delta v = gt = g\Delta h/c$. As a result, due to the Doppler Challenge effect, the frequency of the light arriving at the front has changed. Inserting, one gets

$$\frac{\Delta f}{f} = \frac{g\Delta h}{c^2} = \frac{\Delta\tau}{\tau} \quad . \tag{173}$$

Note that the sign of the frequency change depends on whether the light and the train travel Challenge in the same direction or not. For actual trains or buses, the frequency change is quite small. But before we discuss the consequences of the result, let us check it with a different experiment.

To measure time and space, we use light. What happens to light when gravity is involved? Ref. 5 The simplest experiment is to let light fall or rise in a gravitational field. In order to deduce what must happen, we add a few details. Imagine a conveyor belt carrying masses around See figure 84 two wheels, a low and a high one. The descending, grey masses are slightly larger. Whenever such a larger mass is near the bottom, it converts its mass surplus to light via $E = mc^2$ and sends the light up towards the top.* At the top, one of the lighter, white masses passing by absorbs the light and, due to its added weight, turns the conveyor belt until it reaches the bottom. Then the process repeats.

As the left masses are always heavier, the belt would turn for ever, and this system could continuously *generate* energy. Since energy conservation is at the basis of our definition of See page 120 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 while climbing the height h. The only possibility is that the light arrives at the top with frequency *different* from the one at which it is emitted from the bottom.**

In short, it turns out that rising light is gravitationally redshifted, and that the light descending from the top of a tree down to an observer is *blue shifted*; this gives a darker, older colour to the top in comparison to the bottom of the tree. General relativity thus says that trees have different shades of green along their height.*** How big is the effect?

Challenge The result deduced from the drawing is again that of formula (173). That is expected, as Challenge the two experiments describe equivalent situations, as you might want to check yourself. The formula gives a relative change of frequency f of only $1.1 \cdot 10^{-16}$/m on the surface of the earth. For trees, this so-called gravitational Doppler effect is far to small to be observable, at least using normal light.

* As in special relativity, here and in the rest of our escalation, the term ‘mass’ always refers to rest mass.
** The relation between energy and frequency of light is described and explained in the part on quantum theory, on page 434.
Challenge *** How does this argument change if one includes the illumination by the sun?

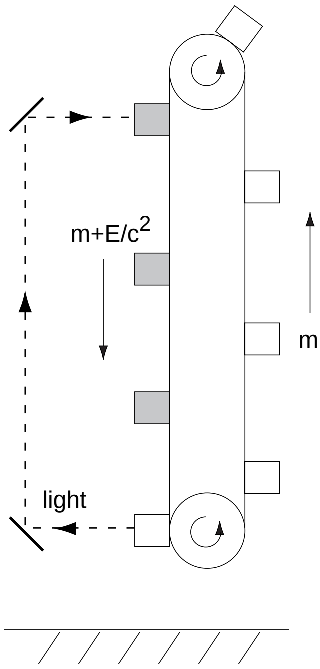


Figure 84 The necessity of blue and redshift of light: why trees are greener at the bottom

In 1911, Einstein proposed to check the change of frequency with height by measuring the redshift of light emitted by the sun, using the famous Fraunhofer lines as colour markers. The first experiments, by Schwarzschild and others, were unclear or even negative, due to a number of other effects that change colours at high temperatures. Only in 1920 and 1921, Grebe and Bachem, as well as Perot showed that careful experiments indeed confirm the gravitational Doppler effect. In later years, technology 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 Doppler shift in their university tower. Ref. 18 Ref. 6

But our two thought experiments tell us much more. Using the same arguments as in the case of special relativity, the colour change also implies that clocks run *differently* at the top and at the bottom, as they do in the front and in the back of a train. Therefore, in gravity, *time is height dependent*, as the definition says. In fact, *height keeps young.* Can you confirm this conclusion? Challenge

In 1972, by flying four precise clocks in an airplane, and keeping an identical one on the ground, Hafele and Keating found that clocks indeed run differently at different altitudes according to expression (173). Subsequently, in 1976, a team around Vessot shot a clock based on a maser, a precise microwave generator and oscillator, upwards on a missile, and again confirmed the expression by comparing it with an identical maser on the ground. And in 1977, Briatore and Leschiutta showed that a clock in Torino indeed goes faster than one on the top of the Monte Rosa. They confirmed the prediction that on earth, for every 100 m of height gained, one stays younger about 1 ns per day. In the meantime, this effect has been confirmed for all gravitational systems for which experiments were performed, such as several other planets, the sun, and many other stars. Ref. 7 Ref. 8 Ref. 9 Challenge

Can you show that the formula describing gravitational redshift complies with the general limit (172) on length to mass ratios?

Note that both an observer at the lower position and one at a higher position *agree* on the result. Both agree that the lower clock goes faster. In other words, when gravity is present, space-time is *not* described by the Minkowski space-time of special relativity, but by some more general space-time. To put it mathematically, whenever gravity is present, one has

$$ds^2 \neq c^2 dt^2 - dx^2 - dy^2 - dz^2 \quad . \tag{174}$$

We will give the correct expression shortly. But is this view of gravity really reasonable? In order to finish our check, we have to show that the fall of stones results from the different running of clocks at different heights.

What is weight?

We saw that a single observer cannot distinguish the effects of gravity from those of acceleration. This property of nature allows to make 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.

Obviously, an accelerating surface of the earth produces a weight for each body which is proportional to its inertial mass, or, as this is usually expressed, the inertial mass of a body is exactly identical to the gravitational one. This is indeed observed, and to the highest precision achievable. Roland van Eötvös* performed many such high-precision experiments throughout his life, without finding any discrepancy. In these experiments, he used the fact that the inertial mass is important for centrifugal effects, and the gravitational mass is for free fall. Can you imagine how exactly he tested the equality?

Ref. 10

Challenge

See page 56

However, this is not a surprise. We remember the definition of mass ratio as negative inverse acceleration ratio, independently of its origin, and we remember that mass measurements cannot be used to distinguish inertial and gravitational mass at all. We saw that both masses are equal by definition already in classical physics, and that the whole discussion is a red herring.

See page 84

In any case, when we step into an elevator in order to move down a few stories, and push the button, the following happens: the elevator is pushed upwards by the accelerating surface of the earth somewhat less than the building; the building overtakes the elevator, which therefore remains behind. Moreover, due to the weaker push, at the beginning everybody inside the elevator feels a bit lighter. When the contact with the building is restored, the elevator is accelerated to catch up with the accelerating surface of the earth. Therefore everybody feels like in a strongly accelerating car, pushed into direction opposite to the acceleration: for a short while, one feels heavier. And of course, during free fall, one feels no weight; this is obvious, since no floor is pushing.

Why do apples fall?

> Vires acquirit eundo.
> Vergilius**

Answering this question is now straightforward. Sitting in an accelerating car, an object thrown forward will soon be caught by the car again. For the same reason, a stone thrown upwards is soon caught up by the surface of the earth, which is continuously accelerating upwards. Similarly, when an an apple detaches from a tree, it stops being accelerated by the branch, and after a short while of rest which we usually call free fall, it is squashed by the approaching surface of the earth.

We are not 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.

* Roland van Eötvös (1848, Budapest–1919, Budapest), hungarian physicist. The university of Budapest is named after him.

** 'Going it acquires strength.' Publius Vergilius Maro, (Andes, 70 BCE–Brundisium, 19 BCE) Aeneis 4, 175.

In other words, if somebody steps out of a window in a skyscraper, he has decided to reject being continuously accelerated by the floor or by the window-sill, and chosen in favour of the calmness of real rest. Unfortunately, the accelerating surface of the earth approaches mercilessly and, depending on the time he stayed at rest, the earth hits the body with a corresponding velocity, leading to more or less severe shape deformation.

Can this reasoning can be continued without limit? One 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. But it is really impossible to distinguish acceleration from gravitation? No; distinction *is* possible. One only has to compare *two* falling observers. Challenge

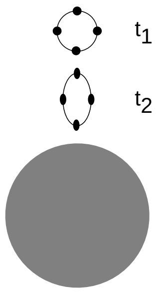


Figure 85 Tidal effects: what bodies feel when falling

Two nearby observers in a gravitational field observe that during the fall, their relative distance changes. As a consequence, a large body in free fall is slightly *squeezed*. The essence of gravity is that free fall is *different* from point to point. One also says that gravity is characterized by its *tidal effects*. In short, gravity is simple only *locally*.

There is a second way to put this issue. Due to its tidal effects, it is impossible to find a single inertial reference frame describing different observers freely falling near a mass. Kittinger could specify an inertial frame for himself, but not one which is also inertial for a colleague falling on the opposite side of the earth!

As it is impossible to find a common inertial frame, that means that there is no such frame. In short, space-time is *not* described by Minkowski-space when gravity is present. In particular, if one wants a description of gravity which works for all observers, incorporates tidal effects, and is consistent with a constant light speed, there is only one conclusion.

Bent space

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, the Times newspaper in London announced the results of a double expedition to South America, which for the first time proved that the theory of universal gravity, essentially given by $a = GM/r^2$, was wrong, and that instead space had been shown to be 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, so that people found the field equations of general relativity, in tensor form and with the greek indices, in the middle of Time magazine. This did not happen to any other physicist before or afterwards.

The expedition had performed an experiment proposed by Einstein himself. Since clocks in those days were not precise enough to detect the variation of time with height, he had

* When a bug walks over the surface of a sphere it probably does not notice that the path it walks is curved. I had the luck to notice it.

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 an eclipse, and compare it with a picture of the same stars at night, when the sun is far away. Einstein predicted a change in position of 1.75' for star images at the border of the sun, a result *twice* as large as the effect predicted by universal gravity. The prediction, corresponding to about $1/40$ mm on the photographs, was confirmed, and universal gravity was ruled out.

See page 84

Does this experiment *imply* that space is warped or *curved*, as physicists prefer to say? The answer is simple: no, it doesn't, but space-time is curved alright. In fact, other explanations could be given for the result of the eclipse experiment, such as a potential differing from the one of universal gravity.

But we know more: we know about the change of time with height.* Experiments show that any two observers at different height measure the same value for the speed of light c near themselves, as experiments confirm. 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: meter bars, like clocks, also *change* with height, and in such a way to yield the same speed of light everywhere.

Challenge

Since meter bars change with height, *space is curved near masses*.

In the twentieth century, many experiments checked whether meter bars indeed behave differently wherever gravity is present. 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 measured. All results confirm the curvature of space and space-time, and confirm the predicted values. In other words, near masses meter bars do indeed change their size from place to place, and even from orientation to orientation. An impression of the situation can be got from figure 86.

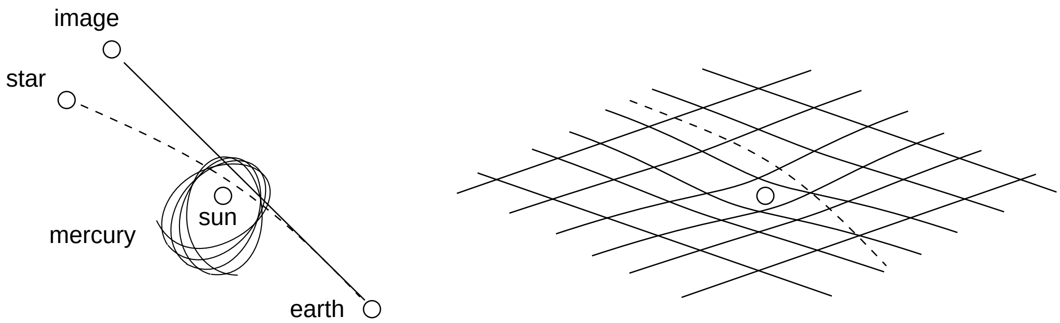


Figure 86 The path of a light beam and of a satellite near a spherical mass

But attention: the right hand figure, even though found in all textbooks, is misleading. It can be easily mistaken to show a *potential* around a body. Indeed, it is impossible to

* Historically, the different running of clocks with height was shown experimentally only forty years *after* the detection of space curvature. We described those experiments above. We thus can explain the situation much more easily nowadays.

draw a graph showing curvature and potential separately. (Why?) We will see that for small curvatures, it is in fact possible to describe the meter bar change with 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, potentials cannot be defined, and thus there is indeed no way to avoid curved space in the general case. We will discuss the way to describe curvature in detail later on. Challenge

If gravity means curved space, one follows that any accelerated observer, like a man standing on the earth, must also observe that space is curved. Obviously, in everyday life we do not note any such effect. How would you devise a precision experiment to check the statement? Challenge

In fact, not only space, but also space-time is curved, even though figure 86 only shows the curvature of space alone. The issue then becomes how to describe the shape of space, e.g. the one shown in figure 86, as well as the shape of space-time.

In the case of the figure, the best description of points is the use of a time t defined as the time measured by a clock located at infinity; that avoids problems with the uneven running of clock with distance from the central mass. For a 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 body divided by 2π.

The shape is described by the behaviour of the space-time distance di, also called wristwatch time, between two neighboring points with coordinates (t,r) and $(d+dt, r+dr)$. We know from above that gravity means that in spherical coordinates one has See page 156 See page 210

$$di^2 \neq dt^2 - dr^2/c^2 - r^2 d\varphi^2/c^2 \quad . \tag{175}$$

This inequality means that space-time is curved. Indeed, the above experiments show that the space-time interval around a spherical mass follows

$$di^2 = \left(1 - \frac{2GM}{rc^2}\right) dt^2 - \frac{dr^2}{1 - \frac{2GM}{rc^2}} - r^2 d\varphi^2/c^2 \quad . \tag{176}$$

This is called the *Schwarzschild metric* after one of its discoverers.* The other discoverer, unknown to Einstein, was the Dutch physicist J. Droste. The metric (176) describes the curvature of space-time around a spherical non-rotating mass, such as well approximated by the earth or the sun. (Why can the rotation be neglected?) Gravity's strength is obviously measured by a dimensionless number h defined as Ref. 12 Challenge

$$h = \frac{2GM}{Rc^2} \quad . \tag{177}$$

It measures the gravitational strain with which lengths and the vacuum are deformed, as well as the amount that clocks go late. On the surface of the earth it has the small value of $1.4 \cdot 10^{-9}$, and on the surface of the sun the larger value of $4.2 \cdot 10^{-6}$. Modern clocks can easily detect these changes. The consequences and uses will be discussed shortly.

* Karl Schwarzschild (1873–1916), important german astronomer; he was one of the first persons to understand general relativity. He published his solution in december 1915, only few months after Einstein had published his field equations. He died prematurely, at age 42, much to Einstein's chagrin. We will deduce the metric again later on, directly from the field equations of general relativity.

One also notes that if a mass gets small, in particular when its radius gets *equal* to its so-called *Schwarzschild radius*

$$R_S = \frac{2GM}{c^2} \quad (178)$$

the above metric behaves strangely: time disappears. What happens precisely will be shown See page 280 below. The situation is not common; the Schwarzschild radius for the earth is 8.8 mm and for the sun 3.0 km; you might want to check that it is quite smaller than the object size for Challenge all systems in everyday life. Bodies who reach this limit are called *black holes*, and we will study them in detail shortly. In fact, as stated above, general relativity states and is based on the fact that *no* system in nature is smaller than its Schwarzschild size, or that h is *never* above unity.

In summary, the results mentioned so far make it clear that curvature is generated by mass. Special relativity then tells us that as a consequence, space should also be curved by energy-momentum. For example, light or neutrinos should also curve space-time. Unfortunately, even the highest energy beams correspond to extremely small masses, and thus to unmeasurably small curvatures. On the other hand, it is still possible to show that energy curves space, since in almost all atoms, a large part of the mass is due to the electrostatic energy among the positively charged protons. For example, in 1968 Kreuzer showed this Ref. 13 with a clever experiment using a floating mass.

It is straightforward to picture that the uneven running of clock is the temporal equivalent Challenge of space-time curvature. The complete statement is therefore to say that in case of gravity, *space-time* is curved. In short, since gravity is like acceleration, since acceleration is position dependent time, and since light speed is constant, one deduces that *energy-momentum tells space-time to curve.* This statement is the first half of general relativity.

In addition, we can deduce that different observers measure different curvatures. The set of transformations from one viewpoint to the other in general relativity is called *diffeomorphism symmetry*. We will study it in more detail below.

Obviously, the next question is *how* space-time curves. One can answer it starting from the principle that the length to mass ratio never, exceeds $4G/c^2$, for any observer whatsoever. Challenge More about this below.

Since matter moves, one 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 spacetime, is *elastic*. However, it is rather stiff, and quite a lot stiffer than steel.* In fact, to curve a piece of space by 1%, one needs an energy density enormously Challenge larger than that required to curve a usual train rail by 1%. The consequences of space-time curvature and of its elasticity are a lot of fun.

Why does a stone thrown into the air fall back? – geodesics

In special relativity, we saw that inertial or free floating motion is that motion which connects two events that requires the *longest* proper time. The motion fulfilling this requirement in the absence of gravity is *straight* motion. On the other hand, we are used to think See page 47 of straightness, as we did near the beginning of our escalation, as the shape of light rays.

* A good book in popular style on the topic is David BLAIR & Geoff MCNAMARA, *Ripples on a cosmic sea*, Allen & Unwin, 1997.

Indeed, we all are used to check the straightness of an edge by looking along it. And whenever we draw the axes of a physical coordinate system, we imagine drawing paths of light rays.

In the absence of gravity, object paths and light paths coincide.

But when gravity is present, objects do not move along light paths, as every thrown stone shows. In the case of gravity, both paths are bent. Light does not define spatial straightness any more. But the other statement remains: even when gravity is present, bodies follow paths of *longest possible proper time*. Such paths are called *(timelike) geodesics* for objects and *(lightlike) geodesics* for light.

This statement can be checked in several ways. One is to use the equivalence of gravity and acceleration. Another is to check it explicitly. Still another is to show that it follows from the limit on length mass ratios.

We saw already above, that fall can be seen as due to earth's surface approaching. But we also saw that it is more useful to use the curvature of space-time to explain the fall of bodies. This can be done in the following way.

– CS – story to be filled in – CS –

Interestingly, the limit on length to mass ratios can also explain the fall of bodies.

– CS – to be filled in – CS –

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 other words, in space-time any *deviation from free fall keeps young*. In short, the motion of any particle falling freely "in a gravitational field" is described by the same variational principle than that of a free particle in special relativity: it follows the path that maximizes the proper time $\int d\tau$. We rephrase this by saying that any particle in free fall from point A to point B maximizes the action S given by

$$S = -mc^2 \int_A^B d\tau \quad . \tag{179}$$

That is all one needs to know about the fall of objects. As we will see below, this description of fall has been tested extremely precisely, and no differences between this expression and experiment has ever been observed. We will also see that for free fall, the predictions of general relativity and of universal gravity differ substantially both for rapid particles and for central bodies near the size to mass limit. All experiments show that whenever the two predictions differ, general relativity is right and universal gravity, as well as any other theory developed so far, are wrong. See page 289 Ref. 14

Above we called free fall the official definition of *rest*; we can thus say that with general relativity everything about rest (of large bodies) is known, as well as everything about the departure from it.

Of course, also here the next question is whether energy falls in the same way as mass. Bound energy does, as is proven by comparing the fall of objects made of different mate-

Challenge rials. They have different percentages of bound energy (why?). For example, on the moon, where there is no air, cosmonauts dropped steel balls and feathers and found that they fell in the same way, alongside each other. The independence on material composition has been checked over and over again, and no difference has ever been found. Ref. 15

What about radiation? Radiation is energy without rest mass, and it moves like extremely fast and extremely light objects. Therefore deviations from universal gravity become most apparent for light. Already long before relativity, in 1801, the prussian astronomer Johann Soldner understood that universal gravity implies that light is *deflected* when passing near a mass. He also calculated the deflection angle, which depends on the mass of the body and the distance. But nobody cared to check the result experimentally. Obviously, light has energy, and energy also has weight; the deflection of light by itself is thus *not* a proof of the curvature of space. Ref. 16 See page 84

General relativity also predicts a deflection of light near masses, but of *twice* the classical value, because the curvature of space around large masses adds to the effect already included by universal gravity. The deflection of light thus only proves the curvature of space if the *value* agrees with the one predicted by general relativity. And indeed, the observations coincide with the prediction. More calculation and experimental details will be given shortly. See page 228

Experiments such as this confirm that the motion of light and of massless neutrinos can be described, by a slight modification of the action for objects, namely Ref. 17

$$S^2_{\text{massless}} = \int_A^B (d\tau/ds)^2 ds = \int g_{ab} \frac{dx^a}{ds}\frac{dx^b}{ds} ds \quad . \tag{180}$$

This action is minimized for motion of massless entities between events A and B. These are the *lightlike geodesics* of space-time.

In short, all experiments show that not only mass, but also energy falls along geodesics, whatever its type, bound or free, and whatever the interaction, electromagnetic, or nuclear. In addition, the motion of radiation confirms that space-time is curved. In summary, we find that *space-time tells matter, energy and radiation how to fall*. This statement is the second half of general relativity.

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, we have some fun.

General relativity in everyday life

> 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

* If you do not take the answer too seriously and take it only for amusement, I can explain it to you in the following way: in the past it was thought that if all things disappear from the world, space and time would remain. But following relativity theory, space and time disappear together with the things.

General relativity is a beautiful topic with numerous interesting aspects. One can learn a lot from its more curious sides.

• The radius of curvature of space-time at the earth's surface is $1.7 \cdot 10^{11}$ m. Are you able to confirm this value? Challenge

• We saw in special relativity that if two twins are identically accelerated in the same direction, the first one ages more than the second one. Is this the same in a gravitational field? What happens, when the field varies with height, as happens on the earth? See page 160 Challenge

• How do cosmonauts weigh themselves? Challenge

• Is a cosmonaut really floating freely? No. It turns out that space stations and satellites are accelerated by several 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. They all lead to accelerations of the order of $10^{-6}\,\mathrm{m/s^2}$ to $10^{-8}\,\mathrm{m/s^2}$, depending on the height of the orbit. When will an apple floating in space will hit the wall of a space station? Challenge

• There is no negative mass in nature. This means that gravitation cannot be shielded, as is possible for electromagnetic interactions. Even antimatter has positive mass. 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 trouble produced by this limitations later on; for example, an *upper limit* for the entropy of systems will appear. See page 518

• 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 of space-time, 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? Yes; however, such a curved space would need to have a negative energy density. Such a situation is in contrast with the definition of energy and with the nonexistence of negative mass. The requirement that this does not happen is also called the *weak energy condition*. Can you say whether it is included in the limit on length to mass ratios? Ref. 19 Challenge

• Like in special relativity, the limit $L/M \geqslant 4G/c^2$ is a challenge to devise experiments to overcome it. Can you explain what happens when a fast observer floats past a mass, so that it is length contracted until the limit is reacged? Challenge

• There is an important mathematical detail which singles out the dimension 3 from all other possibilities. A closed curve can be knotted *only* in $\mathbf{R}^3$, whereas it can be unknotted in any other, i.e. higher dimension. (This fact is also the reason that three is the smallest dimension that allows chaotic particle motion.) However, general relativity does not tell *why* space-time has three plus one dimensions. It is simply based on the fact. This issue will be settled only in the third part of the escalation.

• Henri Poincaré, who died in 1912, shortly before the general theory of relativity was finished, thought that curved space was not a necessity, but only a possibility. He thought that one could simply continue using euclidean space and simply say that light follows curved paths. Can you show why his idea is wrong? Challenge

• Can two atoms circle each other, in their respective gravitational field? What would be the size of this "molecule"? Challenge

• Can two light pulses circle each other, in their respective gravitational field? Challenge

• The various motions of the earth mentioned in the section on galilean physics, such as rotation on its axis, rotation around the sun, etc. give the reason for the use of various types of time in physics and astronomy. The time defined by the best atomic clocks is called 'terrestrial dynamical time'. By inserting leap seconds every now and then to compensate for the bad definition of the second (an earth rotation does not take 86400, but 86400.002 seconds) and, in minor ways, for the slowing of earths rotation, one gets the *universal time coordinate*; then one has the time derived from this by taking into account all those *leap seconds*. One then has the time which would be shown by a nonrotating clock in the centre of the earth. Finally, one has 'barycentric dynamical time', which is the time that would be shown by a clock in the centre of mass of the solar system. Only using this time latter, satellites can be reliably steered through the solar system. Relativity says goodbye to Greenwich mean time, as does british law, in a few cases were the law follows science.

See page 71
See page 730
Ref. 20

• 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 the aimed spots and usually even the whole planet. In fact, space agencies take the safe sided; they use a generalization of general relativity, namely the so-called parametrized post-newtonian formalism, which includes a continuous check whether general relativity is correct. Within measurement errors, no deviation was ever found, so far.*

• General relativity is also used by space agencies around the world to know the exact positions of satellites, and to tune radios to the frequency of radio emitters on satellites. The so-called *global positioning system*, or GPS, is now becoming a standard tool in navigation.** GPS consists of 24 satellites with clocks flying around the world. Why does the system need general relativity to operate? Since both a satellite as well as a person on the surface of the earth travel in circles, we have $dr = 0$ and we can rewrite the Schwarzschild metric (176) as

Ref. 21

$$\left(\frac{di}{dt}\right)^2 = 1 - \frac{2GM}{rc^2} - r^2\left(\frac{d\varphi}{dt}\right)^2 = 1 - \frac{2GM}{rc^2} - v^2 \quad . \tag{182}$$

Challenge

For the relation between satellite and earth time we then get

$$\left(\frac{dt_{\text{sat}}}{dt_{\text{earth}}}\right)^2 = \frac{1 - \frac{2GM}{r_{\text{sat}}c^2} - \frac{v_{\text{sat}}^2}{c^2}}{1 - \frac{2GM}{r_{\text{earth}}c^2} - \frac{v_{\text{earth}}^2}{c^2}} \tag{183}$$

* 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

$$a = \frac{GM}{r^2} + f_2 \frac{GM}{r^2}\frac{v^2}{c^2} + f_4 \frac{GM}{r^2}\frac{v^4}{c^4} + f_5 \frac{Gm}{r^2}\frac{v^5}{c^5} + \dots \tag{181}$$

where the numerical factors f_n are of order one. The first uneven terms are missing because of reversibility, were it not for gravity wave emission, which accounts for the small term f_5; note that it contains the small mass instead of the large one. Nowadays, all factors f_n up to f_7 have been calculated. However, in the solar system, only the term up to f_2 has ever been detected, a situation which might change with future high precision satellite experiments. Higher order effects, up to f_5, have been measured in the binary pulsars, as discussed below.

For a *parametrized* post-newtonian formalism, all factors f_n are fitted through the data coming in, including the even ones; so far all these factors agree with general relativity´s prediction.

** For more information, see the http://www.gpsworld.com web site.

Can you deduce how many microseconds a satellite clock runs fast every day, given that the satellites turn around the earth every twelve hours? Since only three microseconds would give a position error of one kilometer after a single day, the clocks in the satellites are adjusted to run slow by the calculated amount. The results confirm general relativity within experimental errors. Challenge Ref. 22

• The gravitational constant G does not seem to change with time. Present experiments limit its rate of change to less than 1 part in 10^{12} per year. Can you imagine how this can be checked? Ref. 23 Challenge

• Could the idea that we live in 3 space dimensions be due to a limitation of our senses? How? Challenge

• What is the strongest possible gravitational field? The one of black small holes, as already mentioned. The strongest *observed* gravitational field is somewhat smaller 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 up to 30% of the maximum possible value. What is the gravitational acceleration, assuming a mass equal to the sun? Ref. 24 Challenge

• What is the angular size δ of a mass M with radius R at distance d? Light deflection leads to the expression Ref. 25 Challenge

$$\delta = \arcsin\left(\frac{R\sqrt{1-R_S/D}}{D\sqrt{1-R_S/R}}\right) \quad \text{where} \quad R_S = \frac{2GM}{c^2} \tag{184}$$

We will come back to the issue shortly. See page 286

• Much information about general relativity is available on the net. As a good starting point for US-american material, see the http://math.ucr.edu/home/baez/relativity.html web site.

• Is it correct to claim that matter cannot be continuous because inside a closed shell of matter gravity can still be present? Challenge

Many aspects of general relativity can be understood without or with little mathematics. The next section slowly introduces the calculations necessary for the description of motion. If the concepts get too involved for a first reading, just continue with the sections on cosmology and on black holes, which again use very little maths.

8. Motion in general relativity

I have the impression that Einstein understands relativity theory very well.
Chaim Weitzmann, chemist, later first president of Israel

Before we enter the deepest guts of general relativity, we study how the motion of objects and light *differs* from that predicted in universal gravity, and how these differences can be measured.

Weak fields

As mentioned above, one calls *strong* gravity those situations for which universal gravity gives incorrect results. This happens when

$$\frac{2GM}{Rc^2} \approx 1 \tag{185}$$

and applies near black holes, as we will see below, or to extremely high energies, as we will discover in the third part of our escalation. For most of nature, gravity is a *weak* effect and the number just mentioned much smaller than one, despite the violence of avalanches or of falling asteroids. In these cases, gravitation can still be approximated by a field, despite what was said above. These weak field situations are interesting because they are simple to understand, as they only require for their explanation the different running of clocks with height, allowing to mention space-time curvature only in passing to still to think of gravity as an interaction. However, many new and interesting effects appear.

The Thirring effects

In 1918, the german physicist Joseph Thirring published two simple and beautiful predictions of motions, one with his collaborator Hans Lense, which do not appear in universal Ref. 30 gravity, but do appear in general relativity.

In the first example, called the *Thirring effect*, centrifugal accelerations as well as Coriolis accelerations on all masses in the interior of a rotating mass shell are predicted, in contrast Challenge to the classical description. Are you able to deduce this effect from the figure?

The *Thirring-Lense effect* is somewhat more complex. It predicts that the oscillating Foucault pendulum, or a satellite circling the earth in a polar orbit, do not stay precisely in a fixed plane compared to the rest of the universe, but that the earth drags the plane along a tiny bit. This *frame-dragging*, as it is also called, arises from the fact that the earth in vacuum behaves like a ball in honey; when it rotates, it drags some honey with it. Similarly, the earth drags some vacuum with it, and thus moves the plane of the pendulum. Of course, the effect also moves the plane of an orbiting satellite.

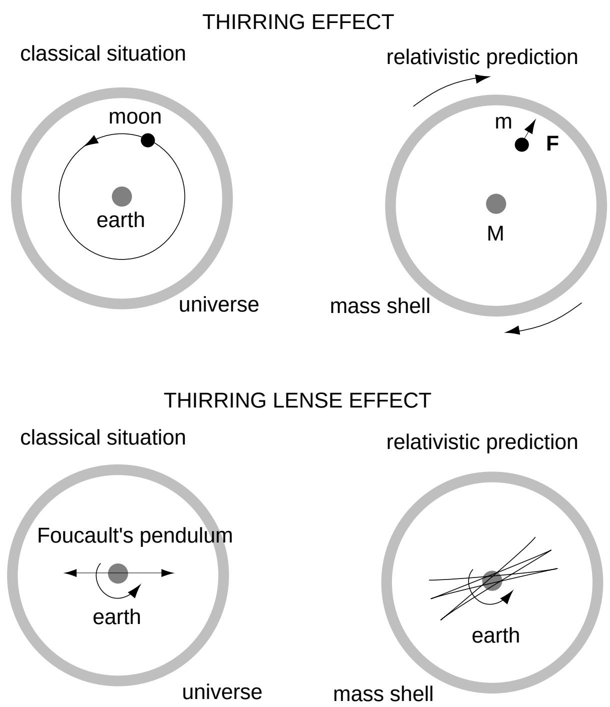


Figure 87 The Thirring and the Thirring-Lense effects

The Thirring-Lense effect has been measured for the first time in 1998 by the italian group led by Ignazio Ciufolini. They followed the motion of two special artificial satellites consisting only of a body of steel and some cat eyes. The group measured the satellite's motion around the earth with extremely high precision using reflected laser pulses. This method allowed this low budget experiment to Ref. 31 beat by many years the efforts of much larger but much more sluggish groups.* The results confirm general relativity within about 25%.

Frame dragging effects have also been measured in binary star systems, which is possible if one of the stars is a pulsar; such stars send out regular radio pulses, e.g. every millisecond. By measuring the exact time when they arrive on earth, one can deduce the way these stars move, and confirm even such subtle effects as frame dragging. Ref. 32

Gravitomagnetism

Frame-dragging and the Thirring-Lense effect can be seen as special cases of gravitomagnetism. This approach to gravity, already studied by Heaviside, has become popular again in recent years, especially for its didactic aspect.

As mentioned above, talking about a gravitational *field* is always an approximation. But for weak gravity, it is a good one. Many relativistic effects can be described with it, without using space curvature and the metric tensor. For a relativistic description of such weak gravity situations, the field can be split into an "electric" and a "magnetic" component, as is done for the electromagnetic field.

Like in the case of electromagnetism, the split depends on the observer; on the other hand, for the standard observer outside the system, we have a good feeling on how the two fields behave. In the case of gravity, one the talks about the *gravitoelectric* and the *gravitomagnetic* fields.*

The frame dragging effects just mentioned can be visualised by this method quite easily. The acceleration of a charged particle in electrodynamics is described by the Lorentz' equation

$$m\ddot{\mathbf{x}} = q\mathbf{E} - e\dot{\mathbf{x}} \times \mathbf{B} \quad . \tag{186}$$

For gravity this becomes, as we will show below,

$$m\ddot{\mathbf{x}} = m\mathbf{G} - m\dot{\mathbf{x}} \times \mathbf{H} \quad . \tag{187}$$

In this expression we already know the field $\mathbf{G}$, given by

$$\mathbf{G} = \nabla\varphi = \nabla\frac{GM}{r} = -\frac{GM\mathbf{x}}{r^3} \tag{188}$$

As usual, the quantity φ is the (scalar) potential. This is the field of universal gravity produced by every mass, and in this context is called the *gravitoelectric field*.

In fact it is not hard to show that if gravitoelectric fields exist, gravitomagnetic fields must exist as well, as they follow whenever one changes from an observer at rest to a moving one. (The same argument is used in electrodynamics.) A particle falling perpendicularly towards an infinitely long rod already makes the point, as shown in figure 88. An observer at rest with respect to the rod can describe the whole situation with gravitoelectric forces alone. A second observer, moving along the rod with constant speed, observes that the momentum of the particle *along the rod* also increases. Equivalently, moving masses produce a gravitomagnetic (3-) acceleration on test masses m given by Ref. 33 Challenge

* Such as the so-called Gravity Probe B satellite experiment, which will drastically increase the measurement precision around the year 2005.

* The approximation requires slow velocities, weak fields, as well as localized and stationary mass-energy distributions.

$$m\mathbf{a} = m\mathbf{v} \times \mathbf{H} \tag{189}$$

where, as in electrodynamics, a static *gravitomagnetic field* obeys

$$\mathbf{H} = \nabla \times \mathbf{h} = 4\pi N \rho \mathbf{v} \tag{190}$$

where ρ is mass density and N is a proportionality constant. The quantity $\mathbf{h}$ is obviously called the *gravitomagnetic (vector) potential*. We see that universal gravity is the approximation of general relativity appearing when all gravitomagnetic effects are neglected.

Challenge When the situation in figure 88 is evaluated, one gets

$$N = \frac{G}{c^2} = 7.4 \cdot 10^{-28}\,\mathrm{m/kg} \tag{191}$$

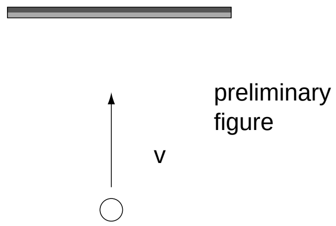


Figure 88 The reality of gravitomagnetism

The constant is extremely small. In addition, a second point makes the observation extremely difficult. In contrast to electromagnetism, in the case of gravity there is

Challenge now way to observe *pure* gravitomagnetic fields (why?); they are always mixed with the usual, gravitoelectric ones. And as in the electrodynamic case, the latter are stronger by a factor of c^2. For these reasons, gravitomagnetic effects have been measured for the first time only in the 1990s.

In short, *if a mass moves, it also produces a gravitomagnetic field*. In this description, *all frame dragging effects are gravitomagnetic effects*.

Obviously, a gravitomagnetic field also appears when a large mass *rotates*. For an angular momentum J it is given by

$$\mathbf{H} = \nabla \times \mathbf{h} = \nabla \times \left(-2 \frac{\mathbf{J} \times \mathbf{x}}{r^3} \right) \tag{192}$$

exactly as in the electrodynamic case. In particular, like in electromagnetism, a spinning test particle with angular momentum $\mathbf{S}$ feels a *torque* if it is near a large spinning mass with angular momentum $\mathbf{J}$. And obviously, this torque $\mathbf{T}$ is given by

$$\mathbf{T} = \frac{d\mathbf{S}}{dt} = \tfrac{1}{2} \mathbf{S} \times \mathbf{H} \quad . \tag{193}$$

Since for a torque one has $\mathbf{T} = \dot{\mathbf{\Omega}} \times \mathbf{S}$, a large rotating mass with angular momentum $\mathbf{J}$ has an effect on an orbiting particle. Seen from infinity one gets, for an orbit with semimajor

Challenge axis a and eccentricity e,

$$\dot{\mathbf{\Omega}} = -\frac{\mathbf{H}}{2} = -\frac{\mathbf{J}}{|\mathbf{x}|^3} + \frac{3(\mathbf{J}\mathbf{x})\mathbf{x}}{|\mathbf{x}|^5} = \frac{2\mathbf{J}}{a^3(1-e^2)^{3/2}} \tag{194}$$

which is the prediction by Lense and Thirring.* The effect is extremely small, giving a change of only 8″ per orbit for a satellite near the surface of the earth. Despite this smallness, and a number of larger effects disturbing it, Ciufolini's team managed to confirm the result.

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

The distinction of gravitoelectric and gravitomagnetic effects corresponds to a description in which only clock effects are taken into account, as we will find out below. Despite this limitation, it is quite useful. For example, it helps to answer questions such as: How can gravity keep the earth around the sun, if gravity needs 8 minutes to get from the sun to us? Challenge To find the answer, thinking about the electromagnetic analog can help.

The split of the gravitational field into gravitoelectric and gravitomagnetic components also allows a simple description of gravity waves.

Gravitational waves

One of the most fantastic predictions of physics is the existence of gravitational waves. Gravity waves prove that empty space-time itself has the ability move. The basic idea is simple. Since space-time is elastic, it is predicted that it it can oscillate in the form of propagating waves, like any other elastic medium.

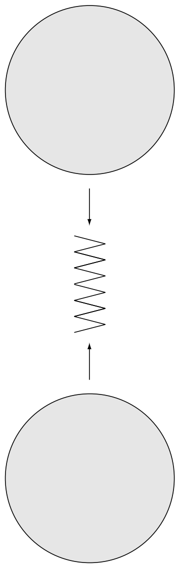


Figure 89 A Gedankenexperiment showing the necessity of gravity waves

Kalckar and Ulfbeck have given a simple argument for the necessity of gravitational waves by studying two equal masses falling towards each other. They simply imagined a spring and a metre bar in their middle. The central spring stores the kinetic energy from the falling masses. That energy can be measured by determining, with the metre bar, the length by which the spring is compressed. When the spring springs back and hurls the masses back into space, the gravitational attraction will decelerate the masses, until they again fall towards each other, thus starting the same cycle again.

However, the energy stored in the spring is smaller with each cycle. When a sphere detaches from the spring, it gets decelerated by the other sphere. Now comes the point. The value of this deceleration is given by the distance to the other mass, but since there is a maximal propagation velocity, it is given by the distance the other mass had when its gravity reached the end of the spring. In short, while going up, the deceleration is *larger* than the one calculated without taking the time delay into account.

Similarly, when the mass falls back gown, it is accelerated by the other mass using the distance it had when its gravity reached the other. Therefore, while going down, the deceleration is *smaller* than without time delay. One sees that the masses arrive with a smaller energy than they depart with.

The difference of these two energies is lost to each mass; it is taken away by space-time; that is an effect of gravitational radiation. As we will see, this effect has already been measured, with the difference that instead of being tied by a spring, the two masses orbit each other.

A simple mathematical description of gravity waves appears with the split into gravitomagnetic and gravitoelectric effects. It does not take much to extend gravitomagnetostatics Ref. 34 and gravitoelectrostatics to *gravitodynamics*. Just as electrodynamics can be deduced from Coulomb's attraction, when one switched to inertial observers, gravitodynamics can be deduced from universal gravity. One gets the four equations Challenge

$$\nabla \mathbf{G} = -4\pi G\rho \quad , \quad \nabla \times \mathbf{G} = -\frac{\partial \mathbf{H}}{\partial t}$$
$$\nabla \mathbf{H} = 0 \quad , \quad \nabla \times \mathbf{H} = -4\pi G\rho + \frac{H}{G}\frac{\partial \mathbf{G}}{\partial t} \tag{195}$$

which are *exactly the same* as those for electrodynamics. One can easily deduce a *wave equation* for the gravitoelectric and the gravitomagnetic fields **G** and **H**. In other words, *gravity can radiate*. All this follows from the expression of universal gravity when applied to moving observers, using the fact that observers cannot move faster than c. The story with the spring and the mathematical story use the same assumptions and come to the same conclusion.

A few manipulations show that the speed of these wave is given by

$$c = \sqrt{\frac{G}{N}} \tag{196}$$

which corresponds to the famous expression from electromagnetism

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \quad . \tag{197}$$

The same letter has been used for the two speeds, as they are identical. Both influences travel with the speed common to all energy moving with vanishing rest mass.

How does one have to imagine these waves? The waves correspond to moving deformations of space-time. It turns out that gravity waves are *transverse*. One finds also that waves can be polarized in two ways. The effect of a gravitational wave in one polarization is shown in figure 90. The effect of the other polarization is the same, rotated by 45 degrees.*

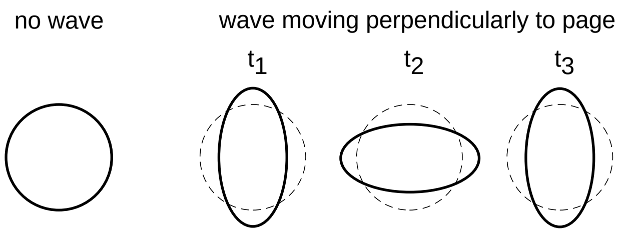


Figure 90 Effects on a circular body of a plane gravitational wave moving vertically to the page

How does one produce such waves? The conservation of energy does not allow changing mass monopoles. Also a spherical mass which changes in radius periodically would not

* A (small amplitude) plane gravity wave travelling in z-direction is described by a metric

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1+h_{xx} & h_{xy} & 0 \\ 0 & h_{xy} & -1+h_{xx} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{198}$$

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

$$h_{ab} = A_{ab} \sin(kz - \omega t + \varphi_{ab}) \tag{199}$$

as in all plane harmonic waves. The dispersion relation resulting from the wave equation

$$\frac{\omega}{k} = c \tag{200}$$

shows that the waves move with the speed of light.

emit gravitational waves. The conservation of momentum does not allow changing mass dipoles. In summary, only *changing quadrupoles* can emit waves. For example, two masses in orbit around each other will emit gravitational waves. Challenge

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

$$h_{ab} = \frac{2G}{c^4}\frac{1}{r}d_{tt}Q^{\text{ret}}_{ab} = \frac{2G}{c^4}\frac{1}{r}d_{tt}Q_{ab}(t-r/c) \tag{202}$$

This shows that the amplitude of gravity waves *decreases only with* $1/r$, in contrast to naive expectations. However, also this feature is the same as for electromagnetic waves. In addition, the small value of the prefactor, $1.6\cdot10^{-44}\,\mathrm{Wm/s}$, shows that one needs gigantic systems to produce quadrupole moment changes of changes yielding any detectable length variations $\delta l/l = h$. To be convinced, just insert a few numbers, keeping in mind that the present detectors are able to measure length changes of $\delta l/l = 10^{-19}$. Challenge

Gravity waves, like all other waves, transport energy.* Specializing the general formula for the emitted power P to the case of two masses m_1 and m_2 in circular orbits around each other at distance l one gets Ref. 2

$$P = -\frac{dE}{dt} = \frac{G}{45c^5}\,\dddot{Q}^{\text{ret}}_{ab}\,\dddot{Q}^{\text{ret}}_{ab} = \frac{32}{5}\frac{G}{c^5}\left(\frac{m_1m_2}{m_1+m_2}\right)^2 l^4\omega^6 \tag{203}$$

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

$$P = \frac{32}{5}\frac{G^4}{c^5}\frac{(m_1m_2)^2(m_1+m_2)}{l^5} \tag{204}$$

For elliptical orbits, the rate is higher.**

Inserting the values in the case of the earth and the sun, one gets a power of about 200 W, and a value of 400 W for the Jupiter-sun system. These values are so small that their effect cannot be detected at all.

The *frequency* of the waves is twice the orbital frequency, as you might want to check. Challenge

As a result, the only observation of effects gravitational waves to date is in binary pulsars. Pulsars are extremely small stars; even with a mass equal to that of the sun, their size is only about 10 km; thus they can orbit each other at very small distances and at high speeds. Indeed, in the most famous binary pulsar system, PSR 1913+16, the two stars orbit each other in an amazing 7.8 h, their semimajor axis is about 700 Mm, less than twice the earth-moon distance. Since their orbital speed is up to 400 km/s, the system is noticeably relativistic.

In another gauge, a plane wave can be written as

$$g = \begin{pmatrix} c^2(1+2\varphi) & A_1 & A_2 & A_3 \\ A_1 & -1+2\varphi & h_{xy} & 0 \\ A_2 & h_{xy} & -1+h_{xx} & 0 \\ A_3 & 0 & 0 & -1 \end{pmatrix} \tag{201}$$

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

* Gravitomagnetism and gravitoelectricity, as in electrodynamics, allow to define a gravitational Poynting vector. It is as easy to work with as in the electrodynamic case. Ref. 33

** See e.g. the explanation by Goenner. Ref. 2

Pulsars have a useful property: do to their rotation, they emit extremely regular radio pulses (hence their name), often in millisecond periods. Therefore it is easy to follow their orbit by measuring the change of pulse arrival time. In a famous experiment, a team of astrophysicists around Joseph Taylor* measured the speed decrease of the binary pulsar Ref. 35 system just mentioned. Eliminating all other effects, and collecting data for 20 years, they found a slowing down of the orbital frequency shown in figure 92. The slow down is due to gravity wave emission. The results exactly fits the prediction by general relativity, *without any adjustable parameter*. (You might want to check that the effect must be quadratic in Challenge time.) This is the only case so far that general relativity has See page 219 been tested up to the $(v/c)^5$ precision. To get an idea of the precision, this measurement detected a reduction of the orbit Ref. 35 diameter of 3.1 mm per orbit, or 3.5 m per year! The measurements were possible only because the two stars in this system are pulsars with small size, large velocities, and purely gravitational interactions.

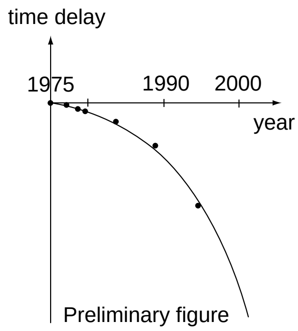


Figure 91 Comparison between measured time delay in the periastron of the binary pulsar PSR 1913+16 and the prediction due to energy loss by gravitational radiation

The *direct* detection of gravitational waves is one of the holy grails of experimental general relativity. The race is on since the 1990s. The basic idea is simple and taken from figure 90: one takes four bodies, the line connecting one pair being perpendicular to the line connecting the other pair. One then measures the distance changes of each pair. If a gravitational wave comes by, one pair will increase in distance and the other will decrease, at the *same* time.

Figure to be included

Figure 92 Detection of gravitational waves

Since gravitational waves cannot be produced in sufficient strength by humans, for detection one first of all needs a lot of time to wait for a strong enough wave to come by. Secondly, one needs a system able to detect length changes of the order of 10^{-22} or better – in other words, a lot of money. Any detection for sure will make the news in television.**

It turns out that even for a body around a black hole, only about 6% of the rest mass can be radiated away; in particular, most of the energy is radiated during the final fall into the black hole, so that only quite violent processes, such as black hole collisions, are good candidates for detectable gravity wave sources.

Actually, gravity waves are even more interesting if, instead of the linear approximation See page 243 described here, the full field equations are used. More about the topic shortly.

∗ He later shared the Nobel prize in physics for his life's work.

Ref. 36 ∗∗ The topic of gravity waves is full of interesting sidelines. For example, can gravity waves be used to power Challenge a rocket? Yes, say Bonnor and Piper. You might ponder the possibility yourself.

Light and radio wave bending

As we now from above, gravity also influences the motion of light. A far away observer measures different values for the light speed near a mass. It turns out that a far away observer measures a lower speed, so that for him, gravity has the same effects as a dense medium. It takes only a little bit of guessing that this effect will *increase* the bending of light already found in universal gravity.

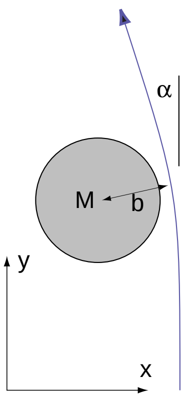


Figure 93 Calculating the bending of light by a mass

To calculate the effect, a simple way is the following. As usual, we use the coordinate system of flat space-time at infinity. The idea is to do all calculations to first order, as the value of the bending is very small. The angle of deflection α, to first order, is simply Ref. 37

$$\alpha = \int_{-\infty}^{\infty} \frac{\partial c}{\partial x} dy \tag{205}$$

The next step is to use the Schwarzschild metric

$$di^2 = (1 - \frac{2GM}{rc^2})dt^2 - \frac{dr^2}{(1 - \frac{2GM}{rc^2})} - r^2 d\varphi^2 \tag{206}$$

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

$$di^2 = (1 - \frac{2GM}{rc^2})dt^2 - (1 + \frac{2GM}{rc^2})(dx^2 + dy^2) \tag{207}$$

which to first order leads to

$$\frac{\partial c}{\partial x} = (1 - \frac{2GM}{rc^2}) \quad . \tag{208}$$

As we know already, far away observers see light slowed down when passing near a mass. One can also call the right hand side a height dependent index of refraction. Constant local light speed leads to a global slowdown. This effect will be discussed shortly.

Inserting this result above and using a smart substitution, one gets Challenge

$$\alpha = \frac{4GM}{c^2} \frac{1}{b} \tag{209}$$

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

This value is *twice* the result we found for universal gravity. For the sun, one gets the famous value of 1.75″, which was confirmed by the measurements of 1919. They made Einstein famous, as they showed that universal gravity is wrong. In fact, Einstein was lucky. The earlier expeditions organized to measure the value had failed. In 1912, it was impossible to take data because of rain, and in 1914 in Crimea, scientists were arrested (by mistake) as spies, due to the beginning of the world war. In 1911, Einstein had published an *incorrect* calculation, giving only the Soldner value; only in 1915, when he completed general relativity, he found the correct result. Therefore Einstein became famous only because the first two expeditions failed.

See page 84 Ref. 38 See page 84

For high precision experiments around the sun, it is better to measure the bending of radio waves, as they encounter fewer problems when they propagate through the corona. Over a dozen independent experiments, using radio sources in the sky which are on the path of the sun, confirmed the prediction within a few percent. Ref. 21, 2

Of course, the bending of light also confirms that in a triangle, the sum of the angles does not add up to π, as was predicted above for curved space. (What is the sign of the curvature?) One thus follows that if light would not be bent near a mass, it would go faster than c! See page 233 Challenge

In the meantime, bending of radiation has also observed near Jupiter, near certain stars, near several galaxies, and near galaxy clusters. For the earth, the angle is at most 3 nrad, too small to be measured, even though this may change in the near future. There is a chance to detect this value if, as Andrew Gould proposes, the data of the satellite Hipparcos, which takes precision pictures of the night sky, is analyzed properly in the future. See page 266

Time delay

The above calculation shows that for a distant observer, light is slowed down near a mass. Constant *local* light speed leads to a *global* light speed slowdown. If light were not slowed down near a mass, it would go faster than c for an observer near the mass! In 1964, I.I. Shapiro had the idea to measure this effect. He proposed two methods. The first was to send radar pulses to Venus, and measure the time for the reflection to get back to earth. If the signals pass near the sun, they will be delayed. The second was to use an artificial satellite communicating with earth. Ref. 39

The first measurement was published in 1968, and confirmed the prediction by general relativity within experimental errors. All further tests up to the present have confirmed the prediction within experimental uncertainties, which are of the order of one part in a 1000. The delay has even been measured in binary pulsars, as there are a few systems in which the line of sight lies almost precisely in the orbital plane. Ref. 40 Ref. 41

Orbits

Astronomy allows the most precise measurements of motion, so that Einstein first of all applied his results to the motion of planets. He later said that the moment he found out that his calculation for the precession of Mercury matched observations was one of the happiest of his life.

The calculation is not difficult. In universal gravity, orbits are calculated by setting $a_{\text{grav}} = a_{\text{centri}}$, in other words, by setting $GM/r^2 = \omega^2 r$ and fixing energy and angular momentum. In general relativity the space curvature needs to be included.

It is straightforward to use the Schwarzschild metric mentioned above to deduce that the initial condition for the energy E, together with its conservation, leads to a relation between proper time and time at infinity: See page 214

$$\left(\frac{dt}{d\tau}\right) = \frac{E}{1 - 2M/r} \tag{210}$$

whereas the initial condition on the angular momentum J, which is also conserved, means that

$$\left(\frac{d\varphi}{d\tau}\right) = \frac{J}{r^2} \quad . \tag{211}$$

Inserting all this into the Schwarzschild metric, one gets that the motion of a particle follows

$$(\frac{dr}{d\tau})^2 + V^2(J,r) = E^2 \tag{212}$$

where the effective potential is given by

$$V^2(J,r) = (1 - \frac{2M}{r})(1 + \frac{J^2}{r^2}) \quad . \tag{213}$$

One only needs to solve for $r(\varphi)$. For *circular* orbits one gets Challenge

$$r = \frac{J(J \pm \sqrt{J^2 - 12M^2})}{2M} \tag{214}$$

where the higher value gives a stable orbit, and the lower one an unstable orbit. If $J < 2\sqrt{3}M$, no stable orbit exists; the object will impact the surface or, for a black hole, be swallowed. There is a stable circular orbit *only* if the angular momentum J is larger than $2\sqrt{3}M$.

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

$$u'' + u = \frac{GM}{J^2} + \frac{3GM}{c^2}u^2 \quad . \tag{215}$$

Without the nonlinear correction on the far right, the solutions are the famous conical sections

$$u_o(\varphi) = \frac{GM}{J^2}(1 + \varepsilon\cos\varphi) \tag{216}$$

which describe ellipses, parabolas, and hyperbolas, depending on the value of the parameter ε. General relativity introduces the nonlinear term in (215). The solutions are not conical sections any more; however, as the correction is small, a good approximation is given by

$$u_o(\varphi) = \ldots \tag{217}$$

which give the famous rosetta path. Such a path is characterized by a periastron shift.

– CS – a few lines to be inserted, sorry – CS –

The effect has been measured for the orbits of Mercury, Icarus, Venus, Mars around the sun, and for several binary star systems. In all cases, it describes the motion within experimental errors. For Mercury, the value is 43″ per century. This was the only effect unexplained by universal gravity in Einstein's time; when he found exactly that value in his calculation, he was overflowing with joy.

In fact, *no* orbit is really stable in general relativity, because of the emission of gravitational waves. However, in the solar system, the power lost this way is is completely negligible even over thousands of millions of years, as we saw above.

The geodesic effect

When a pointed body orbits a central mass m at distance r, the direction of the tip will not be the same after a full orbit. The angle α describing the direction change is given by

$$\alpha = 2\pi 1 - \sqrt{1-\frac{3Gm}{rc^2}}\,) \approx \frac{3\pi Gm}{rc^2} \qquad (218)$$

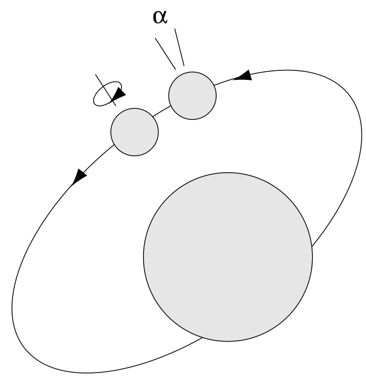


Figure 94 The geodesic effect

The result, called the *geodesic effect*, is also a consequence of the split in gravitoelectric and gravitomagnetic fields. Challenge In the case that the pointing is realized by an intrinsic rotation such as a spinning satellite, the geodesic effect produces a precession of the axis. The effect works like a spin-orbit coupling in atomic theory. (There is also an effect analogous to spin-spin coupling, but it is much smaller, and has not been detected yet.) The geodesic effect was predicted by de Sitter in 1916; it was first detected in 1988 through a combination of radiointerferometry and lunar ranging, making Ref. 42 use of the cat-eyes deposed by the cosmonauts on the moon.

We now return to the general case of relativistic motion, not limited to weak gravity, where curvature cannot be neglected.

How is curvature measured?

We saw that in the precise description of gravity, motion depends on space-time curvature. In order to add numbers to this idea, we first of all need to describe curvature itself as accurately as possible. To clarify the issue, we will start the discussion in two dimensions, and then go over to three and four dimensions.

Obviously, a flat sheet of paper has no curvature. If one rolls it into a cone or a cylinder, it gets what is called *extrinsic curvature*; however, the sheet of paper still looks flat for any two-dimensional animal living on it – as approximated by an ant walking over it. In other words, the *intrinsic curvature* of the sheet of paper is zero even if the sheet as a whole is extrinsically curved. (Can a one-dimensional space have intrinsic curvature? What about a Challenge torus?)

Intrinsic curvature is thus the stronger concept, measuring the curvature which can be observed even by an ant. The surface of the earth, the surface of an island, or the slopes of a mountain are intrinsically curved. Whenever one talks about curvature in general relativity, one always means *intrinsic* curvature, since any observer in nature is by definition in the same situation as an ant on a surface: their experience, acts and plans always only concern their closest neighbourhood in space and time.

But how precisely can an ant determine whether it lives on an intrinsically curved surface?* One way is shown in figure 96. The ant can check whether either the circumference of a circle or its area fits with the measured radius. She can even use the difference between the two numbers as a measure for the local intrinsic curvature, if she takes the limit for vanishingly small circles and if she normalizes the values correctly. In other words, the ant can

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

imagine to cut out a little disk around the point she is on, to iron it flat, and to check whether the disk would tear or produce folds. Any a two-dimensional surface is intrinsically curved whenever ironing is not sufficient to make a street map out of it.

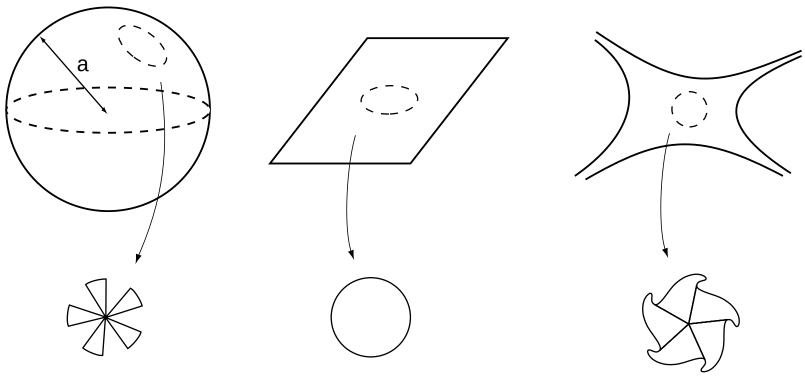


Figure 95 Positive, vanishing and negative curvature in two dimensions

This means that one can recognize intrinsic curvature also by checking whether two parallel lines stay parallel, approach each other, or depart from each other. In the first case, such as lines on a paper cylinder, the surface is said to have *vanishing* intrinsic curvature; a surface with approaching parallels, such as the earth, is said to have *positive* curvature, and a surface with diverging parallels, such as a saddle, is said to have *negative* curvature. In short, positive curvature means that one is locked in, negative that one is locked out. You might want to check this with figure 95.

Let us see how to quantify these ideas. First a question of vocabulary: a sphere with radius a is said, by definition, to have an intrinsic curvature $K = 1/a^2$. Therefore, a plane has vanishing curvature. You might check that for a circle on a sphere, the measured radius r, circumference C, and area A are related by Ref. 26 Challenge

$$C = 2\pi r(1 - \frac{K}{6}r^2 + ...) \quad \text{and} \quad A = \pi r^2(1 - \frac{K}{12}r^2 + ...) \tag{219}$$

where the dots imply higher order terms. This allows to define the intrinsic curvature K, also called the *gaussian* curvature, for a point in two dimensions in either of the following two equivalent ways:

$$K = 6 \lim_{r\to 0}(1 - \frac{C}{2\pi r})\frac{1}{r^2} \quad \text{or} \quad K = 12 \lim_{r\to 0}(1 - \frac{A}{\pi r^2})\frac{1}{r^2} \quad . \tag{220}$$

This expression allows a bug to measure the intrinsic curvature at each point for any smooth surface.* From now on in this text, *curvature* will always mean *intrinsic* curvature. Note that the curvature can be different from place to place, and that it can be positive, like for an egg, or negative as at the inside of any handle. Also a saddle is an example for the latter case, but, like the handle, with a curvature *changing* from point to point. In fact, it is not possible at all to fit a surface of *constant* negative curvature inside three-dimensional space; one needs at least four dimensions, as you can find out if you try to imagine the situation. Challenge

∗ If the n-dimensional volume of a sphere is written as $V_n = C_n r^n$ and the n-dimensional surface as $O_n = nC_n r^{n-1}$, one can generalize the expressions to Ref. 43

$$K = 3n \lim_{r\to 0}(1 - \frac{O_n}{nC_n r^{n-1}})\frac{1}{r^2} \quad \text{or} \quad K = 3(n+2) \lim_{r\to 0}(1 - \frac{V_n}{C_n r^n})\frac{1}{r^2} \quad . \tag{221}$$

Challenge as shown by Vermeil. A famous riddle is to determine C_n.

Note that for any surface, at any point, the direction of maximum curvature and the direction of minimum curvature are always *perpendicular* to each other. This fact was discovered by Leonhard Euler. You might want to check this with a torus, or with a sculpture by Henry Moore, or with any other curved object from your surroundings, such as a Volkswagen beetle. The gaussian curvature is in fact the product of the two corresponding inverse curvature radii. Thus, even though line curvature is not intrinsic, this special product is. Physicists are thus particularly interested in gaussian curvature – and its higher-dimensional analogs.

Challenge

For *three*-dimensional objects, the issue is a bit more involved. Obviously, we have difficulties imagining the situation. But we can still "see" that the curvature of a small disk around a point will depend on its orientation. But let us first look at the simplest case. If the curvature at a point is the same in all directions, the point is called *isotropic*. One then can imagine a small sphere around that point. In this special case, in three dimensions, the relation between the measured radius and the measured sphere surface A leads to define the curvature

Challenge

$$K = 9 \lim_{r\to 0}\left(1-\frac{A}{4\pi r^2}\right)\frac{1}{r^2} = 18 \lim_{r\to 0}\frac{r-\sqrt{A/4\pi}}{r^3} = 18 \lim_{r\to 0}\frac{r_{\text{excess}}}{r^3} \quad . \tag{222}$$

Defining the *excess radius* as $r-\sqrt{A/4\pi}$, one gets that for a three-dimensional space, *the curvature is eighteen times the excess radius of a small sphere divided by the cube of its radius*.A positive curvature is equivalent to a positive excess radius, and similarly for vanishing and negative cases.

Of course, this value is only an average. The precise way requires to define curvature with disks; these values will *differ* from the values calculated by using the sphere, as they will depend on the *orientation* of the disk. However, all possible disk curvatures at a given point are related among each other and must form a tensor. (Why?) For a full description of curvature, one thus has to specify, as for any tensor in three dimensions, the main curvatures in three orthogonal directions.*

Challenge

What are the numbers in practice? Already in 1827, the mathematician and physicist Friedrich Gauss checked whether the three angles between three mountain peaks added up to π. Nowadays we know that the deviation δ from the angle π on the surface of a body of mass M and radius r is given by

$$\delta = \pi - (\alpha+\beta+\gamma) \approx A_{\text{triangle}}\frac{GM}{r^3c^2} \quad . \tag{223}$$

For the case of the earth and typical mountain distances, one gets an angle of the order of 10^{-14} rad. Gauss had no chance to detect any deviation, and in fact he didn't. But Gauss did not know, as we do today, that gravity and curvature go hand in hand. Even with lasers and high precision set-ups, no deviation has been detected yet – on earth. The right-hand factor, which will turn out to be the curvature of the earth, is too small.

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

Curvature and space-time

In nature, with *four* space-time dimensions, the situation requires a more involved approach. First of all, the use of space-time coordinates automatically introduces the speed of light c as limit speed, which is the main requirement leading to general relativity. Furthermore, the number of dimensions being four, we expect a value for an average curvature at a point, defined by comparing the 4-volume of a 4-sphere in space-time and with the one deduced from the measured radius; then we expect a set of 'almost average' curvatures defined by 3-volumes of 3-spheres in various orientations, plus a set of 'low-level' curvatures defined by usual 2-areas of usual 2-disks in even more orientations. Obviously, we need to bring some order in this set, and we need to avoid the double counting we already encountered in the case of three dimensions.

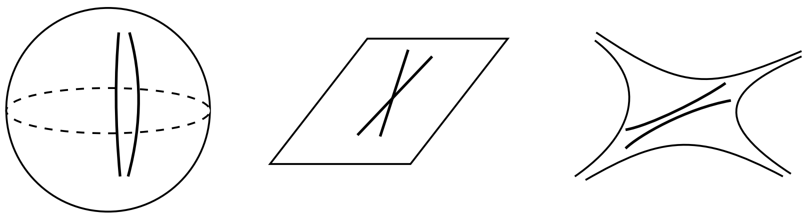


Figure 96 Curvature (in two dimensions) and geodesic behaviour

Fortunately, physics can help to make the mathematics easier. First of all, however, we need to define what we mean by curvature of space-time. Then we will define curvatures for disks of various orientations. To achieve this, we translate the definition of curvature into another picture, which allows to generalize it to time as well. Figure 96 shows that the curvature K also describes how to geodesics *diverge*. Geodesics are the straightest paths on a surface, i.e. those paths that a tiny car or tricycle would follow if it drives on the surface keeping the steering wheel straight.

If a space is curved, the separation s will increase along the geodesics as Challenge

$$\frac{d^2s}{dl^2} = -Ks + \quad \text{higher orders} \tag{224}$$

where l measures the length along the geodesic, and as above K is the inverse square curvature radius.

In space-time, this is extended by substituting proper length with proper time (times the speed of light). Thus the curvature becomes:

$$\frac{d^2s}{d\tau^2} = -Kc^2s + \quad \text{higher orders} \tag{225}$$

This turns out to be the definition of an acceleration. In short, what in the purely spatial case is described by *curvature*, in the case of space-time becomes the *relative acceleration* of two particles freely falling from the same point. But we encountered these accelerations already, in the beginning of our escalation: there they described what we called tidal effects. See page 82 In short, space-time curvature and tidal effects are precisely the same.

Obviously, the value of tidal effects and thus of curvature will depend on the orientation – more precisely on the orientation of the space-time plane formed by the two particle velocities. But the definition also shows that K is a tensor, and that later on we will have to add indices to it. (How many?) The fun is that we can avoid indices for a while by looking at a special combination of spatial curvatures. If we take three planes in space, all orthogonal to each other and through the same point, the *sum* of the three curvature values does *not* depend on the observer. (This corresponds to the tensor trace.) Can you show this, by using the definition of the curvature just given? Challenge Ref. 27

The sum of all three such curvatures defined for mutually orthogonal planes, called *sectional curvatures* in this context and written $K_{(12)}$, $K_{(23)}$, and $K_{(31)}$, is related to the excess radius defined above. Can you find out how? Challenge

In summary, curvature is not such a difficult concept. It describes the *deformation* of space-time. If one imagines space (-time) as a big blob of rubber in which we live, the curvature at a point describes how this blob is squeezed at that point. Since we live *inside* the rubber, we need to use "insider" methods, such as excess radii and sectional curvatures, to describe the deformation. Relativity is only difficult to learn because people often do not like to think about the vacuum in this way, and even less to explain it in this way. (For hundred years it was a question of faith for every physicist to say that the vacuum is empty.) But that is slowly changing.

Curvature and motion in general relativity

As mentioned above, one half of general relativity says that any object moves along paths of *maximum* proper time, i.e. along geodesics. All of the *other* half is contained in a single expression: The sum of all three *proper* sectional *spatial* curvatures at a point is given by

$$K_{(12)} + K_{(23)} + K_{(31)} = \frac{8\pi G}{c^4} W^{(0)} \tag{226}$$

where $W^{(0)}$ is the *proper* energy density at the point, *and* this is valid for *every* observer. That is general relativity in one paragraph.

Challenge An equivalent way to describe the expression is easily found using the excess radius defined above, and introducing the mass M by $M = W^{(0)}/c^2$. One gets

$$-r_{\text{excess}} = \sqrt{A/4\pi} - r = \frac{G}{3c^2} M \quad . \tag{227}$$

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

Note that the expression means that the average space curvature at a point in empty space *vanishes*. As we will see shortly, this means that near a spherical mass the curvature *towards* the mass and twice the curvature *around* the mass exactly compensate each other.

Curvature will also differ from point to point. In particular, the expression implies that if energy *moves*, curvature will move with it. In short, space curvature, and, as we will see shortly, also space-time curvature, *changes* over space and time.

The quantities appearing in expression (226) are *independent* of the observer. But often people want to use observer-dependent quantities. The relation then gets more involved; the single equation (226) must be expanded to ten equations, called *Einstein's field equations*. They will be introduced below. But before we do that, we check that general relativity makes sense. We skip the check that it contains special relativity as limiting case, and directly go to the main test.

Universal gravity

The only reason which keeps me here is gravity.
Anonymous

For small velocities, one finds that the temporal curvatures (224) can be defined as the second spatial derivatives of a single scalar function φ via Challenge

$$K_{(0j)} = \frac{\partial^2 \varphi}{\partial (x^j)^2} \tag{228}$$

Universal gravity is the description for small speeds and small spatial curvature. Both limits imply, taking $W^{(0)} = \rho c^2$ and using $c \longrightarrow \infty$, that

$$K_{(ij)} = 0 \quad \text{and} \quad K_{(01)} + \mathbf{K}_{(02)} + \mathbf{K}_{(03)} = 4\pi G\rho \quad . \tag{229}$$

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

Can you show that expression (226) indeed means that time near a mass depends on the height, as stated in the beginning of this chapter? Challenge

The Schwarzschild metric

What is the curvature of space-time near a spherical mass? Ref. 27

– CS – to be inserted – CS –

The curvature of the Schwarzschild metric is given by

$$\begin{aligned} K_{r\varphi} = K_{r\theta} = -\frac{G}{c^2}\frac{M}{r^3} \quad &\text{and} \quad K_{\theta\varphi} = 2\frac{G}{c^2}\frac{M}{r^3} \\ K_{t\varphi} = K_{t\theta} = \frac{G}{c^2}\frac{M}{r^3} \quad &\text{and} \quad K_{tr} = -2\frac{G}{c^2}\frac{M}{r^3} \end{aligned} \tag{230}$$

everywhere. The dependence on $1/r^3$ follows from the dependence for all tidal effects we calculated in the chapter on universal gravity. The factors c^2 follow from space-time, and only the numerical prefactors need to be calculated from general relativity. As expected, the curvature on the surface of the earth is exceedingly small. Ref. 27 See page 82 Challenge

All observers: heavier mathematics

Jeder Schuljunge in den Straßen unseres mathematischen Göttingen versteht mehr von vierdimensionaler Geometrie als Einstein. Trotzdem vollbrachte Einstein das Werk, und nicht die Mathematiker.
David Hilbert*

* Every schoolboy in the streets of our mathematical Göttingen knows more about four-dimensional geometry than Einstein. Nevertheless, it was Einstein who did the work, not the mathematicians.

Now that we have a feeling for curvature, we want to describe it in a way that allows *any* observer to talk to any *other* observer.* Unfortunately, this means to use formulas with tensors. These formulas look exactly the way that non-scientists imagine: daunting. The challenge is to be able to see in each of the expressions the essential point (e.g. by forgetting all indices for a while) and not to be impressed by those small letters sprinkled all over them.

The curvature of space-time

We mentioned above that a 4-dimensional space-time is described by 2-curvature, 3-curvature, and 4-curvature. Many texts on general relativity start with 3-curvature. These curvatures describing the distinction between the 3-volume calculated from a radius and the actual 3-volume. They are described by the *Ricci tensor*. With an argument we encountered already for the case of geodesic deviation, it turns out that the Ricci tensor describes how the shape of a spherical cloud of freely falling particles is deformed on its path.

– CS – a bit more in the next version – CS –

In short, the Ricci tensor is the general relativistic version of $\Delta\varphi$, or better, of $\Box\varphi$.

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

$$R = -2K = -\frac{2}{r^2_{\text{curvature}}} \qquad (231)$$

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

$$R = R^{\lambda}{}_{\lambda} = g^{\lambda\mu} R_{\lambda\mu} \qquad (232)$$

The Ricci scalar, describing the curvature averaged over space *and* time, always vanishes in vacuum. This allows for example, on the surface of the earth, to relate the spatial curvatures and the changes of time with height. Challenge

Now comes one of the issues discovered by Einstein in two years of hard work. The quantity of importance for the description of curvature in nature is not the Ricci tensor R, but a tensor built from it. This *Einstein tensor* G is defined mathematically as

$$G_{ab} = R_{ab} - \frac{1}{2} g_{ab} R \qquad (233)$$

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

$$G_{00} = K_{(12)} + K_{(23)} + K_{(31)} \qquad (234)$$

* This section might be skipped at first reading; the part on cosmology, on page 247, then is the right point to continue the escalation.

Similarly, the diagonal elements G_{ii} are the sum (taking into consideration the minus signs of the metric) of sectional curvatures in the planes *orthogonal* to the i direction. For example, one has:

$$G_{11} = K_{(02)} + K_{(03)} - K_{(23)} \quad . \tag{235}$$

The other components are defined accordingly. The distinction between the Ricci tensor and the Einstein tensor is thus the way in which the sectional curvatures are combined: disks *containing* the coordinate in question in one case, disks *orthogonal* to the coordinate in the other case. Both describe the curvature of space-time equally, and fixing one means fixing the other. (What is the trace of the Einstein tensor?) Challenge

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

The description of momentum, mass and energy

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

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

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

$$d\mathbf{p} = T\ d\mathbf{A} \quad . \tag{236}$$

The surface is assumed to be characterized by its normal vector $d\mathbf{A}$. Since the energy-momentum density is a proportionality factor between two vectors, T is a *tensor*. Of course, we are talking about 4-flows and 4-surfaces here. Thus the tensor can be split in the following way:

$$T = \left(\begin{array}{c|ccc} w & S_1 & S_2 & S_3 \\ \hline S_1 & t_{11} & t_{12} & t_{13} \\ S_2 & t_{21} & t_{22} & t_{23} \\ S_3 & t_{31} & t_{32} & t_{33} \end{array}\right) = \left(\begin{array}{c|c} \text{energy} & \text{energy flow density, or} \\ \text{density} & \text{momentum density} \\ \hline \text{energy flow or} & \text{momentum} \\ \text{momentum density} & \text{flow density} \end{array}\right) \tag{237}$$

where $w = T_{oo}$ is a 3-scalar, S a 3-vector, and t a 3-tensor. The total quantity T, the so called *energy-momentum tensor*, has two essential properties. It is symmetric, and its divergence vanishes.

The vanishing divergence, often written as

$$\partial_a T^{ab} = 0 \quad \text{or abbreviated} \quad T^{ab}{}_{,a} = 0 \tag{238}$$

expresses that the tensor describes a *conserved* quantity. In every volume, energy can change only through flow through its boundary. Can you confirm that the description of energy-momentum with this tensor follows the requirement that any two observers, differing by position, orientation, speed *and* acceleration, can communicate their results to each other?

Challenge

The energy-momentum tensor gives a full description of the distribution of energy, momentum, and mass over space and time. As an example, let us determine the energy-momentum density for a moving liquid. For a liquid of density ρ, a pressure p and a 4-velocity **u**, one has

$$T^{ab} = (\rho_o + p) u^a u^b - p g^{ab} \quad . \tag{239}$$

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

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

$$p = p(\rho) \tag{241}$$

which is a material property and thus cannot be determined from relativity. It has to be derived from the constituents of matter or radiation and their interactions. The simplest possible case is *dust*, i.e. matter made of point particles with no interactions at all among them. Its energy-momentum tensor is given by

$$T^{ab} = \rho_o u^a u^b \tag{242}$$

Challenge

Can you explain the difference to the liquid case?

Challenge

The divergence of the energy-momentum vanishes, as you may want to check. This property is the same as for the Einstein tensor presented above. But before we elaborate on the issue, a short remark. How does one take count of *gravitational energy*? It turns out that gravitational energy cannot be defined in general. Gravity is *not* an interaction, and does *not* have an associated energy.**

The symmetry of general relativity

– CS – to be written – CS –

* In the comoving frame one thus has

$$T^{ab} = \begin{pmatrix} \rho_o & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \tag{240}$$

** In certain special circumstances, such as weak fields, slow motion, or an asymptotically space-time, one can define the integral over the G^{00} component off the Einstein tensor as negative gravitational energy. This leads to the famous speculation that the total energy of the universe is zero. Do you agree?

Challenge

Mass and ADM

The diffeomorphism invariance of general relativity makes life quite annoying. We will see that it allows to say that we live on the *inside* of a hollow sphere, and that it does not allow to say where energy actually is located. If energy cannot be located, what about mass? It became clear that mass, or energy, can be localized *only* if space-time far away from it is known to be flat. It is then possible to define a localized mass value by the following intuitive idea: the mass is measured by the time a probe takes to orbit the unknown body.

This definition was formalized by Arnowitt, Deser, and Misner, and is since then often called the *ADM mass*. Obviously, this approach *requires* flat space-time at infinity, and cannot be extended to other situations. In short, *mass* is defined only for asymptotically flat space-time. Challenge

Now that we can go on talking about mass without (too much) a bad conscience, we turn to the equations of motion. Ref. 80

Hilbert's lagrangian

When Einstein discussed his work with David Hilbert, Hilbert found a way to do in a few weeks what Einstein had done in years. Hilbert understood that general relativity *in empty space* was described by the lagrangian

$$S = \frac{c^3}{16\pi G}\int R + 2\Lambda dV \quad . \qquad (243)$$

There were few other choices possible, as the Ricci scalar R and the cosmological constant Λ were the only observer invariant quantities appearing in the description of motion. He then deduced the field equations by the usual variational method method.

– CS – to be written – CS –

Einstein's field equations

[Einstein's general theory of relativity]
cloaked the ghastly appearance of atheism.
A US-american sorceror

Do you believe in god? Prepaid reply 50 words.
Subsequent telegram by a competing
sorceror to his hero Albert Einstein

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

At the basis of all these worries were the famous field equations. They contain the full description of general relativity and are simply given by

$$G^a{}_b = \kappa T^a{}_b + \Lambda \delta^a{}_b \tag{244}$$

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

$$\kappa = \frac{8\pi G}{c^4} = 2.1 \cdot 10^{-43}\, 1/\mathrm{N} \tag{245}$$

and its small value reflects the weakness of gravity in everyday life, or better, the difficulty to bend space-time. The constant Λ, the so-called *cosmological constant*, corresponds to a vacuum energy volume density Λ/κ and is harder to measure. The presently favoured value

See page 256 is

$$\Lambda \approx 10^{-52}/\mathrm{m}^2 \quad \text{or} \quad \Lambda/\kappa \approx 0.5\,\mathrm{nJ/m}^3 \quad . \tag{246}$$

In other words, the field equations state that the curvature at a point is equal to the flow of energy-momentum through that point, taking into account the vacuum energy density. In short, *energy-momentum tells space-time how to curve.**

The field equations of general relativity can be simplified for the case that speeds are small. In that case $T_{oo} = \rho c^2$ and all other components of T vanish. Using the definition of

Challenge

κ and setting $\varphi = (c^2/2) h_{oo}$ in $g_{ab} = \eta_{ab} + h_{ab}$, one finds

$$\nabla^2 \varphi = 4\pi\rho \quad \text{and} \quad \frac{d^2x}{dt^2} = -\nabla\varphi \tag{247}$$

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

$$a = G \frac{M}{r^2} \tag{248}$$

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

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

- *Principle of general covariance*: the equations of physics must be stated in tensorial form; even though Ref. 29 it is known today that all equations can be written with tensors, even universal gravity, in many cases they require unphysical "absolute" elements, i.e. quantities which affect others but are not affected themselves. This See page 408 unphysical idea is in contrast with the idea of *inter*action, as explained above.

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

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

- *Mach's principle*: inertia is due to the interaction with the rest of the universe; this principle is correct, even though it is often maintained that it is not fulfilled in general relativity. In any case, it is not the essence of See page 271 general relativity.

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

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

a value which is independent of the mass *m* of the falling body. And indeed, as noted already by Galileo, all bodies fall with the same acceleration, independently of their size, their mass, their colour, etc. Also in general relativity, gravitation is completely democratic.*

To get a feeling for the complete field equations, we have a short walk through their main properties.

First of all, all motion due to space-time curvature is *reversible*, *differentiable* and thus *deterministic*. Note that only the complete motion, of space-time *and* matter *and* energy, has these properties. For particle motion only, motion is in fact *irreversible*, as in most examples of motion, some gravitational radiation is emitted. Challenge

By contracting the field equations one finds for the Ricci scalar the expression

$$R = -\kappa T + \ldots\Lambda \qquad (253)$$

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

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

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

What can we do of interest with these equations? In fact, to be honest, not much that we have not done already. Very few processes need the full equations. Many textbooks on relativity even stop after writing them down!

But studying them is worthwhile. For example, one can show that the Schwarzschild solution is the *only* spherically symmetric solution. Similarly, in 1923, Birkhoff showed that every rotationally symmetric vacuum solution is static. That is the case even if masses themselves move, as for example during the collapse of a star.

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

$$\nabla_e b_a = R_{ceda} v^c v^d \qquad (249)$$

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

$$\nabla_e b^a = \nabla_e \nabla^a \varphi = R^a_{ced} v^c v^d \qquad (250)$$

which implies that

$$\begin{aligned} \Delta\varphi = \nabla_a \nabla^a \varphi &= R^a_{cad} v^c v^d \\ &= R_{cd} v^c v^d \\ &= \kappa (T_{cd} v^c v^d - T/2) \end{aligned} \qquad (251)$$

Introducing $T_{ab} = \rho v_a v_b$ one gets

$$\Delta\varphi = 4\pi G \rho \qquad (252)$$

as we wanted to show.

** For more mathematical details, see the famous three-women-book by Yvonne Choquet-Bruhat, Cecile DeWitt-Morette & Margaret Dillard-Bleick, *Analysis, manifolds, and physics*, North-Holland, 19▷▷.

In fact, maybe the most beautiful application of the field equations are the various *movies* made of relativistic processes. The world wide web provides several of them; they allow to see what happens when two black holes collide, what happens when an observer falls into a black hole, etc. For these movies, the field equations usually need to be solved directly, without approximations.*

– CS – more to be added – CS –

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

We end this section with a side note. Usually, the field equations are read in one sense only, by saying that energy-momentum produce curvature. One can also read them in the other way, calculating the energy-momentum needed to produce a given curvature. When one does this, one discovers that not all curved space-times are possible, as some would lead to *negative* energy (or mass) densities. They would contradict the mentioned limit on size to mass ratio for physical systems. The limit on length to mass ratios thus also restricts the range of possible curvatures of space-time.

How to calculate the shape of geodesics

The other half of general relativity states that bodies fall along geodesics. It is thus useful to be able to calculate their trajectories.** To start with, one needs to know the *shape of space-time*, that is the generalization of the shape of a two-dimensional surface. For a being living on the surface, it is usually described by the metric g_{ab}, which defines the distances between neighboring points through

$$ds^2 = dx_a\, dx^a = g_{ab}(x)\, dx^a\, dx^b \quad . \tag{254}$$

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

Challenge

$$\frac{d}{ds}\left(g_{ad}\frac{dx^d}{ds}\right) = \frac{1}{2}\frac{\partial g_{bc}}{\partial x^a}\frac{dx^b}{ds}\frac{dx^c}{ds} \tag{255}$$

as long as ds is different from zero along the path.*** All bodies in free fall follow such geodesics. We showed above that the geodesic property implies that a stone thrown in the

See page 216

∗ See for example, the ...web sites.

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

∗ ∗ ∗ This is often written as

$$\frac{d^2x^a}{d^2s} + \Gamma^a_{bc}\frac{dx^b}{ds}\frac{dx^c}{ds} = 0 \tag{256}$$

where the condition

$$g_{ab}\frac{dx^a}{ds}\frac{dx^b}{ds} = 1 \quad . \tag{257}$$

air falls back, except if it is thrown with a speed larger then the escape velocity. Expression (255) thus replaces both the expression $d^2x/dt^2 = -\nabla\varphi$ valid for falling bodies and the expression $d^2x/dt^2 = 0$ valid for freely floating bodies in special relativity.

The path does not depend on the mass or on the material of the body. Therefore also *antimatter* falls along geodesics. In other words, antimatter and matter do not repel; they also attract each other. Interestingly, even experiments performed with normal matter can show this, if they are carefully evaluated. Are you able to find out why? Ref. 45 Challenge

For completion, we mention that light follows *lightlike* or *null geodesics*, an affine parameter u exists, and the geodesics follow

$$\frac{d^2x^a}{d^2u} + \Gamma^a_{bc}\frac{dx^b}{du}\frac{dx^c}{du} = 0 \tag{259}$$

with the different condition

$$g_{ab}\frac{dx^a}{du}\frac{dx^b}{du} = 0 \quad . \tag{260}$$

Given all these definitions of various types of geodesics, what are the lines drawn in figure 86 on page 213? Challenge

Is gravity an interaction? – again

Another way to look at the issue is the following. For satellite orbiting Jupiter, the energy-momentum $\mathbf{p}$ is as usual, defined as $\mathbf{p} = m\mathbf{u}$. If one calculates the energy-momentum change along its path s, one gets Challenge

$$\frac{d\mathbf{p}}{ds} = m\frac{d\mathbf{u}}{ds} = m(\mathbf{e}_a\frac{d\mathbf{u}^a}{ds} + \frac{d\mathbf{e}_a}{ds}\mathbf{u}^a) = m\mathbf{e}_a(\frac{d\mathbf{u}^a}{ds} + \Gamma^a_{bd}\mathbf{u}^b\mathbf{u}^c) = 0 \tag{261}$$

where $\mathbf{e}$ describes the unit vector along a coordinate axis. The energy-momentum change vanishes along any geodesic, as you might check. Therefore, the energy-momentum of this motion is conserved. In other words, no force is acting on the satellite. One could reply that in equation (261) the second term alone is the gravitational force. But the term can be made to vanish identically along any given world line. Challenge Ref. 44 Challenge

In short, nothing changes between two bodies in free fall around each other: gravity could be said not to be an interaction. The behaviour of energy confirms this argument. Of course, the conclusion is somewhat academic, as it contradicts daily life. But for the full understanding of motion, in the third part of the escalation, we will need this result. Challenge

Also the behaviour of radiation confirms this result. In vacuum, radiation is always moving freely. In a sense, one can say that radiation always is in free fall. Strangely, since we called free fall the same as rest, we should conclude that radiation always is at rest. This is

must be fulfilled. It simply requires that all the tangent vectors are *unit* vectors, and that $ds \neq 0$ all along the path. The symbols Γ appearing above turn out to be defined as

$$\Gamma^a_{bc} = \left\{ {a \atop bc} \right\} = \frac{1}{2}g^{ad}(\partial_b g_{dc} + \partial_c g_{db} - \partial_d g_{bc}) \quad , \tag{258}$$

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

not wrong! We saw that light cannot be accelerated.* We even saw that gravitational bending is not an acceleration, since light follows straight paths in space-time in this case as well. Even though light seems to slow down near masses far away observers, it always moves at the speed of light locally. In short, even gravitation doesn't manage to move light.

There is another way to show that light is always at rest. A clock for an observer trying to reach the speed of light goes slower and slower. For light, in a sense, time stops: if one prefers, *light does not move*.

Riemann gymnastics

Most books introduce curvature the hard way, namely historically.** They introduce it as history did, using the Riemann curvature tensor. This is a short summary, so that you can understand that old stuff when you get it in your hands.

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

Challenge

The *Riemann-Christoffel curvature tensor* R is then defined as a quantity allowing to calculate the curvature $K(\mathbf{p},\mathbf{q})$ for the surface spanned by $\mathbf{p}$ and $\mathbf{q}$, with area A, through

$$K(\mathbf{p},\mathbf{q}) = \frac{R\,\mathbf{pqpq}}{A^2(\mathbf{p},\mathbf{q})} = \frac{R_{abcd}\,p^a q^b p^c q^d}{(g_{\alpha\delta}g_{\beta\gamma} - g_{\alpha\gamma}g_{\beta\delta})\,p^\alpha q^\beta p^\gamma q^\delta} \qquad (262)$$

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

∗ Refraction, the slowdown of light inside matter is a consequence of light-matter interactions, and is not a counterargument, as strictly speaking, light inside matter is constantly being absorbed and reemitted. In between, it still propagates with the speed of light in vacuum. The whole process only *looks* like acceleration in the macroscopic limit. The same applies to diffraction and to reflection. A list of apparent ways to bend light can be found on page 328; details of the quantum mechanical processes at their basis can be found on page 437.

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

∗∗∗ Bernhard Riemann (1826, Breselenz–1866, Selasca)

∗∗∗∗ Above, we showed that space-time is curved by noting changes in clock rates, in meter bar lengths, and in light propagation. Such experiments most easily provide the metric g. We know that space-time is described by a four-dimensional manifold $\mathbf{M}$ with a metric g_{ab} which locally, at each space-time point, is a Minkowski metric with all its properties. Such a manifold is called a *riemannian manifold*. Only such a metric allows to define a local inertial system, i.e. a local Minkowski space-time at every space-time point. In particular, one has

$$g_{ab} = 1/g^{ab} \quad \text{and} \quad g_a{}^b = g^a{}_b = \delta^a_b \qquad (263)$$

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

$$R^a{}_{bcd} = \frac{\partial \Gamma^a{}_{bd}}{\partial x^c} - \frac{\partial \Gamma^a{}_{bc}}{\partial x^d} + \Gamma^a{}_{ec}\Gamma^e{}_{bd} - \Gamma^a{}_{fd}\Gamma^f{}_{bc} \qquad (264)$$

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

It turns out the the tensor R describes just this relative accelerations, i.e. the tidal effects. Obviously, the relative acceleration $\mathbf{b}$ increases with the separation $\mathbf{d}$ and the square (why?) of the speed $\mathbf{u}$ of the two particles. Therefore one can also define R as a (generalized) proportionality factor among these quantities: Challenge

$$\mathbf{b} = R\,\mathbf{u}\,\mathbf{u}\,\mathbf{d} \quad \text{or, more clearly} \quad b^a = R^a{}_{bcd}\, u^b\, u^c\, d^d \quad . \tag{266}$$

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

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

$$\begin{aligned} g_{ab} = {} & \eta_{ab} + \frac{1}{3} R_{acbd} x^c x^d + O(x^3) \\ = {} & \frac{1}{2} (\partial_c \partial_d g_{ab}) x^c x^d + O(x^3) \end{aligned} \tag{268}$$

The curvature tensor R is a large beast; it has $4^4 = 256$ components at each point of space-time; however, its symmetry properties reduce them to twenty independent numbers.** The

The curvature tensor is built from the second derivatives of the metric. On the other hand, one can also determine the metric if the curvature is known, using

$$g = ...R... \tag{265}$$

In other words, either the Riemann tensor R or the metric g specify the whole situation of a space-time.

∗ This second definition is also called the definition through *geodesic deviation*. It is of course not evident that it coincides with the first. For an explicit proof, see the literature. There is also a third way to picture the tensor R, a more mathematical one, namely the original way Riemann introduced it. If one parallel transports a vector $\mathbf{w}$ around a parallelogram formed by two vectors $\mathbf{u}$ and $\mathbf{v}$, each of length ε, the vector $\mathbf{w}$ is changed to $\mathbf{w} + \delta\mathbf{w}$. One then has Ref. 28

$$\delta\mathbf{w} = -\varepsilon^2 R\,\mathbf{u}\,\mathbf{v}\,\mathbf{w} + \quad \text{higher order tems} \tag{267}$$

∗∗ The second definition indeed shows that the Riemann tensor is symmetric in certain indices and antisymmetric in others: Challenge

$$R^{abcd} = R^{cdab} \quad , \quad R^{abcd} = -R^{bacd} = -R^{abdc} \tag{269}$$

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

$$R^{abcd} + R^{adbc} + R^{acdb} = 0 \quad . \tag{270}$$

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

actual number of importance in physical problems is still smaller, namely only ten. These are the components of the Ricci tensor, which can be defined with help of the Riemann tensor by contraction, i.e. by setting

$$R_{bc} = R^{a}{}_{bac} \quad . \tag{271}$$

Its components, like those of the Riemann tensor, are inverse square lengths.

9. Cosmology – motion in the universe

The soul is a spark of the substance of the stars.
Ref. 46 Heraclitos of Ephesos (ca. 540–ca. 480 BCE)

Interestingly, general relativity allows to explain many of the general properties of the universe, as well as many details of its history. To see how, we start with a short overview of the data collected by modern astronomy.* (Ref. 47)

Table 20 Some observations about the universe

Aspect	main properties	value
Phenomena		
galaxy formation	trigger event	unknown
galaxy collisions		
star formation		
novae	new bright stars	$L > \ldots$
supernovae	new bright star	$L > \ldots$
gamma ray bursts	luminosity	up to $3 \cdot 10^{47}$ W, almost equal to the whole universe
	energy	ca. 10^{46} J
	duration	ca. 0.015 – 1000 s
	obs. numbers	ca. 2 per day
optical bursts		
radio sources		
X-ray sources		
cosmic rays	energy	from 0 eV to 10^{22} eV
gravitational lensing	light bending	
comets	recurrence, evaporation	
meteorites	age	up to $4 \cdot 10^{9}$ a
Observed components		
intergalactic space		
quasars	redshift	up to 5.8
	luminosity	..., about the same as one galaxy
galaxy superclusters	number	ca. 10^{8} inside horizon
our local supercluster		with about 4000 galaxies
galaxy groups		with a dozen up to 1000 galaxies

* Many details about the universe can be found in the beautiful text by W.J. KAUFMANN & R.A. FRIEDMAN, *Universe*, fifth edition, W.H. Freeman & Co., 1999. On the remote history of the universe, see the excellent text by G. BÖRNER, *The early universe – facts & fiction*, 3rd edition, Springer Verlag, 1993.

Aspect	main properties	value
our local group		with 30 galaxies
galaxies	number	ca. 10^{11} inside horizon
	containing	typically 10^{11} stars
our galaxy	speed	600 km/s towards Hydra-Centaurus
nebulae, clouds		
our interstellar cloud	size	20 light years
	composition	atomic hydrogen at 7500 K
star systems	types	orbiting double stars, star plus dwarfs, possibly a few planetary systems
our solar system	speed	370 km/s from Acquarius towards Leo
stars		
giants and supergiants	large size	up to ...
brown, L, and T dwarfs	low temperature	
white dwarfs		
neutron stars	nuclear mass density	
pulsars	radio emission	
magnetars	high magnetic fields	
black holes	horizon	
General properties		
cosmic horizon	distance	ca. 10^{26} m ...
expansion	Hubble's constant	between 59 km s^{-1}Mpc^{-1}and 70 km s^{-1}Mpc^{-1}, or ca. … /s
vacuum energy density	cosmological constant	...
large size shape	space curvature	almost vanishing
large size shape	topology	simple in our galactic environment, unknown at large scales
dimensions	number	3 for space, 1 for time, at low and moderate energies
mass-energy	density	2 to $11 \cdot 10^{-27}$ kg/m^3 or 1 to 6 hydrogen atoms per cubic metre
baryons	density	one sixth of the previous
other	density	five sixth unknown
photons	number density	4 to $5 \cdot 10^8$ /m^3 $= 1.7$ to $2.1 \cdot 10^{-31}$ kg/m^3
neutrinos	number density	not measured
average temperature	photons	2.7 K
	matter	ca. 0 K
	neutrinos	not measured, 2 K predicted
original inhomogeneity	amplitude of radiation anisotropy	
	amplitude of matter clustering	

Research in astrophysics is directed at discovering and understanding all phenomena observed in the skies. In our escalation, we skip most of this fascinating topic, since as usual, we focus on motion. But before we continue we clarify a question of vocabulary.

What is the universe?

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

At least three definitions are possible for the term 'universe'.

- The *(visible) universe* is the totality of all observable mass and energy.
- The *(believer) universe* is the totality of all mass and energy, *including* any parts of them which are not visible. All books of general relativity state that there definitely exists matter or energy beyond the observation boundaries. We explain the origin of this idea below.
- The *(total) universe* is the sum of matter, energy *as well as* space-time itself.

These definitions are often mixed up in physical and philosophical discussions. There is *no* generally accepted consensus, so one has to be careful. In this walk, when we use the term 'universe', we imply the *last* definition only. We will discover repeatedly that without clear distinction between the definitions the complete escalation of motion mountain becomes impossible.

Motion around the universe

Verily, at first chaos came to be ...
Theogony, v. 120, Hesiod.*

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

First of all, on large scales, i.e. averaged over about ten million light years, the matter density in the universe is *homogeneous* and *isotropic*. Obviously, at smaller scales inhomogeneities exist, such as galaxies or cheese cakes. Our galaxy for example is neither isotropic Ref. 51 nor homogeneous. But at large scales the differences average out. This large scale homogeneity of matter position is often called the *cosmological principle*.

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

* The Theogony was finalized about ca. 700 BCE. It can be read in english and greek on the http://perseus.csad.ox.ac.uk/cgi-bin/ptext?lookup=Hes.+Th.+5 web site.

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

he found a relation

$$v = H\,d \quad , \tag{272}$$

where the proportionality constant H, so-called *Hubble constant*, is known today to have a value between 59 km s^{-1}Mpc^{-1}and 70 km s^{-1}Mpc^{-1}. (Hubble's own value was far outside this range.) For example, a star at a distance of 2 Mparsec is moving away from earth with a speed between 118 km/s and 140 km/s, and proportionally more for stars further away.

In fact, this discovery implies that every galaxy moves away from all the others. (Why?) Challenge In other words, the matter in the universe is *expanding*. The scale of this expansion and the enormous dimensions involved are amazing. The motion of all the thousand millions of galaxy groups in the sky are described by the single equation (272)! Actually, some deviations are observed or nearby galaxies, as mentioned above, and for far away galaxies, to be explained below.

In addition, the cosmological principle and the expansion imply that the universe cannot be older than that time when it was of vanishing size; the universe thus has a *finite age*. Including the evolution equations, as explained in more detail below, the Hubble constant points to an age value of around twelve thousand million years, with an error of about a sixth of this value. That also means that the universe has a *horizon*, i.e. a finite distance beyond which no signal reaches us.

Since the universe is expanding, in the past it has been much smaller and thus much denser than it is now. It turns out that it also has been hotter. George Gamow* predicted in 1948 that since hot objects radiate light, the sky cannot be completely black at night, Ref. 48 but must be filled with black body radiation emitted during the times it was in heat. That radiation, called the *background radiation*, must have cooled down due to the expansion of the universe. (Can you confirm this?) Challenge Despite various similar predictions by other authors, in one of the most famous cases of missed scientific communication, the radiation was found only much later, by two researchers completely unaware of all this work. A famous paper in 1964 by Doroshkevich and Novikov had even stated that the antenna used by the Ref. 49 (unaware) later discoverors was the best device to look for the radiation! In any case, only in 1965, Arno Penzias and Robert Wilson discovered the radiation, in one of the most beautiful discoveries of physics, for which both received the Nobel prize for physics. The radiation is Ref. 50 described by the blackbody radiation for a body with a temperature of 2.7 K to about 1 part in 10^4. We will come back to these results later.

But apart from expansion and cooling, the past twelve thousand million years also produced a few other memorable events.

A short history of the universe

The adventures of the universe, or better, of the matter and radiation inside it, are summarized in table 21. Some parts will become clear only in the second part of the escalation. The table has applications no scientist would have imagined. The sequence is so beautiful

* George Gamow (1904, Odessa –1968), russian-american physicist; he explained alpha decay as a tunnel effect, predicted the microwave background, and wrote several very successful popular science texts, such as *1, 2, 3, infinity*, and the *Mr Thompkins*, series.

and impressive that nowadays it is used in certain psychotherapies to point out to people the story behind their existence and to remind them of their own worth. Enjoy.

Table 21 A short history of the universe

Time from now[a]	Time from big bang[b]	Event	Temperature
$\approx 13 \cdot 10^9$ a	$\approx t_{Pl}$ [b]	Time, space, matter, and initial conditions make no sense	10^{32} K $\approx T_{Pl}$
$13 \cdot 10^9$ a	ca. 800 t_{Pl} $\approx 10^{-42}$ s	Distinction of space-time and matter, initial conditions make sense	10^{30} K
	10^{-35} s to 10^{-32} s	*Inflation* & *GUT epoch* starts; strong and electroweak interactions diverge	$5 \cdot 10^{26}$ K
	10^{-12} s	Antiquarks annihilate; electromagnetic and weak interaction separate	10^{15} K
	$2 \cdot 10^{-6}$ s	Quarks get confined into hadrons; universe is a plasma Positrons annihilate	10^{13} K
	0.3 s	Universe becomes transparent for neutrinos	10^{10} K
	a few seconds	*Nucleosynthesis*: D, ^{4}He, ^{3}He and ^{7}Li *nuclei* form; radiation still dominates	10^9 K
	2500 a	*Matter domination* starts; density perturbations magnify	75000 K
$z = 1100$	300 000 a	*Recombination*: during these latter stages of the big bang, H, He and Li *atoms* form, and the universe becomes 'transparent' for light, as matter and radiation decouple, i.e. as they acquire different temperatures; the 'night' sky starts to get darker and darker	3000 K
		Sky is almost black except for blackbody radiation	$T_\gamma = T_o(1+z)$
$z = 10 - 30$		*Galaxy formation*	
$z = 5.8$		Oldest object seen so far	
$z = 5$		Galaxy clusters form	
$z = 3$	10^6 a	First generation of stars (population II) is formed, hydrogen fusion starts; helium fusion produces carbon, silicon, oxygen	
	$2 \cdot 10^9$ a	First stars explode as supernovae[c]; iron is produced	
$z = 1$	$3 \cdot 10^9$ a	Second generation of stars (population I) appears, and subsequent supernova explosions of the aging stars form the trace elements (Fe, Se, ..) we are made of and blow them into the galaxy	
$4.7 \cdot 10^9$ a		Primitive cloud, made from such explosion remnants, collapses; sun forms	
$4.6 \cdot 10^9$ a		Earth and the other planets form	
$4.3 \cdot 10^9$ a		Craters form on the planets	
$4.0 \cdot 10^9$ a		Moon forms from material ejected during the collision of a large asteroid with the earth	
$4.0 \cdot 10^9$ a		*Archeozoic* starts: Earth's crust solidifies, oldest minerals water condenses	
$3 \cdot 10^9$ a		Unicellular (microscopic) life appears	
$2.6 \cdot 10^9$ a		*Protozoic* starts: atmosphere becomes rich in oxygen	
$1 \cdot 10^9$ a		Macroscopic life appears	
$580 \cdot 10^6$ a		*Paleozoic* starts: animals appear; oldest fossils	

Time from now[a]	Time from big bang[b]	Event	Temperature
$400 \cdot 10^6$ a		Land plants appear	
$370 \cdot 10^6$ a		Wooden trees appear	
$220 \cdot 10^6$ a		*Mesozoic* starts: mammals appear	
$150 \cdot 10^6$ a		Continent Pangea splits into Laurasia and Gondwana	
ca. $150 \cdot 10^6$ a		The star cluster of the Plejades forms	
$150 \cdot 10^6$ a		Birds appear	
135 to $65 \cdot 10^6$ a		Golden time of dinosaurs	
$100 \cdot 10^6$ a		Start of formation of Alps, Andes and Rocky mountains	
$65 \cdot 10^6$ a		*Cenozoic* starts: Dinosaurs become extinct due to a comet or asteroid hitting the earth in the Yucatan, primates appear	
$50 \cdot 10^6$ a		Large mammals appear	
$6 - 8 \cdot 10^6$ a		Hominids appears	
$5 \cdot 10^6$ a		Homo appears	
500000 a		Formation of youngest stars in galaxy	
300000 a		Homo sapiens appears	
100000 a		Beginning of last ice age	
90000 a		Homo sapiens sapiens appears	
20– to 11000 a		End of last ice age	
6000 a		First written texts	
2500 a		Physics starts	
500 a		Coffee usage spreads, modern physics starts	
200 a		Electricity usage begins	
100 a		Einstein publishes	
10 – 120 a		You were an unicellular being	
present		You are reading this	$T_\gamma = 2.73$ K, $T_\nu \approx 1.6$ K, $T_b \approx 0$ K
future		You enjoy life; for more details see page 357	

a. The time coordinate used here is the one given by the coordinate system defined by the microwave background radiation, as explained on page 255.
b. This quantity is not exactly defined since the big bang is not a space-time event. More on the issue later, on page 610.
c. The history of the atoms shows that we are made from the leftovers of a supernova. We truly are made of *stardust*.

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

The evolution of space-time

If the universe is full of matter, it cannot be static. Gravity always changes the distances between bodies, as the only possible exceptions are for orbits. Gravity also changes the *average* distances between bodies; gravity always tries to collapse clouds. The biggest cloud of all, the one formed by the matter in the universe, must therefore either be collapsing, or still be in expansion.

The first to dare to take this conlcusion was Aleksander Friedmann.* In 1922 he deduced the detailed evolution of the universe with homogeneous, isotropic mass distribution in the know standard fashion. For a universe which is homogeneous and isotropic for every point,
Challenge
the line element is given by

$$ds^2 = c^2dt^2 - a^2(t)(dx^2 + dy^2 + dz^2) \tag{273}$$

and matter is described by a density ρ_M and a pressure p_M. Inserting all this into the field equations, one gets two equations

$$\left(\frac{\dot a}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho_M + \frac{\Lambda}{3} \tag{274}$$

$$\ddot a = -\frac{4\pi G}{3}(\rho_M + 3p_M)\, a + \frac{\Lambda}{3}\, a \tag{275}$$

which imply

$$\dot\rho_M = -3\frac{\dot a}{a}(\rho_M + p_M) \tag{276}$$

At the present time t_o, the pressure of matter is negligible. In this case, the expression $\rho_M a^3$ is constant in time.

Before we discuss the equation, first a few points of vocabulary. It is customary to relate
Challenge
all mass densities to the so-called *critical mass density* ρ_c given by

$$\rho_c = \frac{3H_o^2}{8\pi G} \approx 8 \pm 2 \cdot 10^{-27}\,\mathrm{kg/m^3} \tag{277}$$

corresponding to about 8, give or take 2, hydrogen atoms per cubic metre. On earth, one would call this value an extremely good *vacuum*. Such are the differences between everyday life and the universe as a whole. In any case, the critical density characterizes a matter distribution leading to an evolution of the universe just between neverending expansion and collapse. In fact, this density is the critical one, leading to a so-called *marginal* evolution, only in the case of *vanishing* cosmological constant. Despite this restriction, the term is now

* Aleksander Aleksandrowitsch Friedmann (1888–1925), russian physicist who predicted the expansion of the universe. Due to his early death of typhus, his work remained almost unknown until Georges A. Lemaître (1894–1966), belgian priest and cosmologist, took it up and expanded it in 1927, focussing, as his job required, on solutions with an initial singularity. Lemaître was one of the propagators of the (erroneous) idea that the big bang was an "event" of "creation" and convinced his whole organisation about it. The Friedman-Lemaître solutions are often erroneously called after two other physicists, who studied them again much later, in 1935 and 1936, namely H.P. Robertson and A.G. Walker.

See page 259, 260

used for this expression in all other cases as well. One thus speaks of dimensionless mass densities Ω_M defined as

$$\Omega_M = \rho_o/\rho_c \quad . \tag{278}$$

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

$$\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c} = \frac{\Lambda c^2}{8\pi G \rho_c} = \frac{\Lambda c^2}{3H_o^2} \quad . \tag{279}$$

A third dimensionless parameter Ω_K describes the curvature of space. It is defined as

$$\Omega_K = \frac{-k}{R_o^2 H_o^2} \tag{280}$$

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

The present time Hubble parameter is defined by $H_o = \dot{a}_o/a_o$. From equation (274) one then gets: Challenge

$$\Omega_M + \Omega_\Lambda + \Omega_K = 1 \tag{281}$$

In the past, when data was lacking, physicists were divided into two camps: the *claustrophobics* believing that $\Omega_K > 0$ and the *agoraphobics* who believe that $\Omega_K < 0$. More details about the measured values of these parameters will be given shortly. The diagram of figure 97 shows the range of sensible parameters with the corresponding behaviour of the universe.

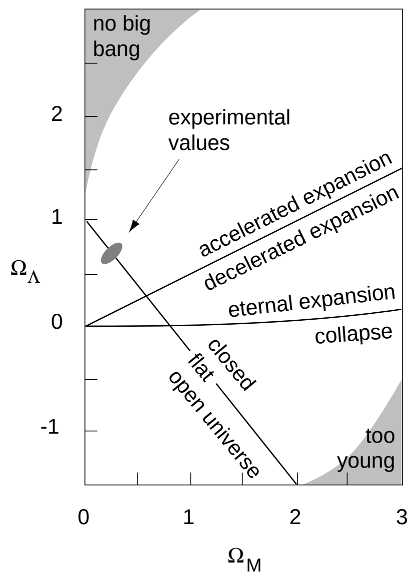


Figure 97 The ranges for the Ω parameters and their consequences

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

$$q_o = -\frac{\ddot{a}_o}{a_o H_o^2} = \frac{1}{2}\Omega_M - \Omega_\Lambda \tag{282}$$

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

An even clearer way to picture the expansion of the universe for vanishing pressure is to rewrite equation (274) using $\tau = t H_o$ and $x(\tau) = a(t)/a(t_o)$, yielding

$$\left(\frac{dx}{d\tau}\right)^2 + U(x) = \Omega_K$$
$$\text{with } U(x) = -\Omega_\Lambda x - \Omega_\Lambda x^2 \tag{283}$$

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

For vanishing Ω_Λ, the universe either expands for ever, or recollapses, depending on the value of the mass-energy density,

For non-vanishing (positive) Ω_Λ, the potential has exactly one maximum; if the particle has enough energy to get over the maximum, it will accelerate continuously. That is the situation the universe seems to be in today.

For a certain time range, the result is shown in figure 98. There are two points to be noted: the set of possible curves is described by *two* parameters, not one. In addition, lines cannot be drawn down to the origin of the diagram. There are two main reasons: we do not know the behaviour of matter at very high energy yet, and we do not know the behaviour of space-time at very high energy. We return to this important issue shortly, and then again in the third part of the escalation.

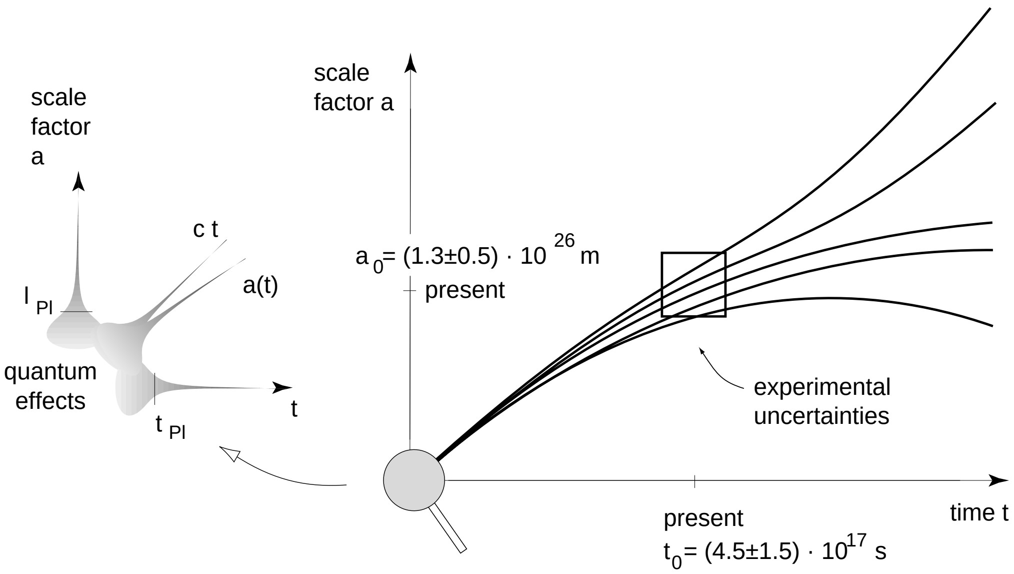


Figure 98 The evolution of the universe's scale R for different values of its mass density

The main result of Friedmann's work was that a homogenous and isotropic universe is *not static*: it either expands or contracts. In either case, it has a *finite age*. This profound result took many years to spread around the cosmology community; even Einstein took a long time to get accustomed to it.

Note that due to its isotropic expansion, in the universe there is a preferred reference frame: the frame defined by average matter. The time measured in that frame is the time listed in table 21 and the one meant when one talks about the *age* of the universe.

An overview of the possibilities for the long time evolution is given in figure 99. The evolution can have various outcomes. In the early twentieth century, people decided among them by personal preference. Albert Einstein first preferred the solution $k = 1$ and $\Lambda = a^{-2} = 4\pi G\rho_M$. It is the unstable solution found when $x(\tau)$ remains at the top of the potential $U(x)$.

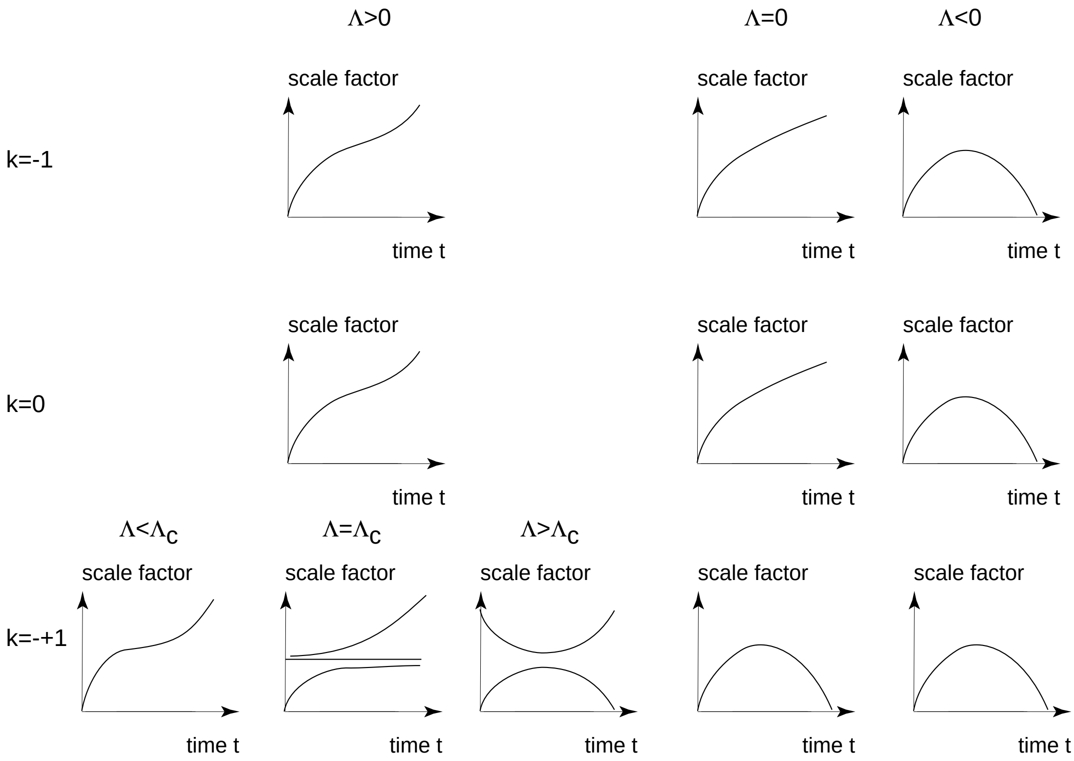


Figure 99 The long term evolution of the universe's scale *a* for various parameter combinations

De Sitter had found, much to Einstein's personal dismay, that an empty universe with $\rho_M = p_M = 0$ and $k = 1$ is possible. This type of universe expands for large times. Challenge

Lemaître had found expanding universes for positive mass, and his results were also contested by Einstein in the beginning. When the first data come in, massive and expanding universes became popular and the standard story in textbooks. However, in a sort of collective blindness that lasted from around 1950 to 1990, almost everybody believed that $\Lambda = 0$.* Only towards the end of the twentieth century experimental progress allowed to make statements free of personal beliefs.

Is the universe open, closed or marginal?

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

Sometimes the history of the universe is summed up in two words: *bang!...crunch*. But will the universe indeed recollapse or will it expand for ever? The parameters deciding its fate are the mass density and cosmological constant, and so far, they point into a different direction.

* In this case, one has the connection that for $\Omega_M \geqslant 1$, the age of the universe follows $t_o \leqslant 2/3(1/H_o)$, where the limits correspond. For vanishing mass density one has $t_o = 1/H_o$. Challenge

The main news of the last decade of the twentieth century astrophysics are the experimental results allowing to determine all these parameters. Several methods are being used. The first method is obvious: determine speed and distance of distant stars. For large distances, this is difficult, since the stars get faint. But it has now become possible to search the sky for supernovae, the bright exploding stars, and to determine their distance through their brightness.

A second method is the measurement of the anisotropy of the cosmic microwave background. From the power spectrum as function of the angle one can deduce the curvature of spacetime.

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

See page 266

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

Ref. 52

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

$$(\Omega_M, \Omega_\Lambda, \Omega_K) \approx (0.3, 0.7, 0.0) \qquad (284)$$

where the errors are of the order of 0.1 or less. The values imply that *the universe is spatially flat, its expansion is accelerating, and there will be no big crunch*. However, no definite statement on the topology is possible. We come back to this issue shortly.

See page 267

In particular, the data show that the density of matter, inclusive all dark matter, is only about one third of the critical value.* Twice that amount is given by the cosmological term. For the cosmological constant Λ one gets the value

$$\Lambda = \Omega_\Lambda \frac{3H_o^2}{c^2} \approx 10^{-52}/\mathrm{m}^2 \qquad (285)$$

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

$$\rho_\Lambda c^2 = \frac{\Lambda c^4}{8\pi G} \approx 0.5\,\mathrm{nJ/m^3} \approx \frac{10^{-46}\,(GeV)^4}{(\hbar c)^3} \qquad (286)$$

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

See page 241

$$\Delta\varphi = 4\pi G(\rho + 3p/c^2) \qquad (287)$$

one gets

$$\Delta\varphi = 4\pi G(\rho_M - 2\rho_\Lambda) \qquad (288)$$

Challenge

one thus for the gravitational acceleration

$$a = \frac{GM}{r^2} - \frac{\Lambda}{3}c^2 r = \frac{GM}{r^2} - \Omega_\Lambda H_o^2\, r \qquad (289)$$

* The difference between the total matter density and the separately measurable baryonic matter density, only about one sixth of the former value, is also not explained yet. It might even be that the universe contains matter of a type unknown so far.

which shows that a *positive* vacuum energy indeed leads to a *repulsive* gravitational effect. Inserting the abovementioned value for Λ one finds that the repulsive effect is small even for the distance between the earth and the sun. In fact, the order of magnitude is so much smaller that one cannot hope for a direct experimental confirmation of this deviation from universal gravity at all. Probably astrophysical determinations will remain the only possible ones. A positive gravitational constant manifests itself through a positive component in the expansion rate, as we will see shortly. Challenge

But the situation is puzzling. The origin of this cosmological constant is *not* explained by general relativity; this mystery will be solved only much later in our escalation, by quantum theory. In any case, the cosmological constant is the first local or quantum aspect of nature detected by astrophysical means.

Predictions of the modern big bang model of cosmology

Above all, the big bang model model states that about twelve thousand million years ago the whole universe was extremely small, a fact that gave the big bang its name. The expression 'big bang' was created in 1950 by Fred Hoyle* who by the way never believed that it gives a correct description of the evolution of the universe. Since the past smallness cannot itself be checked, one needs to look for other, verifiable consequences. The central ones are the following: Ref. 53

- all matter moves away from all other matter;
- there is about 25% helium in the universe;
- there is thermal background radiation of about 3 K;
- the maximal age for any system in the universe is around twelve thousand million years;
- there are background neutrinos with a temperature of about 2 K.**
- for nonvanishing cosmological constant, Newtonian gravity is slightly reduced;

All predictions, except the last two, are confirmed by observations. Technology probably will not allow to check them in the foreseeable future; however, there is also no hint putting them into question.

Competing descriptions of the universe have not been successful in matching these predictions. In addition, theoretical arguments state that with matter distributions such as the observed one, plus some rather weak general assumptions, there is no known way to avoid a period in the *finite* past in which the universe was extremely small. Therefore it is worth having a close look at the situation. Ref. 53 Ref. 54

Was the big bang a big bang?

Was it a kind of explosion? An explosion assumes that some material transforms internal energy into motion of its parts. There has not been any such process in the early history of the universe. The origin for the initial velocity of matter is *unknown* at this point of the escalation. One cannot call the whole phenomenon an explosion at all. And obviously there neither was nor is any air in interstellar space, so that one cannot speak of a "bang" in any sense of the term.

* Fred Hoyle (1915–), british astronomer and astrophysicist.

** The theory states that $T_\nu/T_\gamma \approx (4/11)^{1/3}$. These neutrinos appeared about 0.3 s after the big bang.

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

Was the big bang an event?

The big bang is a description of what happened in the *whole* of space-time. Despite what is often written in bad newspaper articles, at every moment of the expansion, space is always of non-vanishing size; space *never* was a single point. People who pretend this are making at first sight plausible, but false statements. The big bang is a description of the *expansion* of space-time, not of its beginning. Following the motion of matter back in time, general relativity cannot deduce the existence of an inital singularity. The issue of measurement errors is probably not a hindrance; however, the effect of the nonlinearities in general relativity at situations of high energy densities is not clear.

Most importantly, quantum theory shows that the big bang was *not* a singularity, as no observable, neither density nor temperature, reaches an infinitely large or infinitely small See page 619 value, since such values cannot exist in nature.* In any case, it is a general agreement that arguments based on *pure* general relativity alone cannot make correct statements on the big bang. Most newspaper article statements are of this sort.

Was the big bang a beginning?

Asking what was before the big bang is like asking what is north of the north pole. Since nothing is north of the north pole, nothing "was" before the big bang.

This analogy could be misinterpreted to imply that the big bang took its start at a single point in time, which of course is incorrect, as just explained. But the analogy is better than it looks; in fact, there is *no* precise north pole, since quantum theory shows that there is a basic uncertainty on its position, as we will see in the second part of our escalation. There is also a corresponding uncertainty for the big bang.

In fact, it does not take more than three lines to show with quantum theory that time and space are *not* defined either at or near the big bang. We will give this simple argument in See page 584 the first chapter of the third part of the escalation. The big bang therefore cannot be called a "beginning" of the universe. There never was a time when the scale factor $R(t)$ of the universe was zero. This conceptual mistake is frequently encountered. Near the big bang, events can *neither* be ordered *nor* even be defined. More bluntly, there is *no* beginning; there has never been an initial event or singularity, despite the numerous statements pretending the contrary.

Obviously the concept of time is not defined "outside" or "before" the existence of the Ref. 55 universe; this fact was clear to thinkers already over thousand years ago. It is then tempting to conclude that time must have *started*. But as we saw, that is a logical mistake as well: first of all, there is no starting event, and secondly, time does not flow, as clarified already See page 41 in the beginning of our walk.

* Many physicists are still wary to make such strong statements at this point. The first sections of the third part See page 578 of the escalation give the precise arguments leading to them.

A similar mistake lies behind the idea that the universe "had certain initial conditions." Initial conditions *by definition* make only sense for objects or fields, i.e. for entities which can be observed from the outside, i.e. for entities which have an environment. The universe does not comply to these requirements; the universe thus cannot have initial conditions. Nevertheless, many people still insist on thinking about the issue; interestingly, Steven Hawking sold millions of books explaining that a description without initial conditions is the most appealing, overlooking that there is no other possibility anyway. This statement will still lead to strong reactions among physicists; it will be discussed in more detail in the section on quantum theory.

See page 94
Ref. 56
Challenge

In summary, the big bang does not contain a beginning nor does it imply one. In the third part of our escalation we will uncover the correct way to think about it.

Does the big bang imply creation?

[The general theory of relativity produces]
universal doubt about god and his creation
A US-american witch hunter

Creation, i.e. the appearance of something out of nothing, needs an existing concept of space and time to make sense. The concept of 'appearance' makes no sense otherwise. But whatever the description of the big bang, be it classical, as in this chapter, or quantum mechanical, as in later ones, this condition is never fulfilled. Even in the present, classical description of the big bang, which gave origin to its name, there is *no* appearance of matter, nor of energy, nor of anything else. And this situation does not change in any latter, improved description, as time or space are never defined *before* the appearance of matter.

See page 415

In fact, all properties of a creation are missing; there is no "moment" of creation, no appearance from nothing, no possible choice of any "initial" conditions out of some set of possibilities, and as we will see in more detail later on, not even any choice of particular physical "laws" from any set of possibilities.

In summary, the big bang does not imply nor harbour a creation process. The big bang was not an event, not a beginning, and *not* a case of creation. It is impossible to continue the escalation of motion mountain if one cannot accept each of these three conclusions. If one denies them, one has decided to continue in the domain of beliefs, thus effectively giving up on the escalation. But what then is the big bang? We'll find out in the third part.

Challenge

Note that this requirement is not new. In fact, it was already contained in equation (1) at the start of our walk, as well as in all the following ones. However, it appears more clearly at this point.

See page 43

10. Why can we see the stars?

Zwei Dinge erfüllen das Gemüt mit immer neuer und zunehmender Bewunderung
und Ehrfurcht, je öfter und anhaltender sich das Nachdenken damit beschäftigt:
der bestirnte Himmel über mir und das moralische Gesetz in mir.*
Immanuel Kant (1724–1804)

Ref. 57

* Two things fill the mind with ever new and increasing admiration and awe, the more often and persistently thought considers them: the starred sky above me and the moral law inside me.

On clear nights, between two and five thousand stars are visible with the naked eye. Several hundreds of them have names. Indeed, in all parts of the world, the stars and the constellations they form are seen as memories of ancient events, and stories are told about them.* But simply the fact that we can see the stars tells a story much more fantastic than all myths. It touches almost all aspects of modern physics.

Which stars do we see at all?

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

In fact almost all visible stars are from our own galaxy. The only extragalactic object *constantly* visible to the naked eye in the northern hemisphere is the so-called Andromeda nebula, which in fact is a whole galaxy like our own, as Immanuel Kant already had conjectured in 1755. The other extragalactic object visible in the sky with the naked eye is the Tarantula Nebula in the southern hemisphere. Other, temporary exceptions are the rare *novae*, exploding stars which can be seen also if they appear in nearby galaxies, or the even rarer *supernovae*, which can often be seen even in faraway galaxies.

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

Challenge

Is the universe different from our milky way? Yes, it is. There are several arguments. First of all, our galaxy – that is just the greek original of the term 'milky way' – is *flattened*, due to its rotation. If the galaxy rotates, there must be other masses which determine the background with respect to which this rotation takes place. In fact, there is a huge number of other galaxies – about 10^{11} – in the universe, a discovery dating only from the 20th century.

Figure missing

Figure 100 How our galaxy looks in the infrared

Why did this happen so late? Well, people had the same difficulty as when the the shape of the earth had to be determined. One had to understand that the galaxy is not only a milky strip seen in clear nights, but an actual physical system, made of about 10^{11} stars gravitating around each other.** As in the case of the earth, the galaxy was found to have a three-dimensional *shape*; it is shown in figure 100. Our galaxy is a flat and circular structure, with a diameter of ... light years and a bulge in the center. As said before, rotates once in about 200 to 250 million years. Can you guess how this is measured? The rotation quite is slow:

Challenge

* About the myths around the stars and the constellations, see e.g. the text by G. Fasching, *Sternbilder und ihre Mythen*, Springer Verlag, 1993. There are also the beautiful www.astro.wisc.edu/~dolan/constellations/constellations.html and www.astro.uiuc.edu/~kaler/sow/sow.html web sites.

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

since the sun exists, it made only about 20 to 25 full turns around the centre.

It is even possible to measure the *mass* of our galaxy. The trick is to use a binary pulsar on the outskirts of the galaxy. If one observes it for many years, one can deduce its acceleration around the galaxy centre. The pulsar reacts with a frequency shift which can be measured on earth. However, one needs many decades of observations, and one has to eliminate many spurious effects. Nevertheless, such measurements are ongoing. Ref. 58

Figure missing

Figure 101 A few galaxies: the Andromeda nebula M31, a spiral galaxy, the elliptical galaxy NGC 205, and the colliding galaxies M51, also known as NGC 5194

Why is the universe transparent?

Could the universe be filled with water, which is transparent, as maintained by some popular books in order to explain rain? No. Even if it were filled with air, the total mass would never have allowed the universe to reach the present size; it would have recollapsed much earlier and we would not exist. Ref. 59 Challenge

The universe is thus transparent because it is mostly empty. But *why* is it so empty? First of all, in the times when the size of the universe was small, all antimatter annihilated with the corresponding amount of matter, and only a tiny fraction of matter, which was slightly more abundant, was left over. This 10^{-9} fraction is the matter we see now. Therefore the number of photons in the universe is 10^9 larger than that of electrons or quarks. See page 550

If one remembers that the average density of the universe is $10^{-26}\,\mathrm{kg/m^3}$ and then that most of the matter is lumped by gravity in galaxies, one can imagine what an excellent vacuum one has in between. In short, light can travel along large distances without hindrance.

In addition, 300 000 years after antimatter annihilation, all available nuclei and electrons recombined, forming atoms, and the aggregates like stars. No free charges interacting with photons were lurking around any more, so that from that period onwards light could travel through space like it does today, being affected only when it hits some star or some dust particle.

But why is the vacuum transparent? That is a much deeper question and we reserve it for a later stage of our walk. But the answers are not complete yet. See page 509

Why can we see the sun?

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

In fact, even air is not completely transparent; air molecules *scatter* light a little bit. That is why the sky and far away mountains appear blue and sunsets red,* and stars are invisible during daylight.

Secondly, we can see the sun because the sun, like all hot bodies, *emits* light. We describe the details of See page 354 *incandescence*, as this effect is called, below.

Thirdly, we can see the sun because we and our environment and the sun's environment are *colder* than the sun. In fact, incandescent bodies can be distinguished from their background only if the background is colder. This is a consequence of the properties of incandescent light emission, usually called *black body radiation*. The radiation is material independent, so that for an environment with the same temperature as the body, one cannot see anything at all. Just have a look on the photograph of page 356 as a proof.

Figure 102 The absorption of the atmosphere

Finally, we can see the sun because it is not a black hole. If it were, it wouldn't emit (almost) any light, as we will see shortly.

Obviously, each of these conditions applies for stars as well. For example, we can only see them, because the night sky is black. But

Why is the sky dark at night?

First of all, the sky is not black at night. It has the same colour as during the day, as any long exposure photograph shows. But that colour, like to colour of the sky during the day, is not due to the temperature of the sky, but to the light from the stars. If one looks for temperature radiation, one does find some. Measurements show that the sky is not completely cold at night. It is filled with radiation of around 200 GHz; more precise measurements show that the radiation corresponds to the thermal emission of a body of 2.73 K. This *background radiation* is the thermal radiation left over from the big bang.

Ref. 60 The universe is indeed colder than the stars. But why is this so? If the universe were homogenous on large scales and infinitely large, it would have an infinite number of stars. Given any direction one would look at, one would hit the surface of a star. The night sky would be as bright as the surface of the sun! Are you able to convince your grandmother about this? Challenge

In addition, we would effectively live inside a furnace with a temperature of the average star, namely about 6000 K, thus making it effectively impossible to enjoy ice cream. This paradox was most clearly formulated in 1826 by the astronomer Wilhelm Olbers.** Two main effects can avoid the contradiction with observations. First, since the universe is finite

* Air scattering makes the sky blue also at night, as can be proven by long time exposure cameras; however our eyes are not able to perform this trick, and the low levels of light make it black to us.

** Heinrich Wilhelm Matthias Olbers (1758, Arbergen – 1840, Bremen), astronomer. He discovered two planetoids, Pallas and Vesta, and five comets; he invented the method to calculate parabolic orbits for comets still in use today. The paradox is named after him, though others had made similar points before, such as the swiss astronomer de Cheseaux in 1744. He also actively supported F.W. Bessel.

in age, far away stars are shining for less time, so that their share is smaller, and thus the average temperature of the sky is reduced.*

Secondly, one could imagine that the radiation of far away stars is shifted to the red, and the volume the radiation must fill is increasing continuously, so that the average temperature of the sky is also reduced. One needs calculations to decide which effect is the greater one. This issue has been studied in great detail by Paul Wesson; he explains that the first effect is larger than the second by a factor of three. We may thus state correctly that the sky is dark at night *mostly* because the universe has a *finite* age. We can thus add that the sky would be brighter if the universe were not expanding. Ref. 61

In addition, the darkness of the sky is possible only because the speed of light is *finite*. Can you confirm this? Ref. 60 Challenge

Finally, the darkness of the sky also tells us that the universe has a *large* age. Indeed, the 2.7 K background radiation is that cold, despite having been emitted at 3000 K, because it is red shifted due to the Doppler effect. Under reasonable assumptions, the temperature of this radiation changes with the scale factor of the universe as Ref. 62

$$T \sim \frac{1}{R(t)} \quad . \tag{290}$$

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

In summary, the sky is black at night because space-time is of finite, but old age.

As a side issue, a quiz: is there an Olbers' paradox for gravitation? Challenge

Why are the colors of the stars different?

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

Not all stars a good approximations of black bodies, so that the black body radiation law sometimes is not a accurate description for their color. However, most of the stars are good approximations of black bodies. The temperature of a star depends mainly on its size, its mass, its composition, and its age, as the astrophysicists are happy to explain. Orion is a good example of a colored constellation. See page 355 Ref. 63

Table 22 The colour of the stars

Class	temperature	example	position	colour
O	30 kK	Mintaka	δ Orionis	blue-violet
O	31 ± 10 kK	Alnitak	ζ Orionis	blue-violet
B	22(6) kK	Bellatrix	γ Orionis	blue
B	kK	Saiph	$\varkappa$ Orionis	blue-white
B	kK	Rigel	β Orionis	blue-white
B	kK	Alnilam	ε Orionis	blue-white
B	17(5) kK	Regulus	α Leonis	blue-white

* Are you able to explain that the sky is not black because it is painted black or made of black chocolate? Or more generally, that the sky is not a made of or does not contain some dark and cold substance, as Olbers himself suggested, and as J. Herschel proved wrong in 1848? Challenge

Table 22 The colour of the stars

Class	temperature	example	position	colour
A	9.9 kK	Sirius	α Canis Majoris	blue-white
A	8.6 kK	Megrez	δ Ursae Majoris	white
A	7.6(2) kK	Altair	α Aquilae	yellow-white
F	7.4(7) kK	Canopus	α Carinae	yellow-white
F	6.6 kK	Procyon	α Canis Minoris	yellow-white
G	5.8 kK	Sun	ecliptic	yellow
K	3.5(4) kK	Aldebaran	α Tauri	orange
M	2.8(5) kK	Betelgeuse	α Orionis	red

The basic colour determined by temperature is changed by two effects. The first, the

Challenge *Doppler red shift*, depends on the speed v between source and observer following

$$z=\frac{\Delta\lambda}{\lambda}=\frac{f_{\mathrm{S}}}{f_{\mathrm{O}}}-1=\sqrt{\frac{c+v}{c-v}}-1 \qquad (291)$$

Such shifts only play a significant role only for far away, and thus faint stars visible through the telescope. With the naked eye, Doppler shifts cannot be seen. But Doppler shifts can make far away stars shine in the infrared instead of in the visible domain. Indeed, the highest Doppler shifts observed for luminous objects are larger than 5, corresponding to more than

Challenge 94% of the speed of light. Not that in the universe, the red shift is also related to the scale factor R by

$$z=\frac{R(t_{\mathrm{o}})}{R(t_{\mathrm{emission}})}-1 \qquad (292)$$

Light at a red shift of 5 thus was emitted at an age around a quarter of the present.

The other colour changing effect, the *gravitational red shift*, depends on the matter density of the source and is given by

$$z=\frac{\Delta\lambda}{\lambda}=\frac{f_{\mathrm{S}}}{f_{\mathrm{O}}}-1=\frac{1}{\sqrt{1-\frac{2GM}{c^2R}}}-1 \qquad (293)$$

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

Other red shift processes are not known; moreover, such processes would contradict all the properties of nature we know. But the colour issue leads to the next question:

Are there dark stars?

It could be that some stars are not seen because they are dark. This issue is of importance, since it could lead to incorrect matter density estimates for the universe, and thus to incorrect evolution predictions.

This issue is hotly debated. It is known that objects more massive than Jupiter but less massive than the sun can exist in states which do not emit almost any light. They are also called *brown dwarfs*. It is unclear at present, how many such objects exist. Many of the so-called extrasolar “planets” are probably brown dwarfs. The issue is not closed.

The other possibility for dark stars are black holes. They are discussed in detail below.

Why do stars shine?

Do you see the stars shine?
I am the only in the world one who knows why.
Eddington, in conversation.

Stars are hot because of nuclear reactions in their interior. We will discuss them in more detail in the chapter on the nucleus. See page 549

Most of all, since stars shine, they also die. And they are somehow formed. In other words, stars can be seen if they are born but not yet dead at the moment of light emission. That also leads to restrictions on their visibility, especially for high red shifts. Indeed, the objects one observes at large distances, such as quasars, are not stars, but much more massive and bright. These issues are still being studied by astrophysicists.

Why are the stars the same every night?

Stars are long lived onjects. Or so it seems. In fact, every now and then a new star appears in the sky: a *nova*. Especially bright novae are called *supernovae*. Novae and similar phenomena remind us that stars usually live much longer than humans, but that like them, they are born and die. The fascinating details of these processes are part of astrophysics, and will not be explored during this escalation.

Are all stars different? – gravitational lenses

Per aspera ad astra.

Are wesure that at night, two stars are really different? The answer is no. Recently, it was shown that two stars were actually two images of the same object. This was found by comparing the flicker of two different images. It was found that the flicker of one image was exactly the same as the other, just shifted by 423 days. This heroic result was found by Johannes Pelt from Estonia, and his research group, while observing two quasar images of the system Q0957+561. Ref. 64

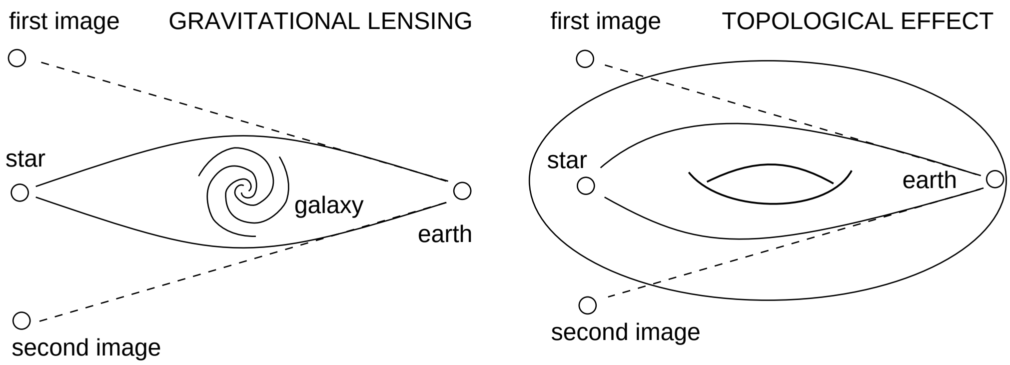


Figure 103 How one star can lead to several images

The two images are the result of *gravitational lensing*. Indeed, a large galaxy can be seen between the two images, at much smaller distance from the earth. This effect was imagined by Einstein; however he did not believe that it was observable. The real father of Ref. 65 gravitational lensing is Fritz Zwicky, who predicted in 1937 that the effect would be quite frequent and easy to observe, as indeed it turned out to be the case.

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

In fact, if the two objects one observes are lined up behind each other, one sees the more distant one as *ring* around the nearer one. Such rings have indeed been observed, and the object B1938+666 is one of the most beautiful ones. Other people even try to find earth-like planets on other stars using this method.

Photographs to be included

Figure 104 The Zwicky-Einstein ring B1938+666 and multiple galaxy images around CL0024+1654

Generally speaking, nearby stars are truly different, but for the far away stars the problem is tricky. For single stars, the issue is not so important, seen overall. Reassuringly, about 40 double star images have been identified so far. But when whole galaxies are seen as several images at once, and several dozens are known so far, one starts to get nervous. In the case of CL0024+1654, the image of the distant galaxy is seen seven times around the image of the nearer mass.

In this case, apart from lensing, also the shape of the universe could play some tricks.

What is the shape of the universe?

There is a standard explanation to avoid some of the just mentioned problems. The universe in its evolution is similar to the surface of an ever increasing sphere: the surface is finite, but it has no boundary. The universe simply has an additional dimension; therefore its volume is also everincreasing, finite, but without boundary. This explanation presupposes that the universe has the same topology, the same “shape” as that of a sphere with an additional dimension.

Ref. 66 But what is the experimental evidence for this statement? Nothing. Nothing is yet known about the shape of the universe. It is extremely hard to determine it, because of its sheer size.

What do experiments say? In the nearby region of the universe, say a few million light years, the topology is simply connected. But for large distances, almost nothing is sure.

Maybe research into gamma ray bursts will provide a way to determine topology, as these bursts often originate from the dawn of time, and thus might tell something about the topology.*

Since little is known, one can ask about the range of possible answers. In the standard model with $k = 1$, space-time is usually assumed to be a product of linear time, with the topology R of the real line, and a sphere S^3 for space. That is the simplest possible shape, corresponding to a *simply connected* universe. For $k = 0$, the simplest topology of space is three-dimensional real space R^3, and for $k = -1$ a hyperbolic manifold H^3.

Also depending on the value of the cosmological constant, space could be finite and bounded, or infinite and unbounded.

Simple connectedness is usually tacitly assumed (but not at all required) in the Friedman-LeMaître-Robertson-Walker calculations.

However, space-time could also be *multiply* connected, like a higher-dimensional version of a torus, or have even more complex topologies.**

In this case, it could even be that the actual number of galaxies is much smaller than the observed number. This situation would correspond to a kaleidoscope, where a few stones produce a large number of images. In addition, topological surprises could also be hidden *behind* the horizon.

In fact, the issue of topology gets an additional and unexpected twist in the third part of our walk. See page 618

What is behind the horizon?

> The universe is a big place;
> perhaps the biggest.
> Kilgore Trout

The horizon is a tricky entity. In fact, all cosmological models show that it moves rapidly away from us. A detailed investigation shows that for a matter dominated universe it moves away from us with a velocity Ref. 68 Challenge

$$v_{\text{horizon}} = 3c \quad . \qquad (294)$$

A pretty result, isn't it? Obviously, since the horizon does not transport any signal, this is not a contradiction with relativity. But what is behind the horizon?

If the universe were *open* or *marginal*, the matter we see at night would only be a – literally – infinitely small part of all existing matter, since an open or marginal universe implies that there is an infinite amount of matter behind the horizon. Is such a statement verifiable? In other words, is such a statement a belief or a fact? Challenge

Unfortunately, a *closed* universe fares only slightly better. Matter is still predicted to exist behind the horizon; however, in this case it is only a finite amount.

In short, the standard model of cosmology states that there is a lot of matter behind the horizon. The description of inflation makes more specific statements about it.

* The story is told from the mathematical point of view by Bob Osserman, *Poetry of the universe*, 1990.

** The FLRW metric is also valid for any quotient of the just mentioned simple topologies by a group of isometries, leading to dihedral spaces and lens spaces in the case $k = 1$, to tori in the case $k = 0$, and to *any* hyperbolic manifold in the case $k = -1$. Ref. 67

Why are there stars all over the place? – Inflation

What were the initial conditions of matter? Obviously it was roughly a constant density over space. How could this happen? The person to have asked this question most thoroughly was Alan Guth.

– CS – to be added later – CS –

Why are there so few stars? The energy and entropy content of the universe

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

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

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

$$\begin{aligned} E &= E_b + E_\gamma + E_\nu \\ &\approx \frac{c^2 M_o}{T_o} + \ldots + \ldots \approx \frac{c^2}{G} + \ldots \end{aligned} \tag{295}$$

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

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

However, the discussion of *entropy* puts a strong question mark behind all these seemingly obvious statements. Some people try to give values for the entropy of the universe. Still other check whether one has the relation

Challenge

$$S = \frac{kc^3}{G\hbar}\frac{A}{4} = \frac{kG}{\hbar c} 4\pi M^2 \quad , \tag{296}$$

as is the case for black holes, assuming that all the matter and all the radiation of the universe can be described by some average temperature. Others even speculate where the entropy of the universe comes from, and whether the horizon is the source for it.

* The energy of the universe is constant. Its entropy tends towards a maximum.
** Except for the case when pressure can be neglected.

But let us be careful. Clausius assumes, without the slightest doubt, that the universe is a closed system, and thus deduces the above statement. Let us check this assumption. Entropy describes the maximum energy one can extract from a hot object. After the discovery of the particle structure of matter, it became clear that entropy is also given by the number of microstates looking like a specific macrostate. But both definitions make no sense if one applies them to the universe as a whole. There is no way to extract energy from it, and no way to say how many microstates of the universe would look like the macrostate.

The basic reason is the impossibility to apply the concept of *state* to the universe. In the beginning, we defined a state as those properties of a system which allow to distinguish it from other systems with the same intrinsic properties, or which differ from one observer to the other. You might want to check for yourself that for the universe, such state properties do exist at all! Challenge

If there is no state of the universe, there is no entropy for it. And neither an energy value. This is in fact the only correct conclusion one can take about the issue.

Why is matter lumped?

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

It turns out that homogeneous mass distributions are *unstable*. If for any reason the density fluctuates, regions of higher density will attract matter and increase in density, whereas regions of lower density will deplete.

But how did the first inhomogeneities form? That is one of the big problems of modern astrophysics, and there is no accepted answer yet.

Modern experiments try to measure the variations of the cosmic background radiation spectrum with angular position and with polarisation; these results, which will be available in the coming years, might provide information towards settling the issue. Ref. 69

Why are stars so small compared with the universe?

Given that the matter density is around the critical one, the size of stars, which contain most of the matter, is a result of the interaction of the elementary particles composing them. Below we will show that general relativity (alone) cannot explain any size appearing in nature. The discussion of this issue is a theme of the second part of our escalation. See page 275

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

Can one distinguish between expanding space and galaxies moving apart? Yes, one can. Are you able to find an argument or to devise an experiment? Challenge

Does the expansion of the universe also apply to the space on the earth? No. The expansion is calculated for a homogeneous and isotropic mass distribution. Matter is not homogeneous nor isotropic inside the galaxy; the approximation of the cosmological principle is not valid down here. It has been checked experimentally by studying atomic spectra in various places in the solar system that on its scale there is *no* Hubble expansion taking place. Ref. 70

Is there more than one universe?

That is another possible direction to study the question whether we see all the stars. In fact, you might check that neither definition of universe given above, be it 'all matter-energy' or 'all matter-energy and all space-time', allows to answer the question positively.

Challenge

There is no way to define a plural for universe: either the universe is everything, and then it is unique, or it is not everything, and then it is not the universe. We will discover in the second part of the escalation that quantum mechanics does not change this conclusion, despite recurring reports of the contrary.

See page 488

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

The two arms of humans played an important role in discussions about motion, and especially in the development of relativity. Looking at the stars at night, one can make a simple observation, if one keeps one's arms relaxed. Standing still, our arms hang down. Then we turn rapidly. Our arms lift up; in fact they do so whenever the stars turn. Some people have spent their lives on this connection. Why?

Ref. 71

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

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

This idea is indeed realized in the description of nature via general relativity. No question about it.

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

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

Ref. 72

Ref. 73

But in fact it is easy to see that Mach *cannot* have meant a mass variation at all: one then would also have to conclude that mass should be distance dependent, and that this should be so even in galilean physics. But this statement is indeed known to be wrong, and nobody in his right mind has ever had any doubts about it.

Challenge

The whole story is due to a misunderstanding of what is meant by 'inertia': one can interpret it as inertial *mass* or as inertial *motion* (like the moving arms under the stars). There is no evidence that Mach believed either in non-isotropic mass nor in distance-dependent mass; the whole discussion is an example of the frequent game consisting of being proud of not making a mistake which is incorrectly imputed to a supposedly more stupid other person. At school one usually hears that Columbus was derided because he thought the earth to be spherical. But he was not derided at all for this reason; there were only disagreements

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

on the *size* of the earth, and in fact it turned out that his critics were right, and that he was wrong with his own, much too small radius estimate.

The same happened with Mach's principle. Obviously, inertial effects do depend on the distribution of mass in the rest of the universe. Mach's principle is correct. Mach made some blunders in his life (he is in-famous for fighting the idea of atoms until he died, against experimental evidence) but his principle is *not* one of them, in contrast to the story told in many textbooks. But it is to be expected that the myth about the incorrectness of Mach's principle will persist, like that of the derision of Columbus. Ref. 73

In fact, Mach's principle is valuable. As an example, take our galaxy. Experiments show that she is flattened and rotating. The sun turns around her centre in about 250 million years. Indeed, if the sun would not turn around the galaxy's centre, we would fall into it in about 20 million years. As the physicist Dennis Sciama pointed out, from the shape of our galaxy we can take a powerful conclusion: there must be a lot of other matter, i.e. a lot of other stars and galaxies in the universe. Are you able to confirm his reasoning? Challenge

Resting in the universe

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

In other words, the night sky is black because we move with almost no speed through background radiation. If the earth had a large velocity relative to the background radiation, the sky would be bright at night, at least in certain directions. Can you confirm this? Challenge

The reason why the galaxy and the solar system move with small speed across the universe has been already studied in our walk. Can you give a summary? Challenge

Does light attract light?

Another reason that we can see stars is that their light reaches us. Do parallel light beams remain parallel? If light is energy, and energy attracts energy through gravitation, light should *attract* light. That could have strange effects on the light emitted by stars.

Interestingly, a precise calculation shows that gravitation does *not* alter the path of two parallel light beams, even though it *does* alter the path of antiparallel light beams. The reason is that for parallel beams moving at light speed, the gravitomagnetic component exactly *cancels* the gravitoelectric component. Ref. 75 Challenge

Since light does not attract light moving along, light will not be disturbed by its own gravity during the millions of years that it takes from distant stars to reach us.

Does light decay?

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

Challenge But any decay would also mean that light would change its direction (why?) and thus produce blurred images for far away objects. However, no blurring is observed. In addition, the soviet physicist ... Bronstein demonstrated in the 1930s that any light decay process would have a larger rate for smaller frequencies. So people checked the shift of radiowaves, in particular the famous 21 cm line, and compared it with the shift of light from the same source. No difference was found for all galaxies tested. Ref. 76

People even checked that Sommerfeld's fine structure constant, the constant of nature which determines the colour of objects, does not change over time. No sizeable effect could be detected over thousand of millions of years. Ref. 77

Of course, instead of decaying, light could also be *hit* by some so far unknown entity. But also this case is excluded by the just presented arguments. In addition, these investigations show that there is no additional red shift mechanism in nature apart from Doppler and gravitational red shifts. Challenge

In summary, the fact that we can see the stars at night yields numerous properties of nature. We now continue our escalation with a more fundamental issue, nearer to our main topic, namely the fundaments of motion.

11. Does space differ from time?

Tempori parce.
Seneca*

People in bad mood say that time is our master. Nobody says that of space. Time and space are obviously different in everyday life. But what is the precise difference between them in general relativity? And do we need them at all? These questions by themselves form an important topic.

In general relativity it is assumed that we live in a (pseudo-riemannian) space-time of variable curvature. The curvature is an observable and is related to the distribution and motion of matter and energy in the way described by the field equations.

However, there is a fundamental problem. The equations of general relativity are invariant under numerous transformations which *mix* the coordinates x_0, x_1, x_2 and x_3. For example, the transformation

$$\begin{aligned} x'_0 &= x_0 + x_1 \\ x'_1 &= -x_0 + x_1 \\ x'_2 &= x_2 \\ x'_3 &= x_3 \end{aligned} \qquad (297)$$

is allowed in general relativity, and leaves the field equations invariant. You might want to find other examples. Challenge

The consequence is clear: diffeomorphism invariance makes it *impossible* to distinguish space from time *inside* general relativity. This surprising conclusion is in sharp contrast with everyday life.

* 'Care about time.' Ep. 88, 39

More explicitly, the coordinate x^0 cannot simply be identified with the physical time t, as implicitly done up to now. This identification is only possible in *special* relativity. In special relativity the invariance under Lorentz (or Poincaré) transformations of space and time singles out energy, linear, and angular momentum as the fundamental observables. In general relativity, there is *no* metric isometry group; consequently, there are *no* basic physical observables singled out by their characteristic of being conserved. But invariant quantities are necessary for communication! In fact, we can *talk* to each other only because we live in an approximately *flat* space-time. If the angles of a triangle would not add up to 180 degrees, we could not communicate, since there would be no invariant quantities.

So how did we sweep this problem under the rug so far? We used several ways. The simplest way was to always require that in some part of the situation under consideration space-time is our usual flat Minkowski space-time, where x_0 can be set equal to t. This requirement can be realized either at infinity, as we did around spherical masses, or in zeroth approximation, as we did for gravitational radiation and for all other perturbation calculations. In this way, the free mixing of coordinates is eliminated and the otherwise missing invariant quantities appear as expected. This pragmatic approach is the usual way out of the problem. In fact, there are otherwise excellent texts on general relativity refusing any deeper questioning of the issue. Ref. 62

A common variation of this trick is to let the distinction 'sneak' into the calculations by the introduction of matter and its properties, or by the introduction of radiation. Both matter and radiation distinguish between space and time simply by their presence. The material properties of matter, for example their thermodynamic state equations, always distinguish space and time. Radiation does the same, by its propagation. Obviously this is true also for those special combinations of matter and radiation called clocks and meter bars. In fact, the method of introducing matter is the same as the one introducing Minkowski space-time, if one looks closely: matter properties are always defined using space-time descriptions of special relativity.*

Still another variation of the pragmatic approach is the use of the cosmological time coordinate. An isotropic and homogeneous universe does have a preferred time coordinate, namely the one used in all the tables on the past and the future of the universe. Also here one is in fact using a combination of the previous two ways.

But we are on a special escalation here. We want to *understand* motion, not only to calculate its details. We want a *fundamental* answer, not a pragmatic one. And for this we need to know how the x_i and time t are connected, and how we can define invariant quantities. This question prepares us for the moment when gravity is combined with quantum theory, as we will do in the third part of the escalation.

A fundamental solution requires to describe clocks together with the system under consideration, and deduce how the reading t of the clock relates to the behaviour of the system in space-time. But we note that any description of a system requires measurements, e.g. to determine the initial conditions. We enter a vicious circle, since that is what we wanted to avoid in the first place.

* We note something astonishing here: the inclusion of some condition at small distances (matter) has the same effect as the inclusion of some condition at infinity. Is this a coincidence? We will come back to this issue in the third part of the escalation. Challenge

We get a suspicion. Does a fundamental difference between space and time exits at all? Let us have a tour of the various ways to investigate the question.

Can space and time be measured?

See page 144

In order to distinguish space and time in general relativity, one must be able to measure them. But already in the section on universal gravity we had mentioned the impossibility of measuring lengths, times and masses with gravitational effects alone. Does this situation change in general relativity? Lengths and times are connected by the speed of light, and in addition lengths and masses are connected by the gravitational constant. Despite this additional connection, it takes only a moment to convince oneself that the problem persists. In fact, one needs *electrodynamics* to solve it. Only using the electromagnetic charge e one can form length scales, of which the most simple one is given by

$$l_{\text{scale}} = \frac{e}{\sqrt{4\pi\varepsilon_o}} \frac{\sqrt{G}}{c^2} \approx 1.4 \cdot 10^{-36}\,\text{m} \quad . \tag{298}$$

In fact, only *quantum mechanics* provides a solution to this issue, as can be seen by rewriting the elementary charge e as the combination of nature's fundamental constants, namely

$$e = \sqrt{4\pi\varepsilon_o c\hbar\alpha} \quad , \tag{299}$$

which changes expression (298) into

$$l_{\text{scale}} = \sqrt{\frac{\alpha\hbar G}{c^3}} = \sqrt{\alpha}\, l_{\text{Pl}} \quad . \tag{300}$$

Challenge

The expression shows that every length measurement is based on the electromagnetic coupling constant and on the Planck length. Of course the same is valid for time and mass measurements. There is no way to define or measure lengths, times and masses in general relativity.* Therefore, the answer to the section title being negative in general relativity, the next question is:

Are space and time necessary?

Ref. 78

Robert Geroch answers this question in a beautiful five-page article. He explains how to formulate the general theory of relativity without the use of space and time, by taking as starting point the physical observables only.

He starts with the set $\{a\}$ of all observables. Among them there is one, called v, standing out, because it allows to say that for any two observables a_1, a_2 there is a third one a_3, for which

$$(a_3 - v) = (a_1 - v) + (a_2 - v) \quad . \tag{301}$$

* In the past, John Wheeler tended to state that his geometrodynamic clock was a counterexample; that is not correct, however.

Such an observable is called the *vacuum*. Once such an observable is known, Geroch shows how to use it to construct the derivatives of observables. Then the so-called Einstein algebra can be built, which comprises the whole of general relativity.

Usually one describes motion by deducing space-time from matter observables, by calculating the evolution of space-time, and then by deducing the motion of matter following from it. Geroch's description shows that the middle step, the use of space and time, is not necessary.

What does one conclude? It is possible to formulate general relativity without the use of space and time. But if they are both unnecessary, it is unlikely that there is fundamental difference between them. Still, one difference between time and space is well-known:

Do closed timelike curves exist?

In other words, is it possible that the time coordinate behaves, at least in some regions, like a torus? Is it possible, like in space, to come back in time from where one has started?

The question has been studied in great detail. The standard reference is the text by Hawking and Ellis; they list the various properties of spacetime compatible with each other or excluding each other. Among others, they find that space-times which are smooth, globally hyperbolic, oriented and time-oriented do not contain any such curves. It is usually assumed that this is the case for the universe, so that nobody expects to observe closed timelike curves. Ref. 54

That seems to point to a difference. But in fact, these investigations do not help: they are based on the behaviour of matter and thus add in from the start what were looking for. In short, also this topic cannot help to decide whether space and time differ. So let us look at the issue in another way.

Is general relativity local? – The hole argument

When Albert Einstein developed general relativity, he quite some trouble with diffeomorphism invariance. Most startling is his famous *hole argument*, better called the *hole paradox*. Take the situation shown in figure 105, in which a mass deforms the space-time around it. Einstein imagined a small region of the vacuum, the *hole*, which is shown with a dotted line. What happens if one changes the

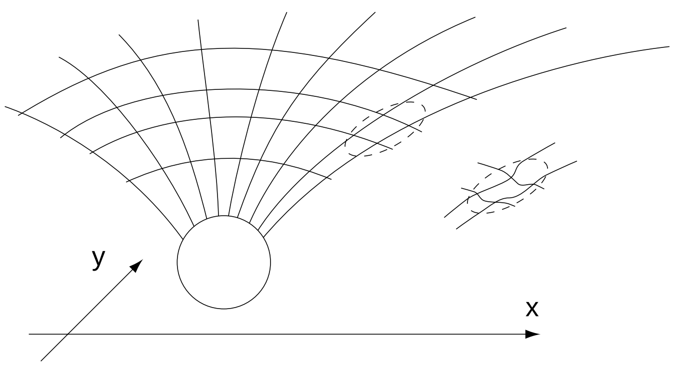


Figure 105 The hole

curvature inside the hole while leaving the situation outside it unchanged, as shown in the inset of the picture? Ref. 79

- One one hand, the new situation is obviously physically different from the original one, as the curvature inside the hole is different. This difference thus implies that the curvature around a region does not determine the curvature inside it. That is extremely unsatisfactory. Worse, if one generalizes this operation to the time domain, one gets the biggest nightmare possible in physics: determinism is lost.

- On the other hand, general relativity is diffeomorphism invariant. The deformation shown in the figure is a diffeomorphism. The situation must be physically equivalent to the original situation.

Who is right? Einstein first favored the first point of view, and therefore dropped the whole idea of diffeomorphism invariance for about a year. Only later he understood that the second assessment is correct, and that the first statement makes a fundamental mistake.

Indeed, the first opinion arrives to the conclusion that the two situations are different because it assumes an independent existence of the coordinate axes x and y shown in the figure. But during that deformation, the coordinates x and y automatically change as well, so that there is *no* physical difference between the two situations.

The moral of the story is that *there is no difference between space and gravitational field*. Space-time is a quality of the field, as Einstein put it, not an entity with separate existence.

This is often difficult to spot. For example, it turns out that many problems in general relativity are equivalent to the Schwarzschild situation, even though they appear completely different. As a result, researchers have "discovered" the Schwarzschild solution (of course with different coordinate systems) over twenty times, often thinking that they had found a new, unknown solution.

Diffeomorphism invariance will play a central role in the third part of our escalation. But already here the topic has a startling consequence:

Is the earth hollow?

The *hollow earth hypothesis*, i.e. the conjecture that we live on the *inside* of a sphere, was popular in paranormal circles around the year 1900, and still is among certain crackpots today. These strange people propagate the idea on the internet, explaining that humans live on the *inside* of a sphere, with the sun somewhere on the way to the center, the stars even nearer to the center, etc. They state that we are fooled by education into the usual description, because we are brought up to believe that light travels in straight lines. Get rid of this belief, they say, and the hollow earth appears in all its glory.

Interestingly, there is *no way* to disprove this sort of description of the universe. In fact, the diffeomorphism invariance of general relativity even proclaims the equivalence between the two views. The fun starts when either of the two camps wants to tell the other than *only* its own description is correct. You might check that any such argument is wrong; it is fun to slip into the shoes of such a crackpot and to defend the hollow earth hypothesis against your friends. Both descriptions are exactly equivalent. Challenge

Are you able to confirm that even quantum theory, with its introduction of length scales into nature, does not change the situation? In fact, in the third part of our walk we will Challenge discover that there is a symmetry in nature quite similar to the change in viewpoint from the hollow earth view of crackpots to the standard view of physicists and back! See page 626

Are space, time and mass independent?

We conclude from this short discussion that there does not seem to be a fundamental distinction between space and time in general relativity. Pragmatic distinctions, using matter, radiation, or space-time at infinity are the only possible ones.

In the third part we will discover that even the inclusion of quantum theory is consistent with this view. We will show explicitly that no distinction is possible in principle. In addition, we will show that the origin of this impossibility is that mass and space-time are on equal footing and that in a sense, particles and vacuum are made of the same substance. The argument for these connections only takes a few lines. All distinctions turn out to be possible only at low, daily life energies. See page 595

In other words, in the beginning of our escalation we saw that we needed matter to define space and time. Now we even found that we need matter to *distinguish* space and time. Similarly, in the beginning we saw that we need space and time to define matter; now we found that we even need *flat* space-time to define it. See page 98

In summary, general relativity does not answer several important questions about motion; it even makes the matter less clear than before! Continuing the escalation is really worth the effort. To increase our understanding, we now tackle a completely different topic.

12. Black holes – falling forever

Why study them?

Black holes are the extreme case of general relativity; they realize the limit of length to mass ratio possible in nature. Therefore, they cannot be studied *without* general relativity. But in addition, black holes are a central stepping stone towards unification and the final description of motion. Strangely enough, for many years their existence was in doubt. The present experimental situation has lead most experts to conclude that there is one at the centre of at least 15 nearby galaxies, including our own; in addition, half a dozen smaller black holes have been identified scattered inside our own galaxy. Interestingly, gravity theory can show that black holes do not really exist! For this and many other reasons, black holes, the most impressive and the most relativistic systems in nature, are a fascinating subject of study.* Ref. 82

Horizons and orbits

An object whose escape velocity is larger than the speed of light c is called a *black hole*. They were first imagined by the british geologist John Michell in 1784 and independently by the french mathematician Pierre Laplace in 1795, long before general relativity. Even if they were hot shining stars, they would appear to be *black*, and not be visible in the sky. It only takes a short calculation to show that light cannot escape form a mass whenever the radius is smaller than a critical value given by Ref. 83 Challenge

* An excellent and entertaining book on the topic, without any formula, but nevertheless accurate and detailed, is the paperback by Igor NOVIKOV, *Black holes and the universe*, Cambridge University Press, 1990.

$$R_S = \frac{2GM}{c^2} \qquad (302)$$

the so-called Schwarzschild radius. The formula is valid both in universal gravity and in general relativity, provided that in general relativity one takes the radius as meaning the circumference divided by 2π. That is exactly the limit value for length to mass ratios in nature. For this and other reasons to be given shortly, we will call R_S also the *size* of the black hole of mass M (although properly speaking it is only half the size). In principle, an object could be imagined to be smaller than this value; but nobody has observed one. As a note, the surface gravity of a black hole is given by

Challenge

$$g_{surf} = \frac{c^4}{4GM} = \frac{c^2}{2R_S} \quad . \qquad (303)$$

Such a black star thus swallows what falls into it, be it matter or radiation, without letting anything out. It acts like a cosmic trash can. In 1967, John Wheeler* called them *black holes*.

As it is impossible to send light from a black hole to the outside world, what happens when a light beam is sent upwards from the horizon? And from slightly above the horizon?

Challenge

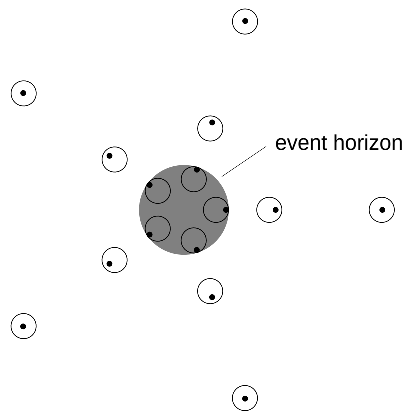


Figure 106 The light cones in the equatorial plane around a non-rotating black hole, seen from above

Black holes, when seen as astronomical objects, are thus different from planets. During the formation of planets, matter lumped together and as soon as it could not be compressed any further, equilibrium was formed, determining the radius of the planet. That is the same mechanism as when a stone is thrown towards the earth: it stops falling when it *hits* the ground thus formed. The bottom is reached when matter hits other matter. In the case of a black hole, there is no ground; everything *continues* falling. This happens, as we will see in the part on quantum theory, when the concentration of matter is so large that the forces which make matter *impenetrable* in daily life are overcome. As british physicist Freeman Dyson says, *a black hole is matter in permanent free fall.* In russian, black holes used to be called *collapsars*. Note that despite this permanent free fall, their radius remains constant. Due to this permanent free fall, black holes are the only state of matter in thermodynamic equilibrium! All other states are metastable. In 1939, Robert Oppenheimer and Hartland Snyder showed theoretically that a black hole forms when a star of sufficient mass stops burning.

Ref. 85

The characterizing property of a black holes is its *horizon*. We have encountered horizons already in special relativity. Here the situation is similar; for an outside observer, the horizon

* John Archibald Wheeler (1911–) US american physicist, important expert on general relativity and author of several excellent textbooks, among them the beautiful John A. Wheeler, *A journey into gravity and spacetime*, Scientific American Library & Freeman, 1990, in which he explains general relativity with passion and in detail, but without any mathematics.

is the surface beyond which he cannot receive signals. For black holes, the horizon is located at the gravitational radius.

The proof comes from the field equations. They lead to a space-time around a rotationally symmetric, thus non-rotating, and electrically neutral mass described by the Schwarzschild metric See page 214

$$di^2 = (1-\frac{2GM}{rc^2})dt^2 - \frac{dr^2}{1-\frac{2GM}{rc^2}} - r^2d\varphi^2/c^2 \quad . \tag{304}$$

As mentioned above, r is the circumference divided by 2π, and t is the time measured at infinity. However, no *outside* observer will ever receive any signal emitted from $r = 2GM/c^2$ (or smaller)! Indeed, as the proper time of an observer at radius r is related to the one of an observer at infinity through

$$di = \sqrt{1-\frac{2GM}{rc^2}}\,dt \quad , \tag{305}$$

one finds that an observer at the horizon would have zero proper time. In other words, at the horizon one has infinite redshift. Everything happening there goes on infinitely slowly for a far away observer. More clearly, for a faraway observer nothing at all happens at the horizon.

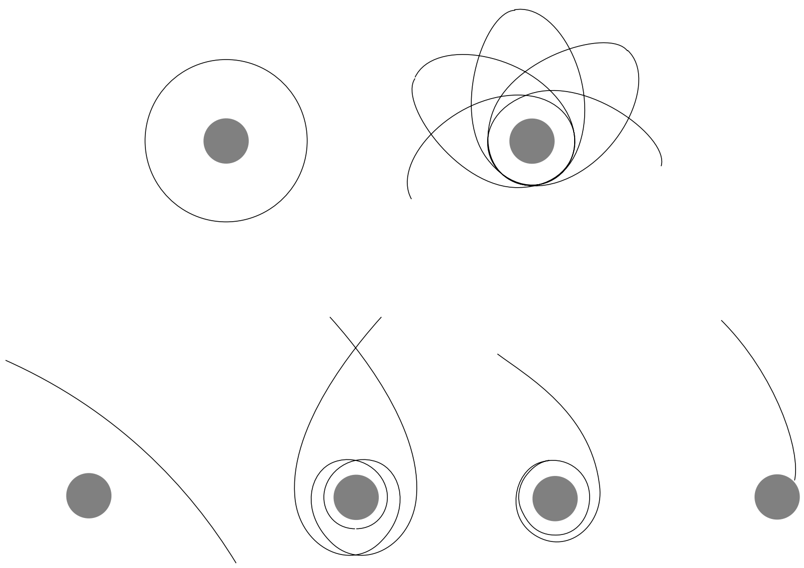


Figure 107 Motion of uncharged objects around a nonrotating black hole

Since black holes curve space-time strongly, a body moving near a black hole behaves in more complicated ways than in the case of universal gravity. In universal gravity, paths are either ellipses, parabolas, or hyperbolas; all these are plane curves. It turns out that paths lie Ref. 76

in a plane only near *nonrotating* black holes.* Around such black holes, circular paths are impossible for radii less than $3R/2$ (can you show why?) and are unstable to perturbations from there up to a radius $3R$. Only at larger radii circular orbits are stable. Around black holes, there are no elliptic paths; the corresponding rosetta path is shown in figure 107. Such a path shows the famous periastron shift in all its glory.

Challenge

Note that the potential around a black hole is not appreciably different from $1/r$ for distances above about fifteen Schwarzschild radii. For a black hole of the mass of the sun, that would be 42 km from its centre; at the distance of the earth, we would not be able to note any difference in the gravitation of the earth.

Challenge

For falling bodies coming from infinity, the situation near black holes is even more interesting. Of course there are no hyperbolic paths, only trajectories similar to hyperbolas for bodies passing far enough. But for small, but not too small impact parameters, a body will make a number of turns around the black hole, before leaving again. The number of turns increases beyond all bounds with decreasing impact parameter, until a value is reached at which the body is captured into an orbit at a radius of $2R$, as shown in figure 107. In other words, this orbit captures incoming bodies if they reach it below a certain critical angle. For comparison, remember that in universal gravity, no capture exists. At still smaller impact parameters, the black hole swallows the incoming mass. In both cases, capture and deflection, a body can make several turns around the black hole, whereas in universal gravity, it is impossible to make more than *half* a turn around a body.

The most absurd looking orbits though are those (purely academic) orbits corresponding to the parabolic case of universal gravity. Relativity changes the motions due to gravity quite drastically.

Challenge

Around *rotating* black holes, the orbits of point masses are even more complex than those shown in figure 107; for bound motion for example, the ellipses do not stay in one plane, but also change – due to the Thirring-Lense effect – the plane in which they lie, leading to extremely involved orbits in three dimensions filling the space around the black hole.

For *light* passing a black hole, the paths are equally interesting, as shown in figure 108. there are only few differences with the case of rapid particles. For a non-rotating black hole, the path lies in a single plane. Of course, there is strong bending of light, as well as capture, if light passes sufficiently nearby. Again, light can also make one or several turns around the black hole before leaving or being captured. The limit between the two cases is the path in which light moves in a circle around a black hole, at $3R/2$, albeit in an unstable orbit. If one would be located on that orbit, one would see the back of one's head by looking forward! The set of all those orbits is called the *photon sphere*. It thus divides path leading to capture from those leading to infinity. As a note, there is no *stable* orbit for light around a black hole. Are there any rosetta paths for light around a black hole?

Challenge

Challenge

* For such paths, Kepler's rule connecting the average distance and the time of orbit

$$\frac{GMt^3}{(2\pi)^2} = r^3 \qquad (306)$$

still holds, provided the proper time and the radius measured by a faraway observer is used.

Challenge

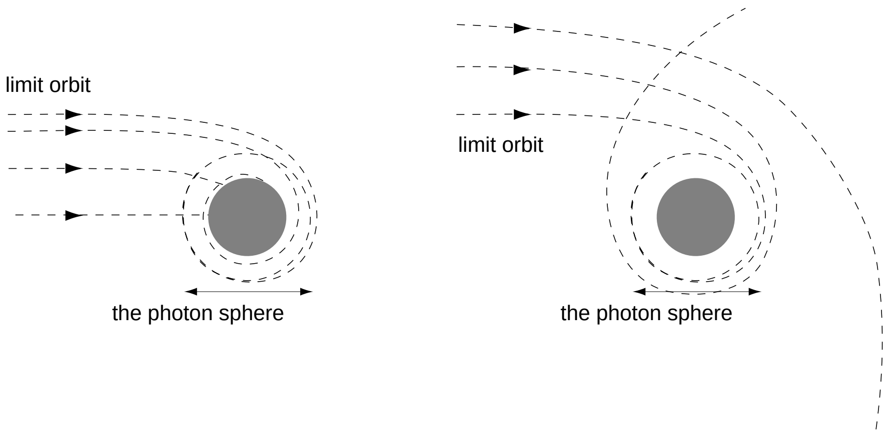


Figure 108 Motion of *light* passing near a non-rotating black hole

Around a *rotating* black hole, paths are much more complex. Already in the equatorial plane there are two possible circular light paths, namely a smaller one in direction of the rotation, and a larger one in the opposite direction. Challenge

For *charged* black holes, the orbits for falling charged particles are even more complex. One has to study the electrical field lines; several fascinating effects appear with no correspondence in usual electromagnetism, such as effects similar to electrical versions of the Meissner effect. The whole field is still mostly unexplored and is one of today´s research themes in general relativity.

But this is enough about orbits. Let us continue with another topic.

Hair and entropy

How is a black hole characterized? It turns out that black holes have no choice for their size, their shape, their colour, their magnetic field and all their material properties to be discussed later on. They all follow from the few properties characterizing them, namely their mass M, their angular momentum J, and their electrical charge Q.* All other properties are uniquely determined by them.** It is as though one could deduce every characteristic of a woman only by her size, her waist, and her hair colour. Physicist thus to say that black holes "have no hair," following Wheeler's colourful language.*** This was shown by Israel, Carter, Robinson and Mazur; they showed that for a black hole with given mass, angular momentum and charges, there is only *one* possible black hole. (The uniqueness theorem is Ref. 89

∗ There are other entities encountered so far with the same reduced number of characteristics: particles. More on this connection will be uncovered in the third part of our escalation.

∗∗ Mainly for marketing reasons, neutral non-rotating and electrically neutral black holes are often called *Schwarzschild* black holes: uncharged and rotating ones are often called *Kerr* black holes, after Roy Kerr,
Ref. 86 who discovered the corresponding solution of Einstein's field equations in 1963. Electrically charged, but non-rotating black holes are also called *Reissner-Nordstrom black holes*, after the german H. Reissner and the danish
Ref. 87 G. Nordström. The general case, charged and rotating, is often named after Kerr and Newman.

∗ ∗ ∗ It is not a secret that Wheeler was inspired by a clear anatomical image when he says that "black holes, in contrast to their surroundings, have no hair."

not valid any more if the black holes carries nuclear quantum numbers, such as weak or
Ref. 90 strong charges.)

In other words, independently of how the black holes has formed, independently of which material and composition was used when building it, the final result does not depend on those details. Black holes all have identical composition, or better, they have no composition at all (at least classically). More about this topic shortly.

The mass of a black hole is not restricted by general relativity. It may be as small as that of a microscopic particle, and as large as several solar masses. But for angular momentum and for for electric charge the situation is different. A rotating black hole has a maximum possible angular momentum and a maximum possible electrical (and magnetic) charge.* The limit in angular momentum appears as its perimeter may not move faster than light.
Challenge Also for the charge there is a limit. The two limits are not independent; they follow from

$$\left(\frac{J}{cM}\right)^2 + \frac{GQ^2}{4\pi\varepsilon_o c^4} \leqslant \left(\frac{GM}{c^2}\right)^2 \qquad (307)$$

Black holes realizing the limit fast are called *extremal* black holes. The limit (307) simply
Challenge follows from the limit on length to mass ratios at the basis of general relativity which implies that the horizon radius of a general black hole is given by

$$r_h = \frac{GM}{c^2}\left(1 + \sqrt{1 - \frac{J^2c^2}{M^4G^2} - \frac{Q^2}{4\pi\varepsilon_o GM^2}}\right) \qquad (308)$$

For example, for a black hole with the mass $2 \cdot 10^{30}$ kg and angular momentum $\ldots \cdot 10^{\cdots}$ kg m^2/s^2 of the sun, the charge limit is about … C.

How does one distinguish rotating from non-rotating black holes? First of all by the *shape*. Non-rotating black holes must be spherical (any non-sphericity is radiated away as gravitational waves) and rotating black holes have a slightly flattened shape, uniquely determined by their angular momentum. Due to their rotation, their surface of infinite gravity or infinite redshift, called the *static limit*, is different from their horizon. The region in between, the *ergosphere* as the name does not say, it is *not* a sphere. (It is called this way because, as we will see shortly, it can be used to extract energy from the black hole.) The motion of bodies between the ergosphere and the horizon can be quite complex. It suffices to mention that rotating black holes drag any infalling body into an orbit around them, in contrast to nonrotating black holes, which swallow them. In other words, rotating black holes are not really ‘holes’ at all, but rather black vortices.

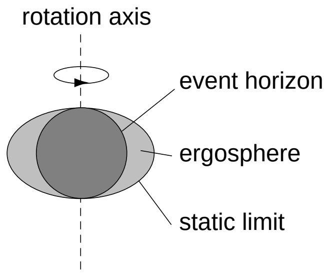


Figure 109 The ergosphere of a rotating black hole

The distinction between rotating and non-rotating black holes also appears in the horizon surface area. The (horizon) surface of a non-rotating and uncharged black hole is obviously
Challenge related to its mass by

* More about the still hypothetical magnetic charge later on. It enters like an additional type of charge into all expressions.

$$A = \frac{16\pi G^2}{c^4} M^2 \quad . \tag{309}$$

The surface-mass relation for a rotating and charged black hole is more complex; it is given by

$$A = \frac{8\pi G^2}{c^4} M^2 \left(1 + \sqrt{1 - \frac{J^2 c^2}{M^4 G^2} - \frac{Q^2}{4\pi\varepsilon_o G M^2}} \right) \tag{310}$$

where J is the angular momentum. In fact, the relation

$$A(M,J,Q) = \frac{8\pi G}{c^2} M r_h \tag{311}$$

is valid for *all* black holes, even if charged and rotating. Obviously, in the case of electrically charged black holes, the rotation also produces a magnetic field around them. This is in contrast with non-rotating black holes; they cannot have a magnetic field.

Can one extract energy from a black hole? Roger Penrose discovered that this is possible for *rotating* black holes. A rocket orbiting a rotating black hole in its ergosphere could switch its engines on and then would get hurled into outer space at tremendous velocity, much greater than what the engines could have produced by themselves. In fact, the same effect is used by rockets on the earth as well and is the reason that all satellites orbit the earth in the same direction; it would cost much more fuel to let them turn the other way.* Anyway, the energy gained by the rocket is lost by the black hole, which thus slows down and would lose some mass; on the other hand, the mass increases due to the exhaust gases falling into the black hole. This increase always is larger or at best equal to the loss due to rotation slowdown. The best one can do is to turn the engines on exactly at the horizon; then the horizon area of the black hole stay constant, and only the rotation is slowed down.** Ref. 91

As a result, for a neutral black hole *rotating* with its maximum possible angular momentum, $1 - 1/\sqrt{2} = 29.3\%$ of its total energy can be extracted through the Penrose process. For black holes rotating more slowly, the percentage is obviously smaller. Challenge

For *charged* black holes, such irreversible energy extraction processes are also possible. Can you think of a way? Inserting value (307), one finds that up to 50% of the mass of a non-rotating black hole can be due to its charge. In fact, in the second part of the escalation we will encounter a process which nature seems to use quite frequently. Challenge Challenge See page 522

The Penrose process allows to determine how angular momentum and charges increase the mass of a black hole. The result is the famous mass-energy relation Ref. 92

$$M^2 = \frac{E^2}{c^4} = (m_{\text{irr.}} + \frac{Q^2}{16\pi\varepsilon_o G m_{\text{irr.}}})^2 + \frac{J^2}{4m_{\text{irr.}}^2}\frac{c^2}{G^2} = (m_{\text{irr.}} + \frac{Q^2}{8\pi\varepsilon_o \rho_{\text{irr.}}})^2 + \frac{J^2}{\rho_{\text{irr.}}^2}\frac{1}{c^2} \tag{312}$$

∗ And it would be much more dangerous, since any small object would hit such an against-the-stream satellite with about 15.8 km/s, thus transforming any small object into a dangerous projectile. In fact, any power wanting to destroy satellites of the enemy would simply have to load a satellites with nuts or bolts, send it into space the wrong way, and distribute the bolts into a cloud. It would make satellites impossible for many decades to come. Challenge

∗∗ It is also possible to extract energy from rotational black holes through gravitational radiation.

which shows how the electrostatic and the rotational energy enter the mass of a black hole. In the expression, $m_{\text{irr.}}$ is the *irreducible mass* defined as

$$m_{\text{irr.}}^2 = \frac{A(M,0,0)}{16\pi} \frac{c^4}{G^2} = \left(\rho_{\text{irr.}} \frac{c^2}{2G} \right)^2 \qquad (313)$$

and $\rho_{\text{irr.}}$ is the *irreducible radius*.

These investigations showed that there is no process which *decreases* the horizon area and thus the irreducible mass or radius of the black hole. People have checked this in all ways possible and imaginable. For example, when two black holes merge, the total area increases. One calls processes which keep the area and energy of the black whole constant *reversible*, and all others irreversible. In fact, the area of black holes behaves like the *entropy* of a closed system: it never decreases. That the area in fact *is* an entropy was first stated in 1970 by Jakob Bekenstein. He deduced that only when an entropy is ascribed to a black hole, it was possible to understand where the entropy of all the material falling into it was collected.

Ref. 93

Again, the value of the entropy is a function only of the mass, the angular momentum and the charge of a black hole. You might want to confirm Bekenstein's deduction that the entropy is proportional to the horizon area. Later it was found, using quantum theory, that

Challenge

$$S = \frac{A}{4} \frac{c^6}{\hbar G} = \frac{A}{4 l_{\text{Pl}}^2} \quad . \qquad (314)$$

This famous relation needs quantum theory for its deduction, as the absolute value of entropy is never fixed by classical physics alone. We will discuss it later on in our escalation.

See page 523

If black holes have an entropy, they also must have a temperature. If they have a temperature, they must emit black body radiation. Black holes cannot be black! The last conclusion was proven by Stephen Hawking in 1974 with extremely involved calculations. However, it could have been deduced already in the 1930s, as the following Gedankenexperiment shows.

– CS – to be completed – CS –

More interesting connections between black holes, thermodynamics, and quantum theory will be presented in the second part of the escalation. Of course, black hole radiation is a further, though tiny (quantum) mechanism for energy extraction, and is applicable even for non-rotating, uncharged black holes.

See page 521

Paradoxes

Tiens, les trous noirs. C'est troublant.*
Anonyme

Black holes show many counterintuitive results.**

* No translation possible.
** Other paradoxes, including the quantum effects, are discussed on page 526.

• What happens to a person falling into a black hole? An outside observer gives a clear answer: the falling person *never* arrives there since it needs an infinite time to reach the horizon. Can you confirm this result? The falling observer however, reaches the horizon in a *finite* amount of his own time. (Can you calculate it?) Challenge Challenge

This is surprising, as it means that for an outside observer in a universe with *finite* age, black holes cannot have formed yet! At best, one can only observe systems busy forming one. In a sense, it is thus correct to say that black holes do not exist.

There is one way out: black holes could have existed right from the start in the fabric of space-time. We will find out later why this is impossible. In other words, it is important to keep in mind that the idea of black hole is an *approximation*.

Independently of this last issue, we can confirm that in nature, the length to mass ration always follows

$$\frac{L}{M} \geqslant \frac{4G}{c^2} \quad . \tag{315}$$

• Interestingly, the *size* of a person falling into a black hole is also experienced in vastly different ways by the falling person and the one staying outside. If the black hole is large, the infalling observer feels almost nothing, as the tidal effects are small. The outside observer makes a startling observation: he sees the falling person *spread* all over the horizon of the black hole. *Infalling, extended bodies cover the whole horizon.* Can you explain the result, e.g. by using the limit on length to mass ratios? Challenge

This strange result will be of importance later on in our walk, and lead to important conclusions.

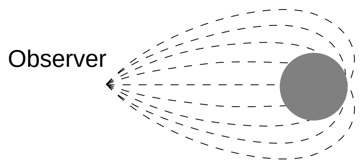


Figure 110 Motion of some light rays from a dense body to an observer

• An observer near a (non-rotating) black hole, or in fact near any object smaller than 7/4 times its gravitational radius can see even the complete *back* side of the object, as shown in figure 110. Can you imagine how the image looks? Note that in addition to the paths shown in figure 110, light can also turn several times around the black hole before hitting its surface! Therefore, such an observer sees an infinite number of images of the black hole. The formula for the angular size of the innermost image was given above. Challenge See page 220

In fact, gravity has the effect to allow the observation of more than half a sphere of any object. In everyday life the effect is not so large, however; for example, light bending allows to see about 50.0002% of the surface of the sun.

• A mass point inside the smallest circular path of light around a black hole, at $3R/2$, cannot stay in a circle, because in that region, something strange happens. A body who circles another in everyday life always feels a tendency to be pushed outwards; this centrifugal effect is due to the inertia of the body. But at values below $3R/2$, a circulating body is pushed *inwards* by its inertia. There are several ways to explain this paradoxical effect. The simplest is to note that near a black hole, the weight increases faster than the centrifugal force, as you may want to check yourself. Only a rocket with engines switched on can orbit at $3R/2$. Ref. 94 Challenge

• By the way, how can gravity or an electrical field come out of a black hole, if no signal and no energy can leave a black hole? Challenge

Formation and search for black holes

How might black holes form? At present, at least three mechanisms are distinguished; the question is still subject of research. First of all, black holes could have formed during the early stages of the universe. These *primordial black holes* might grow through *accretion*, i.e. through the swallowing of nearby matter and radiation, or disappear through one of the mechanisms to be studied later.

Ref. 95

See page 522

Of the *observed* black holes, the so-called *supermassive black holes* are found at the center of every galaxy studied so far. They have masses in the range of 10^6 to 10^9 solar masses. They are conjectured to exist at the center of all galaxies and seem to be related to the formation of galaxies themselves. This is a hot topic of research. Supermassive black holes are supposed to have formed through the collapse of large dust collections, and to have grown through subsequent accretion of matter. There might also be a relation to quasars; it might be that quasars are black holes.

Ref. 96 On the other hand, black holes can form when old massive stars *collapse*. It is estimated that when stars with at least three solar masses burn out their fuel, the matter will collapse into a black holes. Such *stellar* black holes have a mass between one and a hundred solar masses; they can also continue growing through subsequent accretion. This situation provided the first candidate ever, Cygnus X-1, which was discovered in 1971.

Recent measurements suggest also the existence of *intermediate black holes*, with masses around thousand solar masses or more; their formation mechanisms are still unknown.

The search for black holes is a popular sport among astrophysicists. The conceptually simplest way to search for them is to look for strong gravitational fields. But only double stars allow to measure fields directly, and the strongest ever measured gravitational field so far is 30% of the theoretical maximum value.

Ref. 97

Another way is to look for strong gravitational lenses, and try to get a mass to size ratio pointing to a black hole.

Still another way is to look at the dynamics of stars near the center of galaxies. Measuring their motion, one can deduce the mass of the body they orbit.

The most favourite method is to look for extremely intense X-ray emission from point sources through space-based satellites or balloon based detectors. If the distance to the object is known, its absolute brightness can be deduced; if it is above a certain limit, it must be a black hole, since normal matter cannot produce an unlimited amount of light.

Obviously, the direct observation of energy disappearing into a horizon is also a way to search for a black hole. This might have been observed recently.

Ref. 98

At present, measurements show that in all galaxies studied so far – more than a dozen – a supermassive black hole seems to be located at their centre. The masses vary; the black hole at the centre of our own galaxy has about 2.6 million solar masses. The central black hole of the galaxy M87 has 3 thousand million solar masses.

Ref. 82

About a dozen stellar black holes between 4 and 20 solar masses are known in the rest of our own galaxy, all discovered in the years after 1971, when Cygnus X-1 was found. In 2000, a couple of intermediate mass black holes have also been found. This list of discoveries, as well as the related results are expected to expand dramatically in the coming years.

Ref. 82

Singularities

Solving the equations of general relativity, one finds that for many classes of initial conditions, a cloud of dust collapses to a *singularity*, i.e. to a point of infinite density. The same conclusion appears when one follows the evolution of the universe backwards in time. In fact, Roger Penrose and Stephen Hawking have proven several mathematical theorems on the necessity of singularities for many classical matter distributions. These theorems assume only the continuity of space-time and a few rather weak conditions on the matter in it. The theorems state that in expanding systems such as problably the universe itself, or in collapsing systems such as black holes in formation, events with infinite matter density should exist somewhere in the past, respectively in the future. Ref. 99

Researchers distinguish two types of singularities: with and without an horizon. The latter ones, the so-called *naked* singularities, are especially strange; for example, a tooth brush can fall into a singularity and disappear without leaving any trace. Since the field equations are time invariant, one can thus expect that every now and then, naked singularities emit tooth brushes. (Can you explain why dressed singularities are less dangerous?) Challenge

Of course, naked singularities violate the limit on the size of physical systems, and could thus be dismissed as academic. Nevertheless, many people have tried to discover some theoretical principles forbidding the existence of naked singularities. It turns out that there is such a principle, and it is called *quantum theory*.* In fact, whenever one encounters a prediction of an infinite value, one has extended a description to a domain for which it was not conceived. In this case the applicability of pure general relativity to very small distances and very high energies has been assumed. As will become clear in the next two parts of the book, nature is different; quantum theory shows that it makes no sense to talk about “singularities” nor about what happens “inside” a black hole horizon.** See page 619

A quiz: is the universe a black hole?

Could it be that we live inside a black hole? Both the universe and black holes have horizons. Even more interesting, the horizon distance r_o of the universe is about

$$r_o \approx 3ct_o \approx 4 \cdot 10^{26}\,\mathrm{m} \tag{316}$$

and its matter content is about

$$m_o + \frac{4\pi}{3}\rho_o r_o^3 \quad \text{whence} \quad \frac{2Gm_o}{c^2} = 72\pi G\rho_o c t_o^3 = 6 \cdot 10^{26}\,\mathrm{m} \tag{317}$$

for a density of $3 \cdot 10^{-27}\,\mathrm{kg/m^3}$. Thus one has

$$r_o \approx \frac{2Gm_o}{c^2} \tag{318}$$

similar to the black hole relation $r_S = 2Gm/c^2$. Is this a coincidence? The section on cosmology allows to tell. Challenge

∗ There are also attempts to formulate such forbidding principles *inside* general relativity, called *cosmic censorship*, but we do not discuss them here. Ref. 100

∗∗ Many physicists are still wary to make such strong statements at this point. The part on quantum theory and the first sections of the third part of the escalation give the precise arguments leading to these conclusions. See page 578

13. General relativity in ten points

Sapientia felicitas.***

General relativity is the final description of *paths of motion*, or if one prefers, of the *motion of macroscopic bodies*. It describes how the observations of motion of *any* two observers are related to each other, and also describes motion due to gravity. In fact, general relativity is based on the following observations:

• All observers agree that there is a "perfect" velocity in nature, namely a common maximum energy velocity relative to matter. It is realized by massless radiation such as light or radio signals.

• Any system of dimension L and mass M is bound by the limit

$$\frac{L}{M} \geqslant \frac{4G}{c^2} \tag{319}$$

which is realized only for black holes.

From these central facts one deduces:

• Space-time consists of events in $3+1$ continuous dimensions, with a *curvature* varying from point to point. The curvature can be deduced from distance measurements among events. We thus live in a pseudo-riemannian space-time. Measured times, lengths, and curvatures vary from observer to observer.

• Space-time is curved near mass and energy. The average curvature at a point is determined by the energy-momentum density at that point and described by the field equations. When matter and energy move, the curvature moves along with them. A built-in delay renders faster than light transport of energy impossible. The proportionality constant between energy and curvature is so small that the curvature is not observed in everyday life, but only its indirect manifestation, namely gravity.

• Curved space-time is *elastic*; it can wiggle also independently of matter; one then speaks of gravitational radiation or of gravity waves.

• Freely falling matter moves along *geodesics*, i.e. along paths of maximal length in curved space-time; in space this means that light bends when it passes near large masses by twice the amount predicted by universal gravity.

• To describe gravitation one *needs* curved space-time, i.e. general relativity, *at the latest* whenever distances are of the order of the Schwarzschild radius $r_S = 2Gm/c^2$. When distances are much larger, the description by universal gravity, namely $a = GM/r^2$, together with flat Minkowski space-time, will do as approximation.

• Space and time are not distinguished globally, but only locally. *Matter* is required to do so.

In addition, all matter and energy we observe around provide two observations:

• The universe has a finite age; it is the reason for the darkness at night. A horizon limits the measurable space-time intervals to about twelve thousand million years.

• On cosmological scale, everything moves away from everything else: the universe is expanding. This expansion of space-time is also described by the field equations.

*** 'Wisdom is happiness.' It is also the motto of Oxford university.

Measured effect	confirmation	type	reference
equivalence principle	10^{-12}	motion of matter	Ref. 10, 101
$1/r^2$ dependence (dimensionality of space-time)	10^{-10}	motion of matter	Ref. 103
time independence of G	10^{-19} /s	motion of matter	Ref. 101
redshift (light & microwaves on sun, earth, Sirius)	10^{-4}	space-time curvature	Ref. 6, 18, 101
perihelion shift (four planets, Icarus, pulsars)	10^{-3}	space-time curvature	Ref. 101
light deflection (light, radiowaves around sun, stars, galaxies)	10^{-3}	space-time curvature	Ref. 101
time delay (radio signals near sun, near pulsars)	10^{-3}	space-time curvature	Ref. 101
gravitomagnetism (earth, pulsar)	10^{-1}	space-time curvature	Ref. 31
geodesic effect (moon, pulsars)	10^{-1}	space-time curvature	Ref. 42, 101
gravity wave emission delay (pulsars)	10^{-3}	space-time curvature	Ref. 101

Table 23 Present types of tests of general relativity

The accuracy of the description

Was general relativity worth the effort? The discussion of its accuracy is most conveniently split into two sets of experiments. The first set is given by measurements of how *matter moves*. Do objects really follow geodesics? So far, all experiments agree with theory within measurement errors, i.e. at least within 1 part in 10^{12}. In short, the way matter falls is indeed described by general relativity in all details. Ref. 101 Ref. 102

The second set consists of measurements of the dynamics of space-time itself. Does *space-time move* following the field equations of general relativity? In other words, is space-time really bent by matter in the way the theory predicts? Many experiments have been performed, some near and most far from earth, both in weak and in strong fields. All agree with the predictions within errors. However, the best measurements so far have only about 3 significant digits. Note that even though numerous experiments have been performed, there are only few *types* of tests, as table shows; in the past, discovering a new type has always meant fame and riches. Most sought after, of course, is the direct detection of gravitational waves. Ref. 101, 102 Challenge

Another comment of the table is in order. After many decades in which all measured effects were only of order v^2/c^2, several so-called *strong field effects* in pulsars allowed to test order v^4/c^4 effects. Soon a few effects of this order should also be found even inside the solar system, using high-precision satellite experiments. The present crown of all measurements, the gravity wave emission delay, is the only v^5/c^5 effect measured so far. Ref. 101 See page 219

The difficulty to achieve high precision for space-time curvature measurements is the reason that mass is measured with balances, always (indirectly) using the prototype kilogram in Paris, instead of defining some standard curvature and fixing the value of G. Indeed, no terrestrial curvature experiment has ever been carried out. Also in this domain a breakthrough would make the news. At present, any terrestrial curvature method would not even allow to define a kilogram of gold or of oranges with a precision of a single kilogram!

Another possible check of general relativity is the search for alternative descriptions of gravitation. Quite a number of competing theories of gravity have been formulated and studied, but none is in agreement with all experiments.

Ref. 102, 104

In summary, as Thibault Damour likes to say, general relativity is at least 99.999 999 999 9% correct concerning the motion of matter and energy, and at least 99.9% correct about the way matter and energy curve and move space-time. No exceptions, no anti-gravity, and no unclear experimental data are known. All macroscopic motion, on earth and in the skies, is described by general relativity. All these confirmations show the importance of the achievement of Albert Einstein.

Ref. 101

Research in general relativity and cosmology

Despite all these successes, research in general relativity is more intense than ever.*

Ref. 106

• The description of collisions and of many body problems, as in the motion of stars, neutron stars, and black holes, with its richness of behaviour, helps astrophysicists to improve their understanding of signals they observe in their telescopes.

Ref. 107

• The study of the early universe and of elementary particle properties, with topics such as *inflation*, a short period of accelerated expansion during the first few seconds, is still an important topic of investigations.

Ref. 109

• The study of chaos in the field equations is of fundamental interest in the study of the early universe, and may be related to the problem of galaxy formation, one of the biggest open problems in physics.

Ref. 108

• The determination of the cosmological parameters, such as the matter density, the curvature, and the vacuum density, is a central effort of modern astrophysics.

Ref. 52

• Astrophysicists regularly discover new phenomena in the skies. For example, despite what will be said later, gamma ray bursts are still not completely understood. The longest and most energetic so far, several hours long with photons of 25 GeV energy, was observed in February 1994.

See page 522

Ref. 110

• A computer database of all solutions of the field equations is being built. Among others, researchers are checking whether they really are all different from each other.

Ref. 111

• The inclusion of torsion into field equations, a possible extension of the theory, is one of the promising attempts to include particle spin into general relativity.

Ref. 112

• Studying solutions with nontrivial topology, such as wormholes and particle-like solutions, is a fascinating field of enquiry, related to string theory.

Ref. 113

• Other formulations of general relativity, describing space-time with quantities other than the metric, are continuously developed; the hope is to clarify the relation to the quantum world. The unification of quantum physics and general relativity, the topic of the third part of this escalation, will occupy researchers for many years to come.

Ref. 114

• Finally, the teaching of general relativity, which for many decades has been hidden behind greek indices, differential forms and other antididactic methods, will benefit greatly from future improvements focusing on the physics, and less on the formalism.

Ref. 115

In short, general relativity is still an extremely interesting field of research, and important discoveries are still expected.

* There is even a free and excellent internet based research journal, called *Living Reviews in Relativity*, to be found at the http://www.livingreviews.org web site.

The limits of general relativity

Even though successful, the description of motion presented so far is unsatisfactory; maybe you already have some stomach feeling about certain unresolved issues. First of all, even Challenge though the speed of light is the starting point of the whole theory, we still do not know what light actually *is*. Finding out will be the next topic.

Secondly, we saw that everything falls along geodesics. But a mountain does not fall. Somehow the matter below prevents it from falling. How does it achieve this? General relativity does not provide an answer; in fact, it does not describe matter *at all*. Einstein said that the left-hand side of the field equations, describing the curvature of space-time, was granite, the right-hand side, describing matter, was sand. As already remarked, to change the sand into rock one first needs quantum theory and then, in a further step, its unification with relativity. This is also the program for the rest of the escalation.

We also saw that matter is necessary to clearly distinguish space and time, and in particular, to understand the working of clocks, meter bars, and balances. In particular, one question remains: why are there units of mass, length and time in nature *at all*? This deep question will also be addressed in the third part of our escalation.

Additionally, we found that we do not know enough about the vacuum. We need to understand the magnitude of the cosmological constant and the number of spacetime dimensions to answer the question: Why the sky is so far away? General relativity does not help here. Worse, the smallness of the cosmological constant contradicts the simplest version of quantum theory, and is one of the reasons why we still have quite some height to escalate before we reach the top of motion mountain. See page 578

In short, to describe motion well, we realize that we need a more precise description of light, of matter, and of the vacuum! Otherwise we cannot hope to answer questions about mountains, clocks and stars. In a sense, it seems that we achieved quite little. Worse, for the See page 494 following topic we are forced to go backwards, to situations *without* gravity, i.e. back to the framework of special relativity. That is the next, middle section of our escalation. On the other hand, a lot of fun is waiting there.

> It's a good thing we have gravity, or else when birds died
> they'd just stay right up there. Hunters would be all confused.
> Steven Wright, comedian.

References

1 Albert Einstein, *Die Grundlage der allgemeinen Relativitätstheorie*, Annalen der Physik **49**, pp. 769–822, 1916. A selection of english language textbooks for deeper study should include, in ascending order of depth and difficulty:
– Almost no formulas, but tons of insight are found in the enthusiastic text by John A. Wheeler, *A journey into gravity and spacetime*, 1990. – Beauty, simplicity and shortness is

the characteristic of Malcolm LUDVIGSEN, *General relativity, a geometric approach*, Cambridge University Press, 1999.

– A good overview of experiments and theory is given in James FOSTER & J.D. NIGHTINGALE, *A short course in general relativity*, Springer Verlag, 2nd edition 1998.

– A pretty text is Sam LILLEY, *Discovering relativity for yourself*, Cambridge University Press, 1981.

– A modern text is by Ray D'INVERNO, *Introducing Einstein's relativity*, Clarendon Press, 1992. It includes an extended description of black holes and gravitational radiation, and regularly refers to present research.

– A classic is Steven WEINBERG, *Gravitation and Cosmology*, Wiley, 1972.

– The mightiest text will remain the one by Charles MISNER, Kip THORNE & John A. WHEELER, *Gravitation*, Freeman, 1973.

– The toughest and most professional text is Robert M. WALD, *General relativity*, University of Chicago Press, 1984.

In other languages, see the next reference. Cited on page 205.

2 A beautiful teaching text is the classic G. FALK & W. RUPPEL, *Mechanik, Relativität, Gravitation – ein Lehrbuch*, Springer Verlag, Dritte Auflage, 1983.

Modern references are Torsten FLIESSBACH, *Allgemeine Relativitätstheorie*, Akademischer Spektrum Verlag, 1998.

Excellent is Hubert GOENNER, *Einführung in die spezielle und allgemeine Relativitätstheorie*, Akademischer Spektrum Verlag, 1996.

In french, there is ...

Cited on page 226, 226, 229.

3 P. MOHAZZABI & J.H. SHEA, *High altitude free fall*, American Journal of Physics **64**, pp. 1242–1246, 1996. As a note, due to a technical failure Kittinger had his hand in (near) vacuum during his ascent, without incurring any permanent damage. See the http://www.sff.net/people/geoffrey.landis/vacuum.html web site. Cited on page 205.

4 The speed of light has been measured for observers with emitters which were accelerated and moving with over 99.9% of the speed of light. This was observed in in synchrotons and in star explosions. See ... Cited on page .

5 H. BONDI, *Gravitation*, European Journal of Physics **14**, pp. 1–6, 1993. Cited on page 209.

6 R.V. POUND & G.A. REBKA, The famous paper is *Apparent weight of photons*, Physical Review Letters **4**, pp. 337–341, 1960. A higher prcision version was published by R.V. POUND & J.L. SNIDER, Physical Review Letters **13**, p. 539, 1964, and R.V. POUND & J.L. SNIDER, Physical Review B **140**, p. 788, 1965, Cited on page 210, 290.

7 J.C. HAFELE & Richard E. KEATING, *Around-the-world atomic clocks: predicted relativistic time gains*, Science **177**, pp. 166–167, and *Around-the-world atomic clocks: observed relativistic time gains*, pp. 168–170, 14 July 1972. Cited on page 210.

8 R.F.C. VESSOT & al., *Test of relativistic gravitation with a space-borne hyrogen maser*, Physical Review Letters **45**, pp. 2081–2084, 1980. The experiment was performed in 1976; there are more than a dozen coauthors involved in this work, which involved shooting a maser into space with a scout missile to a height of ca. 10000 km. Cited on page 210.

9 L. BRIATORE & S. LESCHIUTTA, *Evidence for the earth gravitational shift by direct atomic-time-scale comparison*, Il Nuovo Cimento **37B**, pp. 219–231, 1977. Cited on page 210.

10 The equality was first tested with precision by Roland VON EÖTVÖS, Annalen der Physik & Chemie **59**, p. 354, 1896, and by R. VON EÖTVÖS & al., Annalen der Physik 4, Leipzig **68**, p. 11, 1922. He found agreement to 5 parts in 10^9. More experiments were performed by P.G. ROLL, R. KROTKOW & R.H. DICKE, Annals of Physics (NY) **26**, pp. 442–517, 1964, and by V.B. BRAGINSKY & V.I. PANOV, Soviet Physics - JETP **34**, pp. 463–466, 1971. Modern results, with less than one part in 10^{12}, are by Y. Su, Physical Review **D50**, p. 3614, 1994. Several experiments have been proposed to test the equality in space to less than one part in 10^{16}. Cited on page 211, 290.

11 This story is told e.g. by W.G. Unruh, *Time, gravity, and quantum mechanics*, gr-qc/9312027. Cited on page .

12 J. Droste, *Het veld van een enkel centrum in Einstein's theorie der zwaartekracht, en de beweging van een stoffelijk punt*, Verslag gew. Vergad. Wiss. Amsterdam **25**, pp. 163–180, 1916. Cited on page 214.

13 L.B. Kreuzer, *Experimental measurement of the equivalence of active and passive gravitational mass*, Physical Review **169**, pp. 1007–1012, 1968. With a clever set-up, he showed that the gravitational mass of fluorine and of bromine are equal. Cited on page 215.

14 That bodies fall along geodesics has been checked by ... Cited on page 216.

15 So far, the experiments confirm that electrostatic and (strong) nuclear energy fall like matter to within one part in 10^8, and weak (nuclear) energy to within a few percent. This is summarized in ... Cited on page 217.

16 J. Soldner, *Berliner Astronomisches Jahrbuch auf das Jahr 1804*, 1801, p. 161. Cited on page 217.

17 See the excellent explanation in N. Dragon, ..., downloadable from the http://www.itp.uni-hannover.de/ dragon web site. Cited on page 217.

18 J.W. Brault, Princeton University PhD thesis, 1962. See also J.L. Snider, Physical Review Letters **28**, pp. 853–856, 1972, and for the star Sirius see J.L. Greenstein & al., Astrophysical Journal **169**, p. 563, 1971. Cited on page 210, 290.

19 See for example K.D. Olum, *Superluminal travel requires negative energies*, Physical Review Letters **81**, pp. 3567–3570, 1998. Cited on page 218.

20 See the *Astronomical Almanac*, and its *Explanatory supplement*, H.M. Printing Office, London and U.S. Government Printing Office, Washington, 19▷▷. For the information about various time coordinates used in the world, such as barycentric coordinate time, the time at the barycentre of the solar system, see also the http://tycho.usno.navy.mil/systime.html webpage. It also contains a good bibliography. Cited on page 219.

21 An overview is given in C. Will, *Theory and Experiment in Gravitational Physics*, chapter 14.3, revised edition, Cambridge University Press, 1993. See also C. Will, *Was Einstein right? – Putting general relativity to the test*, Oxford University Press, 1993. See also his paper gr-qc/9811036. Cited on page 219, 229.

22 Edwin F. Taylor, *"The boundaries of nature: special and general relativity and quantum mechanics, a second course in physics" – Edwin F. Taylor's acceptance speech for the 1998 Oersted Medal presented by the American Association of Physics Teachers, 6 January 1998*, American Journal of Physics **66**, pp. 369–376, 1998. The calculation in the text omits several effect, such as rotation of the earth, redshift, and others. Cited on page 220.

23 A.G. Lindh, *Did Popper solve Hume's problem?*, Nature **366**, pp. 105–106, 11 November 1993, Cited on page 220.

24 The measurement is presented in the Astrophysical Journal, in 1998 or 1999. Some beautiful graphics on the models of this star system can be found at the http://www.physics.uiuc.edu/groups/tastro/movies/spm/ web site. Cited on page 220.

25 R.J. Nemiroff, .., American Journal of Physics **61**, p. 619, 1993. Cited on page 220.

26 Wolfgang Rindler, *Essential relativity – special, general, and cosmological*, Revised second edition, Springer, 1977. Cited on page 232.

27 M. Santander, L.M. Nieto & N.A. Cordero, *A curvature based derivation of the Schwarzschild metric*, American Journal of Physics **65**, pp. 1200–1209, 1997. Cited on page 235, 236, 236.

28 The equivalence of the various definitions of the Riemann tensor is explained in ... Cited on page 246.

29 Arguments for the emptyness of general covariance are given by John D. Norton, *General covariance and the foundations of general relativity*, Reports on Progress in Physics **56**, pp. 791–858, 1993. The opposite point, including the discussion of "absolute elements," is made in the book by ... Anderson, *Principles of relativity physics*, chapter 4, 19▷▷. Cited on page 241.

30 The Thirring effect was predicted in H. Thirring, Physikalische Zeitschrift **19**, p. 33, 1918, and also studied in H. Thirring, Physikalische Zeitschrift **22**, p. 29, 1921. The Thirring-Lense effect was predicted in H. Thirring & J. Lense, Physikalische Zeitschrift **19**, p. 156, 1918. Cited on page 221.

31 The feat, using the LAGEOS and LAGEOS II satellites, is told in Ignazio Ciufolini, *The 1995-99 measurements of the Thirring-Lense effect using laser-ranged satellites*, Classical and Quantum Gravity **17**, pp. 2369–2380, 2000. Cited on page 221, 290.

32 The detection of Thirring-Lense effects in binary pulsars ... Cited on page 222.

33 D. Bedford & P. Krumm, *On relativistic gravitation*, American Journal of Physics **53**, p. 889›890, 1985. P. Krumm & D. Bedford, *The gravitational Poynting vector and energy transfer*, American Journal of Physics **55**, p. 362›363, 1987. Cited on page 222, 226.

34 See for example, K.T. McDonald, *Answer to question #49. Why c for gravitational waves?*, American Journal of Physics **65**, pp. 591–592, 1997, and section III of V.B. Braginsky, C.M. Caves & K.S. Thorne, *Laboratory experiments to test relativistic gravity*, Physical Review D **15**, pp. 2047–2068, 1992. Cited on page 224.

35 The beautiful summary by Daniel Kleppner, *The gem of general relativity*, Physics Today pp. 9–11, April 1993, appeared half a year before the authors of the cited work, Joseph Taylor and Russel Hulse, received the Nobel prize for the discovery of millisecond pulsars. A more detailed review article is J.H. Taylor, *Pulsar timing and relativistic gravity*, Philosophical Transactions of the Royal Society, London A **341**, pp. 117–134, 1992. The original paper is J.H. Taylor, J.M. Weisberg, *Further experimental tests of relativistic gravity using the binary pulsar PSR 1913+16*, Astrophysical Journal **345**, pp. 434–450, 1989. See also J.M. Weisberg, J.H. Taylor & L.A. Fowler, *Pulsar PSR 1913+16 sendet Gravitationswellen*, Spektrum der Wissenschaft Dezember 1981. Cited on page 227, 227.

36 W.B. Bonnor & M.S. Piper, *The gravitational wave rocket*, gr-qc/9702005. Cited on page 227.

37 L. Lerner, *A simple calculation of the deflection of light in a Schwarzschild gravitational field*, American Journal of Physics **65**, pp. 1194–1196, 1997. Cited on page 228.

38 A. Einstein, Annalen der Physik **35**, p. 898, 1911. Cited on page 228.

39 I.I. Shapiro, Physical Review Letters **13**, pp. 789–792, 1964. Cited on page 229.

40 I.I. Shapiro & al., Physical Review Letters **20**, pp. 1265–1268, 1968. Cited on page 229.

41 J.H. Taylor, *Pulsar timing and relativistic gravity*, Proceedings of the Royal Society, London A **341**, pp. 117–134, 1992. Cited on page 229.

42 I.I. Shapiro & al., *Measurement of the de Sitter precession of the moon: a relativistic three body effect*, Physical Review Letters **61**, pp. 2643–2646, 1988. Cited on page 231, 290.

43 This is told (without the riddle solution) on p. 67, in *W. Pauli*, *Relativitätstheorie*, Springer Verlag, Berlin, 2000, the edited reprint of a famous text originally published in 1921. The reference is H. Vermeil, *Notiz über das mittlere Krümmungsmaß einer n-fach ausgedehnten Riemannschen Mannigfalktigkeit*, Gött. Nachr., math.-phys. Kl. p. 334, 1917. Cited on page 232.

44 See for example K. Dalton, *Gravity, Geometry, and Equivalence*, http://xxx.lanl.gov/abs/gr-qc/9601004 and L. Landau & E. Lifshitz, *The Classical Theory of Fields*, Pergamon, 4th edition, 1975, p. 241. Cited on page 244.

45 A recent overview on the experimental tests of the universality of free fall is given by R.J. Hughes, *The equivalence principle*, Contemporary Physics **4**, pp. 177–191, 1993. Cited on page 244.

46 Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 61, 1991. Cited on page 247.

47 See for example A.N. Cox, editor *Allen's Astrophysical Quantities*, AIP Press and Springer Verlag, 2000. Cited on page 247.

48 G. Gamow, Physical Review **174**, p. 505, 1948. Cited on page 250.

49 A.G. Doroshkevich & I.D. Novikov, Dokl. Akad. Nauk. SSSR **154**, p. 809, 1964. It appeared translated into english a few months later. The story of the prediction was told by Penzias in his Nobel lecture. Cited on page 250.

50 Arno A. Penzias & Rober W. Wilson, Astrophysical Journal **142**, p. 419, 1965. Cited on page 250.

51 An opposite idea is defended by ... Cited on page 249.

52 The experiments are discussed in detail in the excellent review by D. Giulini & N. Straumann, *Das Rätsel der kosmischen Vakuumenergiedichte und die beschleunigte Expansion des Universums*, Physikalische Blätter **556**, pp. 41–48, 2000. See also N. Straumann, *The mystery of the cosmic vacuum energy density and the accelerated expansion of the universe*, European Journal of Physics **20**, pp. 419–427, 1999. Cited on page 257, 291.

53 Fred Hoyle still defends his idea that the universe is static, e.g. in ... Physics Today ... Cited on page 258, 258.

54 Stephen W. Hawking & G.F.R. Ellis, *The large scale structure of space-time*, Cambridge University Press, 1973. Cited on page 258, 276.

55 Augustine ... Cited on page 259.

56 S. Hawking, *A brief history of time – From the big bang to black holes*, 1988. Cited on page 260.

57 This famous quote is the first sentence of the final chapter, the 'Beschluß', of Immanuel Kant, *Kritik der praktischen Vernunft*, 1797. Cited on page 260.

58 Galaxy mass measurments are described by ... Cited on page 262.

59 The author of the bible explains *rain* in this way, as can be deduced from its very first page, Genesis 1, 6-7. Cited on page 262.

60 H. Knutsen, *Darkness at night*, European Journal of Physics **18**, pp. 295–302, 1997. Cited on page 263, 264.

61 Paul Wesson, ..., Astrophysical Journal **367**, p. 399, 1991. Cited on page 264.

62 Steven Weinberg, *Gravitation and cosmology*, John Wiley, 1972. An excellent book written with a strong personal touch, and stressing most of all the relation with experimental data. It does not develop a strong feeling for space-time curvature, and does not address the basic problems of space and time in general relativity. Excellent for learning how to actually calculate things, but less for the aims of our escalation. Cited on page 264, 274.

63 Star details are explained in many texts. See for example ... Cited on page 264.

64 Johannes PELT, ..., Astronomy and Astrophysics 1995. Cited on page 266.

65 F. ZWICKY, Physical Review Letters **51**, p. 290, and p. 679, 1937. The negative view by Einstein is found in A. EINSTEIN, Science **84**, p. 506, 1936.

A review on gravitational lensing can be found online, in the paper by J. WAMBSGANSS, Living Reviews in Relativity **1,12**, pp. 1–80, 1998, on the http://www.livingreviews.org/Articles/Volume1/1998-12wamb/ web site.

There is also the book by P. SCHNEIDER, J. EHLERS & E.E. FALCO, *Gravitational lenses*, Springer Verlag, Berlin, 1992. Cited on page 267.

66 M. LACHIÈZE-REY, J.-P. LUMINET, *Cosmic topology*, Physics Reports **254**, pp. 135–214, 1995. Cited on page 267.

67 Thanks to Steve Carlip for clarifying this point. Cited on page 268.

68 G.F.R. ELLIS, T. ROTHMANN, *Lost horizons*, American Journal of Physics **61**, pp. 883–893, 1993. Cited on page 268.

69 C.L. BENNET, M.S. TURNER & M. WHITE, *The cosmic rosetta stone*, Physics Today **50**, pp. 32–38, November 1997. The cosmic background radiation differs from black hole radiation by less than 0.005%. Cited on page 270.

70 The lack of expansion in the solar system is shown in ... Cited on page 270.

71 A pretty article explaining how one can make experiments to find out how the human body senses rotation even when blindfolded and earphoned is described by M.-L. MITTELSTAEDT & H. MITTELSTAEDT, *The effect of centrifugal force on the perception of rotation about a vertical axis*, Naturwissenschaften **84**, pp. 366–369, 1997. Cited on page 271.

72 The independence of inertia has been tested ... Cited on page 271.

73 The present status is given in the conference proceedings by Julian BARBOUR & Herbert PFISTER, (editors) *Mach's principle: from Newton's bucket to quantum gravity*, Birkhäuser, 1995. Various formulations of Mach's principle – in fact, 21 different ones – are compared on page 530. Cited on page 271, 272.

74 Information on the rotation of the universe is given in J.D. BARROW, R. JUSZKIEWICZ & D.H. SONODA, *Universal rotation: how large can it be?*, Monthly Notices of the Royal Astronomical Society **213**, pp. 917–943, 1985. See also J.D. BARROW, R. JUSZKIEWICZ & D.H. SONODA, *Structure of the cosmic microwave background*, Nature **309**, pp. 397–402, 1983. Cited on page .

75 The issue has beed discussed within linearized gravity by Richard TOLMAN, in his textbook *Relativity, thermodynamics, and cosmology*, Clarendon Press, 1934, on pp. 272–290. The exact problem has been solved by A. PERES, *Null electromagnetic fields in general relativity theory*, Physical Review **118**, pp. 1105–1110, 1960, and by W.B. BONNOR, *The gravitational field of light*, Commun. Math. Phys. **13**, pp. 163–174, 1969. See also N.V. MITSKIEVIC & K.K. KUMARADTYA, *The gravitational field of a spinning pencil of light*, Journal of Mathematical Physics **30**, pp. 1095–1099, 198. Cited on page 272.

76 For a delightful popular account from a black hole expert, see Igor NOVIKOV, *Black holes and the universe*, Cambridge University Press 1990. Cited on page 273, 280.

77 ... CARILLI & al., *Limits on the change of fine structure constant ...*, Physical Review Letters 25 december 2000. Cited on page 273.

78 R. GEROCH, *Einstein Algebras*, Commun. Math. Phys. **26**, pp. 271–275, 1972. Cited on page 275.

79 A. MACDONALD, *Einstein's hole argument*, American Journal of Physics **69**, pp. 223–225, 2001. Cited on page 277.

80 See for example, the definition of the ADM mass in ... Cited on page 240.

81 For a historical introduction, see the paper ... R. Ruffini, *The physics of gravitationally collapsed objects*, ... Cited on page .

82 The observations of black holes at the centre of galaxies and elsewhere are summarized by R. BLANDFORD & N. GEHRELS, *Revisiting the black hole*, Physics Today **52**, June 1999. Cited on page 278, 287, 287.

83 J. MICHELL, Philosophical Transactions of the Royal Society London **74**, p. 35, 1784, reprinted in S. DETWEILER, *Black holes – selected reprints*, American Association of Physics Teachers, 1982. Cited on page 278.

84 The name black hole was introduced *in print* in the article by R. RUFFINI & J.A. WHEELER, *Introducing the black hole*, January, 1971. Cited on page .

85 The beautiful paper is R. OPPENHEIMER & H. SNYDER, ..., 1939. Cited on page 279.

86 R. KERR, Physical Review Letters **11**, p. 237, 1963. Cited on page 282.

87 E.T. NEWMAN, E. COUCH, R. CHINNAPARED, A. EXTON, A. PRAKASH & R. TORRENCE, Journal of Mathematical Physics **6**, p. 918, 1965. Cited on page 282.

88 M.A. ABRAMOWICZ & J.P. LASOTA, *On traveling round without feeling it and uncurving curves*, American Journal of Physics **54**, pp. 936–939, 1986. Cited on page .

89 For a summary, see P.O. MAZUR, *Black hole uniqueness theorems*, pp. 130–157, in M.A.H. MAC CALLUM, editor, *General relativity and gravitation*, 1987. Cited on page 282.

90 H.P. KÜNZLE, A.K.M. MASOOD-UL-ALAM, *Spherically symmetric static SU(2) Einstein-Yang-Mills fields*, Journal of Mathematical Physics **31**, pp. 928–935, 1990. Cited on page 282.

91 R. PENROSE & R.M. FLOYD, *Extraction of rotational energy from a black hole*, Nature **229**, pp. 177–179, 1971. Cited on page 284.

92 The mass energy relation for a rotating black hole is due to D. CHRISTODOULOU, *Reversible and irreversible transformations in black hole physics*, Physical Review Letters **25**, pp. 1596–1597, 1970. For a general, charged and rotating black hole it is due to D. CHRISTODOULOU & R. RUFFINI, *Reversible transformations of a charged black hole*, Physical Review D **4**, pp. 3552–3555, 1971. Cited on page 284.

93 J.D. BEKENSTEIN, *Black holes and entropy*, Physical Review **D7**, pp. 2333–2346, 1973. Cited on page 285.

94 The paradox is discussed in M.A. ABRAMOWICZ, *Black hole and the centrifugal force paradox*, Scientific American pp. 23–31, March 1993, and in the comment by Don N. PAGE, *Relative alternatives*, Scientific American p. 5, August 1993. See also M.A. ABRAMOWICZ & E. SZUSZKIEWICZ, *The wall of death*, American Journal of Physics **61**, pp. 982–991, 1993. Cited on page 286.

95 About black holes in the early universe, see ... Cited on page 286.

96 About black holes formation via star collapse, see ... Cited on page 287.

97 Frederick LAMB, APS meeting 1998, binary star 4U1820-30, 20 000 lightyears from earth, Physics News Update, April 27, 1998. Cited on page 287.

98 The first evidence for the observation of matter falling into a black hole was publicized in early 2001. ... Cited on page 287.

99 For a readable summary of the Penrose-Hawking singularity theorems, see ... Cited on page 288.

100 For an overview of cosmic censorship, see T.P. SINGH, *Gravitational collapse, black holes, and naked singularities*, gr-qc/9805066 or R.M. WALD, *Gravitational collapse and cosmic*

censhorship, gr-qc/9710068. The original idea is due to R. PENROSE, *Rivista del Nuovo Cimento*, 1 p. 252, 1969. Cited on page 288.

101 Thibault DAMOUR, *Experimental tests of relativistic gravity*, gr-qc/9904057. It is the latest in a series of his papers which started with Thibault DAMOUR, *Was Einstein 100% right?*, gr-qc/9412064. Cited on page 289, 290, 290, 290, 290, 290, 290, 290, 290, 290, 290, 290.

102 H. DITTUS, F. EVERITT, C. LÄMMERZAHL & G. SCHÄFER, *Die Gravitation im Test*, Physikalische Blätter **55**, pp. 39–46, 1999. Cited on page 290, 290, 290.

103 The inverse square dependence has been checked down to 0.2 mm, as reported by ... Physics Today ...2000. Cited on page 290.

104 For theories competing with general relativity, see ... Among others, the absence of the Nordvedt effect, a hypothetical 28-day oscillation in the earth moon distance, looked for by laser ranging experiments without any result, killed several competing theories. For a summary of the measurements, see J. MÜLLER, M. SCHNEIDER, M. SOFFEL & H. RUDER, Astrophysical Journal Letters **382**, p. *L*101, 1991. Cited on page 290.

105 See S. BÄSSLER & al., *Improved test of the equivalence principle for gravitational selfenergy*, Physical Review Letters **83**, pp. 3585–3588, 1st November 1999. See also C. WILL, ..., Physics Today **..**, p. ▹▹▹, October 1999. Cited on page .

106 Almost all of imortance in general relativity is published in the Journal *Classical and quantum gravity*. For the US american side of research, see the webzine *Matters of Gravity*, published on http://vishnu.nirvana.phys.psu.edu/mog.html. Cited on page 291.

107 Collisions and many body problems ... Cited on page 291.

108 The study of chaos in Einstein's field equations is in its beginnings. See e.g. L. BOMBELLI, F. LOMBARDO & M. CASTAGNINO, *Chaos in Robertson-Walker Cosmology*, gr-qc/9707051. Cited on page 291.

109 Inflation and early universe ... Cited on page 291.

110 The best introduction into the topic of gamma ray bursts is S. KLOSE, J. GREINER & D. HARTMANN, *Kosmische Gammastrahlenausbrüche – Beobachtungen und Modelle*, Teil I und II, Sterne und Weltraum March and April 2001. Cited on page 291.

111 The field solution database is built around the work of A. Karlhede, which allows to distinguish solutions with a limited amount of mathematical computation. ... Cited on page 291.

112 Torsion ... Cited on page 291.

113 Wormholes and nontrivial topologies...A basic approach is the one by Mark HADLEY, Classical and Quantum Gravity October 2000. Cited on page 291.

114 Other formulations of relativity ... Cited on page 291.

115 Teaching will benefit in particular from new formulations, from concentration on principles and their consequences, as has happened in special relativity, from simpler descriptions at the weak field level, and from future research in the theory of general relativity. The newer textbooks cited above all are steps in these directions. Cited on page 291.

116 A beautiful introduction into modern astronomy is Paolo MAFFEI, *I mostri del cielo*, Mondadori Editore, 1976. Cited on page .

117 Michael H. SOFFEL, *Relativity in astronomy, celestial mechanics and geodesy*, Springer Verlag, 1989. Cited on page .

118 M. NAGANO & A.A. WATSON, *Observations and implications of the ultrahigh energy cosmic rays*, Reviews of Modern Physics **72**, pp. 689–732, 2000. Cited on page .

119 The speed of light as function of height has been checked e.g. by ... The change is smaller than ... Cited on page .

120 Steven WEINBERG, *The cosmological constant problem*, Reviews of Modern Physics **61**, pp. 1–23, 1989. An influential review article on the views on cosmological constant problem at that time. Cited on page .

121 A. ASHTEKAR, 1986, *New variables for classical and quantum gravity*, Physical Review Letters **57**, pp. 2244–2247, Cited on page .

122 The new book by R. WALD, ... Cited on page .

123 A well written text on the connections between the big bang and particle physics is by I.L. ROZENTAL, *Big bang – big bounce, how particles and fields drive cosmic evolution*, Springer, 1988. Cited on page .

Typeset in April 2001

Chapter IV Classical Electrodynamics

What is light? The study of relativity left us completely in the dark, even though we had embarked in it for precisely that aim. True, we have learned how the motion of light compares to that of objects, but we haven't learned anything about its own nature. The answer to this old question emerges only from the study of those types of motion which are *not* related to gravitation.

14. Liquid electricity, light, and levitation

See page 91

Returning to the list of of motors one finds in this world, one remarks that gravitation does not describe almost any of them. Neither the motion of see waves, fire, and earthquakes, nor that of a gentle breeze are due to gravity. The same applies to the motion of muscles. Have you ever listened to your own heart beat, e.g. with a stethoscope? Nobody can pretend to have experienced the mystery of motion without this experience. You have about 3000 million beats in your lifetime. Then they stop.

Challenge

It was one of the most astonishing discoveries of science that all these and most other cases of everyday motion, as well as the nature of light itself, are connected to observations performed already thousands of years ago with two strange stones. These stones show that all examples of motion which are called *mechanical* in everyday life, are, without exception, of *electrical* origin.

In particular, the solidity and softness of matter as well as its impenetrability is due to electricity; also light is. As these aspects are part of everyday life, we leave aside all complications due to gravity and curved space-time. To understand mechanics, the most productive way to proceed is, to study, as in the case of gravity, those types of motion which are generated without any contact between the involved bodies.

Amber and lodestone

Any fool can ask more questions
than seven sages can answer.

The story of electricity starts with trees. Trees have a special relation to electricity. When one cuts a tree, a viscous resin appears. With time it solidifies, and after millions of years it forms *amber*. When one rubs amber with a cat fur, it acquires the ability to attract small

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?

Challenge - Did you find any mistakes?

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

Search	Magnetic charge
Smallest magnetic charge suggested by quantum theory	$g = \frac{h}{e} = \frac{eZ_0}{2\alpha} = 4.1\,\mathrm{pWb}$
Search in minerals	none Ref. 1
Search in meteorites	none
Search in cosmic rays	none
Search with high energy accelerators	none

Table 24 Some searches for magnetic monopoles, i.e., for magnetic charges

objects, such as saw dust or pieces of paper. This was already known to Thales of Miletus, one of the original seven sages, in the sixth century BCE. The same observation can be made with many other polymer combinations, for example with combs and hair, with shoe soles and carpets, or with TV screens and dust. Children are still surprized at the effect a rubbed comb has on a running water tap.

The other part of the story is about an iron mineral found in certain caves around the world, e.g. in Greece, in the province of Thessalia, in a region (still) called Magnesia, or in China. When one puts two stones of this mineral near each other, they attract or repel each other, depending on their relative orientation. In addition, these stones attract objects made of cobalt, nickel, or iron.

Today one also finds various little objects in nature with more sophisticated properties. Some are able to switch one televisions, others unlock car doors, still others allow to talk with far away friends.

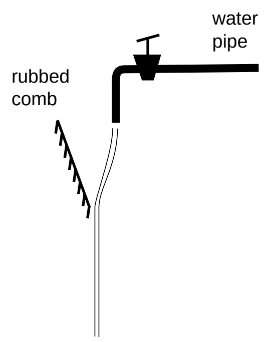


Figure 111 How to amaze kids

All these observations show that in nature there are situations where bodies exert influence on other, *distant* bodies. The space surrounding a body in which such an influence acts is said to contain a field. A *(physical) field* is thus an entity which manifests itself through a force on other bodies in that region of space. The field surrounding the mineral found in Magnesia is called a *magnetic field* and the stones themselves *magnets*.* The field around amber – ελεκτρον in greek, from a root meaning 'brilliant, shining' – is called an *electric field*. The name is due to a proposal by the famous english part-time physicist William Gilbert (1544–1603) who was the physician of Queen Elizabeth. Objects surrounded by a permanent electric field are called *electrets*. They are much less common than magnets; among others, they are used in certain loudspeaker systems.

Fields influence other bodies over a distance, without any material support. For a long time, this was quite rare in everyday life, as laws in most countries have strict upper limits for machines using and producing such fields. For any device which moves, produces sounds, or creates moving pictures, the fields are usually required to remain inside them. For this reason magicians moving an object on a table via a hidden magnet still continue to

* A pretty book about the history of magnetism and the excitement it generates is James D. Livingston, *Driving force – the natural magic of magnets*, Harvard University Press, 1996.

Observation	Magnetic field
Lowest measured value	ca. 1 fT
Produced by brain currents	ca. 0.1 pT
Intergalactic magnetic fields	1 pT to 10 pT
Magnetic field of our galaxy	0.5 nT
Magnetic field of earth	$20\,\mu$T to $70\,\mu$T
Magnetic field below high voltage power line	ca. $10^{-..}$ T
Magnetic field inside modern home	10^{-7} T to 10^{-4} T
Magnetic field near mobile phone	ca. $10^{-..}$ T
Magnetic field in light beam	... T
Magnetic fields near iron magnet	100 mT
Solar spots	ca. 1 T
Magnetic fields near hight tech permanent magnet	max 1.3 T
Magnetic fields in particle accelerator	ca. 10 T
Highest long time static magnetic fields produced in laboratory	ca. 14 T
Highest pulsed magnetic fields produced in laboratory	ca. 40 T
Magnetic fields produced in laboratory with implosions of coils	ca. 1000 T
Field on neutron star	ca. 10^{6} T to 10^{11} T
Quantum critical magnetic field	ca. $6 \cdot 10^{9}$ T
Highest field ever measured, on magnetar SGR-1806-20	$0.8 \cdot 10^{11}$ T
Maximum (Planck) magnetic field	$2.2 \cdot 10^{53}$ T

Table 25 Some observed magnetic fields

surprise and entertain their public. To feel the fascination of fields more strongly, a deeper look into a few experimental results is worthwhile.

How can one make lightnings?

Everybody has seen a lightning, and the effect it can have when hitting a tree. Obviously lightning is a moving phenomenon. Photographs show that their tips advance with a speed of over $10^5\,\mathrm{m/s}$. But what is moving? To find out, one has to find a way to make lightnings oneself.

Figure yet to be included

Figure 112 Lightning: a picture taken with a moving camera, showing the multiple strokes it consists of

In 1995, the car company General Motors accidentally rediscovered an old method for achieving this. They had inadvertently build a spark generating mechanism into their cars; when filling the tank with fuel, the sparks generated sometimes lead to the Ref. 2 explosion of the fuel. They had to recall 2 million vehicles of its Opel brand. What had they done?

The engineers had unknowingly copied the conditions for a electrical device which everybody can build at home and which was originally invented by William Thomson.* Repeating his experiment today, one would take a few water taps, four empty bean or coffee cans, of which two have been opened at both Ref. 3 sides, some nylon rope and some metal wire.

Putting all together as shown in picture 113 and letting the water flow, one finds a strange effect: strong sparks periodically jump between the two copper wires at the point where they are nearest to each other, making loud bangs. Can you guess what condition for the flow has to be realized for this to work? And what Opel Challenge did to repair the cars?

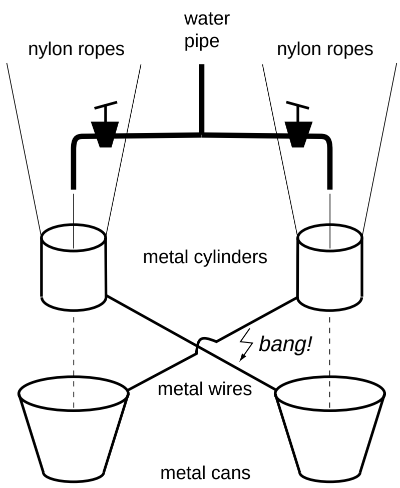


Figure 113 A simple Kelvin generator

If one stops the water flow just before the next spark is due, one finds that both buckets attract sawdust and pieces of paper. The generator thus does the same that

∗ William Thomson (1824–1907), important unionist irish physicist, professor in Glasgow. He worked on the determination of the age of the earth, showing that it was much older than 6000 years, as several sects believed; he was influential in the development of the theory of magnetism and electricity, on the description of the aether, and on thermodynamics. He propagated the use of the term 'energy' as it is common today, instead of the unclear older terms. He was one of the last scientists propagating mechanic analogies for the explanation of phenomena, and thus strongly opposed Maxwell's description of electromagnetism. Probably for this reason he did not receive a Nobel prize. He also was one of the minds behind the laying of the first transatlantic telegraphic cable. Victorian to his bones, when he was was made a Lord, he chose the name of a small brook near his home as his new name; thus he became Lord Kelvin of Largs. Therefore the temperature unit got its name from a small english river.

rubbing amber does, just with more bang for the buck(et). Both buckets are surrounded by electric fields. The field increases with time, until the spark jumps. Just after the spark, the buckets are (almost) without electric field. Obviously, the flow of water somehow builds up an entity on each bucket, today called *electric charge*, which can flow in metals and, when the fields are high enough, through air. One also finds that the two buckets are surrounded by two different types of electric fields: bodies which are attracted by one bucket are repelled by the other. All other experiments confirm that there are two types of charges. The US politician and part-time physicist Benjamin Franklin (1706–1790) called the electricity created on a glass rod rubbed with a dry cloth *positive*, the one on a piece of amber *negative*. (Before him, the two types of charges used to be called called 'vitreous' and 'resinous'.) Bodies with charges of the same sign repel each other, bodies with opposite charges attract each other.

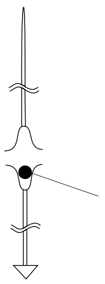


Figure 114 Franklin's personal lightning rod

In summary, electric fields start at bodies, if they are charged. Charging is possible by rubbing and similar processes. Charge can flow, and then is called electric current. The worst conductors of current are polymers; they are often called insulators. Metals are the best conductors, especially silver and copper. This is the reason that at present, after hundred years of use of electricity, the highest concentration of copper in the world is below the surface of Manhattan.

Of course, one has to check whether real thunderstorm lightnings actually are electrical in origin. In 1752, experiments performed in France, following a suggestion of Benjamin Franklin published in London in 1751, showed that one can indeed draw electricity from thunderstorms via a long rod.* These french experiments rendered Franklin world famous; they also started the use of lightning rods throughout the world. Later on, Franklin himself had a lightning rod built through his house, but of a somewhat unusual type, as shown in figure 114. Can you guess what it did in his hall during bad weather, all parts being made of metal? Challenge

What is electric charge?

If all experiments with charge can be explained by calling the two charges positive and negative, the implication is that some bodies have more, and some less charge than the uncharged, *neutral* ones. Electricity only flows when two differently charged bodies are brought into contact. Now, if charge can flow and accumulate, one must be able to somehow measure its amount. Obviously, the *amount* of charge on a body, usually abbreviated q, is defined via the influence the body, say a piece of saw dust, feels when subjected to a field. Charge is thus defined by comparing it to a standard reference charge. For a charged body of mass m accelerated in a field, its unknown charge q is determined by the relation

$$\frac{q}{q_{\text{ref}}} = \frac{ma}{m_{\text{ref}}a_{\text{ref}}} \quad , \tag{320}$$

* There is still research going on into the details of how lightnings are generated and how they propagate. A little about this topic is said on page 341.

Charges	Physical property	Mathematical name (see later for definitions)
can be compared	distinguishability	set
can be ordered	sequence	order
can change gradually	continuity	completeness
can be stored	accumulability	additivity
don't change	conservation	invariance
can be divided	separability	positive or negative

Table 26 Properties of classical electric charge

Observation	Charge
Smallest known non-vanishing charge	$0.5 \cdot 10^{-19}$ C
Charge per bit in computer memory	10^{-13} C
Charge in small capacitor	10^{-7} C
Charge flow in average lightning stroke	1 C to 100 C
Charge stored in a full car battery	0.2 MC
Charge of planet earth	ca. 1 MC
Charge separated by modern power station in one year	ca. $3 \cdot 10^{11}$ C
Total charge of one sign observed in universe	ca. 10^{62} C

Table 27 Some observed charges

i.e., by comparing it to the acceleration and mass of the reference charge. This definition reflects the observation that mass alone is not sufficient for a complete characterization of a body. For a full description of motion one needs to know its electric charge; charge is therefore the second intrinsic property of bodies we discover in our walk.

By the way, the unit of charge, the *coulomb*, is nowadays defined through a standard flow through metal wires, as explained in appendix B. This is possible because all experiments show that charge is conserved, that it flows, that it flows continuously, and that it can accumulate. Charge thus behaves like a fluid substance. Therefore one is forced to use for its description a scalar quantity q, which can take positive, vanishing, or negative values.

In everyday life these properties of charge, listed also in table , describe observations with sufficient accuracy. But as in the case of all previously encountered classical concepts, these experimental results about electrical charge will turn out to be only approximate. More precise experiments will require a revision of several properties.

Experiments show that the entity which accelerates charged bodies, the *electric field*, behaves like a little arrow fixed at each place in space; its length and its direction does not depend on the observer. In short, the electric field $\mathbf{E}(\mathbf{x})$ is a *vector* field. Experiments show that it is best defined by the relation

$$q\mathbf{E}(\mathbf{x}) = m\mathbf{a}(\mathbf{x}) \qquad (321)$$

taken at every point in space $\mathbf{x}$. The definition of the electric field is thus indeed based on how it *moves* charges. The field is measured in multiples of the unit N/C or the identical V/m.

Challenge

Observation	Electric field
Cosmicnoise	ca. $10\,\mu\mathrm{V/m}$
Field of a 100 W FM radio transmitter at 100 km distance	$0.5\,\mathrm{mV/m}$
Field in solar wind	...
Field in clouds	...
Field inside conductors	$0.1\,\mathrm{V/m}$
Field of a 100 W bulb at 1 m distance	$50\,\mathrm{V/m}$
Ground field in earth's atmosphere	100 to $300\,\mathrm{V/m}$
Maximum electric field in air before sparks appear	$1\,\mathrm{MV/m}= 1\,\mathrm{kV/mm}$
Electric fields in biological membranes	$10\,\mathrm{MV/m}$
Electric fields inside capacitors	up to $1\,\mathrm{GV/m}$
Maximum electric field in vacuum, limited by pair production	$1.3\,\mathrm{EV/m}$
Planck electric field	$6.5 \cdot 10^{61}\,\mathrm{V/m}$

Table 28 Some observed electric fields

To describe motion due to electricity completely, one also needs a relation explaining how charges *produce* electric fields. This relation was first established with precision by Charles-Augustin de Coulomb in his private estate, during the french revolution.* He found that around a small or spherical charge Q *at rest* there is an electric field given by

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_o}\frac{Q}{r^2}\frac{\mathbf{r}}{r} \quad \text{where} \quad \frac{1}{4\pi\varepsilon_o} = 8.9\,\mathrm{GVm/C} \quad . \tag{322}$$

Later on we will extend the relation for a charge in motion. The strange proportionality constant is due to the historical way the unit of charge was defined first.** The essential point of the formula is the decrease of the field with the square of the distance; can you imagine the reason? Challenge

The two previous equations allow to write the interaction between two charged bodies as

$$\frac{d\mathbf{p}_1}{dt} = \frac{1}{4\pi\varepsilon_o}\frac{q_1 q_2}{r^2}\frac{\mathbf{r}}{r} = -\frac{d\mathbf{p}_2}{dt} \tag{323}$$

where $\mathbf{p}$ is the momentum change, and $\mathbf{r}$ is the vector connecting the two centers of mass. This famous expression for electrostatic attraction, also due to Coulomb, is valid only for small or for spherical charged bodies *at rest*.

The strength of this interaction is considerable. For example, it is the basis for the force of our muscles. Their force is a macroscopic effect of this equation. Another example is provided by the strength of steel or diamond. As we will discover, its atoms, and those of any other material, are kept together by electrostatic attraction. As a final example to

* Charles-Augustin de Coulomb (1736, Angoulême–1806, Paris), french engineer and physicist. His careful experiments on electric charges provided the basis for the study of electricity.

** Other definitions of this and other proportionality constants to be encountered later are possible, leading to unit systems different from the SI system used here. It is presented in detail in appendix B. Among the older competitors, the gaussian unit system often used in theoretical calculations, the Heaviside-Lorentz unit system, the electrostatic unit system, and the electromagnetic unit system are the most important ones. For more details, see the standard text by J.D. JACKSON, *Classical electrodynamics*, 3rd edition, Wiley, 1998, convince yourself of the strength of electrostatic attraction, answer the following: What is the force between two boxes with a gram of protons each, located on the two poles of the earth? Try to guess the result, before you calculate the astonishing value. Challenge

Due to the strength of electromagnetic interactions, separating charges is not an easy task. This is the reason that electrical effects are in common use only for about a hundred years. One had to wait for the invention of practical and efficient devices for separating charges and putting them into motion. Of course this implies the use of energy. Batteries, as used e.g. in portable phones, use chemical energy to do the trick,* thermoelectric elements, as used in some watches, use the temperature difference between the wrist and the air to separate charges, solar cells use light, and dynamos or the Kelvin generator use kinetic energy.

What then is electricity? The answer is simple: *electricity* is *nothing in particular*. It is the name for a field of inquiry, but not the name for any specific observation or effect. Electricity is neither electric current, nor electric charge, nor electric field. It is not a specific term; it applies to all of these phenomena. One has to be a little careful when using it. In fact the vocabulary issue hides a deeper question, which was unanswered in the twentieth century: what is the nature of electric charge? Since it flows, one can start by asking:

Can one feel the inertia of electricity?

If electric charge really is something *flowing* through metals, one should be able to observe the effects shown in figure 115, as already Maxwell predicted. And indeed, each of these effects has been observed.** For example, when a long metal rod is kept vertically, one can measure an electrical potential difference, a voltage, between the top and the bottom. In other words, one can measure the *weight* of electricity this way. Similarly, one can measure Ref. 4 potential differences between the ends of an accelerated rod. In particular, one can measure a potential difference between the center and the rim of a rotating metal disk. This latter experiment was in fact the way in which the ratio q/m for currents in metals was first measured with precision. The result is

$$q/m = 1.8 \cdot 10^{11}\,\mathrm{C/kg} \qquad (324)$$

for all metals, with small variations. In short, electrical current has mass. Therefore, whenever one switches on an electrical current, one gets a *recoil*. This simple effect can easily be Ref. 5 measured and confirms the mass to charge ratio just given. Also the emission of current into air or into vacuum is observed; in fact, every television tube uses this principle to generate Ref. 6 the beam producing the picture. It works best when metal objects have a sharp, pointed tip. The rays created this way – one could say that they are “free” electricity – are called *cathode rays*. Within a few percent, they show the same mass to charge ratio as expression (324). The experiment shows that charges in metals move almost freely; that is the reason metals are such good conductors.

If electric charge falls inside vertical metal rode, one can take the astonishing deduction that cathode rays – as we will see later, they consist of free electrons – should not be able to fall through a vertical metal tube. This is due to the fact that the electrical field generated

Challenge * By the way, are batteries sources of charges?

** Maxwell tried to detect these effects (apart from the last one, which he did not predict), but his apparatuses where not sensitive enough.

by the displaced electricity in the tube precisely compensates the acceleration of gravity, so that electrons should not be able to fall through long thin cylinders. The experiment has indeed been performed, and a reduction of the acceleration of free fall for electrons of 90% has been observed. Can you imagine why the ideal value of 100% is not achieved? Ref. 7

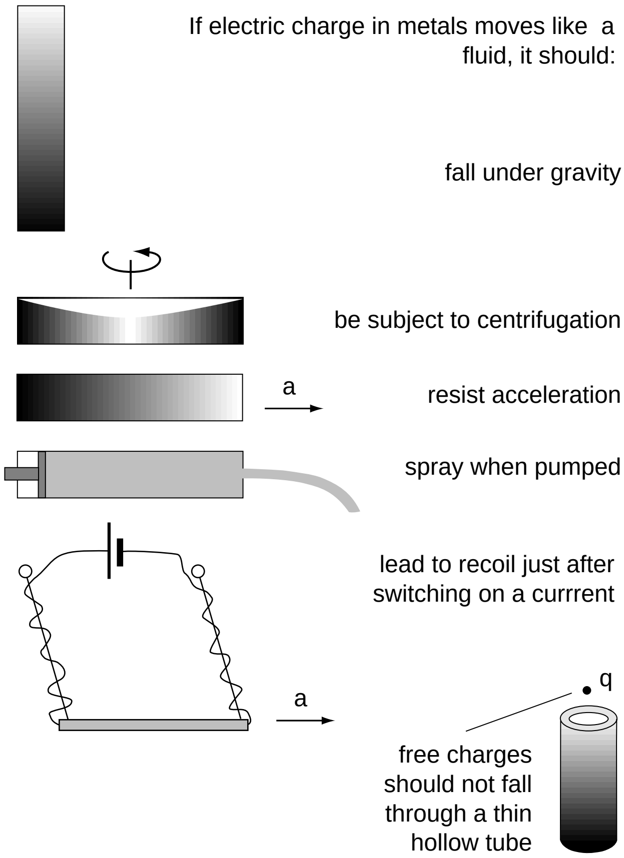


Figure 115 Consequences of the flow of electricity

How fast do charges move?

In vacuum, such as inside a colour television, charges accelerated by a tension of 30 kV move with a third of the speed of light. In modern particle accelerators charges move so rapidly that their speed is indistinguishable from that of light for all practical purposes. Challenge

In metals, electric signals move roughly with speeds around the speed of light. (Actually, the precise value depends on the capacity of the cable, and is usually in the range $0.3c$ to $0.5c$.) But when one measures the speed of charges inside metals one gets the same value as for ketchup inside its bottle, namely around 1 mm/s. Are you able to explain this apparent contradiction? Challenge

In atoms, electrons* behave even more strangely. One tends to think that they turn around the nucleus (as we will see later) at rather high speed, as the orbit is so small. However, it turns out that in most atoms many electrons do not turn around the nucleus at all. The strange story behind this fact will be told in the second part of this escalation.

Inside fluids, charges move with different speed than inside metals, and their charge to mass ratio is also different. We all

∗ The name ‘electron’ is due to Johnstone Stoney.

know that from direct experience. Our *nerves* work by using electric signals and take (only) a few milliseconds to respond to stimuli, even though they are meters long. A similar speed is observed inside semiconductors and inside batteries. The details will become clear in the next part of the escalation.

How can one make a motor?

Communism is soviets plus electricity.
Lenin (1870, Simbirsk-1924, Gorki)

The reason for this famous statement were two discoveries made in 1820 by the danish physicist Hans Christian Oersted (1777–1851) and in 1831 by the english physicist Michael Faraday.* The consequences of these experiments changed the world completely in less than one century. Oersted found that when a current is sent through a wire, the wire and a nearby magnet are attracted or repelled. In other words, he found that the flow of electricity can move bodies.

Similar experiments show that *two* wires in which charges flow attract or repel each other, depending on whether the currents are parallel or antiparallel. These and similar experiments show that wires in which electricity flows behave the like magnets.** In particular, magnetic fields have always two poles, usually called the north and the south pole. Opposite poles attract, similar poles repel each other. As is well known, the earth is itself a large magnet, with its magnetic north pole near the geographic south pole, and vice versa.

* Michael Faraday (1791, Newington, Surrey–1867) born in a simple family, without schooling, of deep and simple religious ideas, as a boy he became assistant of the most famous chemist of his time, Humphry Davy. Without mathematical training, at the end of his life he became member of the Royal society. A modest man, he refused all other honours in his life. He worked on chemical topics, the atomic structure of matter, and most of all, developed through all his experimental discoveries, such as induction, paramagnetism, diamagnetism, electrochemistry, the Faraday effect, and the idea of (magnetic) field and field lines. Fields were described mathematically by Maxwell, who at his time was the only one in Europe who took over the concept.
** In fact, if one imagines tiny currents moving in circles inside magnets, one has the same description for all magnetic fields observed in nature.

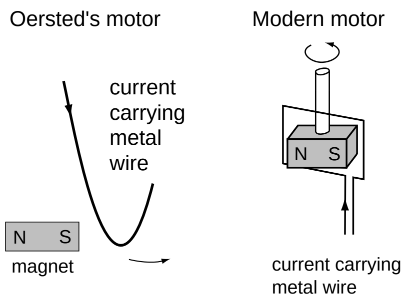


Figure 116 An old and a modern version of an electric motor

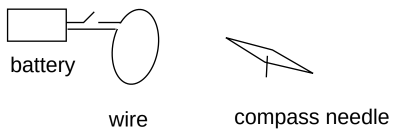


Figure 117 An electrical current always produces a magnetic field

Experiments show that the magnetic field turns out always to have a given direction in space, and to have a magnitude common to all (resting) observers. One is tempted to describe it by a vector. However, this is wrong, since a magnetic field does not behave like an arrow when placed before a mirror. It turns out that a magnetic field pointing towards a mirror does not change direction for the mirror set up. Are you able to confirm this using what was told about magnetic fields up to know? Challenge

In other words, it is not completely correct to describe a magnetic field by a vector $\mathbf{B} = (B_x, B_y, B_z)$; the precise way is to describe it by the quantity*

$$\mathrm{B} = \begin{pmatrix} 0 & -B_z & B_y \\ B_z & 0 & -B_x \\ -B_y & B_x & 0 \end{pmatrix} \quad , \tag{325}$$

called an *antisymmetric tensor*. (It is also called a *pseudovector*; note that also angular momentum and torque are examples of such quantities.) Magnetic fields are defined by the acceleration they impart on moving charges. This acceleration turns out to follow

$$\mathbf{a} = \frac{e}{m}\mathbf{v}\mathrm{B} = \frac{e}{m}\mathbf{v} \times \mathbf{B} \tag{326}$$

a relation which is often called *Lorentz acceleration* after the important dutch physicist Hendrik A. Lorentz (Arnhem, 1853–Haarlem, 1928) who first stated it clearly. It describes the basic effect in any electric motor. The motor is a device using magnetic fields efficiently for acceleration of charges flowing in a wire; through their motion the wire – or the interacting magnet – is moved as well.

Like in the electric case, we now need to know how the strength of magnetic fields is determined. Experiments show that the magnetic field is due to moving charges, and that a

* The quantity B was not called 'magnetic field' until recently. We follow here the modern, logical definition, which is superseding the traditional one, in which B was called the 'magnetic flux density' or 'magnetic induction' and a different quantity, **H**, was called – incorrectly – the magnetic field. That quantity will not appear in this walk, but is important for the description of magnetism in materials.

charge moving with velocity **v** produces a field given by

$$\mathrm{B}(\mathbf{r}) = \frac{\mu_o}{4\pi} q \frac{\mathbf{v} \times \mathbf{r}}{r^3} \quad \text{where} \quad \frac{\mu_o}{4\pi} = 10^{-7}\,\mathrm{N/A^2} \tag{327}$$

It is easy to see that the field has an intensity given by $\mathbf{v}\mathbf{E}/c^2$, where **E** is the electric field measured by an observer moving with the charge. It looks as if magnetism is a relativistic effect; later on we will confirm this conclusion.* Challenge

In 1831, Michael Faraday discovered an additional piece of the puzzle. He found that a moving magnet could cause a current flow in an electrical circuit. This allowed the production of electrical current flow with generators, using water power, wind power or steam power, thus starting the modern use of electricity in our world.

Additional experiments show that magnetic fields also lead to electric fields when one changes to a moving viewpoint, as you might check on any of the examples of the figures 116 to 120. *Magnetism indeed is relativistic electricity*. Electric and magnetic fields are partly transformed into each other when switching from one inertial reference frame to the other. The theory of special relativity then tells us that there must be a single concept, an *electromagnetic field*, describing them both. Investigating the details, one finds that the electromagnetic field F surrounding charged bodies has to be described by an antisymmetric 4-*tensor*

$$\mathrm{F}^{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix} \quad \text{or} \quad \mathrm{F}_{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix} \tag{328}$$

Actually, the expression for the field contains everywhere the expression $1/\sqrt{\mu_o\varepsilon_o}$ instead of the speed of light c. We will explain the reason for this substitution shortly. The total influence of electric and magnetic fields on fixed or moving charges is then given by the following expression:

$$\begin{aligned} & m\mathrm{b} = \mathrm{F} \wedge u \quad \text{or} \\ & m\frac{du^\mu}{d\tau} = q\mathrm{F}^\mu{}_\nu u^\nu \quad \text{or} \\ & m\frac{d}{d\tau}\begin{pmatrix} \gamma c \\ \gamma v_x \\ \gamma v_y \\ \gamma v_z \end{pmatrix} = q \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix} \begin{pmatrix} \gamma c \\ \gamma v_x \\ \gamma v_y \\ \gamma v_z \end{pmatrix} \quad \text{or} \\ & W = q\mathbf{E}\mathbf{v} \quad \text{and} \quad d\mathbf{p}/dt = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \end{aligned} \tag{329}$$

In fact, this extended Lorentz relation is the *definition* of the electromagnetic field, since the field is defined as that "stuff" which accelerates charges. In particular, all devices which put charges into motion, such as batteries and dynamos, as well as all devices which are put into motion by flowing charges, such as electric motors and muscles, are described by this relation. That is why it is usually studied already in high school. The Lorentz relation

* Equation (327) is valid only for small velocities and accelerations.

describes all cases in which the motion of objects can be seen by the naked eye or felt by our senses, such as the movement of electrical motors in high speed trains, in elevators and in dental drills, the motion of the picture generating electron beam in television tubes, or the traveling of electrical signals in cables and in the nerves of the body. Ref. 8, 9

The electromagnetic field tensor F is an antisymmetric 4-tensor. (Can you write down the relation between $F^{\mu\nu}$, $F_{\mu\nu}$, and $F^{\mu}{}_{\nu}$?) Like any such tensor, it has two invariants, i.e., two properties which are the same for every observer: the expression $B^2 - E^2/c^2 = \frac{1}{2}\mathrm{tr}\,F^2$ and the product $4\mathbf{EB} = -c\,\mathrm{tr}\,F^*F$. (Can you confirm this?) meaning will become clear shortly. The first expression turns out to be the lagrangian of the electromagnetic field. The first invaraint, a scalar, implies that if E is larger, smaller, or equal cB for one observer, it is also for all others. The second invariant, a pseudoscalar, describes whether the angle between the electric and the magnetic field is acute or obtuse for all observers.* Challenge Challenge

The application of electromagnetic effects to daily life has opened up a whole new world which did not exist before. Electrical light, electric motors, radio, telephone, X-rays, television, and computers changed human life completely in less than one century. For example, the installation of electric lighting in city streets has almost eliminated the previously so common night assaults. These and all other electrical devices use the fact that charges can flow in metals, that one can translate electromagnetic energy into mechanical energy (sound, motors), into light (lamps), into heat and coldness (ovens, refrigerators), that one can send electromagnetic fields across the air (radio and television, remote controls), and that one can use electric or magnetic fields to store information (computers).

How motors prove relativity right

All electric motors are based on the result that electric currents interact with magnetic fields. The simplest example is the attraction of two wires carrying parallel currents. This observation alone, made in 1820 by Ampère,** is sufficient to make motion larger than a certain maximal speed impossible. Ref. 11

* There is in fact a third Lorentz invariant, much less known. It is specific to the electromagnetic field and is a combination of the field and its vector potential:

$$\begin{aligned}\kappa_3 &= \frac{1}{2}A_\mu A^\mu F_{\rho\nu}F^{\nu\rho} - 2A_\rho F^{\rho\nu}F_{\nu\mu}A^\mu \\ &= (\mathbf{AE})^2 + (\mathbf{AB})^2 - |\mathbf{A}\times\mathbf{E}|^2 - |\mathbf{A}\times\mathbf{B}|^2 + 4a^4(\mathbf{AE}\times\mathbf{B}) - a^8(E^2+B^2)\end{aligned} \tag{330}$$

Ref. 10 This expression is Lorentz (but not gauge) invariant; knowing it can help clarify unclear issues, such as the lack of existence of waves in which the electric and the magnetic field are parallel. Indeed, for plane monochromatic waves all three invariants *vanish* in the Lorentz gauge. Also the quantities $\partial_\mu J^\mu$, $J_\mu A^\mu$ and $\partial_\mu A^\mu$ are invariants. (Why?) The latter, the frame idependence of the divergence of the four potential, reflects the invariance of gauge choice. The gauge in which the expression is set to zero is called the *Lorentz gauge*. Challenge

** André-Marie Ampère (1775, Lyon–1836, Marseille), french physicist and mathematician. Autodidact, he read the famous *encyclopédie* as a child; in a life full of personal tragedies, he wandered from maths to chemistry and physics, worked as a high school teacher, and published nothing of importance until 1820. Then the discovery of Oersted reached all of Europe: current can deviate magnetic needles. Ampere worked for years on the problem, and published in 1826 the summary of his findings, which lead Maxwell to call him the Newton of electricity. He named and developed many parts of electrodynamics. The unit of electrical current is named after him.

The argument is beautifully simple. We change the original experiment and imagine two long, electrically charged rods, moving with velocity v and separation d in the same direction. An observer moving with the rods would see a repulsion between the rods due to electrostatic forces, given by

Challenge

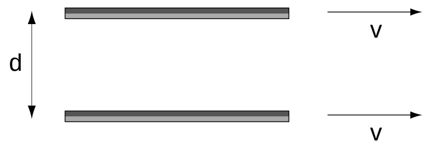


Figure 118 The relativistic aspect of magnetism

$$F_e = -\frac{1}{4\pi\varepsilon_o}\frac{2\lambda^2}{d} \tag{331}$$

where λ is the charge per length of the rods. A second observer at rest sees two effects: the electrostatic repulsion and the attraction discovered by Ampère. He therefore observes a force

Challenge

$$F_{em} = -\frac{1}{4\pi\varepsilon_o}\frac{2\lambda^2}{d} + \frac{\mu_o}{2\pi}\frac{\lambda^2 v^2}{d} \tag{332}$$

It is easy to check that the second observer sees a repulsion, as the first one does, only if

$$v^2 < \frac{1}{\varepsilon_o \mu_o} \quad . \tag{333}$$

This maximum speed is thus valid for any object carrying charges. Are you able to expand the argument for neutral particles as well?

Challenge

In summary, magnetism is due to moving electric charges. (It is not due to magnetic charges.) In short, any running electric motor proves relativity right. But or description is not complete yet; we need the final description of the way charges *produce* the electromagnetic field.

The description of electromagnetic field evolution

In the years between 1861 and 1865, pondering the details of all experiments known to him, James Clerk Maxwell produced a description of electromagnetism which forms one of the pillars of physics.* Maxwell took all experimental results and extracted their common basic principles, shown in figures 119 and 120. Twenty years later, Heaviside and Hertz extracted the main points and called the summary *Maxwell's theory of the electromagnetic field*. They consist of two equations (four in the nonrelativistic case).

See page 752

∗ James Clerk Maxwell (1831, Edinburgh–1879, Cambridge), scottish physicist; founded electromagnetism by unifying electricity and magnetism theoretically, as described in this chapter. His work on thermodynamics forms a second pillar of his activity. In addition, he also studied the theory of colour and developed the now standard horseshoe colour diagram; he was one of the first persons to make a colour photograph. He is often seen as the greatest physicist ever. Clerk and Maxwell were both his family names.

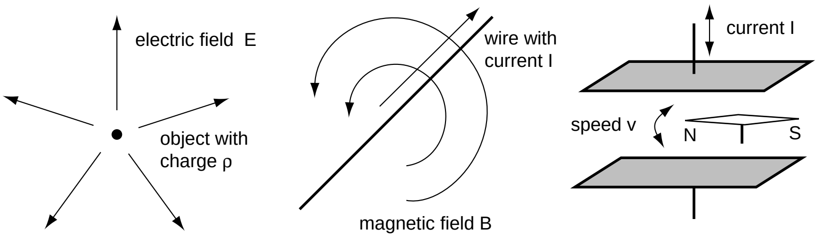


Figure 119 The first of Maxwell's equations

The first result is the precise description for the fact that electromagnetic fields originate at charges. The corresponding equation is*

$$dF = j\sqrt{\frac{\mu_o}{\varepsilon_o}} \quad \text{or}$$

$$d^\nu F_{\mu\nu} = j^\mu\sqrt{\frac{\mu_o}{\varepsilon_o}} \quad \text{or} \tag{334}$$

$$(\partial_t/c, -\partial_x, -\partial_y, -\partial_z)\begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix} = \sqrt{\frac{\mu_o}{\varepsilon_o}}(\rho, j_x/c, j_y/c, j_z/c) \quad \text{or}$$

$$\nabla\mathbf{E} = \frac{\rho}{\varepsilon_o} \quad \text{and} \quad \nabla\times\mathbf{B} - \frac{1}{c^2}\frac{\partial\mathbf{E}}{\partial t} = \mu_o\mathbf{j} \quad .$$

and puts in many signs a simple statement: *electrical charge carries the fields*.

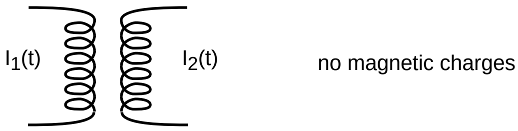


Figure 120 The second of Maxwell's equations

The second result by Maxwell is the precise description of how changing electric fields create magnetic fields, and vice versa. In particular, the electric field can have vortices only when there is a changing magnetic field. In addition it expresses the observation that in nature there are no magnetic charges, i.e. that magnetic fields have no sources. All these

* Maxwell generalized this equation to cases that the charges are not surrounded by vacuum, but located inside matter. We do not explore these situations in our walk; as we will see during our escalation, the apparently special case of vacuum in fact describes all of nature.

results are described by the the relation

$$
\begin{aligned}
& d\,{}^*\mathrm{F} = 0 \quad \text{with} \quad {}^*\mathrm{F}^{\rho\sigma} = \tfrac{1}{2}\varepsilon^{\rho\sigma\mu\nu}\mathrm{F}_{\mu\nu} \quad \text{or} \\
& \varepsilon_{\mu\nu\rho}\partial_\mu \mathrm{F}_{\nu\rho} = \partial_\mu \mathrm{F}_{\nu\rho} + \partial_\nu \mathrm{F}_{\rho\mu} + \partial_\rho \mathrm{F}_{\mu\nu} = 0 \quad \text{or} \\
& \begin{pmatrix} \gamma\frac{1}{c}\partial_t \\ \gamma\partial_x \\ \gamma\partial_y \\ \gamma\partial_z \end{pmatrix} \begin{pmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & -E_z/c & E_y/c \\ -B_y & E_z/c & 0 & -E_x/c \\ -B_z & -E_y/c & E_x/c & 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
& \nabla \mathbf{B} = 0 \quad \text{and} \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}
\end{aligned}
\qquad (335)
$$

expressing the *lack of sources* for the *dual field tensor* ${}^*\mathrm{F}$. The two evolution equations have the same central status as the expression $a = GM/r$ for gravitation: they describe how the electromagnetic field moves given the motion of the charges. Together with Lorentz' evolution equation (329), which describes how charges move given the motion of the fields, Maxwell's evolution equations (335) and (336) describe all electromagnetic phenomena at everyday scales, from portable phones, car batteries, to personal computers, lasers, lightnings, and rainbows.

The gauge field: the electromagnetic vector potential

The study of moving fields is called *field theory*, and electrodynamics is the major example. (The other classical example is fluid dynamics.) Field theory is a beautiful topic; field lines, equipotential lines, and vortex lines are some of the concepts introduced in this domain. They fascinate many.* However, in this escalation we keep the discussion focussed on motion.

We have seen that fields force us to extend our concept of motion. Motion is not only the state change of objects and space-time, but also the *state change of fields*. We therefore need, also for fields, a complete and precise description of their state. The observations with amber and magnets have shown us that fields possess energy and momentum, which can be imparted to particles. The experiments with motors have shown that objects can add energy and momentum to fields. One therefore has to define a *state function* which allows to define energy and momentum for electric and magnetic fields.

Maxwell defined the state function in two standard steps.

The first step is the definition of the *(magnetic) vector potential*, which describes the
Ref. 12 momentum per charge:

$$\mathbf{A} = \frac{\mathbf{p}}{q} \quad . \qquad (336)$$

When a charged particle moves through a magnetic potential $\mathbf{A}(\mathbf{x})$, its momentum changes by $q\mathbf{A}$. Due to this definition, the vector potential has the property that

$$\mathrm{B} = \nabla \times \mathbf{A} = \text{curl}\, \mathbf{A} \qquad (337)$$

∗ See for example the *free* textbook by Bo THIDÉ, *Electromagnetic Field Theory*, on his
Ref. 8 http://www.plasma.uu.se/CED/Book web site. And of course, in english, the text by Jackson.

i.e. that the magnetic field is the curl (or rotation) of the magnetic potential.* For example, the vector potential for a long straight current carrying wire is parallel to the wire, and has the value Challenge

$$A(r) = -\frac{\mu_o I}{2\pi} \ln(r) \tag{338}$$

depending on the distance r from the wire. For a solenoid, the vector potential 'circulates' around it. Inside the solenoid, the vector potential increases from the centre. Similarly, for a constant and uniform magnetic field B one finds the vector potential

$$\mathbf{A}(\mathrm{r}) = -\frac{1}{2}\mathbf{B} \times \mathbf{r} \quad . \tag{339}$$

The magnetic potential is *not* uniquely defined. If $\mathbf{A}(\mathbf{x})$ is a vector potential, then

$$\mathbf{A}'(\mathbf{x}) = \mathbf{A}(\mathbf{x}) + \mathrm{grad}\Lambda \tag{340}$$

where $\Lambda(t, \mathbf{x})$ is some scalar function, is a also vector potential for the same situation. What happens to the corresponding momentum values? They also change.

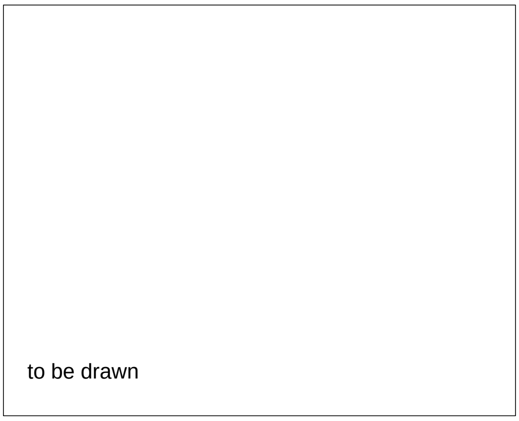


Figure 121 Vector potentials for selected situations

One is more accustomed to the fact that like momentum, also the energy of the magnetic field is defined ambiguously. Indeed, the second step is definition of the *electric potential* as the energy U per charge: Ref. 12

$$\varphi = \frac{U}{q} \tag{341}$$

In other words, the potential $\varphi(\mathbf{x})$ at a point $\mathbf{x}$ is the energy needed to move a unit charge to the point $\mathbf{x}$ starting from a point where the potential vanishes. The potential energy is thus given by $q\varphi$. Due to this definition, the electric field $\mathbf{E}$ is simply the *change* of the potential with position, i.e.

$$\mathbf{E} = -\nabla\varphi - \frac{\partial}{\partial t}\mathbf{A} \quad , \tag{342}$$

Obviously, there is a freedom in the choice of the definition of the potential. If $\varphi(\mathbf{x})$ is a potential, then

$$\varphi'(\mathbf{x}) = \varphi(\mathbf{x}) - \frac{\partial}{\partial t}\Lambda \tag{343}$$

is also a potential function for the same situation. This freedom is the generalization of the fact that energy is defined only up to a constant.

* The curl is called the *rotation* and abbreviated rot in most languages.

In relativistic 4-vector notation, the state function of the electromagnetic field becomes

$$A^{\mu} = (\varphi/c, \mathbf{A}) \quad . \tag{344}$$

It is easy to see that it is a complete description of the field, since one has

$$\mathrm{F} = dA \quad \text{or} \quad \mathrm{F}^{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} \tag{345}$$

which means that the electromagnetic field is completely specified by the 4-potential A. As just said, the 4-potential is not uniquely defined. Indeed any other gauge field A' related to A by the *gauge transformation*

$$A'^{\,\mu} = A^{\mu} + \partial^{\mu}\Lambda \tag{346}$$

where $\Lambda = \Lambda(t,x)$ is any arbitrarily chosen scalar field, leads to the *same* electromagnetic field, and to the same accelerations and evolutions. The gauge 4-field A is thus an *overdescription* of the physical situation as several *different* A correspond to the *same* physical situation. Therefore one has to ensure that all measurement results are independent of gauge transformations, i.e. that all observables are gauge invariant quantities. Such gauge invariant quantities are, as we just saw, the fields F and *F, and in general all classical quantities.

There is a simple image, due to Maxwell, to help overcoming this conceptual difficulty of the vector potential. It turns out that the closed line integral over $\mathbf{A}$ is gauge invariant, because

$$\oint A_{\mu}dx^{\mu} = \oint (A_{\mu} + \partial_{\mu}\Lambda)dx^{\mu} = \oint A'_{\mu}dx^{\mu} \quad . \tag{347}$$

In other words, if one pictures the vector potential as a quantity allowing to associate a number to a tiny ring at each point in space, one gets a good, gauge invariant picture of the vector potential.*

We note that many theoretical physicists use the term 'electromagnetic field´ for the quantity A_{μ}. Now that we have defined a state function which describes energy and momentum, let us see what happens in more detail when electromagnetic fields move.

Colliding charged particles

A simple experiment elucidates the just defined properties of electromagnetic fields. When two charged particles collide, their total momentum is *not* conserved.

Imagine two particles of identical mass and charge just after a collision, when they move from each other. Imagine also that the two masses are large, so that the acceleration due to their electrical repulsion is small. For an observer in the centre of gravity of the two, each particle feels an acceleration from the electrical field of the other, given by the so-called *Heaviside formula*

$$E = \frac{q(1-u^2/c^2)}{4\pi e_{o} r^2} \quad . \tag{348}$$

In other words, the total system has a vanishing total momentum.

Ref. 13 * In the part on quantum mechanics we will see that the exponent of this expression, namely $\exp(iq \oint A_{\mu}dx^{\mu})$, usually called the *phase factor*, can indeed be directly observed in experiments.

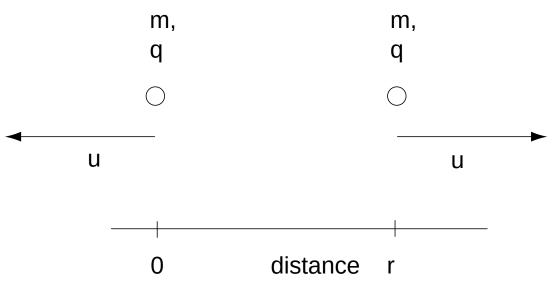


Figure 122 Charged particles after a collision

Take a second observer, moving with respect to the first with velocity u, so hat the first charge will be at rest. The expression of the electrical field leads to two different values for the electric fields at the position of each particle. In other words, the system of the two particles is not in inertial motion, as we would expect; the total momentum is not conserved. Where did it go? Ref. 14 Challenge

This at first surprising effect has even been put in form of a theorem, by Van Dam and Wigner. They showed that for a system of particles interacting at a distance the total energy-momentum cannot remain constant in all inertial frames. Ref. 15

The total momentum of the system is conserved only because the electromagnetic field carries momentum. In other words, electromagnetic fields are able to *hit* objects and be hit by them. As we will show below, light is an electromagnetic field. One then should be able to move objects by shining light onto them. One should even be able to suspend particles in mid air by shining light onto them from below. Both predictions are correct, and a few experiments describing them will be presented shortly.

One notes that all sort of fields leading to particle interactions must carry energy and momentum, as the argument applies to all such cases, in particular to the nuclear interactions. In the second part of our escalation we will also discover that all fields are in fact made of particles.

The lagrangian of electromagnetism

The motion of charged particles and the motion of the electromagnetic field can also be described using a lagrangian instead of using the three equations given above. It is not hard to see that the action S_{CED} for a particle in classical electrodynamics can be symbolically defined by* Challenge

$$S_{\text{CED}} = mc^2 \int d\tau - \int \tfrac{1}{2\mu_o} \mathrm{F} \wedge * \mathrm{F} - c \int j \wedge A \qquad (349)$$

which in index notation is given by

$$S_{\text{CED}} = mc^2 \int_{-\infty}^{\infty} \sqrt{\eta_{\mu\nu} \frac{dx_n^{\mu}(\theta)}{dp} \frac{dx_n^{\nu}(\theta)}{dp}}\, d\theta - \int_{\mathbf{M}} \left(\tfrac{1}{4\mu_o} \mathrm{F}_{\mu\nu} \mathrm{F}^{\mu\nu} + j_\mu A^\mu \right) d^4x \quad .$$

* The symbol $\wedge$, 'wedge', in fact has a precise mathematical meaning; but its background, the concept of *(mathematical) form*, carries us too far from our walk. An electrodynamics text completely written with forms is Kurt Meetz & Walter L. Engl, *Elektromagnetische Felder – Mathematische und physikalische Grundlagen*, Springer, 1980.

The action S_{CED} leads to the evolution equations by requiring that it be stationary under variations δ, δ' of the positions and of the fields which vanish at infinity, i.e. that

$$\begin{aligned} \delta S = 0 \quad & \text{when} \quad x_\mu = x_\mu + \delta_\mu \quad \text{and} \quad A_\mu = A_\mu + \delta'_\mu \quad , \\ & \text{provided} \quad \delta x_\mu(\theta) \to 0 \quad \text{for} \quad |\theta| \to \infty \\ & \text{and} \quad \delta A_\mu(x_\nu) \to 0 \quad \text{for} \quad |x_\nu| \to \infty \quad . \end{aligned} \tag{350}$$

See page 103
Challenge

In the same way as in the case of mechanics, using the variational method for the two variables A and x, one recovers the evolution equations for particle and fields

$$b^\mu = \frac{q}{m} \mathrm{F}^\mu_\nu u^\nu \quad , \quad \partial_\mu \mathrm{F}^{\mu\nu} = j^\nu \sqrt{\frac{\mu_o}{\varepsilon_o}} \quad , \quad \text{and} \quad \varepsilon^{\mu\nu\rho\sigma} \partial_\nu \mathrm{F}_{\rho\sigma} = 0 \tag{351}$$

which we know already. Obviously, they are equivalent to the variational principle based on S_{CED}. Both have to be completed by specifying *initial conditions* for the particles and the fields, as well as *boundary conditions* for the latter. One needs first and zeroth derivatives of the position of the particles, and zeroth derivative for the electromagnetic field.

Note that this result also implies that electromagnetism is *time reversible*. That means that every example of motion due to electric or magnetic causes can also take place backwards. This is easily deduced from the properties of the lagrangian. On the other hand, everyday life shows many electric and magnetic effects which are not time invariant, such as electromagnetic breaking, electric light bulbs, etc. Can you explain how this fits together?

Challenge

With lagrangian (349), all of classical electrodynamics is described and understood. For the rest of this chapter, we look at some specific topics from this vast field.

Symmetries: the energy-momentum tensor

We know from classical mechanics that we get the definition of energy and momentum tensor by using Noether's theorem, if we determine the conserved quantity from the Lorentz symmetry of the lagrangian. For example, we fond that relativistic particles have an energy-momentum *vector*. At the point at which the particle is, it describes the energy and momentum.

Since the electromagnetic field is not a localised entity like a point particle, but extended, one needs to know the flow of energy and momentum at a point separately *for each direction*. This makes a description with a *tensor* necessary.

– CS – to be continued – CS –

In summary, electrodynamic motion, like all other examples of motion encountered so far, is deterministic, is conserved, and is reversible. No big news. But a somewhat special symmetry deserves a special mention.

What is the difference between electric and magnetic fields?

Obviously, the standard answer is that electric fields have sources, and magnetic fields do not; that moreover magnetic fields are small relativistic effects of importance only when charge velocities are high or when electrical fields cancel out.

For situations with matter, this clear distinction is correct. Up to the present day, no particle with a magnetic charge, called a *magnetic monopole*, has ever been found, even though its existence is possible in several unified models of nature. If found, the action (349) would have to be modified by the addition of a fourth term, namely the magnetic current density. No such particle has yet been detected, despite intensive search efforts. See page 553

But in vacuum, when matter is not around, it is possible to take a completely different view. In vacuum the electric and the magnetic field can be seen as two faces of the same quantity, since a transformation such as

$$\begin{aligned} \mathbf{E} &\rightarrow c\mathbf{B} \\ \mathbf{B} &\rightarrow -\mathbf{E}/c \end{aligned} \tag{352}$$

called (electromagnetic) *duality* transformation, transforms each vacuum Maxwell equation into the other. (In fact, there are even more such transformations; can you spot them?) In fact, the duality transformation transforms F into *F. In other words, in vacuum one cannot really distinguish electric from magnetic fields. Challenge

Matter would be symmetric under duality only if magnetic charges, also called magnetic monopoles, would exist. In that case the transformation (352) could be extended to

$$c\rho_e \rightarrow \rho_m \quad , \quad \rho_m \rightarrow -c\rho_e \quad . \tag{353}$$

However, magnetic monopoles have not been found, despite intensive searches.

It was one of the great discoveries of theoretical physics that even though classical electrodynamics with matter is not symmetric under duality, nature is. In 1977, Claus Montonen and David Olive showed that quantum theory allows duality transformations, even with the inclusion of matter. It was already known since the 1930s that quantum theory allows magnetic monopoles. We will discover the important ramifications of this result in the third part of the text. This duality turns out to be one of the essential stepping stones leading to a unified description of motion. (A more difficult question: Extending this duality to quantum theory, can you deduce what a transformation is found for the fine structure constant, and why it is so interesting?) See page 691 Challenge

What is light?

An important consequence of the equations of electrodynamics was deduced by Maxwell in 1865. He found that in the case of vacuum, the equations of the electrodynamic field could be written as

$$\Box\mathbf{A} = 0 \quad \text{or} \quad \varepsilon_o\mu_o\frac{\partial^2\varphi}{dt^2} + \frac{\partial^2 A_x}{dx^2} + \frac{\partial^2 A_y}{dy^2} + \frac{\partial^2 A_z}{dz^2} = 0 \quad . \tag{354}$$

This is called a *wave equation*, because it admits solutions of the type Challenge

$$\mathbf{A}(t,x) = \mathbf{A}_0 \sin(\omega t - \mathbf{kx} + \delta) \tag{355}$$

which are commonly called *waves*. Such a wave satisfies equation (354) for any value of the *amplitude* A_0, of the *phase* δ, and of the *circular frequency* ω, provided the *wave vector*

k satisfies the relation

$$\omega(\mathbf{k}) = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} k \quad \text{or} \quad \omega(\mathbf{k}) = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \sqrt{\mathbf{k}^2} \qquad (356)$$

The relation $\omega(\mathbf{k})$ between the circular frequency and the wave vector, the so-called *dispersion relation*, is the main property of any type of wave, be it a sound wave, a water wave, an electromagnetic wave, or any other kind. Relation (356) specifically characterizes electromagnetic waves in vacuum, and distinguishes them from all other types of waves.*

Equation (354) for the electromagnetic field is *linear* in the field; this means that the sum of two situations allowed by it is itself an allowed situation. Mathematically speaking, any *superposition* of two solutions is a solution as well. For example, this means that two waves can travel through each other without disturbing each other, and that waves can travel through static electromagnetic fields. Linearity also means that any electromagnetic wave can be described as a superposition of pure sine waves, each of which is described by expression (355).

After Maxwell had predicted the existence of electromagnetic waves, in the years between 1885 and 1889 Heinrich Hertz** discovered and studied them, by fabricating a very simple transmitter and receiver for 2 GHz waves. Waves around this frequency are used in the last generation of mobile telephones. These waves are now called *radio waves*, since physicists tend to call all moving force fields *radiation*, recycling an old term which originally meant 'light emission.'

Hertz also measured the speed of these waves; today everybody can do that by himself by telephoning to a friend on another continent using a satellite line (just use a low cost provider). There is about half a second additional delay between the end of a sentence and the answer of the friend, compared to normal day-to day situations. In this half second, the signal goes up to the geostationary satellite, down again and back. This gives a speed of $c \approx 4 \cdot 36000\,\mathrm{km}/0.5\,\mathrm{s} \approx 3 \cdot 10^5\,\mathrm{km/s}$, which is close to the precise value.

But Maxwell did more. He also predicted that *light* itself is a solution of equation (355) and therefore an electromagnetic wave, only with a much higher frequency. This famous prediction can be checked in many ways.

It is easy to confirm the *wave* properties of light, and indeed they were known already long before Maxwell. In fact, the first to suggest that light is a wave was, around the year 1678, the important dutch physicist Christiaan Huygens (1629, 's Gravenhage –1695, Hofwyck). One can confirm this fact with one's own fingers. Simply put your hand one or two centimeters from the eye, look towards the sky through the gap between the middle finger and the index, and let the two fingers almost touch. You will see a number of dark lines dividing the gap. These lines are the interference pattern formed by the light behind the slit created by the fingers. *Interference* is the name given to those amplitude patterns which appear when several waves superpose.*** This experiment therefore also allows to estimate the wavelength of light, and thus if you know its speed, also its frequency. Are you able to do so? Challenge

* Just to be complete, a *wave* in physics is any propagating imbalance. Other types of waves, such as sound, water waves, earthquakes, etc., will not be studied much in this escalation.

** Heinrich Rudolf Hertz (1857, Hamburg–1894, Bonn), important hamburger theoretical and experimental physicist. The unit of frequency is named after him. Despite his early death, Hertz was a central figure in the development of electromagnetism and of radiocommunication technology. See also page 94

Challenge *** Where does the energy go in interference patterns?

Historically, a similar effect was central in convincing everybody that light was a wave: the supernumerary rainbows, the additional bows below the main rainbow. If one looks carefully at a rainbow, below the main reed yellow green blue violet bow, one observes weaker, additional green blue and violet bows. Depending on the intensity of the rainbow, several of these supernumerary rainbows can be observed. They are due to an interference effect, as Thomas Young showed around 1803.* (More about the rainbow below.) It seems that in those times scientists either did not trust their own fingers, or did not have any. Ref. 16 See page 332

There are many other ways that the wave character of light becomes apparent. Maybe the most beautiful is an experiment carried out by a team of dutch physicists in 1990. They simply measured the light transmitted through a slit in a metal plate. It turns out that the transmitted intensity depends on the slit of the hole. Their surprising result is shown in figure 123. Can you explain the origin of the unexpected steps in the curve? Ref. 17 Challenge

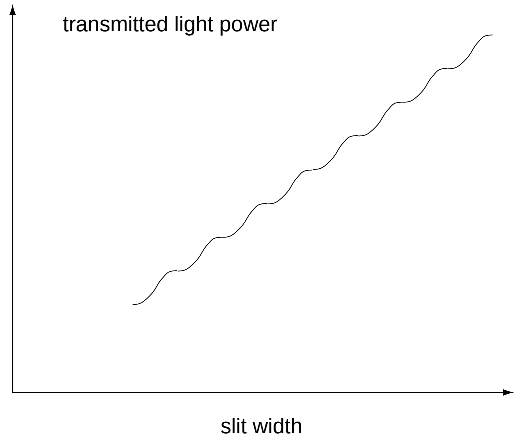


Figure 123 The light power transmitted through a slit as function of its width

Numerous other experiments on the creation, detection and measurement of electromagnetic waves have been performed in the twentieth century. The result of all these experiments is that electromagnetic waves can be distinguished first of all by their frequency or wavelength. The main categories are listed in table 29. See page 325

At the end of the twentieth century the final confirmation has become possible. Using quite sophisticated experiments it became possible to measure the oscillation frequency of light directly. The value, between 375 and 750 THz, is so high that detection was impossible for many years. But with these modern experiments the dispersion relation (356) of light has finally been confirmed completely. Ref. 18

An additional property of electromagnetic waves is hidden in the parameter $\mathbf{A}_0$ in expression (355). Even for identical frequency and phase, waves can still differ: they can have different polarization. The *polarization* is the direction in which the electromagnetic field oscillates with respect to the propagation direction. Polarization determines whether the radio antennas of receivers have to be kept horizontal or vertical. Polarization of light is easily achieved, e.g. by shining it through stretched plastic films. When the polarization of light was discovered in the nineteenth century, it definitively established its wave nature.

* Thomas Young (1773, Milverton–1829), read the bible at two, spoke latin at four; doctor of medicine, he became professor of physics. He introduced the concept of *interference* into optics, explaining the newtonian rings and rainbow, and was the first person to determine light's *wavelength*, a concept that he also introduced, and its dependence on colour. He was the first to deduce the three colour vision explanation of the eye and after reading of the discovery of polarization, explained light as a transverse wave. In short he discovered most what people learn at high school about light. He was a universal talent: he also worked on the deciphering of hieroglyphs, on ship building, and on engineering problems. In Britain his ideas on light were not accepted, since Newton's followers crushed all opposite ideas. He collaborated with Fraunhofer and Fresnel; and finally his ideas were made famous by Fresnel and Helmholtz.

By the way, the human eye is unable to detect polarization, in contrast to many insects. As is well known honey bees use polarization to deduce the position of the sun even when Ref. 19 it is hidden behind clouds, and many insects use polarization to distinguish water surfaces Challenge from mirages. Can you find out how?

So far it is clear that light is a wave. To confirm that the nature of light is indeed *electromagnetic* is more difficult. The first argument was by Maxwell. From equation (356), Maxwell deduced a prediction for the speed of electromagnetic waves, namely the celebrated expression

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \tag{357}$$

Challenge which you should be able to confirm. When he inserted the values in the right hand side, he found, within measurement errors, complete correspondence with the measured speed of light. Note that the right hand side contains electromagnetic values, and the left hand side is an optically measured entity. The expression thus unifies electromagnetism with optics.

Of course, people were not yet completely convinced. They looked for more ways to show that light is electromagnetic in nature. Since Maxwell's evolution equations are linear, electric or magnetic fields alone do not influence the motion of light. On the other hand, since electromagnetic waves are emitted only by accelerated charges, and since all light is emitted from matter, one follows that matter is full of electric charges. This implies that the influence of matter on light could be understood from its internal electric fields, and in particular, that subjecting matter to electromagnetic fields should change its emission of light or its interactions with light.

For example, it is indeed found that electric fields can influence the light transmission of oil, an effect discovered and named after Kerr. With time, many more influences between matter in fields on light were found, and a more extensive list is given in the table on page 349. It turns out that with a few exceptions the effects can *all* be described by the electromagnetic lagrangian S_{CED} (349), or equivalently, by Maxwell's equations (351). Classical electrodynamics indeed unifies the description of electricity, of magnetism, and of optics; all phenomena in these fields, from the rainbow to radio, from lightnings to electric motors, are found to be different aspects of the evolution of the electromagnetic field F.

Table 29 The electromagnetic spectrum

Frequency	wavelength	name	main properties	appearance	use
$3 \cdot 10^{-18}$ Hz	10^{26} m	**lower frequency limit**		see section on cosmology	
< 10 Hz	> 30 Mm	**quasistatic fields**		intergalactic, galactic, stellar, and planetary fields, brain, electrical fish	power transmission, accelerating and deflecting cosmic radiation
		radio waves		electronic devices	
10 Hz-50 kHz	30 Mm-6 km	ELW	go round the globe, penetrate into water	nerve cells, electromechanical devices	power transmission, communication with submarines

Frequency	wavelength	name	main properties	appearance	use
50 -500 kHz	6 km-0.6 km	LW	follow earth curvature, felt by nerves ("bad weather nerves")	emitted by thunderstorms	radio communications, telegraphy, inductive heating
500 - 1500 kHz	600 m-200 m	MW	reflected by night sky		radio
1.5 -30 MHz	200 m-10 m	SW	circle world if reflected by the ionosphere, destroy hot air balloons	emitted by stars	radiotransmissions, radioamateurs
15 -150 MHz	20 m-2 m	VHF	allow battery operated transmitters		remote controls, closed networks, tv, radio amateurs, radio navigation, military, police, taxi
150 - 1500 MHz	2 m-0.2 m	UHF	idem, line of sight propagation		radio, walkie-talkies, tv, cellular phones, internet via cable, satellite communication, bicycle speedometers
		microwaves			
1.5 -15 GHz	20 cm-2 cm	SHF	idem, absorbed by water	night sky, emitted by hydrogen atoms	radio astronomy, used for cooking (2.45 GHz), telecommunications, radar
15 -150 GHz	20 mm-2 mm	EHF	idem, absorbed by water		
		infrared	go through clouds	emitted by every warm object	satellite photography of earth
3 -100 THz	1000 -3 μm	IRC or far infrared		sunlight	
100 -210 THz	3 μm-1.4 μm	IRB or medium infrared		sunlight	used for optical fibre communications for telephone and cable TV
210 -385 THz	1400-780 nm	IRA or near infrared	penetrates for several cm into human skin	sunlight, radiation from hot bodies	healing of wounds, rheumatism, sport physiotherapy, hidden illumination
375 -750 THz	800-400 nm	**light**	not absorbed by air, detected by the eye (up to 850 nm at sufficient power)	heat ("hot light"), lasers & chemical reactions e.g. phosphor oxidation, fireflies ("cold light")	definition of straightness, enhancing photosynthesis in agriculture, photodynamic therapy, hyperbilirubinaemia treatment

Frequency	wavelength	name	main properties	appearance	use
375 -478 THz	780-627 nm	red	penetrate flesh	blood	alarm signal, used for breast imaging
	700 nm	pure red		rainbow	colour reference for printing, painting, illumination and displays
478 -509 THz	627-589 nm	orange		various fruit	attracts birds and insects
	600 nm	standard orange			
509 -530 THz	589-566 nm	yellow		majority of flowers	same
	580 nm	standard yellow			
530 -606 THz	566-495 nm	green	maximum eye sensitivity	algae and plants	
	546.1 nm	pure green		rainbow	colour reference
606 -688 THz	495-436 nm	blue		sky, gems, water	
	488 nm	standard cyan			
	435.8 nm	pure blue		rainbow	colour reference
688 -789 THz	436-380 nm	indigo, violet		flowers, gems	
		ultraviolet			
789 -952 THz	380-315 nm	UVA	penetrate ca. 1 mm into skin, darken it, produce vitamin D, suppress immune system, cause skin cancer, destroy eye lens	emitted by sun & stars	seen by certain birds, integrated circuit fabrication
0.95 - 1.07 PHz	315-280 nm	UVB	idem, destroy DNA, cause skin cancer	idem	idem
1.07 -3.0 PHz	280-100 nm	UVC	form oxygen radicals from air, kill bacteria, penetrate ca. 10 μm into skin	idem	disinfection, water purification, waste disposal, integrated circuit fabrication
3 -24 PHz	100-13 nm	EUV			sky maps, maybe silicon lithography
		X-rays	penetrate materials	emitted by stars and by black holes	imaging human tissue
24 -240 PHz	13-1.3 nm	soft X-rays	idem	synchrotron radiation	idem
$>$ 240 PHz or $>$ 1 keV	$<$ 1.2 nm	hard X-rays	idem	emitted when fast electrons hit matter	crystallography, structure determination

Frequency	wavelength	name	main properties	appearance	use
> 12 EHz or > 50 keV	< 24 pm	**γ-rays**	idem	radioactivity, cosmic rays	chemical analysis, disinfection, astronomy
$1.9 \cdot 10^{43}$ Hz	$\approx 10^{-35}$ m	**Planck limit**		see part three of this text	

The expression of the speed of light does not depend on the proper motion of the observer measuring the electromagnetic fields involved. This strange result was the first hint that the speed of light is a universal constant. However, it took several decades before the consequences we realized and relativity was developed.

As a note, it is often told that the teenager Albert Einstein asked himself what would happen if an observer would move at the speed of light, and in particular, what kind of electromagnetic field he would observe. He once explained that this Gedankenexperiment convinced him already at that young age that *nothing* could travel at the speed of light, since the field observed would have a property not found in nature. Can you guess which one? Challenge

Does light travel straight?

Usually this is the case, since we even use light to *define* 'straightness.' However, there are a few exceptions and every expert on motion should know them. Ref. 20

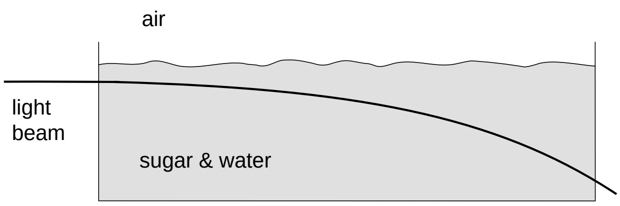


Figure 124 Sugar water bends light

In syrup, light curves, as shown in figure 124. In fact, light bends at any material interface. This effect, called *refraction*, is the same making aquaria seem less deep than they actually are. Refraction is a consequence of the change of speed of light at boundaries, and is indirectly due to its wave properties. Can you explain it, and with it thus explain the syrup effect? Challenge

Refraction in water droplets is also the basis of the rainbow, as shown on page 332, and refraction in ice crystals in the atmosphere is at the basis of the halos and the many other light patterns often seen around the sun and the moon. Ref. 21

A second important observation is that light goes around corners, and the more so the more they are sharp. This effect is called *diffraction* and is also due to the wave nature of light. You probably remember it from high school. In fact, light goes around corners in the same way that sound does.

Because of diffraction, it is impossible to produce strictly parallel light beams. For example, every laser beam diverges by a certain minimum amount, called the *diffraction limit*. Maybe you know that the world's most expensive cat-eye is on the moon, where it has been deposited by the Apollo 11 astronauts. Can you determine how wide a laser beam with minimum divergence has become when it arrives at the moon, assuming that it was 1 m wide when sent to the moon? How wide would it come back if it had been 1 mm wide at the start? Ref. 22 Challenge

Diffraction implies that there are no perfectly sharp images: there exists a *limit on resolution*. This is true for the eye as well, where the resolution is between one and two minutes of arc, i.e. between 0.3 and 0.6 mrad. The limit is due to the limited size of the pupil. Therefore for example, there is a maximum distance at which one can distinguish the two car lights of a car. Can you estimate it? Challenge

For the same reason it is impossible to see the Great Wall in northern China from the moon, contrary to what is often claimed. In the few parts which are not yet in ruins, the wall is about 6 metres wide, and even if it casts a wide shadow during the evening, the angle it subtends is way below a second of arc, so that it is completely invisible to the human eye. In fact, three different astronauts who went to the moon performed careful searches and confirmed that the claim is absurd. (Ref. 23) The story is one of the most tenacious urban legends. (Is it possible to see the wall from the space shuttle?) (Challenge) In fact the largest man-made objects are the polders of reclaimed land in the Netherlands; they *are* visible from outer space. So are most large cities as well as the highways in Belgium at night; their bright illumination makes them stand out clearly from the dark side of the earth.

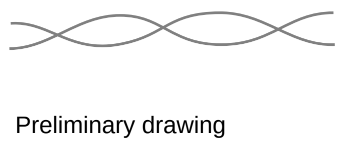


Figure 125 Light beams can spiral around each other

Diffraction also means that behind a small disk illuminated along its axis, the center of the shadow shows, against all expectations, a bright spot. This spot was predicted in 1819 by Denis Poisson in order to show to what absurd consequences the wave theory of light would lead. He had just read the mathematical description of diffraction developed by Augustin Fresnel on the basis of the wave description of light. But shortly afterwards, François Arago* actually observed Poisson's point, making Fresnel famous, and the wave properties of light started to be generally accepted.

Electromagnetic fields do not influence light directly, since light has no charge, and since Maxwell's equation are linear. But in some materials the equations are non-linear, and the story changes. For example, in certain photorefractive materials, two nearby light beams can even *twist* around each other, as shown by Segev and coworkers in 1997. (Ref. 24)

A final way to bend light is gravity, as discussed already in the chapters on universal gravity and on general relativity. Also the effect of gravity between two light beams was discussed there. See page 272

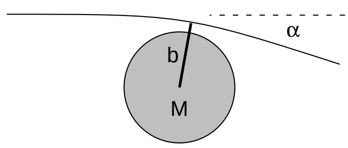


Figure 126 Masses bend light

In summary, light travels straight only if it travels *far from other matter*. In everyday life, 'far' simply means more than a few millimeters, because all gravitational and electromagnetic effects are negligible at these distances, mainly due to lights' truly supersonic motion.

Can one touch light?

If one takes a little glass bead and poses it on top of a powerful laser, the bead remains suspended in mid air, as shown in figure 127. That means that light has momentum. Therefore,

* François Arago (1786–1853), french physicist. Augustin Jean Fresnel (1788–1827), engineer and part time physicist; he pulished in 1818 his great paper on wave theory for which he got the price of the french academy of sciences in 1819. To improve his finances, he worked in the comission responsible for lighthouses, for which he developed the well-known Fresnel lens. He died prematurely, partly of exhaustion due to overwork.

contrary to what we said in the beginning of our escalation, images *can* be touched! In fact, the ease with which objects can be pushed has even a special name. For stars, it is called the *albedo*, and for general objects it is called the *reflectivity* r.

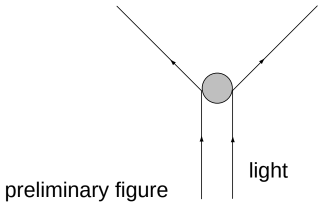


Figure 127 Levitating a small glass bead with a laser

Like each type of electromagnetic field, and like each kind of wave, light carries energy; the energy flow per surface and time is Challenge

$$\mathbf{P} = \frac{1}{\mu_o}\mathbf{E}\times\mathbf{B} \quad \text{giving an average} \quad <P>= \frac{1}{2\mu_o}E_{max}B_{max} \tag{358}$$

Therefore the pressure p exerted by light onto a body is given by Challenge

$$p = \frac{P}{c}(1+r) \tag{359}$$

where for black bodies $r = 0$ and for mirrors $r = 1$; other bodies have values in between. What is your guess for the amount of pressure due to sunlight on a black surface of one square metre? Is that the reason that we feel more pressure during the day than during the night? Challenge

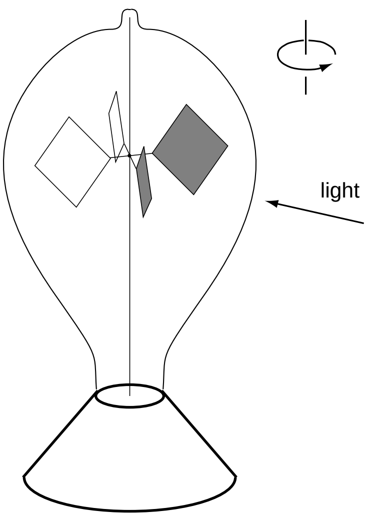


Figure 128 A commercial light mill turns *against* the light

In fact, one needs rather delicate equipment indeed to detect the momentum of light. In order to achieve this, in 1873, William Crookes * invented the light mill radiometer. He had the intention to demonstrate the radiation pressure of light. The light mill consists of four thin plates, black on one side and shiny on the other, which are mounted on a vertical axis, as shown in figure 128. However, when Crookes finished building it – it was similar to those sold in shops today – he found, like everybody else, that it turned in the wrong direction! (Why is it wrong?) You can check it by yourself by pointing a laser pointer onto it. The behaviour has been a puzzle for quite some time. Explaining it involves the tiny amount of gas left over in the glass bulb and takes us too far from the topic of our escalation. Only in 1901, with the advent of much better pumps, it was possible to create a sufficiently good vacuum that allowed to measure the light pressure with such an improved, true radiometer. Ref. 25 Challenge Ref. 26 Ref. 27

In fact, it turns out that the tail of a comet exists only because the light of the sun hits the small dust particles which detach from the comet. For the same reason, the tail always points away from the sun, a well-known fact that you might want to check at the next opportunity. Challenge

But light cannot only touch and be touched, it can also *grab*. In the 1980s, Arthur Ashkin and his research group developed actual *optical tweezers* which allow to grab, suspend, and move small transparent spheres of 1 to 20 μm diameter using laser beams. It is possible to do Ref. 28

∗ William Crookes (1832, London–1919, London) english chemist and physicist, president of the Royal Society, discoverer of Thallium.

this through a microscope, so that one can also observe what one is doing. This technique is now routinely used in biological research around the world, and has been used for example to measure the force of single muscle fibers, by chemically attaching their ends to glass or teflon spheres and then pulling them with such an optical tweezer.

But that is not all. In the last decade of the twentieth century, several groups even managed to *rotate* objects, thus realizing actual *optical spanners*. They are able to rotate particles at will in one direction or the other, by changing the optical properties of the laser beam used to trap the particle. Ref. 28

In fact, it does not take much to deduce that if light has linear momentum, circularly polarized light also has *angular* momentum. In fact, for such a wave the angular momentum L is given by

$$L = \frac{E_{\text{nergy}}}{\omega} \qquad (360)$$

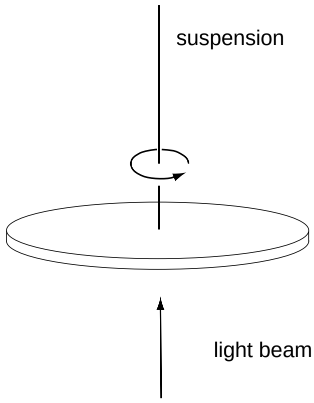


Figure 129 Light can rotate objects

Equivalently, the angular momentum of a wave is $\lambda/2\pi$ times its linear momentum. For light, this result has been confirmed already in the early 20th century: a light beam can put certain materials (which ones?) into rotation, as shown in the figure. Of course, the whole thing works even better with a laser beam. In the 1960s the experiment was performed with microwaves. A circularly polarized microwave beam from a maser can put a metal piece absorbing it into rotation. For a beam with cylindrical symmetry, depending on the sense of rotation, the angular momentum is either parallel or antiparallel to the direction of propagation. All these experiments confirm that light also carries angular momentum, an effect which will play an important role in the second part of our escalation. Challenge

In summary, light can touch and be touched. Obviously, if light can rotate, it can also *be* rotated. Could you imagine how this can be achieved? Challenge

War, light, and lies

From the tiny effects of the equation (359) for light pressure one deduces that light is not an efficient tool for hitting objects. However, light is able to heat up objects, as one can feel on the skin if it is touched by a laser beam of about 100 mW or more. For the same reason even cheap laser pointers are dangerous to the eye.

In the 1980s, a group of people who liked science fiction novels managed to persuade the military – who also indulge in this habit – that lasers could be used to shoot down missiles, and that a lot of tax money should be spent to develop such lasers. Using the definition of the Poynting vector and a hitting time of about 0.1 s, are you able to estimate the weight and size of the battery necessary for such a device to work? What would happen in cloudy or rainy weather? Challenge

Other people tried to persuade NASA to study the possibility to propel a rocket with light instead of ejected gas. Are you able to estimate whether this is feasible? Challenge

What is colour?

We saw that radio waves of certain frequencies are visible. Within that range, different frequencies correspond to different colours. (Are you able to convince a friend about this?) Challenge But the story is not finished here. Numerous colours can be produced either by a single wavelength, i.e. by *monochromatic* light, or by a *mixture* of several other, different colours. For example, standard yellow can be, if it is pure, a beam of 600 nm, or it can be a mixture of standard green of 546.1 nm and standard red of 700 nm. The eye cannot distinguish between the two cases. In everyday life, all colours are mixed, with the exception of those of yellow street lamps, of most laser beams and of the rainbow. You can check this yourself, using an umbrella or a compact disk. Challenge

In particular, white light is a mixture of a continuous range of colours with a given intensity per wavelength. To check that white light is a mixture of colours, simply hold figure 130 so near to your eye that you cannot focus the stripes any more. The unsharp border of the white stripes have a pink and green shade. These colors are due to the imperfections of the lens in the human eye, the so-called *chromatic aberrations*. Aberrations have the consequence that not all light frequencies follow the same path in the lens of the eye, and therefore that they hit the retina at different spots. This is the same effect that occurs in prisms or in water drops showing a rainbow. By the way, the shape of the rainbow tells something about the shape of the water droplets. Can you deduce the connection? Challenge

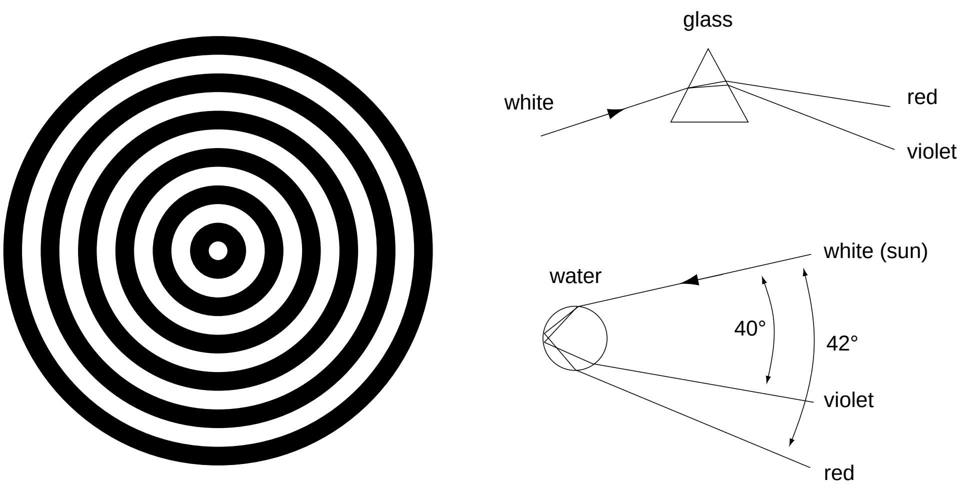


Figure 130 Proving that white light is a mixture of colours

Even pure air splits white light. This is the reason that the sky or far away mountains are blue and that the sun is red at sunset and at dawn. You can repeat this effect by looking through water at a black surface or at a lamp. Challenge Adding a few drops of milk to the water makes the light yellow and then red, and makes the black surface blue (like the sky seen from the earth as compared to the sky seen from the moon). More milk increases the effect. For the same reason, sunsets are especially red after volcanic eruptions.

By the way, at sunset the atmosphere itself acts as a prism as well; that means that the sun is split into different images, one for each colour, which are slightly shifted against each other, a bit like a giant rainbow in which not only the rim, but the whole disk is coloured. The total shift is about 1/60th of the diameter. If the weather is favourable, the air clear and quiet up to and beyond the horizon, for a few seconds it is possible to see, after the red, orange and yellow images have set, the rim of the green-blue image of the sun. That is the
Ref. 29 famous 'rayon vert' described by Jules Verne in his novel of the same title.*

To see the difference between colours in physics and colour in language, a famous discovery deserves to be mentioned: colours in language have a natural *order*. (Colours which point to objects, such as aubergine or sepia, or colours which are not generally applicable, such as blond, are excluded in this discussion.) Colours are ordered by all people in the following order: 1. black and white, 2. red, 3. green and yellow, 4. blue, 5. brown; 6. come mauve, pink, orange, gray and sometimes a twelfth term different from language to language. The result states that if a particular language has a word for any of these colours, it also has a word for all the preceding ones. It also implies that people use these basic colour classes even if their language does *not* have a word for each of them. These strong
Ref. 30 statements have been confirmed for over 100 languages.

What is the speed of light? – again

Physics is talking about motion. Talking is the exchange of sound; and sound is an example of a signal. A *(physical) signal* is the transport of information using transport of energy. There are no signals without motion of energy. This is also obvious from the fact that there is no way to store information without storing energy. To any signal one can thus ascribe a propagation speed. The highest possible signal speed is also the maximal velocity of general influences, or, using sloppy language, the maximal velocity with which effects spread causes.

If the signal is carried by matter, such as by the written text in a letter, the signal velocity is then the velocity of the material carrier, and experiments show that it is limited by the speed of light.

For a wave carrier, such as water waves, sound, light or radio waves, the situation is less evident. What is the speed of a wave? The first answer that comes to mind is the speed with which wave crests of a sine wave move. This already introduced *phase velocity* is given by the ratio between the frequency and the wavelength of a monochromatic wave, i.e. by

$$v_{ph} = \frac{\omega}{k} \quad . \tag{361}$$

For example, the phase velocity determines interference phenomena. Light in vacuum has the same phase velocity $v_{ph} = c$ for all frequencies. Are you able to imagine an experiment
Challenge to test this to high precision?

On the other hand, there are cases where the phase velocity is larger than c, most notably when light travels through an absorbing substance, and the frequency is near to an absorption maximum. In these cases, experiments show that the phase velocity is *not* the signal
Ref. 31 velocity. For such situations, a better approximation to the signal speed is the *group velocity*,

* About this and many other topics on colours in nature, such as e.g. the colour of shadows, the halos around the moon and the sun, and many others, see the beautiful book by Marcel Minnaert mentioned on page 54.

i.e. the velocity at which a group maximum will travel. This velocity is given by

$$v_{gr} = \left.\frac{d\omega}{dk}\right|_{k_0} \tag{362}$$

where k_0 is the central wavelength of the wave packet. One observes that $\omega = c(k)k = 2\pi v_{ph}/\lambda$ implies the relation

$$v_{gr} = \left.\frac{d\omega}{dk}\right|_{k_0} = v_{ph} + \lambda\frac{dv_{ph}}{d\lambda} \quad . \tag{363}$$

This means that the sign of the last term determines whether the group velocity is larger or smaller than the phase velocity. For a traveling group, as shown by the dotted line in figure 131, this means that new maxima either appear at the end or at the front of the group. Experiments show that for light *in vacuum*, the group velocity has the same value $v_{gr} = c$ for all values of the wave vector k.

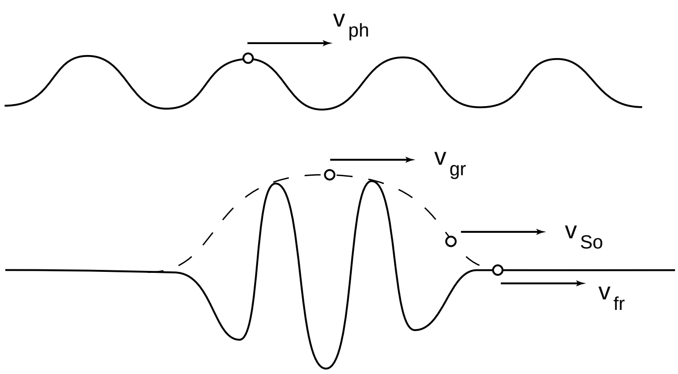


Figure 131 The definition of important velocities in wave phenomena

One should be warned however that still many publications propagate the myth that the group velocity *in materials* is never larger than c, the speed of light in vacuum. Actually, the group velocity in materials can be zero, or infinite, or even negative; this happens when the light pulse is very narrow, i.e. when it includes a wide range of frequencies, or again when one is near an absorption transition. In many (but not all) cases the group is found to widen substantially or even to split, making it difficult to define precisely the group maximum and thus its velocity. Many experiments have confirmed these predictions. For example, the group velocity in certain materials has been measured to be *ten times* that of light. However, in all these cases the group velocity is *not* the same as the signal speed.* Ref. 32

What then is the best velocity describing signal propagation? The german physicist Arnold Sommerfeld ** almost solved the main problem in the beginning of the twentieth century. He defined the signal velocity as the velocity v_{So} of the front slope of the pulse, as shown in figure 131. The definition cannot be summarized in a formula, but it does have Ref. 31

* In quantum mechanics, Schrödinger proved that the velocity of an electron is given by the group velocity of its wavefunction. Therefore the same discussion reappeared in quantum theory, as we will find out in the second part of the escalation.

** Arnold Sommerfeld (1868, Königsberg–1951, München) was a central figure in the spread of special and general relativity, of quantum theory, and of their applications. Professor in Munich, an excellent teacher and text book writer, he worked on atomic theory, on the theory of metals, on electrodynamics, and was the first to understand the importance and the mystery around "Sommerfeld's famous fine structure constant."

the property that it describes signal propagation for practically all experiments, in particular those in which the group and phase velocity are larger than the speed of light. When studying its properties, one finds that for no material Sommerfeld's signal velocity is larger than the speed of light in vacuum.

One might think that it is conceptually easier to describe signal propagation with help of the energy velocity. As mentioned before, every signal transports energy. The *energy velocity* v_{en} is defined as the ratio between the power flow density $\mathbf{P}$, i.e. the Poynting vector, and the energy density W, both taken in the direction of propagation. For electromagnetic fields – the only ones fast enough to be interesting for eventual superluminal signals – this ratio is

$$\mathbf{v}_{en} = \frac{\mathrm{Re}(\mathbf{P})}{W} = \frac{2c^2\,\mathbf{E}\times\mathbf{B}}{\mathbf{E}^2 + c^2\mathbf{B}^2} \quad . \tag{364}$$

However, like in the case of the front velocity, also in the case of the energy velocity one has to specify if one means the energy transported by the main pulse or that of the forerunner; and neither is ever larger than the speed of light in vacuum.* (In general, the energy velocity
Ref. 31 in matter has a value slightly different from Sommerfeld's signal velocity.)

In recent years, the progress in light detector technology has forced a final change in the choice of the best concept to describe signal velocity. Modern detectors allow to send signals with a slightly larger speed than that of the front slope of the pulse. Using detectors with the highest possible sensitivity one can use as signal the first point of the wave train whose amplitude is different from zero, i.e. the first tiny amount of energy arriving. This point's velocity, conceptually similar to Sommerfeld's signal velocity, is commonly called the *front velocity*, or, to distinguish it even more clearly from Sommerfeld's case, the *forerunner*
Challenge *velocity*. It is simply given by

$$v_{fr} = \lim_{\omega\to\infty} \frac{\omega}{k} \quad . \tag{365}$$

The forerunner velocity is never larger than the speed of light in vacuum; in fact it is precisely c, because for extremely high frequencies, the ratio ω/k is independent of the material, and vacuum properties take over. This is true signal velocity, or if one wants, the true *velocity of light*.

Which of all these velocities is measured in experiments determining the velocity of light,
Challenge e.g. when light is sent to the moon and reflected back? And now a more difficult question:
Challenge why is the signal speed of light slower inside matter, as all experiments show?

Signals and predictions

When somebody reads a text through the phone to a neighbour, who listens to it maybe repeats it, one speaks of communication. For any third person, the speed of communication is always smaller than the speed of light. But if the neighbour already knows the text, he can say it without waiting to hear the readers' voice. To the third observer such a situation looks

∗ Signals not only carry energy, they also carry negative entropy ("information"). The entropy of a transmitter increases during transmission. The receiver decreases in entropy (but less than the increase at the transmitter,
Ref. 33 of course).

like faster than light (superluminal) communication. Prediction can thus *mimic* communication, and in particular, it can mimic faster than light communication. Such a situation has been demonstrated most spectacularly in 1994 by Günter Nimtz, who seemingly transported music – all music is predictable for short time scales – through a "faster-than-light" system. To distinguish between the two situations, one notes that in the case of prediction, no energy transport takes place, in contrast to the case of communication. In other words, the definition of a signal as a transport of information is not as useful and clear-cut as the definition of a signal as *transport of energy*. In the mentioned experiment, no energy was transported faster than light. The same distinction between prediction on one hand and signal or energy propagation on the other hand will be used later on to clarify some famous experiments in quantum mechanics. Ref. 34

> If the rate at which physics papers are being published continues to increase,
> physics journals will soon be filling library shelves faster than the speed of light.
> This does not violate relativity since no useful information is being transmitted.

Why can we talk to each other? – Huygens' principle

The properties of our environment often appear in the full importance only when one asks simple questions. Why can we use the radio? Why can we talk on portable phones? Why can we listen to each other? It turns out that a central part of the answer is given by the fact that we live in a space of uneven dimensions.

In spaces of even dimensions, one cannot talk, because messages do not stop. This is an important result which is easily checked by throwing a stone into a lake: even after the stone has disappeared, waves are still emitted from the point at which it entered the water. On the contrary, when we stop talking, no waves are emitted any more.

– CS – text to be added – CS –

One can also say that Huygens' principle holds if the wave equation is solved by a circular wave leaving no amplitude behind it. Mathematicians translate this by saying that the delta function $\delta(c^2t^2 - r^2)$ satisfies the wave equation, i.e. that $\partial_t^2\delta = c^2\Delta\delta$. The delta function is that strange "function" which is zero everywhere except at the origin, where it is infinitely high. A few more properties, not mentioned here, fix the precise way this happens. If one generalizes this to higher dimensions, it turns out that the fundamental solution of the wave equation is zero everywhere only if the space dimension is odd and larger or equal to three.

In summary, when we switch off the light, a room gets dark only because we live in a space of uneven dimensions.

How does the world look when riding on a light beam?

This was the question the teenager Albert Einstein tried to answer.* The situation would have strange consequences.

- One would have no mirror image, like a vampire.

* He took the question from a book on the sciences by Aaron Bernstein which he read at that time.

- Light would not be a wave, but a static field.
- Nothing would move, like in the tale of sleeping beauty.

But also at speeds *near* the velocity of light observations would be interesting. One would

- see a lot of light coming towards one and almost no light from behind. The sky would be blue/white in front and red/black in the back;
- observe that everything around happens very very slowly;
- experience the smallest dust particle as deadly bullet.

Challenge Can you think of more strange consequences? It is rather reassuring that our planet moves rather slowly through its environment.

Does the aether exist?

Gamma rays, light, and radio waves are moving electromagnetic waves. All exist in empty space. What is oscillating when the light comes along? Maxwell himself called the 'medium' in which this happens the *aether*. The properties of the aether found in experiments are listed in table 30.

Physical property	experimental value
permeability	$\mu_o = 1.3\,\mu\mathrm{H/m}$
permittivity	$\varepsilon_o = 8.9\,\mathrm{pF/m}$
wave impedance/resistance	$Z_o = 376.7\,\Omega$
conformal invariance	applies
curvature	varying
spatial dimensionality	3
topology	R^3
mass and energy content	not detectable
friction on moving bodies	not detectable
own motion	not detectable

Table 30 Experimental properties of the aether and of flat vacuum

Ref. 35 The last item of the table is the most important: despite intensive efforts, nobody has been able to detect any *motion* of the aether. In other words, even though the aether oscillates, it does not move. Together with the other data, all these results can be summarized in one sentence: there is no way to distinguish the aether from the vacuum: both are one and the same.* Later in our escalation we will even find out that the ability of the vacuum to allow the propagation of light and its ability to allow the propagation of particles are equivalent:

* One sometimes hears that relativity or certain experiments show that the aether does not exist. This is incorrect. All the data only show that the aether is indistinguishable from the vacuum. Of course, if one uses the change of curvature as definition for motion of the vacuum, vacuum can move; but aether still remains indistinguishable from it.

Ref. 36 In fact, the term 'aether' has been used as an expression for several different ideas, depending on the author. First of all it was used for the idea that vacuum is not empty, but *full*; secondly, that this fullness can be described by *mechanical models*, such as gears, little spheres, vortices, etc.; thirdly, that vacuum is *similar to matter*, being made of the same substrate. Interestingly, some of these issues will reappear in the third part of our escalation.

both require the same properties. Therefore the aether remains *indistinguishable* from vacuum in the rest of our walk. In other words, the aether is a superfluous concept; we drop it Ref. 36 from our walk from now on. However, we are not finished with the study of the vacuum; it will keep us busy for a long time, starting with the intermezzo following this chapter. Moreover, quite a few of the aspects in table 30 will require some amendments later on.

Curiosities

Electromagnetism and light are almost endless topics. A few points to ponder are given here.

• How does one wire a light bulb, the mains, and three switches so that it can be switched on at any of the switches and off at any other switch? And in case of four switches? Nobody Challenge will take a physicist serious who can write Maxwell's equations but cannot solve this little problem.

• Can you make a mirror that does not exchange left and right? In two different ways? Challenge

• How would you measure the speed of the tip of a lightning? What range do you expect? Challenge

Figure to be added

Figure 132 The Farady disk

• One of the simplest possible electric motors is the so-called Faraday disk, a metal disk with a few more components around it, as shown in the figure. It starts turning when the current is switched on. Can you explain how it works? Challenge

• Cosmic radiation consists of charged particles hitting the earth. (We will discuss it in more detail later.) Astrophysicists explain that these particles are accelerated by the magnetic fields around the galaxy. However, the expression of Ref. 37 the Lorentz acceleration shows that magnetic fields can only change the direction of the velocity of charges, not its magnitude. How can one get acceleration nevertheless? Challenge

• The magnetic field of the earth, much higher than that of other planets because of the moon, with a dipole strength of $7.835 \cdot 10^{22}\,\mathrm{Am}^2$, shields us from lethal solar wind and cosmic Ref. 52 radiation particles. We owe it our life.

• If one calculates the Poynting vector for a charged up magnet – or simpler, a point charge near a magnet – one gets a surprise: the electromagnetic energy flows in circles around the magnet. Where does this angular momentum come from? Challenge

Worse, any atom is an example of such a system – actually of two such systems. Why is this effect not taken into account in calculations in quantum theory? Ref. 38

• Ohm's law, the observation that for almost all materials the current is proportional to the voltage, is due to a high school teacher. Georg Simon Ohm explored the question in great depth; at those times, such measurements were difficult to perform.* This has changed now. Recently, the electrical resistance of a single atoms has been measured: in the case of xenon it turned out to be about $10^5\,\Omega$. It was also found that lead atoms are ten times more Ref. 39 conductive than gold atoms. Can you imagine why? Challenge

* Georg Simon Ohm (1789, Erlangen–1854, München), bavarian school teacher and physicist. His efforts were recognized only late in his life, and he eventually was promoted to professor at the University in München. Later the unit of electrical resistance was named after him.

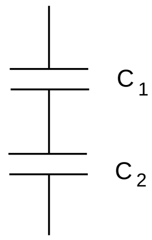


Figure 133 Capacitors in series

Ref. 40 • The charges on two capacitors in series are not generally equal, as naive theory states. For perfect, leak-free capacitors the voltage ratio is given by the inverse capacity ratio $V_1/V_2 = C_2/C_1$, due to the equality of the electric charges stored. However, in practice this is only correct for a few up to a few dozen minutes. Why? Challenge

• Does it make sense to write Maxwell's equations in vacuum? Both electrical and magnetic fields require charges in ordered to be measured. But in vacuum there are no charges! In fact, only quantum theory solves this apparent contradiction. Are yo able to imagine how? Challenge

Ref. 41 • Grass is usually greener on the other side of the fence. Can you give an explanation based on observations for this statement? Challenge

• The *electric polarizability* is the property of matter responsible for the deviation of water flowing from a faucet by a charged comb. (Try it!) See page 303 It is defined as the strength of electric dipole induced by an applied electric field. The definition simply translates the observation that many objects acquire charges when an electric field is applied. Incidentally, why combs get charged when used is still one of the mysteries of modern science.

Ref. 42 • Researchers are trying to detect tooth decay with help of electric currents, using the fact that healthy teeth are bad conductors, in contrast to teeth with decay. How can this effect be used? Challenge

• When solar plasma storms are seen on the sun, astronomers first of all phone the electricity company. They know that about 24 to 48 hours later, the charged particles ejected by the storms will arrive on earth, making the magnetic field on the surface fluctuate. Since power grids often have closed loops of several thousands of kilometers, additional electric currents are induced, which can make transformers in the grid overheat and then switch off. Then other transformers have to take over the additional power, which can lead to their overheating etc. The electricity companies avoid the problems by disconnecting the various grid sections, by avoiding large loops, by reducing the supply voltage to avoid saturation of the transformers, and by disallowing load transfer from failed circuits to others. Ref. 43

• Can you explain to a non-physicist how a microscope works?* Challenge

• Is it really possible to see stars from the bottom of a deep pit or of a well during daytime, as often stated in print? Challenge

What is a mirror?

We will study the strange properties of mirrors several times during our walk. We start with the simplest one first. Everybody can observe, by painting his hands in two different colours, that a mirror does *not* exchange right and left, as little as it exchanges up and down; however, a mirror does exchange right and left *handedness*. In fact, it does so by exchanging front and back.

But is it always possible to distinguish left from right? This seems easy: this text is rather different from a mirrored version, as are many other objects in our surroundings. But take a simple landscape. Are you able to say which of the two pictures of figure 134 is the original?

* If not, read the beautiful text by Elizabeth M. SLATER & Henry S. SLATER, *Light and electron microscopy*, Cambridge University Press, 1993.

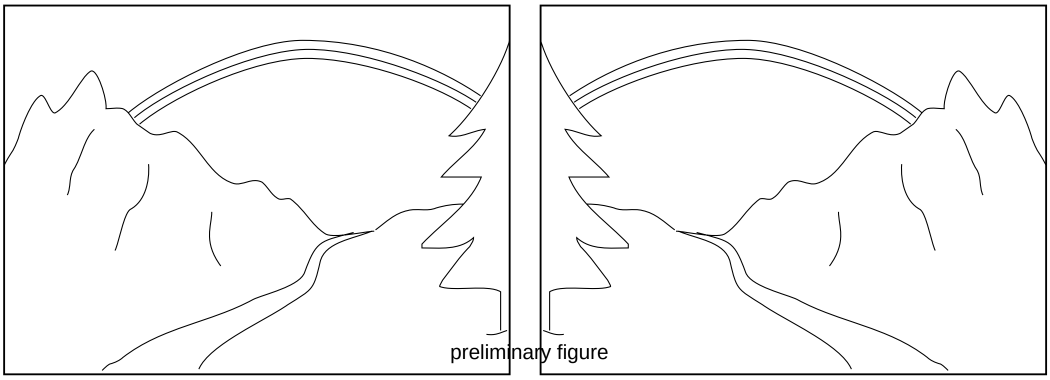


Figure 134 Which one is the original landscape?

Astonishingly, it is actually impossible to distinguish a picture of nature from its mirror image if it does not contain any human traces. In other words, everyday nature is somehow left-right symmetric. This observation is so common that all candidate exceptions, from the jaw movement of ruminating cows to the helical growth of plants, have been studied extensively.* Can you name a few more? See page 549 Challenge

The left-right symmetry of nature is a consequence of the fact that everyday nature is described by gravitation and, as we will see, by electromagnetism. Both interactions share an important property: by substituting all coordinates in their equations by the negative of their values the equations remain unchanged. This means that for any solution of these equations, i.e. for any naturally occurring system, the mirror image is also a possibility which can occur naturally. Everyday nature thus cannot distinguish between right and left. Indeed, there are right *and* left handers, people with their heart on the left *and* others with their heart on the right side, etc.

To explore further this strange aspect of nature, try the following experiment: imagine you are exchanging radio messages with a martian; are you able to explain him what right and left are, so that when you will meet, you are sure you are talking about the same thing? Challenge

Actually, the mirror symmetry of everyday nature – also called its *parity invariance* – is so pervasive that most animals cannot even distinguish left from right in a deeper sense. Ref. 45 Most animals react to mirror stimuli with mirror responses. It is hard to teach them different ways to react, and it is possible almost only for mammals. The many experiments performed on this topic gave the result that animals have symmetrical nervous systems, and possibly only humans show *lateralization*, i.e. a preferred hand and a different use for the left and the right part of the brain.

* The most famous is the position of the heart. The mechanisms leading to this disposition are still being investigated. Most recent research is suggesting that the oriented motion of the cilia on embryos, probably Ref. 44 in the region called *node*, determine the right left asymmetry. The deep origin of this asymmetry is not yet elucidated, however.

Another asymmetry of the human body is the hair whirl on the back of the head; the majority of humans having only one, and in 80% of the cases it is left turning.

Is lighting a discharge? – electricity in the atmosphere

Looking carefully, the atmosphere is full of electrical effects. The most impressive electrical phenomenon we observe, the lightning, is now reasonably well understood. Inside a thunderstorm cloud, especially inside tall *cumulonimbus* clouds,* charges are separated by collision between the falling large 'graupel' ice crystals falling due to their weight and Ref. 46 the small 'hail' ice crystallites rising due to thermal upwinds. Since the collision takes part in an electric field, charges are separated in a way similar to the mechanism in the Kelvin See page 305 generator. Discharge takes place when the electric field becomes to high, taking a strange path influenced by ions created in the air by cosmic rays. It seems that cosmic rays are at least partly responsible for the zigzag shape of lightnings.** By the way, you have a 75% survival chance after being hit by lightning; rapid reanimation is essential to help somebody to recover after a hit.

As a note, everybody knows how to measure the distance of a lightning by counting the seconds between the lightning and the thunder and multiplying by the speed of sound, ca. 330 m/s; it is less well known that one can estimate the *length* of the lightning bolt by measuring the duration of the thunder, and multiplying by it the same factor.

In the nineteen nineties, more electrical details about thunderstorms become known. Airline pilots sometime see weak and coloured light emissions spreading from the top of thunderclouds. There are various types of such emissions, blue *jets* and mostly red *sprites* and *elves*, which are somehow due to electric fields between the cloud top and the ionosphere. The details are still under investigation, and the mechanisms are not yet clear.***

All these details are part of the electrical circuit around the earth. This fascinating part of geophysics would lead us too far from the aim of our escalation. But every physicist should know that there is a vertical electric field of between 100 and 300 V/m on a clear day, as discovered already in 1752. (Can you guess why it is not noticeable in everyday Challenge life?) The field is directed from the ionosphere downwards to the ground; in fact the earth is permanently charged negatively, and on clear weather current flows downwards through the clear atmosphere, trying to *discharge* our planet. The current of about 1 kA is spread over the whole planet; it is possible due to the ions formed by cosmic radiation. (The resistance between the ground and the ionosphere is about 200 Ω, so that the total voltage drop is about 200 kV.) At the same time, the earth is constantly *charged* by several effects, of which the most important one turns out to be the lightning. In other words, contrary to what one may think, lightnings do not discharge the ground, they actually charge it up!**** Of

* From latin 'cumulus,' meaning heap, and 'nimbus', meaning big cloud. The various types of clouds all have latin names.

Ref. 47 ** There is no ball lightning even though there is a Physics Report about them. Ball lightnings are one of the favourite myths of modern pseudo-science. Actually, they would exist if we lived in a giant microwave oven. To show this, just stick a toothpick into a candle, light the toothpick, and put it into (somebody else's) microwave at maximum power.

*** For images, have a look at the interesting http://sprite.gi.alaska.edu/html/sprites.htm web site.

Challenge **** The earth is thus charged to about −1 MC. Can you confirm this? To learn more about atmospheric currents, you may want to have a look at the popularizing review of US work by Edgar Bering, Arthur Few & James Benbrook, *The global electric circuit*, Physics Today **51**, pp. 24–30, October 1998, or the more technical overview by Edgar Bering, Reviews of Geophysics (supplement) **33**, p. 845, 1995.

course, lightnings do discharge the cloud to ground potential difference, but by doing so, they actually send negative charge down to the earth.

The electric field is an important quantity. When helicopters save people on a raft in high sea, the rope must first be earthed by hanging it in the water, otherwise the people die from electrical shock when they first touch the rope, as happened a few times in the past. Can you explain why? Challenge

Why are sparks and lightnings blue? This turns out to be a material property; the colour is given by the material that happens to be excited by the energy of the discharge. This excitation is due to the temperature of 30 kK inside the channel of a typical lightning stroke. For sparks, the temperature can be much smaller. Depending on the situation, the colour may arise from the gas between the two electrodes, such as oxygen or nitrogen, or it may due to the material evaporated from the electrodes by the discharge. For an explanation of such colours, like for the explanation of all material related colours, we need to wait for the next part of our walk.

But not only electric fields, also electromagnetic fields can be dangerous. In 1997, with beautiful weather, a dutch hot air balloon approached the powerful radio transmitter in Hilversum. But after a few minutes near the antenna, the gondola suddenly detached from the balloon, killing all passengers inside.

An investigation team reconstructed the facts a few weeks later. In all modern balloons the gondola is suspended by high quality nylon ropes. To avoid damage by lightning and in order to avoid electrostatic charging problems, such as Kelvin generator effects from the burning flame, all these nylon ropes contain thin metal wires which form a large equipotential surface around the whole balloon. Unfortunately, in front of the radio transmitter these thin metal wires absorbed the radio energy from the transmitter, became red hot, and melted the nylon wires. It was the first time that this was ever observed.

How to prove you're holy

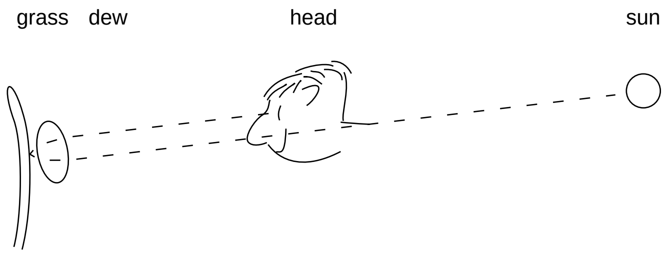


Figure 135 The path of light for dew on grass responsible for the aureole

Light reflection and refraction are responsible for many effects. The originally indian symbol of holiness, now used throughout most of the world, is the *aureole*, also called *halo* or *Heiligenschein*, a ring of light surrounding the head. You can easily see it around your own head. It is sufficient to get up early one morning and to look into the wet grass while turning your back to the sun. You will see an aureole around your shadow.

The effect is due to the morning dew on the grass, which reflects back the light mainly into the direction of the light source, as shown in the figure. The fun part is that if one does this in a group, one sees the aureole only around one's *own* head. Ref. 48

Retroreflective paint works in the same way; it contains tiny glass spheres which play the role of the dew. A large surface of retroreflective paint, a traffic sign for example, can also show one's halo, if the light source is sufficiently far away. Also the so-called 'glow' of the eyes of cats at night is due to the same effect; it is visible only if one looks at the cat with a light source in one's back. By the way, does a cat-eye work like a cat's eye?

Ref. 49

Challenge

Do we see what exists?

Sometimes we see *less* than there is. Close the left eye, look at the white spot in figure 136, approach the page slowly to your eye, and pay attention to the middle lines. At a distance of about 15 to 20 cm the middle line will seem uninterrupted. Why?

Challenge

Figure 136 Another limitation of the eye

Sometimes we see *more* than there is, as the next two figures show.

Our eyes also sees things *differently:* the retina sees an *inverted* image of the world. There is a simple method to show this, due to Helmholtz.* You only need a needle and a piece of paper, e.g. this page of text. Use the needle to make two holes inside the two letters "oo". Then keep the page as near to your eye as possible, look through the two holes towards the wall, keeping the needle vertical, a few centimeters behind the paper. You will see two images of the needle. If you now cover the *left* hole with your finger, the *right* needle will disappear, and vice versa. This shows that the image inside the eye, on the retina, is inverted. Are you able to complete the proof?

Challenge

We thus have to be careful when maintaining that seeing means observing. Examples such as these should make one ponder whether there could be other limitations of our senses which are less evident. And our walk will indeed uncover quite a few more.

* See Hermann von Helmholtz, *Handbuch der physiologischen Optik*, 1867. This famous classic is available in english as *Handbook of physiological optics*, Dover, 1962. The prussian physician, physicist, and science politician born as Hermann Helmholtz (1821, Potsdam–1894) was famous for his works on optics, on acoustics, electrodynamics, thermodynamics, epistemology, and geometry. He founded several physics institutions across Germany. He was one of the first to talk about the conservation of energy. His other important book, *Die Lehere von den Tonempfindungen*, published in 1863, describes the basis of acoustics, and like the handbook, is still worth to be read.

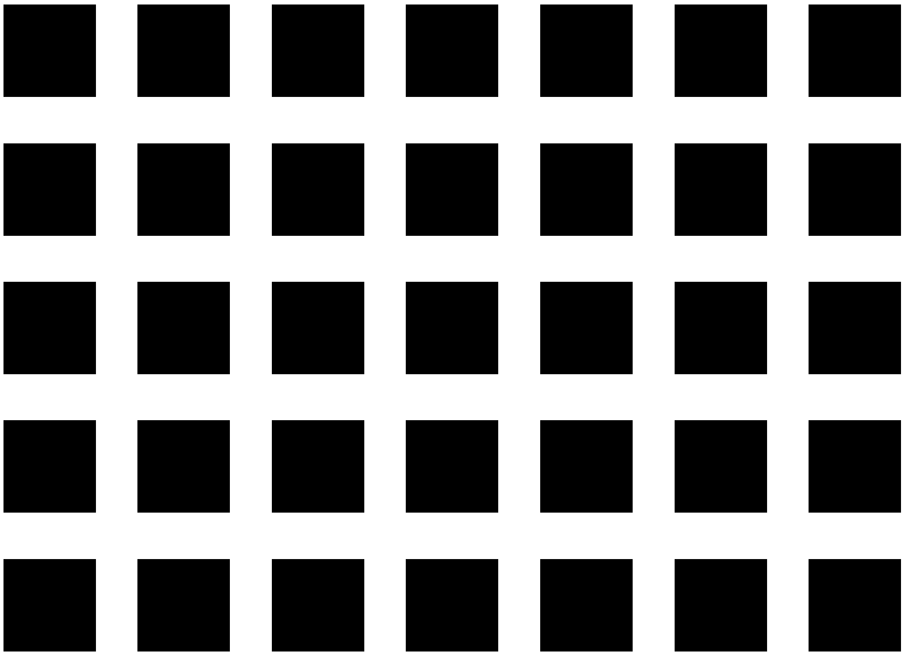


Figure 137 What is the shade of the crossings?

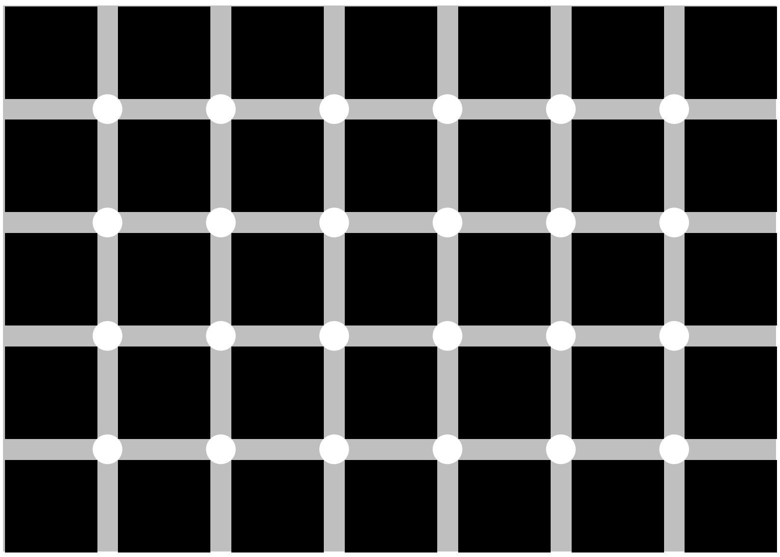


Figure 138 Do you see white or black dots?

How does one make pictures of the inside of the eye?

The most beautiful pictures so far of a *living* human retina, such as that of figure 140, were made by the group of David Williams at the University at Rochester in New York. They used adaptive optics, Ref. 50

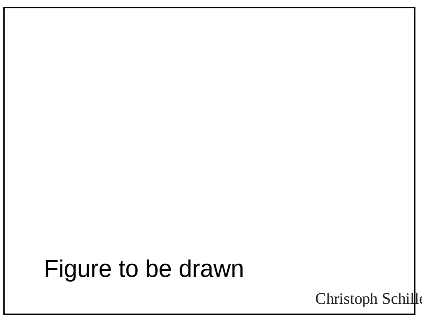


Figure 139 Eyes see inverted images

a technique which changes the shape of the imaging lens in order to compensate for the shape variations of the lens in the human eye.

The eyes see colour by averaging the intensity arriving at the red, blue and green sensitive cones. This explains the possibility, mentioned above, to get the same impression of colour, e.g. yellow, by a pure yellow laser beam, or by the mixture of red and green light. See page 332

But if the light is focussed onto one cone only, the eye makes mistakes. If, using this adaptive optics, a red laser beam is focussed such that it hits a green cone only, a strange thing happens: even though the light is *red*, the eye sees a *green* colour!

picture to be included

Figure 140 A high quality photograph of a live human retina

By the way, figure 140 is quite astounding. In the human eye, the blood vessels are located in front of the light sensitive cones. Why they do not appear in the picture? And why don't they disturb us in everyday life? Challenge

Amongst mammals, only primates can see *colours*. Bulls for example, don't; they cannot distinguish red from blue. On the other hand, the best colour seers overall are the birds. They have receptors for red, blue, green, UV, and depending on the bird, for up to three more sets of colours. Some birds have also much better eye resolution than humans have.

Does gravity make charges radiate?

We learned in the section on general relativity that gravitation has the same effects as acceleration. This means that a charge kept fixed at a certain height is equivalent to a charge accelerated by $9.8\,\mathrm{m/s^2}$, which would imply that it radiates, since all accelerated charges radiate. However, the world around us is full of charges at fixed heights, and there is no such radiation. How is this possible?

Ref. 51 The question has been a pet topic for many years. It turns out that the answer depends on whether the observer detecting the radiation is also in free fall or not, and on the time this started to be the case.

– CS – to be filled in – CS –

Research questions

The classical description of electrodynamics is coherent and complete; nevertheless there are still many subjects of research. Here are a few.

The origin of magnetic field of the earth, the other planets, the sun, and even of the galaxy is a fascinating topic. The way that the convection of fluids inside the planets generates magnetic fields, an intrinsically three dimensional problem, the influence of turbulence, of nonlinearities, of chaos etc. makes it a surprisingly complex question.

The details of the generation of the magnetic field of the earth, usually called the *geodynamo*, began to appear only in the second half of the twentieth century, when the knowledge of the earth's interior reached a sufficient level. The earth's interior is divided into the mantle – the first 2900 km from the surface – and the core. The core is made if a liquid outer core, 2300 km thick, and a solid inner core of 1215 km radius. It seems that the liquid and electrically conducting outer core acts as a dynamo which keeps the magnetic field going. The magnetic energy comes from the kinetic energy of the outer core, which rotates with respect to the earth's surface; the fluid can act as a dynamo because, apart from rotating, it also *convects* from deep inside the earth to more shallow depths, driven by the temperature gradients between the hot inner core and the cooler mantle. Huge electric currents flow in complex ways through these liquid layers, due to friction, and create the magnetic field. Understanding why this field switches orientation at irregular intervals of between a few tens of thousands and a few million years, is one of the central questions. The answers are difficult; experiments are not possible, 150 years of measurements is a short time when compared to the last transition, about 700 000 years ago, and computer simulations are extremely involved. Since the field measurements started, the dipole moment of the magnetic field has steadily diminished, and the quadrupole moment has steadily increased. Maybe we are heading towards a surprise. By the way, the study of *galactic* magnetic fields is even more complex, and still at its beginning. Ref. 52

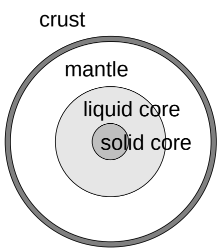


Figure 141 The structure of our planet

Another puzzle results from the equivalence of mass and energy. It is known from experiments that the size d of electrons is surely smaller than 10^{-22} m. That means that the electric field surrounding it has an energy content E given by at least Ref. 53 Challenge

$$E_{\text{nergy}} = \frac{1}{2}\varepsilon_o \int E^2_{\text{lectric field}}\, dV = \frac{1}{2}\varepsilon_o \int_d^\infty \Big(\frac{1}{4\pi\varepsilon_o}\frac{q}{r^2}\Big)^2 4\pi r^2 dr = \frac{q^2}{8\pi\varepsilon_o}\frac{1}{d} > 1.2\,\mu\text{J} \quad . \qquad (366)$$

On the other hand, the *mass* of an electron, usually given as $511\,\text{keV}/c^2$, corresponds to an energy of only 82 fJ, ten million times *less* than the value just calculated. In other words, classical electrodynamics has difficulties describing electrons. In fact, a consistent description of charged point particles within classical electrodynamics is still not completely achieved. This pretty topic receives only a rare interest nowadays – but then often passionate – because the puzzle is solved in the upcoming, second part of our escalation. See page 507

Even though the golden days of materials science are over, the various electromagnetic properties of matter and their applications in devices do not seem to be completely explored yet. About once a year a new effect is discovered which merits to be included in the list of electromagnetic matter properties of table 31. Among others, some newer semiconductor technologies will still have an impact on electronics, such as the recent introduction of low cost light detecting integrated circuits built in *CMOS* (complementary metal oxide silicon) technology. See page 349

Levitation

We have seen that it is possible to move certain objects without touching them, using a magnetic or an electric field, or of course, using gravity. One naturally asks if it is also possible, without touching an object, to keep it fixed, floating in mid air? Does this type of rest exist?

It turns out that there are several methods to levitate objects. They are commonly divided Ref. 54 into two groups: those which consume energy, and those who do not. Among the methods consuming energy one has the floating of objects on a jet of air or of water, the floating of objects through sound waves, e.g. on top of a siren, or through a laser beam coming from below, and the floating of conducting material, even of liquids, in strong radiofrequency fields. The levitation of liquids or solids by strong ultrasound waves is presently becoming Ref. 55 popular in laboratories. These methods give *stationary* levitation. Another group of energy consuming methods sense the way a body is falling and kick it up again in the right way via a feedback loop ; these methods are *non-stationary*, and usually use magnetic fields to keep the objects from falling. The magnetic train being built in Shangai by a german consortium Ref. 56 is levitated this way. The whole train, including the passengers, is levitated and then moved forward with electromagnets. It is thus possible, using magnets, to levitate many tens of tons of material.

For levitation methods which do *not* consume energy – all such methods are necessarily stationary – a well-known limitation can be found studying Coulomb's "law" of electrostatics: no static, i.e. time-independent arrangement of electric fields can levitate a *charged* object in free space or in air. The same result is valid for gravitational fields and *massive* objects;* in other words, one cannot produce a local minimum of potential energy in the middle of a box using electric or gravitational fields. This impossibility is called *Earn-* Ref. 57 *shaw's theorem*. Speaking mathematically, the solutions of the Laplace equation $\Delta\varphi = 0$, the so-called *harmonic functions*, have minima or maxima only at the border, and never inside the domain of definition. (You proved this yourself on page 78.) The theorem can also be proved by noting that given a potential minimum in free space, Gauss' "law" for a sphere around that minimum requires that a source of the field be present inside, which is in contradiction with the original assumption.

One can see easily that it is also impossible to use electric fields to levitate an electrically *neutral* body in air: the potential energy U of such a body, with volume V and dielectric constant ε, in an environment of dielectric constant ε_o, is given by

$$\frac{U}{V} = -\frac{1}{2}(\varepsilon - \varepsilon_o)E^2 \quad . \qquad (367)$$

Challenge Since the electric field E never has a maximum in the absence of space charge, and since for all materials $\varepsilon > \varepsilon_o$, there cannot be a minimum of potential energy in free space for a neutral body.**

In summary, using static electric or static gravitational fields it is impossible to keep an object from falling; neither quantum mechanics, which incorporates phenomena such as antimatter, nor general relativity, including phenomena such as black holes, change this basic result.

* To the disappointment of many science-fiction addicts, this would also be true in case that negative mass would exist, as happens for charge. See also page 58. And even though gravity is not really due to a field, the result still holds in general.

Ref. 58 ** It is possible, however, to 'levitate' gas bubbles in liquids – 'trap' them to prevent them from rising would be a better expression – because in such a case the dielectric constant of the environment is higher than that of Challenge the gas. Can you find a liquid-gas combination where bubbles fall instead of rising?

For static *magnetic* fields, the argument is analogous to electrical fields: the potential energy U of a magnetizable body of volume V and permeability μ in a medium with permeability μ_o containing no current is given by Challenge

$$\frac{U}{V} = -\frac{1}{2}\left(\frac{1}{\mu} - \frac{1}{\mu_o}\right)B^2 \tag{368}$$

and due to the inequality $\Delta B^2 \geqslant 0$, isolated maxima of a static magnetic field are not possible, only isolated minima. Therefore, it is impossible to levitate paramagnetic ($\mu > \mu_o$) or ferromagnetic ($\mu \gg \mu_o$) materials such as steel, including bar magnets, which are all attracted, and not repelled to magnetic field maxima. Challenge

There are thus two ways to get magnetic levitation: levitating a diamagnet or using a time dependent field. Diamagnetic materials ($\mu < \mu_o$) can be levitated by static magnetic fields because they are attracted to magnetic field minima; the best-known example is the levitation of superconductors, which are, at least those of type I, perfects diamagnets ($\mu = 0$). Strong forces can be generated, and this method is also being tested for the levitation of passenger trains in Japan. In some cases, superconductors can even be *suspended* in mid-air, below a magnet. Single atoms with a magnetic moment are also diamagnets; they are routinely levitated this way and have also been photographed in this state. Ref. 59 Ref. 56 Ref. 60

Also single neutrons, which have a magnetic dipole moment, have been kept in magnetic bottles in this way, until they decay. Recently, people have levitated pieces of wood, of plastic, strawberries, water droplets, grasshoppers, fish, and frogs (all alive an without any harm) in this way. They are, like humans, all made of diamagnetic material. Humans themselves have not yet been levitated, but the feat is being planned. Ref. 61

Diamagnets levitate if $\nabla B^2 > 2\mu_o \rho g/\chi$, where ρ is the mass density of the object and $\chi = 1 - \mu/\mu_o$ its magnetic susceptibility. Since χ is typically about 10^{-5} and ρ of order $1000\,\mathrm{kg/m^3}$, one requires field gradients of about $1000\,\mathrm{T^2/m}$, in other words, fields changes of 10 T over 10 cm, nowadays typical for high field laboratory magnets. Challenge

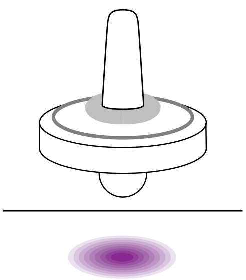


Figure 142 Floating magic nowadays available in toy shops

Finally, *time dependent* electrical or magnetic fields, e.g. periodic fields, can lead to levitation in many different ways without any consumption of energy. This is one of the methods which is used in the magnetic bearings of turbomolecular vacuum pumps. Single charged particles, such as ions and electrons, are now regularly levitated with Penning traps. Even free electrons can be levitated, letting them float above the surface of fluid helium. Figure 142 shows a toy allowing to let one personally levitate a spinning top in mid air, a quite impressive demonstration of levitation. Ref. 54 Ref. 54 Ref. 62

In the most recent twist of the science of levitation, in 1995 Stephen Haley predicted that the suspension height of small magnetic particles above a superconducting ring should be quantized. However, the prediction has not been checked by experiment yet. Ref. 63

For the sake of completeness we mention that the nuclear forces cannot be used for levitation in everyday life, as their range is limited to a few femtometres. However, we will see later that the surface matter of the sun is prevented from falling into the centre by these interactions; one could thus say that it is indeed levitated by nuclear interactions. See page 549

Matter, levitation and electricity

Levitation used by magicians mostly falls into another class. When David Copperfield, a magician performing for the MTV generation at the end of the twentieth century, "flies" during his performances, he does so by being suspended on thin fishing lines kept invisible by clever lighting arrangements. In fact, if one wants to be precise, one should count fishing lines as well as any table as levitation devices. Contrary to impression, a hanging or lying object is not really in contact with the suspension, if one looks at the critical points with a microscope. More about this in the next part of our walk. Ref. 64

But if this is the case, why don't we fall through a table or through the floor? We started the study of mechanics by stating as key property of matter its *solidity*, i.e. the impossibility to have more than one body at the same place at the same time. But what is the origin of solidity? Again, we will be able to answer the question only in the part on quantum mechanics, but we can collect the first clues already at this point. See page 494

Many experiments show that matter is constituted of charged particles; indeed, matter can be moved and influenced by electromagnetic fields in many ways. Over the years, material scientists have produced a long list of such effects, all of which are based on the existence of charged constituents.*

Table 31 Selected matter properties related to electromagnetism, showing among others the role it plays in the constitution of matter; at the same time a short overview of atomic, solid state, and fluid physics. Almost all effects have applications in technical products, and give work to many people.

Name of property	example	definition
thermal radiation or heat radiation or incandescence	every object	temperature dependent radiation emitted by any macroscopic amount of matter
Interactions with charges and currents		
electrification	separating metals from insulators	spontaneous charging
triboelectricity	glass rubbed on cat fur	charging through rubbing
barometer light	mercury slipping along glass	gas discharge due to triboelectricity Ref. 65
insulation	air	no current flow below critical voltage drop
semiconductivity	diamond, silicon or gallium arsenide	current flows only when material is impure ("doped")
conductivity	copper, metals	current flows easily
superconductivity	niobium	current flows indefinitely
ionisation	fire flames	current flows easily
localization (weak, Anderson)	disordered solids	
resistivity, Joule effect	graphite	heating due to current flow
thermoelectric effects: Peltier effect, Seebeck effect, Thomson effect	ZnSb, PbTe, PbSe, BiSeTe, etc.	cooling due to current flow, current flow due to temperature difference, or due to temperature gradients
acoustoelectric effect	CdS	sound generation by currents, and vice versa

* Detailed descriptions of many of these effects can be found in the excellent overview edited by Manfred von Ardenne, Gerhard Musiol & Siegfried Reball, *Effekte der Physik und ihre Anwendungen*, Harri Deutsch, 1997.

Name of property	example	definition
magnetoresistance	iron, metal multilayers	resistance changes with applied magnetic field Ref. 66
recombination	fire alarms	charge carriers combine to neutral atoms or molecules
annihilation	positron tomography	particle and antiparticle, e.g. electron and positron, disappear into photons
Penning effect	Ne, Ar	ionisation through collision with metastable atoms
Richardson effect, thermal emission	BaO_2, W, Mo, used in tv and electron microscopes	emission of electrons from hot metals
skin effect	Cu	high current density on exterior of wire
pinch effect	InSb, plasmas	high current density on interior of wire
Josephson effect	Nb-Oxide-Nb	tunnel current flows through insulator between two superconductors
Sasaki-Shibuya effect	n-Ge, n-Si	anisotropy of conductivity due to applied electric field
switchable magnetism	InAs:Mn	voltage switchable magnetization Ref. 67
Interactions with magnetic fields		
Hall effect	silicon, used in magnetic field measurement	voltage perpendicular to current flow in applied magnetic field
Zeeman effect	Cd	change of emission frequency with magnetic field
Paschen-Back effect	atomic gases	change of emission frequency in strong magnetic fields
ferromagnetism	Fe, Ni, Co, Gd	spontaneous magnetization; material strongly attracted by magnetic fields
paramagnetism	iron	induced magnetization parallel to applied field; attracted by magnetic fields
diamagnetism	water	induced magnetization opposite to applied field; repelled by magnetic fields
magnetostriction	CeB_6, $CePd_2Al_3$	change of shape or volume in magnetic field
magnetoelastic effect	Fe, Ni	change of magnetization by tension or pressure
spin valve effect	metal multilayers	electrical resistance depends on spin direction of electrons with respect to applied magnetic field
magnetooptical activity or Faraday effect or Faraday rotation	flint glass	polarization angle is rotated with magnetic field; different refraction index for right and left circularly polarized light, as in magnetooptic (MO) recording
magnetic circular dichroism	gases	different absorption for right and left circularly polarized light
photoelectromagnetic effect	Insb	current flow due to light irradiation of semiconductor in a magnetic field
Voigt effect	vapours	birefringence induced by applied magnetic field

Name of property	example	definition
Cotton-Mouton effect	liquids	birefringence induced by applied magnetic field
Hanle effect	Hg	change of polarization of fluorescence with magnetic field
Shubnikov-de Haas effect	Bi	periodic change of resistance with applied magnetic field
thermomagnetic effects: Ettinghausen effect, Righi-Leduc effect, Nernst effect, magneto-Seebeck effect	BiSb alloys	relation of temperature, applied fields, and electric current
Ettingshausen-Nernst effect	Bi	appearance of electric field in materials with temperature gradients in magnetic fields
photonic Hall effect	CeF_3	transverse light intensity depends on the applied magnetic field Ref. 68
magnetocaloric effect	gadolinium, GdSiGe alloys	material cools when magnetic field is switched off Ref. 69
cyclotron resonance	semiconductors, metals	selective absorption of radio waves in magnetic fields
magnetoacoustic effect	semiconductors, metals	selective absorption of sound waves in magnetic fields
magnetic resonance	most materials, used for imaging in medicine for structure determination of molecules	selective absorption of radio waves in magnetic fields
magnetorheologic effect	liquids, used in advanced car suspensions	change of viscosity with applied magnetic fields
Meissner effect	type 1 superconductors, used for levitation	expulsion of magetic field from superconductors
Interactions with electric fields		
polarizability	all matter	polarization changes with applied electric field
ionization, field emission, Schottky effect	all matter, tv	charges are extracted at high fields
paraelectricity	$BaTiO_3$	applied field leads to polarization in same direction
dielectricity	water	in opposite direction
ferroelectricity	$BaTiO_3$	spontaneous polarization below critical temperature
piezoelectricity	like the quartz lighter used in the kitchen	polarization appears with tension, stress, or pressure
pyroelectricity	$CsNO_3$, turmaline, crystals with polar axes; used for infrared detection	change of temperature produces charge separation

Name of property	example	definition
electroosmosis or electrokinetic effect	many ionic liquids	liquid moves under applied electric field Ref. 70
electrowetting	salt solutions on gold	wetting of surface depends on applied voltage
electrolytic activity	sulfuric acid	charge transport through liquid
liquid crystal effect	watch displays	molecules turn with applied electric field
electrooptical activity: Kerr effect, Pockels effect	liquids (e.g. oil), crystalline solids	material in electric field rotates light polarisation, i.e. produces birefringence
Freederichsz effect, Schadt-Helfrichs effect	nematics	electrically induced birefringence
Stark effect	hydrogen, mercury	colour change of emitted light in electric field
field ionisation	helium near tungsten tips in field ion microscope	ionisation of gas atoms in strong electric fields
Zener effect	Si	energy-free transfer of electrons into conduction band at high fields
field evaporation	W	evaporation under strong applied electric fields
Interactions with light		
absorption	coal, graphite	transformation of light into heat or other energy forms
blackness	coal, graphite	complete absorption in visible range
colour, metallic shine	ruby	absorption depending on light frequency
photostriction	PbLaZrTi	light induced piezoelectricity
photography	AgBr, AgI	light precipitates metallic silver
photoelectricity	Cs	current flows into vacuum due to light irradiation
internal photoelectric effect, photoeffect	Si p-n junctions, solar cells	voltage generation and current flow due to light irradiation
photon drag effect	p-Ge	current induced by photon momentum
emissivity	every body	ability to emit light
transparency	glass, quartz, diamond	100% less reflection less absorption
reflectivity	metals	light bounces on surface
polarization	pulled polymer sheets	light transmission depending on polarization
optical activity	sugar dissolved in water, quartz	rotation of polarization
birefringence	feldspat	refraction index depends on direction, light beams are split into two beams
dichroism	feldspat	absorption depends on polarisation
optically induced anisotropy, Weigert effect	AgCl	optically induced birefringence and dichroism
second harmonic generation	$LiNbO_3$, KPO_4	light partially transformed to double frequency
luminescence: general term for the opposite of incandescence	GaAs, tv	cold light emission

Name of property	example	definition
fluorescence	CaF_2, X ray production, light tubes, cathode ray tubes	light emission during light and after absorption or other energy input
phosphorescence	$TbCl_3$	light emission due to light, electrical or chemical energy input, continuing *long after* stimulation
electroluminescence	ZnS	emission of light due to alternating electrical field
also photo-, chemo-, tribo-, bio-, thermoluminescence		
thermoluminescence	quartz, feldspat	light emission during heating, used e.g. for archeological dating of pottery Ref. 71
Bremsstrahlung	X ray generation	radiation emission through fast deceleration of electrons
Compton effect	momentum measurements	change of wavelength of light, esp. X rays and gamma radiation, colliding with matter
Cerenkov effect	water, polymer particle detectors	light emission in a medium due to particles, e.g. emitted by radioactive processes, moving faster than the speed of light in that medium
transition radiation	any material	light emission due to fast particles moving from one medium to a second with different refractive index
electrochromicity	wolframates	colour change with applied electric field
Raman effect or Smekal-Raman effect	molecular gases	scattered light changes frequency
laser activity, superradiation	beer, ruby, He-Ne	emission of stimulated radiation
sonoluminescence	air in water	light emission during cavitation
gravitoluminescence	fake – does not exist	
switchable mirror	LaH	voltage controlled change from reflection to transparency Ref. 72
radiometer effect	bi-colored windmills	mill turn due to irradiation (page 330)
luminous pressure	idem	opposite to the previous one (page 330)
solar sail effect	future satellites	motion due to solar wind
acoustooptic effect	$LiNbO_3$	diffraction of light by sound in transparent materials
photorefractive materials	$LiNbO_3$, GaAs, InP	light irradiation changes refractive index
Auger effect	Auger electron spectroscopy	electron emission due to atomic reorganisation after ionisation by X rays
Bragg reflection	crystal structure determination	X ray diffraction by atomic planes
Mößbauer effect	Fe, used for spectroscopy	recoil-free resonant absorption of gamma radiation
pair creation	Pb	transformation of a photon in a charged particle-antiparticle pair
photoconductivity	See, CdS	change of resistivity with light irradiation
optoacoustic affect, photoacoustic effect	gases, solids	creation of sound due to light absorption
optogalvanic effect	plasmas	change of discharge current due to light irradiation

Name of property	example	definition
optical nonlinear effects: parametric amplification, frequency mixing, saturable absorption, n-the harmonic generation, optical Kerr effect, etc.		
phase conjugated mirror activity	gases	reflection of light with opposite phase
solidity	floors, columns, ropes, buckets	at most one object per place at a given time
Interactions with vacuum		
Casimir effect	metals	attraction of uncharged bodies

All matter properties in the list can be influenced by electric or magnetic fields or directly depend on them. This shows that all these material properties are electromagnetic in nature. In other words, charges and their interactions are an essential and fundamental part of the structure of objects. The table shows so many different electromagnetic properties that one can easily deduce that what is going on inside each material is very complex indeed. Most effects are the topic of solid state physics.*

Solid state physics is by far the most important part of physics, when measured by the impact it had on society. Can you find a product or business applications faor any randomly chosen effects from the table? Challenge

In our escalation we however, we look only at one example from the above list: thermal radiation, the emission of light by hot bodies.

Earnshaw's theorem about the impossibility of a stable equilibrium for charged particles at rest implies that the charges inside matter must be *moving*. For any charged particle in motion, Maxwell's equations for the electromagnetic field show that it radiates energy by emitting electromagnetic waves. In short, classical mechanics thus predicts that matter must radiate electromagnetic energy.

Interestingly, everybody knows from experience that this is indeed the case. Hot bodies light up depending on their temperature; the fact that light bulbs work thus proves that metals are made of charged particles. *Incandescence*, as it is called, requires charges. Actually, *every* body emits radiation, even at room temperature. This radiation is called *thermal radiation*; at room temperature it lies in the infrared. Its intensity is rather weak in everyday life; it is given by the general expression Ref. 73

$$I(T) = fT^4\,\frac{2\pi^5 k^4}{15c^2h^3} \quad \text{or} \quad I(T) = f\sigma T^4 \quad \text{with} \quad \sigma = 56.7\,\mathrm{nW/K^4m^2} \tag{369}$$

where f is a material, shape, and temperature dependent factor, with a value between zero and one, and called the *emissivity*. A body whose emissivity is given by the ideal case $f = 1$ is called a *black body*, because at room temperature such bodies also have ideal absorption coefficient and thus appear black. (Can you see why?) The heat radiation they emit is called *black body radiation*. Challenge

By the way, which object radiates more energy: a human body or a piece of the sun of the same mass? Try to guess first. Ref. 74 Challenge

* Probably the best and surely the most entertaining introductory english language book on the topic is the one by Neil Ashcroft & David Mermin, *Solid state physics*, Holt Rinehart & Winston, 1976.

Why can we see each other?

This use of the term 'black' is rather strange, since it turns out that most bodies at temperatures at which they are red hot or even hotter are good approximations of black bodies! For example, the tungsten in incandescent light bulbs, at around 2000 K, emits almost pure black body radiation; however, the glass then absorbs much of the ultraviolet and infrared components. Black bodies are also used to define the colour *white*. What we commonly call pure white is the colour emitted by a black body of 6500 K, namely the sun. This definition
Ref. 75 is used throughout the world, e.g. by the Commission Internationale d'Eclairage. Hotter black bodies are bluish, colder ones are yellow, orange or red.* The stars in the sky are classified in this way, as summarized on page 108.

Let us have a quick summary of black body radiation. Black body radiation has two important properties; first, the emitted power increases with the fourth power of the temperature. With this power relation alone you can check the just mentioned temperature of the sun simply by comparing the size of the sun with the width of your thumb when the
Challenge arm is stretched away from the face. Are you able to do this? (Hint: use the – excellent – approximation that the earth's temperature of about 300 K is due to the sun's irradiation.)**

The precise expression for the emitted energy density u per frequency ν can be deduced from the radiation law for black bodies discovered by Max Planck in 1899:

$$u(\nu,T) = \frac{8\pi h}{c^3}\frac{\nu^3}{e^{h\nu/kT}-1} \quad . \tag{370}$$

He made this important discovery, which we will discuss in more detail in the second part of our escalation, simply by comparing this curve with experiment.*** The new constant h, *Planck's quantum of action* or *Planck's constant*, turns out to have the value $6.6\cdot10^{-34}$ Js, and is central to all quantum theory, as we will see.

Challenge The radiation law gives for the total emitted energy density the expression

$$u(T) = T^4\,\frac{8\pi^5 k^4}{15c^3h^3} \tag{371}$$

Challenge from which equation (369) is deduced using $I = uc/4$. (Why?)

The second property of black body radiation is the value of the peak wavelength, i.e. the wavelength emitted with the highest intensity. This wavelength determines their colour; it
Challenge is deduced from equation (370) to be

$$\lambda_{max} = \frac{hc}{4.956\,k}\frac{1}{T} = 2.9\,\mathrm{mm\,K}/T \quad \text{but} \quad \hbar\nu_{max} = 2.82\,kT = 3.9\cdot10^{-23}\,\mathrm{J/K}\,T \quad . \tag{372}$$

* Most bodies are not black, because colour is not only determined by emission, but also by absorption of light.

Ref. 76 ** The actual average temperature of the earth is 14.0 °C.

*** Max Planck (1858–1947), professor of physics in Berlin, was a central figure in thermostatics. His introduction of the quantum hypothesis was the birth date of quantum theory. He also made the works of Einstein known in the physical community, and later organized a job for him in Berlin. He received the Nobel prize for physics in 1918. He was an important figure in the german scientific establishment; he also was one of the very few who had the courage to tell Hitler *face to face* that it was a bad idea to fire jewish professors. Famously modest, with many tragedies in his personal life, he was esteemed by everybody who knew him.

Either of these expressions is called *Wien's colour displacement* after its discoveror.* For 37 °C, human body temperature, it gives a peak wavelength of 9.3 μm, which is thus the colour of the bulk of the radiation emitted by every human being. (Note that the peak wavelength does not correspond to the peak frequency. Why?) On the other hand, following the telecommunication laws of many countries, any radiation emitter needs a licence to operate; as a consequence in Germany only dead people are legal, and only if their bodies are at absolute zero temperature. Challenge

Above, we predicted that any material made of charges emits radiation. Are you able to find a simple argument showing whether heat radiation is or is not this classically predicted radiation? Challenge

But let us come back to the question in the section title. The existence of thermal radiation implies that any hot body will cool, even if it is left in the most insulating medium there is, namely in vacuum. More precisely, if the vacuum is surrounded by a wall, a body in the vacuum will gradually approach the same temperature as the wall.

Interestingly, when the temperature of the wall and of the body inside have become the same, something strange happens. The effect is difficult to check at home, but impressive photographs exist in the literature. Ref. 77

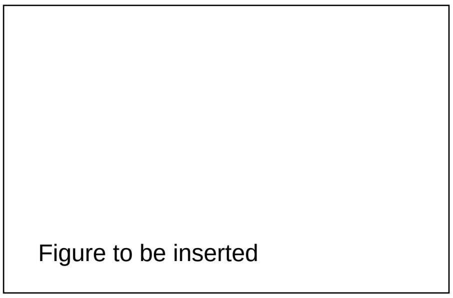


Figure 143 Hot bodies in a hot oven

An arrangement in which the walls and the objects inside are at the same temperature is called an *oven*. It turns out that one cannot *see* objects in an oven using the light coming from thermal radiation. For example, if an oven and all its contents are red hot, taking a picture of the inside of the oven (without a flash!), does not reveal anything; no contrast nor brightness changes exist which allow to distinguish the objects from the walls or their surroundings. Can you explain the finding? Challenge

In short, we are able to see each other only because the light sources we use are at a *different* temperature than ourselves. We can see each other only becuase we do *not* live in thermal equilibrium with our environment.

Could electrodynamics be different?

Any interaction like Coulomb's law which acts, for one given observer, between two particles independently of 3-velocity, must depend on 3-velocity for other inertial observers.** It turns out that such an interaction cannot be independent of the 4-velocity either. Such an interaction, even though it would indeed be 3-velocity dependent, would change the rest mass, since the 4-acceleration would not be 4-orthogonal to the 4-velocity.

The next simplest case is the one in which the acceleration is proportional to the 4-velocity. Together with the request that the interaction leaves the rest mass constant, one then recovers electrodynamics. Ref. 78

* Wilhelm Wien (1864, Gaffken–1824, München), east prussian physicist, received the Nobel prize for physics in 1911 for the discovery of this relation.

** This can be deduced from the special relativity in various ways, e.g. from the reasoning of page 314, or the formula in the footnote of page 174.

In fact, also the requirement of gauge symmetry and of relativity symmetry make it impossible to modify electrodynamics. In short, it does not seem possible to have a behaviour different from $1/r^2$ for a classical interaction.

A non-vansihing mass for the photon would change electrodynamics a bit. Experiments pose tight limits on the mass value, but the inclusion of a tiny mass poses no special problems, and the corresponding lagrangian has already been studied in the literature, just in case.

A summary of classical electrodynamics and its limits

In general, classical electrodynamics can be summarized in a few main ideas.

– the electromagnetic field is a physical observable, as shown e.g. by compass needles;

– its sources are the (moving) charges, as described by Maxwell's evolution equations, as shown e.g. by the properties of amber, lodestone, batteries, and remote controls;

– the electromagnetic field changes the motion of electrically charged objects via the Lorentz expression, as e.g. shown by electric motors;

– it behaves like a continuous quantity, a distribution of little arrows, and propagates as a wave, as shown e.g. by the radio;

– it can exist and move also in empty space, as shown e.g. by the stars.

However, there is quite some fun ahead; even though this description is correct in everyday life, during the rest of our escalation we will find that *each* of these ideas is in fact wrong. A simple example shows the trouble ahead.

At a temperature of zero Kelvin, when matter does not radiate thermally, one has the paradoxical situation that the charges inside matter cannot moving, since no emitted radiation is observed, but they cannot be at rest either, due to Earnshaw's theorem. In short, the fact that matter actually exists shows that classical electrodynamics is wrong.

See page 349

In fact, table 31 makes the same point even more strongly; classical electrodynamics can describe many of the effects, *but it cannot explain the origin of any of them*. Even though few of the effects will be studied in our walk – they are not essential for our adventure – the general concepts necessary for their description will be the topic of the second part of this escalation, that on quantum theory.

15. Classical physics in a nutshell – and the future of planet earth

The description of general relativity and classical electrodynamics concludes our walk across classical physics,* even though classical physics studies many other interesting questions, such as possible future disasters. Some are listed in table 32: but we do not pursue this strange topic in our escalation.

Table 32 Some examples of disastrous motion of possible future importance

Situation	time scale in years from now
• end of physics	ca. 50 (ca. year 2050)
• ozone shield reduction	ca. 100

∗ Others prefer to include in classical physics only *special* relativity; this is a matter of personal preference.

Situation	time scale in years from now
• ocean level increase due to greenhouse warming	ca. 100-1000
• several magnetic north and south poles, allowing solar storms to disturb radio and telecommunications, to interrupt electricity supplies, to increase animal mutations, and to disorient migrating animals such as wales, birds and tortoises	ca. 800
• our interstellar cloud detaches from the solar systems, changing the size of the heliosphere, and thus auroras and solar megnetic fields	ca. 3000
• subsequent reversal of earth's magnetic field, with increased cosmic radiation levels and thus skin cancers and miscarriages	unknown
• atmospheric oxygen depletion due to forest reduction or excess fuel consumption	> 1000
• upcoming ice age	ca. 50000
• gamma ray burst from our own galaxy, causing radiation damage to many living beings	between 0 and $5 \cdot 10^6$
• asteroid hitting the earth, generating tsunamis or darkening sunlight	between 0 and $50 \cdot 10^6$
• neighbouring star approaching, starting comet shower through destabilization of Oort cloud	$> 1 \cdot 10^6$
• instability of solar system	$> 100 \cdot 10^6$
• low atmospheric CO_2 content stops photosynthesis	$> 100 \cdot 10^6$
• ocean level increase due to earth rotation slowing/stop	$> 10^9$
• temperature rise/fall (depending on location) due to earth rotation stop	$> 10^9$
• sun becomes red giant, engulfs earth	$5.0 \cdot 10^9$
• sun stops burning, becomes white dwarf	$5.2 \cdot 10^9$
• earth core solidifies, removing magnetic field and thus earth's cosmic radiation shield	$10.0 \cdot 10^9$
• nearby nova (e.g. Betelgeuse) bathes earth in annihilation radiation	unknown
• nearby supernova (e.g. Eta Carinae) blasts over solar system	unknown
• galaxy center destabilizes rest of galaxy	unknown
• universe recollapses – if ever (see chapter 175)	$> 20 \cdot 10^9$
• matter decays into radiation – if ever (see appendix C)	$> 10^{33}$
• problems with naked singularities	unknown
• the vacuum becomes unstable	unknown

But before we leave classical physics behind us, we summarize what we have learned about motion so far, in order to be prepared for the next legs.

In every example of motion, we distinguish the moving and localized entity, the object, from the extended environment. For either of them we distinguish the fixed *intrinsic properties* from the *varying state*.

Looking for all the fixed *intrinsic* aspects of objects, we find that all sufficiently small objects or particles are described completely by their mass and their electric charge. There is no magnetic charge. Mass and electric charge are the only localized intrinsic properties of objects in classical physics. Extended objects are described by continuous mass and charge distributions. Both mass and electric charge are defined by the accelerations they produce around them. Both quantities can vary continuously, are conserved, and can be added. They are thus described by real numbers. Mass, in contrast to charge, is always positive. Mass and charge describe the interaction of particles with the environment, i.e. with fields, and thus indirectly with other particles.

Looking for all *varying* aspects of objects, namely their state, we find that we can describe it completely using only two basic aspects, the velocity and the position, at each instant of time. Position and velocity can vary continuously in strength and orientation; observing how these aspects are described by different observers, we find that they are completely characterized by three-dimensional vectors. The set of all possible states is called the phase space. The phase space is described by continuous manifolds. The state of large objects made of more than one constituent is given by the states of all its constituent particles. These particles make up all objects and somehow interact electromagnetically.

The states of the same particle described by different observers are related. The relations are called the "laws" of motion. For example, for different times they are called *evolution equations*, for different places and orientations they are called *transformation properties*, and for different gauges they are called *gauge transformations*.

Apart from the motion of massive objects, we observe motion of a massless entity: *radiation*. Radiation such as light, radio waves, and its related forms, are traveling electromagnetic waves, and are described by same equations that describe the interaction of charged or magnetic objects. The speed of massless entities is the maximum possible speed and is the same for all observers. The *state* of radiation is described by the electromagnetic field strength. The *intrinsic properties* of radiation are the field strength, its polarization, and its coupling to matter. The motion of radiation describes the motion of images.

The *environment* is described by space and time coordinates, and turns out to be able to move as well, in form of gravity waves. Space and time are described by entities which are continuous, extended, and which allow to define distances. Their intrinsic properties are the number of dimensions, its signature, and its topology. Their state is given by the metric, which describes the local warpedness. The warpedness can change, so that it is fair to say that empty space can move like a wave.

We learned that our environment is finite in age. We learned the main lines of its history, and the fact that on large scales, the matter in the universe moves away from the surrounding matter. Finally we discovered that we do not know yet the large scale topology of our environment, nor do we know what happens at its spatial and temporal limits.

The motion of *objects* is described by several simple relations. First of all, no two objects can be at the same point at the same time. Secondly, masses move the way space-time tells them, and space moves the way masses tell it. This relation describes the motion of the stars, of thrown stones, of the tides, etc. Thirdly, mass is needed to break the conformal symmetry, and to distinguish space from time.

We learned that electromagnetism is necessary to define length and time intervals, that light travels at the maximum possible velocity, that gravity is not an interaction, and that rest and free fall is the same concept. In summary, we learned that of the two naive types of object motion, namely motion due to interaction with space-time curvature and motion due to the electromagnetic field, only the latter is genuine.

We also saw that speeds in natur are bound from above by a universal constant c, and that length to mass ratios are bound from below by a universal constant $4G/c^2$.

Above all, we learned that motion, be it linear or rotational, is conserved. It is similar to a continuous substance: it is never destroyed, never created, and always only redistributed.

Due to conservation, all motion, that of objects, of images, and of empty space, is predictable and reversible. Due to conservation of motion, time and space can be defined. In summary, despite everyday experience, there are no surprises in nature.

Why is our escalation not finished yet?

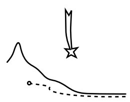
No problem is so formidable that
you can't walk away from it.
Charles Schultz

At the end of the 19th century, both Albert Michelson and Oliver Lodge – two mainly experimental physicists working on electrodynamics – claimed that electrodynamics and galilean physics meant that the major laws of physics were well known. Their statements are often quoted as examples of flawed predictions, especially since their very own experiments lead to the development of relativity, which they did not anticipate.

In addition, these victorian physicists overlooked another contradiction between electrodynamics and nature. In our walk so far we found that clocks and meter bars are necessarily based on matter and electromagnetism. But as we just saw, we do not understand the stability of matter yet. Matter is made of small particles, but the relation between these particles and electricity is not clear. This implies that we do not yet understand space and time, since both are defined with measurement devices made of matter. There is a challenge waiting – the second part of our escalation. The prize is to understand interactions; only the study of interactions allows to settle another question the 19th century overlooked: if motion is conserved in collisions, what exactly is exchanged between the colliding bodies? The fascinating path towards the answer is almost purely a sequence of them.

Subsequently, we need to rethink electromagnetism, as well as the other interactions we will discover, in the presence of space-time curvature. This challenge forms the third and final part of our escalation. There the adventure becomes truly mind boggling and almost incredible.

The reason is simple: both remaining parts of our escalation require a tool for the description of motion which we have not encountered yet: quantum theory. To be well prepared, we first take a break.



References

1 A.S. Goldhaber & W.P. Trower, *Resource letter MM-1: magnetic monopoles*, American Journal of Physics **58**, pp. 429–439, 1990. Cited on page 303.

2 R. Edwards, *Filling station fires spark cars' recall*, New Scientist, pp. 4–5, 4 March 1995. Cited on page 305.

3 S. Desmet, F. Orban & F. Grandjean, *On the Kelvin electrostatic generator*, European Journal of Physics **10**, pp. 118–122, 1989. One can find construction plans for it also in various places on the internet. Cited on page 305.

4 See the old but beautiful papers by Richard C. TOLMAN & T. DALE STEWART, *The electromotive force produced by the acceleration of metals*, Physical Review **8**, pp. 97–116, 1916, Richard C. TOLMAN & T. DALE STEWART, *The mass of the electric carrier in copper, solver and aluminium*, Physical Review **9**, pp. 164–167, 1917, and the later but much more precise experiment by C.F. KETTERING & G.G. SCOTT, *Inertia of the carrier of electricity in copper and aluminum*, Physical Review **66**, pp. 257–267, 1944. (Obviously the american language dropped the 'i' from aluminium during that period.) The first of these papers is also a review of the preceding attempts, and explains the experiment in detail. The last paper shows what had to taken into consideration to achieve sufficient precision. Cited on page 309.

5 This effect has first been measured by S.J. BARNETT, Philosophical Magazine **12**, p. 349, 1931. Cited on page 309.

6 See for example C. SCHILLER, A.A. KOOMANS, T.L. van ROOY, C. SCHÖNENBERGER & H.B. ELSWIJK, *Decapitation of tungsten field emitter tips during sputter sharpening*, Surface Science Letters **339**, pp. $L925$–$L930$, 1996. Cited on page 309.

7 L.I. SCHIFF & M.V. BARNHILL, *Gravitational-induced electric field near a metal*, Physical Review **151**, p. 1067, 1966. F.C. WITTEBORN & W.M. FAIRBANK, *Experimental comparison of the gravitational force on freely falling electrons and metallic electrons*, Physical Review Letters **19**, pp. 1049–1052, 1967. Cited on page 310.

8 See J.D. JACKSON, *Classical electrodynamics*, 3rd edition, Wiley, 1998, or also R.F. HARRINGTON, *Time harmonic electromagnetic fields*, McGraw-Hill, New York, 1961. Cited on page 314, 317.

9 The http://suhep.phy.syr.edu/courses/modules/MM/Biology/biology2.html web site gives an introduction into brain physiology. Cited on page 314.

10 N. SALINGAROS, *Invariants of the electromagnetic field and electromagnetic waves*, American Journal of Physics **53**, pp. 361–363, 1985. Cited on page 314.

11 A.C. de la TORRE, *$V \leqslant c$ in 1820?*, European Journal of Physics **20**, pp. $L23$–$L24$, March 1999. Cited on page 314.

12 Mark D. SEMON & John R. TAYLOR, *Thoughts on the magnetic vector potential*, American Journal of Physics **64**, pp. 1361–1369, 1996. Cited on page 317, 318.

13 T.T. WU & C.N. YANG, 1975, *Concept of nonintegrable phase factors and global formulation of gauge fields*, Physical Review D **12**, pp. 3845–3857, Cited on page 319.

14 Oleg D. JEFIMENKO, *A relativistic paradox seemingly violating conservation of momentum law in electromagnetic systems*, European Journal of Physics **20**, pp. 39–44, 1999. Of course, the missing momentum goes into the electromagnetic field. Given that the electromagnetic momentum is given by the vector potential, are you able to check whether everything comes out right? Cited on page 320.

Challenge

15 H. VAN DAM & E.P. WIGNER, ... Cited on page 320.

16 See the text by Raymond L. LEE & Alistair B. FRASER, *The rainbow bridge: rainbows in art, myth, and science*, Pennsylvania State University Press, 2000. A chapter can be found at the http://www.nadn.navy.mil/Oceanography/RainbowBridge/Chapter_8.html web site. Cited on page 324.

17 The beautiful slit experiment was published by E.A. MONTIE, E.C. COSMAN, G.W. 't HOOFT, M.B. van der MARK & C.W.J. BEENAKKER, *Observation of the optical analogue of quantized conductance of a point contact*, Nature **350**, pp. 594–595, 18 April 1991, and in the longer version E.A. MONTIE, E.C. COSMAN, G.W. 't HOOFT, M.B. van der MARK & C.W.J. BEENAKKER, *Observation of the optical analogue of the quantized conductance of a point contact*, Physica B **175**, pp. 149–152, 1991. The result was also publicized in numerous other scientific magazines. Cited on page 324.

18 A recent measurement of the frequency of light is presented in Th. UDEM, A. HUBER, B. GROSS, J. REICHERT, M. PREVEDELLI, M. WEITZ & T.W. HAUSCH, *Phase-coherent measurement of the hydrogen 1S-2S transition frequency with an optical frequency interval divider chain*, Physical Review Letters **79**, pp. 2646–2649, 6 October 1997. Another is C. SCHWOB & al., *Optical frequency measurement of the 2S-12D transitions in hydrogen and deuterium: Rydberg constant and Lamb shift determinations*, Physical Review Letters **82**, pp. 4960–4963, 21 June 1999. Cited on page 324.

19 See for example Gábor HORVÁTH, József GÁL & Rüdiger WEHNER, *Why are water-seeking insects not attracted by mirages? The polarization pattern of mirages*, Naturwissenschaften **83**, pp. 300–303, 1997. Cited on page 325.

20 The standard reference on the propagation of light is Max BORN & Emil WOLF, *Principles of Optics – Electromagnetic theory of propagation, interference and diffraction of light*, Pergamon Press, 6th edition, 1998. Cited on page 328.

21 An introduction into the topic of the 22° halo, the 46° halo, sun dogs, and the many other arcs and bows which can be seen around the sun, see the beautifully illustrated paper by R. GREENLER, *Lichterscheinungen, Eiskristalle und Himmelsarchäologie*, Physikalische Blätter **54**, pp. 133–139, 1998, or his book *Rainbows, halos, and glories*, Cambridge University Press 1980. Cited on page 328.

22 James E. FALLER & E. Joseph WAMPLER, *The lunar laser reflector*, Scientific American pp. 38–49, March 1970. Cited on page 328.

23 Neil Armstrong of Apollo 11, Jim Lovell of Apollo 8 and Apollo 13, and Jim Irwin of Apollo 15 extensively searched for it and then made negative statements, as told in Science News p. 423, 24 & 31 december 1994. From the space shuttle however, which circles only a few hundred kilometers above the earth, the wall *can* be seen when the sun is low enough such that the wall appears wider through its own shadow, as explained in Science News **149**, p. 301, 1996. Cited on page 329.

24 M. SHIH, M. SEGEV & G. SALAMO, Physical Review Letters **78**, pp. 2551–2554, 1997. See also the more readable paper by Mordechai SEGEV & George STEGEMAN, *Self-trapping of optical beams: spatial solitons*, Physics Today **51**, pp. 42–48, August 1998. Cited on page 329.

25 The value is 4 μPa for black surfaces, and the double for mirrors. Cited on page 330.

26 The first correct explanation of the light mill was given by Osborne REYNOLDS, *On certain dimensional properties of matter in the gaseous state*, Royal Society Philosophical Transactions Part 2, 1879. The best discussion is the one given on the web by Phil GIBBS, in the frequently asked question (FAQ) list of the `sci.physics` news group; it is available at the http://www.desy.de/user/projects/Physics/light-mill.html web site. Cited on page 330.

27 The first experimental proof of the pressure of light was given by Peter LEBEDEV, *Untersuchungen über die Druckkräfte des Lichtes*, Annalen der Physik **6**, pp. 307–458, 1901. He is also the first one who understood that this effect is the basis for the change of direction of the tails of comets when they circle around the sun. Cited on page 330.

28 A short overview is given by Miles PADGETT & Les ALLEN, *Optical tweezers and spanners*, Physics World pp. 35–38, September 1997. The original papers by Ashkin's group are A. ASHKIN, J.M. DZIEDZIC, J.E. BJORKHOLM & S. CHU, *Observation of a gradient force optical trap for dielectric particles*, Optics Letters **11**, p. 288, 1986, and A. ASKIN, J.M. DZIEDZIC & T. YAMANE, *Optical trapping and manipulation of single cells using infrared laser beams*, Nature **330**, p. 769, 1987. A pedagogical explanation on optical spanners, together with a way to build one, can be found in D.N. MOOTHOO, J. ARLT, R.S. CONROY, F. AKERBOOM, A. VOIT & K. DHOLAKIA, *Beth's experiment using optical tweezers*, American Journal of Physics **69**, pp. 271–276, 2001. Cited on page 330, 331.

29 There are also other ways to see the green ray, for longer times, namely when a fata morgana appears at sunset. An explanation with colour photograph is contained in M. VOLLMER, *Gespiegelt in besonderen Düften ...– Oasen, Seeungeheuer und weitere Spielereien der Fata Morgana*, Physikalische Blätter **54**, pp. 903–909, 1998. Cited on page 333.

30 This famous discovery is by Brent BERLIN & Paul KAY, *Basic color terms: their universality and evolution*, 1969. Their ongoing *world colour survey* is eagerly awaited. Of course there also are ongoing studies to find possible exceptions; the basic structure is solid, as shown in the conference proceedings C.L. HARDIN & Luisa MAFFI, *Color categories in thought and language*, Cambridge University Press, 1997. Cited on page 333.

31 For a thorough discussion of the various velocities connected to wave trains, see the classic text by Louis BRILLOUIN, *Wave propagation and group velocity*, Academic Press, New York, 1960. It expands in detail the theme discussed by Arnold SOMMERFELD, *Über die Fortpflanzung des Lichtes in dispergierenden Medien*, Annalen der Physik, 4th series, **44**, pp. 177–202, 1914. See also Arnold SOMMERFELD, *Optik*, Dietrichssche Verlagsbuchandlung, Wiesbaden 1950, section 22. An english translation Arnold SOMMERFELD, *Lectures on theoretical physics: optics*, 1954, is also available. Cited on page 333, 334, 335.

32 Such an experiment was carried out by S. CHU & S. WONG, *Linear pulse propagation in an absorbing medium*, Physical Review Letters **48**, pp. 738–741, 1982. See also S. CHU & D. STYER, *Answer to question #52. Group velocity and energy propagation*, American Journal of Physics **66**, pp. 659–661, 1998. Another example was described in 1993 by the group of Raymond Chiao for the case of certain nonlinear materials in Raymond CHIAO, Paul G. KWAIT & Aephraim M. STEINBERG, *Faster than light?*, Scientific American **269**, p. 52, August 1993, and R.Y. CHIAO, A.E. KOZHEKIN & G. KURIZKI, *Tachyonlike excitations in inverted two-level media*, Physical Review Letters **77**, pp. 1254–1257, 1996. On still another experimental setup using anomalous dispersion in cesium gas, see L.J. WANG, A. KUZMICH & A. DOGARIN, *Gain-assisted superluminal light propagation*, Nature **406**, pp. 277–279, 20 July 2000. Cited on page 334.

33 Y.P. TERLETSKII, *Paradoxes in the theory of relativity*, Plenum Press, New York, 1968, Cited on page 335.

34 G. NIMTZ, A. ENDERS & H. SPIEKER, Journal de Physique I (Paris) **4**, p. 565, 1994. Unfortunately, Nimtz himself seems to believe that he transported energy or signals faster than light; he is aided by the often badly prepared critics of his quite sophisticated experiments. See A. ENDERS & G. NIMTZ, Physikalische Blätter **49**, p. 1119, Dezember 1993, and the weak replies in Physikalische Blätter **50**, p. 313, April 1994. See also A.M. STEINBERG, Journal de Physique I (Paris) **4**, p. 1813, 1994, A. STEINBERG & al., Physical Review Letters **71**, pp. 708–711, 1993, and A. RANFAGNI & al., Physical Review E **48**, p. 1453, 1993. Cited on page 336.

35 A summary of all evidence about the motion of the aether is given by R.S. SHANKLAND & al., *New analysis of the interferometer observations of Dayton C. Miller*, Review of Modern Physics **27**, p. 167, 1955. An older text is H. WITTE, Annalen der Physik **26**, p. 235, 1908. Cited on page 337.

36 The history of the concept of vacuum can be found in the book by E. GRANT, *Much ado about nothing*, Cambridge University Press, 1981, and in the extensive text by Edmund T. WHITTAKER, *A history of the theories of aether and electricity*, Volume 1: *The classical theories*, Volume 2: *The modern theories*, Tomash Publishers, American Institute of Physics 1951, 1987. Cited on page 337, 338.

The various *aether models* – gears, tubes, vortices – proposed in the 19th century were dropped for various reasons. Since many models used to explain electric and magnetic fields as motion of some entities, it was concluded that the speed of light would depend on electric or magnetic fields. One type of field was usually described by linear motion of the entities, the other by rotatory or twisting motion; both assignments are possible. As a consequence, aether must be a somewhat strange fluid which flows perfectly, but which resists rotation of volume elements, as McCullogh deduced in 1839. However, experiments show that the speed of light in vacuum does not depend on electromagnetic field intensity. Vortices were dropped because real world vortices were found out to be unstable. All models received their final blow when they failed to meet the requirements of special relativity.

37 A.D. Erlykin & A.W. Wolfendale, *The origin of cosmic rays*, European Journal of Physics **20**, pp. 409–418, 1999, Cited on page 338.

38 D. Singleton, *Electromagnetic angular momentum and quantum mechanics*, American Journal of Physics **66**, pp. 697–701, 1998, Cited on page 338.

39 A. Yazdani, D.M. Eigler & N.D. Lang, *Off-resonance conduction through atomic wires*, Science **272**, pp. 1921–1924, 28 June 1996. For aluminium, gold, lead, niobium, as well as the influence of chemical properties, see Elke Scheer, *The signature of chemical valence in the electric conduction through a single-atom contact*, Nature **394**, pp. 154–157, 9 July 1998. Cited on page 338.

40 See L. Kowalski, *A myth about capacitors in series*, The Physics Teacher **26**, pp. 286–287, 1988, and A.P. French, *Are the textbook writers wrong about capacitors?*, The Physics Teacher **31**, pp. 156–159, 1993. Cited on page 339.

41 This problem was suggested by Vladimir Surdin. Cited on page 339.

42 Philip Cohen, *Open wide, this won't hurt a bit*, New Scientist p. 5, 3 February 1996. Cited on page 339.

43 The prussian explorer Alexander von Humboldt extensively checked this myth in the 19th century. He visited many mine pits, and asked countless mine workers in Mexico, Peru, and Siberia about their experiences. He also asked numerous chimney-sweeps. Neither him nor anybody else had ever seen the stars during the day. Cited on page 339.

44 J. Travis, *Twirl those organs into place – getting to the heart of how a heart knows left from right*, Science News **156**, 21 August, 1999. Cited on page 340.

45 See for example the discussion by M.C. Corballis & I.L. Beale, *On telling left from right*, Scientific American **224**, pp. 96–104, March 1971. Cited on page 340.

46 J. Latham, *The electrification of thunderstorms*, Quartely Journal of the Royal Meteorological Society **107**, pp. 277–289, 1981. For a more recent and wider review, see Earle R. Williams, *The tripole structure of thunderstorms*, Journal of Geophysical Research **94**, pp. 13151–13167, 1989. See also the book by the National Research Council Staff, *The earth's electrical environment*, Studies in Geophysics, National Academy Press, 1986. Cited on page 341.

47 B.M. Smirnov, *Physics of ball lightning*, Physics Reports **224**, pp. 151–236, 1993. See also D. Finkelstein & J. Rubinstein, *Ball lightning*, Physical Review **135**, pp. 390–396, 1964. For more folklore on the topic, just search the world wide web. Cited on page 341.

48 This happened to Bellini, the great renaissance painter, who even put this experience into writing, thus producing one of the greatest 'gaffes' ever. Cited on page 342.

49 S.R. Wilk, *How retroreflectors really work*, Optics & Photonics News, pp. 6–7, December 1993. Cited on page 343.

50 David R. Williams, *Supernormal Vision*, Science News **152**, pp. 312–313, 15 November 1997. See also http://www.cvs.rochester.edu/people/d_williams/d_williams.html as well as the photographs of the interior of living human eyes shown on http://www.cvs.rochester.edu/people/~aroorda/ao_research.html Cited on page 344.

51 See e.g. A. Kovetz & G.E. Tauber, American Journal of Physics **37**, p. 382, 1969. Cited on page 345.

52 On the geodynamo status, see the article by P.H. ROBERTS & G.A. GLATZMAIER, *Geodynamo theory and simulations*, Reviews of Modern Physics **72**, pp. 1081–1123, 2000. An older article is R. JEANLOZ & B. ROMANOWICZ, *Geophysical dynamics at the center of the earth*, Physics Today pp. 22–27, August 1997. Cited on page 338, 346.

53 This is deduced from the $g-2$ measurements, as explained in his Nobel-prize talk by Hans DEHMELT, *Experiments with an isolated subatomic particle at rest*, Reviews of Modern Physics **62**, pp. 525–530, 1990, and in Hans DEHMELT, *Is the electron a composite particle?*, Hyperfine Interactions **81**, pp. 1–3, 1993. Cited on page 346.

54 An excellent review is E.H. BRANDT, *Levitation in Physics*, Science **243**, pp. 349–355, 1989. Cited on page 347, 348, 348.

55 See the article by ... ... in Physik in unserer Zeit February 2001. Cited on page 347.

56 F.C. MOON & P.Z. CHANG, *Superconducting Levitation – Applications to bearings and magnetic transportation*, Wiley & Sons, 1994. Cited on page 347, 348.

57 W.T. SCOTT, *Who was Earnshaw?*, ... pp. 418–419, 1959. Cited on page 347.

58 The trick is to show that $\operatorname{div}\mathbf{E}=0$, $\operatorname{curl}\mathbf{E}=0$, thus $E\nabla^2E=0$, and from this, $\nabla^2E^2 \geqslant 0$; there are thus no local electric field maxima in the absence of free charges. The same proof works for the magnetic field. However, bodies with dielectric constants lower than their environment can be levitated in static electric fields, e.g. gas bubbles in liquids, as shown by T.B. JONES & G.W. BLISS, *Bubble dielectrophoresis*, Journal of Applied Physics **48**, pp. 1412–1417, 1977. Cited on page 347.

59 However, it is possible to levitate magnets if one uses a combination containing diamagnets. See A.K. GEIM, M.D. SIMON, M.I. BOAMFA & L.O. HEFLINGER, *Magnet levitation at your fingertips*, Nature **400**, pp. 323–324, 1999. Cited on page 348.

60 The first photographs of a single *ion* was in W. NEUHAUSER, M. HOHENSTATT, P.E. TOSCHEK & H. DEHMELT, Physical Review A **22**, p. 1137, 1980. See also D.J. WINELAND & W.M. ITANO, Physics Letters A **82**, p. 75, 1981, as well as F. DIETRICH & H. WALTER, Physical Review Letters **58**, p. 203, 1987.

For single *atoms*, see photographs in Z. HU & H.J. KIMBLE, Optics Letters **1**, p. 1888, 1994, F. RUSCHEWITZ, D. BETTERMANN, J.L. PENG & W. ERTMER, Europhysics Letters **34**, p. 651, D. HAUBRICH, H. SCHADWINKEL, F. STRAUCH, B. UEBERHOLZ, R. WYNANDS & D. MESCHEDE, Europhysics Letters **34**, p. 663, 1996. Cited on page 348.

61 See for example Mark BUCHANAN, *And God said...let there be levitating strawberries, flying frogs and humans that hover over Seattle*, New Scientist pp. 42–43, 26 July 1997, or C. WU, *Floating frogs*, Science News **152**, pp. 632–363, 6 December 1997, and *Molecular Magnetism takes off*, Physics World April 1997, page 28. The experiments by Andre Geim, Jan Kees Maan, Humberto Carmona and Peter Main was made public by P. RODGERS, Physics World **10**, p. 28, 1997. Some of the results can be found in M.V. BERRY & A.K. GEIM, *Of flying frogs and levitrons*, European Journal of Physics **18**, pp. 307–313, 1997. See also their http://www-hfml.sci.kun.nl/hfml/levitate.html web site. Cited on page 348.

62 The well-known toy allows levitation without the use of any energy source and is called the 'Levitron'. It was *not* invented by Bill Hones of Fascination Toys & Gifts in Seattle, as the http://www.levitron.com web site explains. The toy is discussed by Ron EDGE, *Levitation using only permanent magnets*, Physics Teacher **33**, p. 252, April 1995. It is also discussed in M.V. BERRY, *The Levitron^{TM}: an adiabatic trap for spins*, Proceedings of the Royal Society A **452**, pp. 1207–1220, 1996, (of Berry's phase fame) as well as by M.D. SIMON, L.O. HEFLINGER & S.L. RIDGEWAY, *Spin stabilized magnetic levitation*, American Journal of Physics **65**, pp. 286–92, 1997, and by T.B. JONES, M. WASHIZU & R. GANS, *Simple theory for the Levitron*, Journal of Applied Physics **82**, pp. 883–889, 1997. Cited on page 348.

63 The prediction about quantized levitation is by Stephen B. HALEY, *Length quantization in levitation of magnetic microparticles by a mesoscopic superconducting ring*, Physical Review Letters **74**, pp. 3261–3264, 1995. The topic is discussed in more detail in Stephen B. HALEY, *Magnetic levitation, suspension, and superconductivity: macroscopic and mesoscopic*, Physical Review B **53**, p. 3506, 1996, reversed in order with Stephen B. HALEY, *Quantized levitation of superconducting multiple-ring systems*, Physical Review B **53**, p. 3497, 1996, as well as Stephen B. HALEY, *Quantized levitation by multiply-connected superconductors*, LT-21 Proceedings, in Czechoslovak Journal of Physics **46**, p. 2331, 1996. In 1998, there was not yet an experimental confirmation (Stephen Haley, private communication). Cited on page 348.

64 All the illusions of the flying act look as if the magician is hanging on lines, as observed by many, including the author. (Photographic flashes are forbidden, a shimmery background is set up to render the observation of the lines difficult, no ring is ever actually pulled over the magician, the aquarium in which he floats is kept open to let the fishing lines pass through, always the same partner is 'randomly' chosen from the public, etc.) Information from eyewitnesses who have actually seen the fishing lines used by David Copperfield explains the reasons for these setups. The internet group `alt.magic.secrets` was central in clearing up this issue in all its details, including the name of the company which made the suspension mechanism. Cited on page 349.

65 R. BUDDAKIAN, K. WENINGER, R.A. HILLER & Seth J. PUTTERMAN, *Picosecond discharges and stick-slip friction at a moving meniscus of mercury in glass*, Nature **391**, pp. 266–268, 15 January 1998. See also Science News **153**, p. 53, 24 January 1998. Cited on page 349.

66 Henk SWAGTEN & Reinder COEHOORN, *Magnetische tunneljuncties*, nederlands tijdschrift voor natuurkunde **64**, pp. 279–283, november 1998. Cited on page 350.

67 H. OHNO & al., Nature 21-28 December 2000. Cited on page 350.

68 This effect was discovered by Geert RIKKEN, Bart VAN TIGGELEN & Anja SPARENBERG, *Lichtverstrooiing in een magneetveld*, nederlands tijdschrift voor natuurkunde **63**, pp. 67–70, maart 1998. Cited on page 351.

69 Vitalij PECHARSKY & Karl A. GSCHNEIDNER, *Giant magnetocaloric effect in Gd5(Si2Ge2)*, Physical Review Letters **78**, pp. 4494–4497, 1995, and, from the same authors, *Tunable magnetic regenerator alloys with a giant magnetocaloric effect for magnetic refrigeration from ∼20 to* ∼2990 K, Applied Physics Letters **70**, p. 3299, 1997. Cited on page 351.

70 A. AJDARI, *Electro-osmosis on inhomogeneously charged surfaces*, Physical Review Letters **75**, pp. 755–758, 1995. Cited on page 352.

71 M.J. AITKEN, *Thermoluminescence dating*, Academic Press, London, 1985. The precision of the method is far worse that C14 dating, however, as shown by H. HUPPERTZ, *Thermolumineszenzdatierung: eine methodologische Analyse aufgrund gesicherter Befunde*, Peter Lang Verlag, 2000. Cited on page 353.

72 This effect was discovered by J.N. HUIBERTS, R. GRIESSEN, J.H. RECTOR, R.J. WIJNGARDEN, J.P. DEKKER, D.G. DE GROOT & N.J. KOEMAN, *Yttrium and lanthanum hydride films with switchable optical properties*, Nature **380**, pp. 231–234, 1996. Cited on page 353.

73 See any book on thermostatics, such as Linda REICHL, *A modern course in statistical physics*, Wiley, 2nd edition, 1998. Cited on page 354.

74 The sun emits about $4 \cdot 10^{26}$ W from its mass of $2 \cdot 10^{30}$ kg, about 0.2 mW/kg; a person with an average mass of 75 kg emits about 100 W (you can check this in bed at night) i.e. about 500 times more. Cited on page 354.

75 See its http://www.cie.co.at/cie web site. Cited on page 355.

76 P.D. Jones, M. New, D.E. Parker, S. Martin & I.G. Rigor, *Surface air temperature and its changes over the past 150 years,* Reviews of Geophysics **37**, pp. 173–199, May 1999. Cited on page 355.

77 A picture of objects in a red hot oven and at room temperature is shown in C.H. Bennett, *Demons, engines, and the second law*, Scientific American **255**, pp. 108–117, November 1987. Cited on page 356.

78 Wolfgang Rindler, *Essential relativity – special, general, and cosmological*, revised second edition, Springer Verlag, 1977, page 247. There is also the beautiful paper by Levy-Leblond on galilean electrodynamics, which shows the problems one would get only when trying to define the theory non-relativistically. Cited on page 356.

Intermezzo: The Brain, its Language and its Humans

Physic ist wahrlich das eigentliche Studium des Menschen.*
Georg Christoph Lichtenberg (1742–1799)

Alles was überhaupt gedacht werden kann, kann klar gedacht werden.**
Ludwig Wittgenstein, *Tractatus*, 4.116

In our quest for increased precision in the description of all motion around us, it is time to take a break, sit down and look back. In our walk so far, which led through general relativity and electrodynamics, we used several concepts without defining them. Such undefined concepts were, for example, 'information', 'memory', 'measurement', 'set', 'number', 'infinity', 'existence', 'universe', or 'explanation'. They are common and important terms. In this intermezzo, we take a look at these concepts and try to give some simple, but sufficiently precise definitions, keeping them as provocative and entertaining as possible. For example: can you explain to your parents what a concept is? Challenge

Typeset in April 2001

The reason for doing this is simple. We need these clarifications in order to get to the top of motion mountain. Many have lost their way because of lack of clear concepts. Physics has a special role in this case. All sciences have one result in common: every type of change observed in nature is a form of motion. In this sense, but in this sense *only*, physics, focusing on motion itself, forms the basis for all the other sciences. In other words, the search for the famed "theory of everything" is an arrogant expression for the search for a *theory of motion*. Even though the knowledge of motion is basic, its precise description does not imply a description of "everything": just try to solve a marriage problem using the Schrödinger equation to experience the difference.

Anyway, given the basic importance of motion, it is necessary that in physics all statements on observations be as precise as possible. For this reason, many thinkers have investigated physical statements with particular care, using all criteria imaginable. The list of criteria appears once one asks: physics being detailed prattle by curious people about moving things, which abilities does this task require? You might want to fill in the list yourself. Challenge

All necessary abilities have been and still are investigated by researchers. The way the human species acquired the ability to chat about motion is studied by evolutionary biologists. Child psychologists study how the ability develops in a single human being. Physiologists, neurologists and computer scientists are concerned with the way the brain and the senses make this possible; linguists focus on the properties of language we use, while logicians, mathematicians and philosophers of science study the general properties of statements about nature. All of them investigate tools which are essential for the development of physics, for understanding motion, and for specifying the undefined concepts listed above. Their fields structure this intermezzo.

* Physics truly is the proper study of man.

** 'Everything that can be thought at all can be thought clearly.' This and other sentences in this chapter by Ludwig Wittgenstein are from the equally short and famous *Tractatus logico-philosophicus*, written in 1918, first published in 1921; it has now been translated in many other languages.

Evolution

A hen is only an egg's way of making another egg.
Samuel Butler, *Life and habit*, 1877.

The evolution of the human species is a long story, which has been told in many excellent books. A summarizing table on the history of the universe is given in the chapter on general relativity. It is worth remembering the incredible history which has lead to one's own existence, starting with the formation of atoms, of the galaxies, the stars, the planets, the moon, the atmosphere, the first cells, the water animals, the land animals, the mammals, the hominids, the humans, the ancestors, the family, and to oneself. Ref. 1 See page 250 Ref. 2

The way the particles we are made of moved during this sequence, being blown through space, being collected on earth, becoming organized to form people, is one of the most awe-inspiring examples of motion. Remembering this fantastic sequence of motion every now and then can be an enriching experience. Challenge

Biological evolution* in particular tells us a few important things. Without biological evolution, we wold not be able to talk about motion; only moving bodies can study moving bodies. And evolution invented childhood. In this intermezzo we will discover that most concepts of classical physics are introduced already by little children, in the experiences they make while growing up.

Children and physics

Physicists also have a shared reality. Other than that, there isn't really
a lot of difference between being a physicist and being a schizophrenic.
Richard Bandler

During childhood, everybody was a physicist. When one follows one's own memories backwards in time as far as one can, one reaches a certain stage, situated before birth, which forms the starting point of one's human experience. In that magic moment, one sensed somehow that apart from oneself, there is something else. The first observation one makes about the world is thus the recognition that one can distinguish two parts in it: oneself and the rest. This distinction is an example – perhaps the first – of a large number of "laws of nature" one stumbles upon in one's lifetime. By discovering more and more distinctions one brings structure in the chaos of one's experience. One quickly finds out that the world is made of related parts, such as mama, papa, milk, earth, toys, etc. Ref. 3

Later, when one learns to speak, one becomes fond of more difficult words, and one calls the surroundings the *environment*. Depending on the context, one calls the whole formed by oneself and the environment together the (physical) *world*, the (physical) *universe*, *nature*, or the *cosmos*. These concepts are not distinguished from each other in this walk;** they are

* An informative overview over the results of evolution, with the many-branched family tree that it produced, is given on the http://phylogeny.arizona.edu/tree web site. About the results of evolution for human beings, see the informative text by K. Kusch & S. Kusch, *Der Mensch in Zahlen*, Spektrum Akademischer Verlag, 2. Auflage, 2000.

** The differences in usage can be deduced from their linguistic origins. 'World' is derived from old germanic 'wer' – person – and 'ald' – old – and originally means 'lifetime'. 'Universe' is from latin, and designates the one – 'unum' – which one sees turning – 'vertere', and refers to the starred sky at night which turns around all taken to designate the sum of all parts and their relations; they are simply taken here to designate the *whole*.

From the moment of the first distinction onwards, one is ready to extract the numerous distinctions possible in the environment, the various parts of oneself, and the various types of interactions between all these. Distinguishing is the central ability which allows us to change our view from that of the world as *chaos*, i.e. as a big mess, to that of the world as a *system*, i.e. a structured set, in which parts are related in specific ways. (If you like precision, you may ponder whether the two choices of 'chaos' and 'system' are the only possible ones. We will return to this issue in the third part of our escalation.)

Challenge

In particular, the observation of difference between oneself and the environment goes hand in hand with the recognition that one is not independent of the environment, but that one is firmly tied to it in various inescapable ways; one can fall, get hurt, feel warm, cold, etc. Such relations are called *interactions*. Interactions express the observation that even though the parts of nature can be distinguished, they cannot be isolated. In other words, interactions describe the difference between the whole and the sum of its parts. No part can be defined without its relation to its environment.

See page 408

Interactions are not arbitrary; just take touch, smell, or sight as examples. They differ in reach, in strength, an in consequences. We call the characteristic aspects of interactions *patterns of nature*, or *properties of nature*, or *rules of nature*, or equivalently, with their historical but unfortunate name, *"laws" of nature*. The term "law" stresses their general validity but also implies design, aim, coercion and punishment for infringement; however, no design, no aim, nor any coercion is implied in the properties of nature, nor is infringement possible. The ambiguous term "law of nature" was made popular by René Descartes (1596–1650) and has been adopted enthusiastically because it in turn gave more weight to the laws of the state – which were far from perfect at that time – and to those of other organizations – which rarely are. The expression is an anthropomorphism coined by an authoritarian world view, suggesting that nature is "governed." We will therefore use the term as rarely as possible in our walk, and then, always between double, ironical, quotes. Nature cannot be forced in any way. The "laws" of nature are not obligations for nature or its parts, they are obligations only for physicists and all other people: the patterns of nature oblige us to use certain descriptions and to discard others. Whenever one says "laws govern nature" one is babbling nonsense; the correct expression is *rules describe nature*.

During childhood one learns to distinguish among interactions with the environment (or *perceptions*); some are shared with others, and called *observations*, others are uniquely personal ones, and called *sensations*.*

Often a slightly different criterion of 'sharedness' is used to divide the world into 'reality' and 'imagination' or 'dreams'. Our walk will show that this distinction is not essential, provided that one stays faithful to the quest for ever increasing precision: we will find that the description of motion we are looking after does not at all depend on whether the world is 'real' or 'imagined,' 'personal' or 'public.'

the polar star. 'Nature' is also from latin, and means 'what is born'. 'Cosmos' is from greek κόσμος and originally means 'order'.

* A child not able to make this distinction among perceptions – and thus unable to lie – almost surely develops or already suffers from *autism* as recent psychological research showed.

Ref. 4

Humans grow fond of their ability to distinguish parts, which in other contexts they also call *details*, *aspects*, or *entities*, and of their ability to associate them, i.e. to observe the *relations* between them. Human call this activity *classification*. Colours, shapes, objects, mother, places, people, ideas, are some of the entities one discovers first. Ref. 5

Our anatomy provides a handy tool to make efficient use of these relations: *memory*. A lot of input gets stored in it and is then called *experience*. Memory is a tool used by the young child to organize its world, and to achieve security in the chaos of life.

Jean Piaget was the first to describe the influence of the environment on the concepts a child forms. Step by step, children learn that objects are localized in space, that space has three dimensions, that objects fall, that collisions produce noise, etc. In particular, Piaget showed that space and time are not a priori concepts, but result from the interactions of any child with its environment.*

Around the time a child goes to school, it starts to understand the idea of *permanence of substances*, e.g. liquids, and the concept of *contrary*. Only at that stage its subjective experience becomes *objective*, with abstract comprehension. Later on, around puberty, the description of the world by children stops to be animistic: before, the sun, a brook, a cloud are *alive*. In short, only after puberty a human is ready for physics. Ref. 7

Even though everybody was a physicist in his youth, most people only remain *classical* physicists. In this adventure we go on, using all possibilities of a toy that nature provides us: the brain.

Experience is the name everyone gives to their mistakes.
Oscar Wilde (1854, Dublin–1900, Paris), *Lady Windermere's Fan*.

* An overview of the origin of developmental psychology is given by J.H. FLAVELL, *The developmental psychology of Jean Piaget*, 1963. This work summarizes the observations by the french speaking swiss Jean Piaget (1896–1980), the central figure of the field. He was one of the first researchers to look at child development in the same manner that a physicist looks at nature: carefully observing, taking notes, making experiments, extracting hypotheses, testing them, deducing theories. His astonishingly numerous publications, based on his extensive observations, cover almost all stages of child development. His central contribution is the detailed description of the stages of development of the cognitive abilities of humans, the formation of basic concepts, from his way of thinking, his ability to talk, etc., result from the continuous interaction between the child and the environment.

In particular, Piaget described the way children first learn that they are different from the external environment, and how they then learn about the physical properties of the world. Of his many books related to physical concepts, two especially related to the topic of this walk are J. PIAGET, *Les notions de mouvement et de vitesse chez l'enfant*, Presses Universitaires de France, 1972 and *Le developpement de la notion de temps chez l'enfant*, Presses Universitaires de France, 1981, this last book being born from a suggestion by Albert Einstein. These texts should be part of the reading of every physicist and science philosopher interested in these questions.

Piaget also describes how in children the mathematical and verbal intelligence derives from sensomotorial, practical intelligence, which itself stems from the habits and acquired associations to construct new concepts. Practical intelligence requires the system of reflexes provided by the anatomical and morphological structure of our organism. Thus his work shows in detail that our faculty for mathematical description of the world is based, albeit indirectly, on the physical interaction of one's organism with the world.

Some of his opinions on the importance of language in the development are now being revised, notably Ref. 6 through the rediscovery of the work of Lev Vigotsky, who argues that all higher mental abilities, emotions, recollective memory, rational thought, voluntary attention and self-awareness, are not innate, but learned. This learning takes place through language and culture, and in particular through the process of talking to oneself.

Why a brain?

Denken is bereits Plastik.*
Ref. 8 Joseph Beuys (1920–1986), sculptor.

Numerous observations show that sense input is processed, i.e. classified, stored, and retrieved in the brain. Notably, lesions of the brain can lead to loss of part or all of these functions. Ref. 9 Among the important consequences of these basic abilities of the brain are thought and language. All such abilities result from the construction, from the 'hardware' Ref. 10 of the brain.

Systems with the ability to deduce classifications from the input they receive are called *classifiers*, and are said to be able to *learn*. Our brain shares this property with many complex systems; the brain of many animals, but also certain computer algorithms, e.g. the so-called 'neural networks', are examples of such classifiers. Such systems are studied in several fields, from biology to neurology, mathematics and computer science.** Classifiers have the double ability to discriminate and to associate; both are fundamental to thinking.

Machine classifiers have a lot in common with the brain. As an example, following an Ref. 11 important recent hypothesis in evolutionary biology, the necessity of cooling the brain in an effective way is responsible for the upright, bipedal walk of humans. The brain needs a powerful cooling system to work well. In this it resembles modern computers, which usually have powerful fans or even water cooling systems built into them. It turns out that the human species has the most powerful cooling system of all mammals. The upright posture allowed the air to cool the body more effectively in the tropical environment where humans evolved. To allow even better cooling, humans also lack most of their body hair, except on their head, where it protects the brain from direct heating by the sun.

All classifiers are built from smallest classifying entities, sometimes large numbers of them. Usually, the smallest units can classify input into only two different groups. The larger the number of these entities, often called 'neurons' by analogy to the brain, the more sophisticated classifications can be produced by the classifier.*** Classifiers thus work by applying more or less sophisticated combinations of 'same' and 'different'. The distinction by a child of red and blue objects is such a classification; the distinction of compact and non-compact gauge symmetry groups in quantum theory is a more elaborate classification, but relies on the same fundamental ability.

In all classifiers, the smallest classifying units interact with each other. Often these interactions are channeled via connections, and the set is then called a *network*. In these connections, signals are exchanged, via moving objects, such as electrons or photons. Thus one arrives at the conclusion that the ability of the brain to classify the physical world, for example to distinguish moving objects interacting with each other, is a consequence of the fact that it itself consists of moving objects interacting with each other. Without a powerful classifier, humans wold not have become such a successful animal species. And only the motion inside our brain allows us to talk about motion in general.

* Thinking is already sculpture.
** A good introduction into the study of classifiers is ...
*** A good introduction into neural nets is J. Hertz, A. Krogh & R. Palmer, *Introduction to the theory of neural computation*, Addison Wesley, Redwood City, USA, 1991.

Numerous researchers are identifying the parts of the brain used when different intellectual tasks are performed. The experiments become possible with the technique of magnetic resonance imaging and other methods. Other researchers study how thought processes can be modeled from the brain structure. Neurology is still making regular process. In particular, it is steadily destroying the belief that thinking is *more* than a physical process. This belief results from personal fears, as you might want to test by introspection. It will disappear as time goes by.

See page 538

Challenge

What is information?

These thoughts did not come in any verbal formulation.
I rarely think in words at all.
A thought comes, and I may try to express it in words afterward.
Albert Einstein (1879–1955).

We started by saying that studying physics means to talk about motion. To talk means to transmit information. Can information be measured? Can one measure the progress of physics in this way? Is the universe made of information?

Information is the result of classification. A classification is the answer to one or to several yes-no questions. Such yes-no questions are the simplest classifications possible; they provide the basic *units* of classification, from which all others can be built. The simplest way to measure information is therefore to count the implied yes-no questions, the *bits*, leading to it. Are you able to say how many bits are necessary to define the place where you live? Obviously, the number of bits depends on the set of questions one starts with; that could be the names of all streets in a city, the set of all coordinates on the surface of the earth, the names of all galaxies in the universe, the set of all letter combinations in the address.* A variation of the latter method is used in computers. For example, the story of this walk required about fifty million bits. But since the amount of information in a normal letter depends on the set of questions one starts with, it is not possible to define a precise measure for information in this way.

Challenge

The only way to measure information precisely is to take the largest possible set of questions one can ask about a system. In that case, the amount of unknown information is called entropy, a concept we encountered already. Now you are able to deduce yourself whether it is really possible to measure the advance of physics.

See page 134

Challenge

Since categorization is an activity of the brain and other, similar classifiers, information as defined here is a concept that applies to the result of activities by people and by other classifiers. In short, information is produced when talking about the universe – the universe itself is *not the same* as information. There is an increasing number of publications based on the opposite of this view, a more and more frequent conceptual short circuit. Any transmission of information implies an interaction; physically speaking, this means that any information needs *energy* for transmission and *matter* for storage. Without either of them, there is no information. In other words, the universe, with its matter and energy, has to exist *before* transmission of information is possible. Saying that the universe is made of information is as sensible as saying that it is made of toothpaste.

∗ The number of required yes-no questions is rather different for the different cases. What is the most efficient one you can think of?

Challenge

The aim of physics is to give a *complete* classification of all types and examples of motion, in other words, to know everything about motion. Is this possible? Are you able to find
Challenge an argument against this endeavor?

What is memory?

The brain is my second favorite organ.
Woody Allen

Memory is the collection of records of perceptions. The production of such records is the essential aspect of observation. The storage of records can take place in human memory, i.e. in the brain, or in machine memory, as in computers, or in object memory, such as notes on paper. Without memory, there is no science, no life – since life is based on the records
Ref. 9 inside the DNA – and especially, no fun, as proven by the sad life of those who loose it.

Obviously every record is an object. But when does an object qualify as a record? A signature can be the record of the agreement on a commercial transaction. A single small dot of ink is not a record, because it could have appeared by mistake, for example by handling a pen. In contrast, it is improbable that a quantity of ink falls on paper exactly in the shape of a signature – except of course for the signatures of physicians. Simply speaking, a *record* is any object, which, in order to be copied, has to be forged. More precisely, a record is an object or a situation which cannot arise nor disappear by mistake or by chance. Our personal memories, be they images or voices, have the same property; we usually can trust them, because they are so detailed that they cannot have arisen by chance or by uncontrolled processes in our brain.

Can one estimate the probability for a record to appear or disappear by chance? Yes, one can. Every record is made of a characteristic number N of small entities, for example the number of the possible ink dots on paper, the number of iron crystals in a cassette tape, the electrons in a bit of computer memory, the silver iodide grains in a photographic negative, etc. The chance disturbances in any memory are due to internal fluctuations, also called *noise*. Noise makes the record unreadable; it can be dirt on a signature, thermal magnetization changes in iron crystals, electromagnetic noise inside a solid state memory, etc. Noise is found in all classifiers, since it is inherent in all interactions and thus in all information processing.

It is a general property that internal fluctuations due to noise decrease as the size, i.e. the number of components of the record is increased. In fact, the probability p_{mis} for a misread-
Challenge ing or miswriting of a record changes as

$$p_{\text{mis}} \sim 1/N \quad . \tag{373}$$

This relation is a consequence of the fact that for large numbers, the normal distribution is a good approximation of almost any process, and that the width of the normal distribution, which determines the probability of record errors, grows less rapidly than its integral when
Challenge the number of entities is increased. (Are you able to confirm this?)

We conclude that any good record must be made from a *large* number of entities. The larger the number is, the less sensitive the memory is to fluctuations. Now, a system of large size with small fluctuations is called a *(physical) bath*. Only baths make memories possible.

In other words, every record contains a bath. We conclude that any *observation* of a system is the interaction of that system with a bath. This connection will be used several times in the following, in particular in quantum theory. When the record is produced by a machine, one usually calls the 'observation' a *(generalized) measurement*. Are you able to specify the bath in the case of a person watching a landscape? Challenge

From the preceding discussion we can deduce the following surprising statement: since we have such a prodigious memory at our disposition, we can deduce that we are made of many small parts. And since records exist, the world must be made of a large number of small parts as well. No microscope is needed to confirm the existence of molecules or similar small entities; these tools are only needed to determine the *sizes* of these particles. Their existence can be deduced simply from the fact that we have memory. (Of course, another argument proving that matter is made of smallest parts is the ubiquity of noise itself.) See page 126

A second conclusion was popularized in the latex 1980s. Writing a memory does not produce entropy; it is possible to store information into a memory without increasing entropy. However, entropy is produced in every case that the memory is *erased*. It turns out that the (minimum) entropy created by erasing one bit is given by Challenge

$$S_{\text{per erased bit}} = k \ln 2 \quad . \tag{374}$$

As is well known, one needs energy to reduce the entropy of a system. By the way, *dreaming* is connected with the erasing and reorganization of information. Could that be a reason explaining that when one is very tired, without any energy left, one does not dream as much as in other cases? Challenge

Entropy is thus necessarily created when one forgets. This is evident when one imagines that forgetting is similar to the deterioration of an ancient manuscript. Entropy increases when the manuscript is not readable any more, since the process is irreversible and dissipative.* Another way to see this is to recognize that to clear a memory, e.g. a magnetic tape, one has to put energy into it, and thus to increase its entropy. Conversely, *writing* into a memory can often *reduce* entropy; we remember that signals, the entities which write memories, carry negative entropy. For example, the writing of analog magnetic tapes usually reduces their entropy. Ref. 12

The capacity of the brain

Computers are boring. They can give only answers.
Attributed to Pablo Picasso

The human brain is built in a way that its fluctuations cannot destroy its contents. The brain does this by literally growing connections, called *synapses*, between its various *neurons*, which are the cells doing the signal processing, i.e. the basic classification. The neuron is the basic processing element of the brain. It can do only two things: to fire and not to fire. (It

Ref. 13 * As Woycek Zurek explains so clearly, the entropy created inside the memory is the main reason that a Maxwell's demon cannot reduce the entropy of two volumes of gases by opening a door between them in such a way that fast molecules accumulate on one side, and slow molecules accumulate on the other.

Want to play demon? Click on the http:// www.wolfenet.com/ zeppelin/maxwell.htm web site.

is possible that the time at which a neuron fires also carries information; this question is not settled yet.) The neuron fires depending on its input, which usually comes from hundreds of other neurons, via the synapses. A neuron is thus an element which can distinguish the inputs it receives into two cases: those leading to firing and those which do not. Neurons are thus classifiers of the simplest type, able to distinguish between two situations only.

Ref. 14, 15

Every time we store something in our long term memory, like the phone number of a friend, new synapses are grown or the connection strength of existing synapses is changed. The connections between the neurons are much stronger than the fluctuations in the brain. Only strong disturbances, such as a blocked blood vessel or a brain lesion, can destroy neurons, and then lead to loss of memory.

As a whole, the brain is an extremely efficient memory. Despite intense efforts, engineers have not yet been able to build a memory with the capacity of the brain in the same volume. This memory capacity is easily estimated. By multiplying the number of neurons, about 10^{11}, by the average number of synapses per neuron, about 100, and also by the estimated number of bits stored in every synapse, about 10, one arrives at a storage capacity for the brain of about

$$M_{\text{rewritable}} \approx 10^{14}\,\text{bit} \approx 10^{4}\,\text{GByte} \quad . \qquad (375)$$

Note that evolution has managed to put as many neurons in the brain as there are stars in the galaxy, and that if one adds all the synapse lengths, one gets a total length of about 10^{11} m, which corresponds to the distance to from the earth to the sun. Our brain truly is *astronomically* complex.

The large storage capacity* of the brain shows that human memory is filled by the environment and is not inborn: one human ovule plus one sperm have a mass of about 1 mg, which corresponds to about $3 \cdot 10^{16}$ atoms. Obviously, fluctuations make it impossible to store 10^{16} bits in it. In fact, nature stores only about $3 \cdot 10^{9}$ bits in the genes of an ovule, using 10^{7} atoms per bit. In contrast, a typical brain has a mass of 1.5 to 2 kg, containing about 5 to $7 \cdot 10^{25}$ atoms, which makes it as efficient as the ovule. The difference between the number of bits in human DNA and of those in the brain nicely shows that practically all information stored in the brain is taken from the environment, and cannot be of genetic origin, even allowing for smart decompression of stored information.

* In practice, the capacity seems almost without limit, since the brain frees memory every time it needs some, by *forgetting* older data, e.g. during sleep. Note that this standard estimate of 10^{14} bits is not really correct! It assumes that the only component storing information in the brain is the synapse strength. Therefore it only measures the *erasable* storage capacity of the brain. In fact, information is also stored in the structure of the brain, i.e. in the exact configuration in which cell is connected to other cells. Most of this structure is fixed at the age of about two years, but continues at a smaller level for the rest of one's life. Assuming that for each of the N cells with n connections there are $f\,n$ connection possibilities, this *write once* capacity of the brain can be estimated as roughly $N\sqrt{fn}\,fn \log fn$ bits. For $N = 10^{11}$, $n = 10^{2}$, $f = 6$, this gives

Challenge

$$M_{\text{writeonce}} \approx 10^{16}\,\text{bit} \approx 10^{6}\,\text{GByte} \quad . \qquad (376)$$

By the way, even though the brains of sperm whales and of elephants can be five to six time as heavy as those of humans, the number of neurons and of their connections, an thus the capacity, seems to be highest for humans.

Sometimes it is claimed that people use only between 5% or 10% of their brain capacity. This myth, which goes back to the 19th century, would imply that one is able to measure the actually stored data in the brain and compare it with its capacity to an impossible accuracy. It also implies that nature would develop and maintain an organ with 90% overcapacity, wasting all the energy and material to build, repair, and maintain it.

In total, all these tricks used by nature result in the most powerful classifier yet known. Are there any limits to the brain's capacity to memorize and to classify? With the tools that humans have developed to expand the possibilities of the brain, such as paper, writing, and printing to help memory, and the numerous tools to simplify and to abbreviate classifications explored by mathematicians, the practical limit to brain classification is given by the time spent practicing it.* Ref. 16

The brain is unparalleled also in its processing capacity. This is most clearly demonstrated by the most important consequence deriving from memory and classification: thought and language. Indeed the many types of thinking or talking we use, such as comparing, distinguishing, remembering, recognizing, connecting, describing, deducing, explaining, imagining etc., all describe different ways to classify memories or perceptions. In the end all thinking and talking directly or indirectly classify observations. But how far are computers from achieving this! To talk to a computer program, such as to the famous program Eliza which mimics a psychoanalyst or to its improvements, is still a disappointing experience. To understand the reasons, one might ask:

What is language?

> Reserve your right to think, for even to think wrongly
> is better than not to think at all.
> Hypatia of Alexandria (ca. 355–415)

> Ein Satz kann nur sagen, *wie* ein Ding ist, nicht *was* es ist.**
> Ludwig Wittgenstein, *Tractatus*, 3.221

Language possibly is the most fantastic gift of human nature. Using their ability to produce sounds and to put ink on paper, people attach certain *symbols*,***also called *words* or *terms* in this context, to the many partitions they specify with help of their thinking. Such a categorization is then said to define a *concept*, or *notion*, and is then set in *italic typeface* in this text. A standard set of concepts forms a language.**** In other words, a *(human) language* is a standard way of symbolic interaction between people.***** Languages can be based on

* Without tools, there are strict limits, of course. The two-millimetre thick cerebral cortex of humans has a surface of about four sheets of A4 paper, a chimpanzee's yields one sheet, and a monkey's the size of a postcard. It is estimated that the total intellectually accessible memory is of the order of 1 GB, though with a large error.

** Propositions can only say *how* things are, not *what* they are.

*** A symbol is a type of *sign*, i.e. an entity associated by some convention to the object it refers. Following Charles Peirce (1839–1914) – see http://www.peirce.org – the most original philosopher born in the United States, a symbol differs from an *icon* (or *image*) and from an *index*, which are also attached to objects by convention, in that it does not resemble the object, as an icon does, and in that it has no contact with the object, as is the case for an index.

**** The recognition that language is a based on a partition of ideas, using the various differences between them to distinguish them from each other, goes back to the swiss Ferdinand de Saussure (1857–1913), who is regarded as the founder of linguistics. His textbook *Cours de linguistique générale*, Editions Payot, Paris, 1985, has been the reference work of the field for over half a century. Note that Saussure, in contrast to Peirce, prefers the term 'sign' to 'symbol', and that his definition of the term 'sign' includes also the object it refers to.

***** For a slightly different definitions, and a wealth of other interesting information about language, see the beautiful book by David Crystal, *The Cambridge encyclopedia of language*, Cambridge University Press, 1987.

facial expressions, on gestures, on spoken words, on whistles, on written words, etc. The use of *spoken* language is considerably younger than the human species; it seems that it appeared only about one hundred thousand years ago. Written language is even younger, namely only about six thousand years old. But the set of concepts used, the *vocabulary*, is still expanding. For single humans, the understanding of language begins soon after birth (perhaps even before), its active use begins at around a year of age, the ability to read can start as early as two, and personal vocabulary continues to grow as long as curiosity is alive.

Physics being lazy chat about motion, it needs language as essential tool. Of the many aspects of language, from literature to poetry, from jokes to military orders, from expressions of encouragement, dreams, love and emotions, physics uses only a small and rather special segment. This segment is defined by the inherent restriction to talk about motion. Since motion is an observation, i.e. an interaction with the environment that other people experience in the same way, this choice puts a number of restrictions on the contents – the vocabulary – and on the form – the grammar – of such discussions.

For example, from the definition that observations are shared by others, one gets the requirement that the statements describing them must be translatable into all languages. But when can a statement be translated? On this question two extreme points of view are possible; the first maintains that *all* statements can be translated, since it follows from the properties of human languages that each of them can express every possible statement. In this view, only sign systems which allows to express the complete spectrum of human messages form a *human language*. This property distinguishes spoken and sign language from animal languages, such as the signs used by apes, birds or honey bees, and also from computer languages, such as Pascal or C. With this meaning of language, all statements can be translated by definition.

It is more challenging for a discussion to follow the opposing view, namely that precise translation is possible only for those statements which use terms, word types and grammatical structures found in *all* languages. Linguistic research has invested considerable effort in the distillation of phonological, grammatical, and semantic *universals*, as they are called, from the around 7000 languages thought to exist today.*

The investigations into the *phonological* aspect, which showed for example that every language has at least two consonants and two vowels, does not provide any material for the discussion of translation.** Studying the *grammatical* (or *syntactic*) aspect, one finds that all languages use smallest elements, called 'words', which they group into sentences.

* A comprehensive list with 6700 languages (and with 39000 language and dialect names) can be found on the world wide web site by Barbara Grimes, *Ethnologue – languages of the world*, to be found at the address http://www.sil.org/ethnologue or in the printed book of the same name.

It is estimated that 15000 ± 5000 languages have existed in the past.

In today's world, and surely in the sciences, it is often sufficient to know one's own language plus english. But never let the failure of good command of english hamper your curiosity. A well known physics journal has been published with an english language mistake in its title for dozens of years: "Progress of [sic] theoretical physics"; it provides a prime example of collective stubbornness. We don't want to follow it; we want to improve our description of nature continuously, and continuously correct any mistakes: in this way we can enjoy life more and more. Since english is the language with the largest number of words, learning it well is also a greater challenge than for most of the others. Ref. 17

** Studies center on topics such as the fact that in many languages the word for 'little' contains an 'i' (or high pitched 'e') sound: petit, piccolo, klein, tiny, pequeño, chiisai; exceptions are: small, parvus.

They all have pronouns for the first and second person, 'I' and 'you', and always contain nouns and verbs. All languages use *subjects* and *predicates*, or, as one usually says, the three entities *subject*, *verb* and *object*, though not always in this order. Just check the languages you know. Challenge

On the *semantic* aspect, the long list of lexical universals, i.e. words that appear in all languages, such as 'mother' or 'sun', has recently been given a structure by the discovery of semantic primitives. The list of *universal semantic primitives* is the result of the search of the building blocks from which all concepts can be built, in the sense that the definition of any concept can be given using only previously defined concepts, and these in turn can be defined with previously defined concepts and so forth, until one reaches a fundamental level consisting only of the primitives themselves. The list thus results from the study of the many existing languages and from the way concepts are built upon each other. In November 1992, it contained the following primitives:

Table 33 Semantic primitives, following Anna Wierzbicka

I, you, someone, something; people	[substantives]
this, the same, two, all, much (many); one	[determiners and quantifiers]
know, want, think, feel, say	[mental predicates]
do, happen	[agent, patient]
good, bad	[evaluative]
big, small	[descriptors]
very	[intensifier]
can, if (would)	[modality, irrealis]
because	[causation]
no (not)	[negation]
when, where, after (before), under (above)	[time and place]
kind of, part of	[taxonomy, partonomy]
like	[hedge/prototype]

Following the life-long research of Anna Wierzbicka, all these concepts exist in all languages of the world she and her research group was able to study. She showed that in every language all other concepts can be defined with help of this basic set. We note that 'true' and 'false' are not included in the list, because they are seen as composite concepts. We also note that 'motion' is implicitly contained in the verbs 'do' and 'happen'. Also the other verbs in the list can be seen as examples of motion. Also linguistically, motion is at the basis of human experience. Ref. 18 See page 397

The definition of language given above, namely a means of communication which allows to express everything one wants to say, can thus be refined: a *human language* is any set of concepts which includes the semantic primitives. For a physicist, with his aim to talk in as few words as possible, obviously such a long list is not satisfying, especially when he notes that all these concepts are about interactions between different parts of nature. One of the aims of our walk is to arrive at a list consisting of only one or two basic concepts. To appreciate this aim, try to define what 'no' means, or what an 'opposite' is. Or simply try whether you are able to reduce the list. Challenge

We can summarize all these results of linguistics by saying that by constructing a statement made only of subject, verb and object, consisting only of nouns and verbs, using only concepts built from the semantic primitives, one is sure that it can be translated into all languages. This explains why science texts often are so boring: the authors are often too afraid to depart from this basic scheme!

Are there semantic primitives in physics?

Is there a basic set of other concepts on which all physics concepts are based? In classical physics, the concepts of space-time, mass and charge form such a set. Two questions arise straight away. Is the set complete? Can it be reduced to a smaller amount of concepts? Both questions will stay with us for a large part of the escalation. Since the answer to the first is negative, we need to be prepared for a longer escalation, in order to find the complete set. This will happen in the middle section. Then the question of the smallest possible set will arise, and keep us busy in the third part.

What is a concept?

> Alles, was wir sehen, könnte auch anders sein.
> Alles, was wir überhaupt beschreiben können, könnte auch anders sein.
> Es gibt keine Ordnung der Dinge a priori.*
> Ludwig Wittgenstein, *Tractatus*, 5.634

There is a group of people which has taken the strict view on translation and on precision to the extreme. They build all concepts from an even smaller set of primitives, namely only two: 'set' and 'relation', and explore the various possible combinations of these two concepts, studying their classifications. Step by step, this group of radicals, commonly called *mathematicians*, arrived to define with full precision concepts such as numbers, points, curves, equations, symmetry groups etc. The construction of these concepts is summarized partly in this chapter and partly in appendix D.

However, despite their precision, in fact precisely because of it, no mathematical concept talks about nature or about observations.** Therefore the study of motion needs other, more useful concepts. What properties must a useful concept have? An example: What is 'freedom' or what is a 'parachute'? Obviously, a useful concept implies a list of its parts, its aspects and their internal relations, as well of their relation to the exterior world. Thinkers in various fields, from philosophy to politics, agree that the definition of any *concept* requires:

- explicit and fixed content,
- explicit and fixed limits,
- explicit and fixed domain of application.

* Whatever we see could be other than it is. Whatever we can describe at all could be other than it is. There is no a priori order of things.

** Insofar as one can say that mathematics is based on the concepts of 'set' and 'relation', which are based on experience, one can say that mathematics explores a section of reality, and that its concepts are *derived* from experience. This and similar views of mathematics are called platonism. More concretely, platonism is the view that the concepts of mathematics exist *independently* of people, and that they are discovered, and not created by mathematicians.

In short, since mathematics makes use of the brain, which is a physical system, actually *mathematics is applied physics*.

The inability to state these properties or keep them fixed is often the easiest way to distinguish *crackpots* from more reliable thinkers. Unclearly defined terms, which thus do not qualify as concepts, regularly appear in myths, e.g. ‘dragon’ or ‘sphinx’, or in ideologies, e.g. ‘worker’ or ‘soul’. Even physics is not immune. For example, we will discover later that neither ‘universe’ nor ‘creation’ are concepts. Are you able to argue the case? Challenge

But the three defining properties of any concepts are interesting in their own right. Explicit content means that concepts are built onto each other. In particular, the most fundamental concepts should be those which have no parts and no internal relations, but only external ones. Can you think of one? Only in the last part of this walk we will discover the final words on the topic. Challenge

Explicit limits, together with the explicit contents, also imply that all concepts describing nature are *sets*, since sets obey the same requirement. In addition, explicit domains of applications imply that all concepts also are *relations*.* Since math is based on the concepts of ‘set’ and of ‘relation’, one follows directly that mathematics can provide the *form* for any concept, especially whenever high precision is required, as in the study of motion. Obviously, the *content* of the description is only provided by the study itself; only then concepts become useful.

In the case of physics, the search for sufficiently precise concepts can be seen as the single theme structuring the long history of the field. Regularly, new concepts were proposed, explored in all their properties, tested, and finally rejected or adopted, in the same way that children reject or adopt a new toy. Children do this unconsciously, scientists do it consciously, using language.** That is why such concepts are universally intelligible.

Note that concept ‘*concept*’ itself is not definable independently of experience; a concept is something that helps us act and react to the world we live in. Moreover, concepts do not live in a separate world form the physical one: every concept requires memory form its user, since the user has to remember the way it was formed; therefore every concept needs a material support for its use and application. Insofar all thinking and thus every science is fundamentally based on experience.

In conclusion, all concepts are based on the idea that nature is made of related parts. The complementing couples that follow from this idea, such as ‘noun – verb’ in linguistics, ‘set – relation’ and ‘definition – theorem’ in mathematics, and ‘aspect of nature – pattern of nature’ in physics, always guide human thinking, even during childhood, as developmental psychology can testify.

Concepts are merely the results, rendered permanent
by language, of a previous process of comparison.
William Hamilton

* One sees that every physical concept, is an example of a (mathematical) *category*, i.e. a combination of objects and mappings. For more details about categories, with a precise definition of the term, see page 387.

** Concepts formed unconsciously in our early youth are the most difficult to define precisely, i.e. with language. Some people who were unable to do so, like the prussian philosopher Immanuel Kant (1724–1804) used to call them “a priori” concepts (such as ‘space’ and ‘time’) to contrast them with the more clearly defined “a posteriori” concepts. Today, this distinction has been found to be unfounded both by the study of child psychology (see the footnote on page 372) and by physics itself, so that these qualifiers are thus not used in our walk.

The axioms of ZFC set theory

- Two sets are equal if and only if they have the same elements. (Axiom of extensionality)
- The empty set is a set. (Axiom of the null set)
- If x and y are sets, then the unordered pair $\{x, y\}$ is a set. (Axiom of unordered pairs)
- If x is a set of sets, the union of all its members is a set. (Union or sum set axiom)
- The entity $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\{\{\emptyset\}\}\}, ...\}$ is a set – in other words, infinite collections such as the natural numbers are sets. (Axiom of infinity)
- An entity defined by all elements having a given property is a set, provided this property is reasonable – some important technicalities defining 'reasonable' being necessary. (Axiom of replacement)
- The entity y of all subsets of x is also a set, called the power set. (Axiom of the power set)
- A set is not an element of itself – plus some technicalities. (Axiom of regularity)
- Picking elements from a list of sets allows to construct a new set – plus technicalities. (Axiom of choice)

Table 34 The defining properties of a set

What are sets? What are relations?

Defining sets and defining relations are fundamental actions of our thinking. This can be seen most clearly in any book about mathematics; such a book is usually divided in paragraphs labeled 'definition' and others labeled 'theorem', 'lemma' or 'corollary'. The first type of paragraph defines concepts, i.e. defines sets, and the other three types of paragraphs express relations, i.e. connections between these sets. *Mathematics* is thus the exploration of the possible symbolic concepts and their relations – it is the science of symbolic necessities.

Sets and relations are tools of classification; that is why they are also the tools of any bureaucrat. This class of human beings is characterized by heavy use of paper clips, files, metal closets, archives – which all define various types of sets – and by the extensive use of numbers, such as letter reference numbers, customer numbers, passport numbers, account numbers, law article numbers – which define various types of relations between the items, i.e. between the elements in the sets.

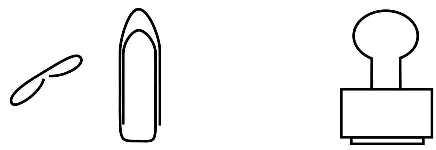


Figure 144 Devices for the definition of sets (left) and of relations (right)

Both the concepts of set and of relation express, in different ways, the fact that nature can be *described*, i.e. that it can be classified into parts which form a whole. The act of grouping together aspects of experience, i.e. the act of classifying them, is expressed in formal language by saying that a set is defined. In other words, a *set* is a collection of *elements* of our thinking. Every set distinguishes the elements from each other, and from the set itself.

This definition of 'set' is called the *naive* definition. For physics, the definition is sufficient, but you won't find anybody admitting this. In fact, mathematicians have refined the definition of the concept 'set' several times, because it does not work well for infinite sets. A famous example is the story about sets which do not contain themselves. Any set is of two sorts: either it contains itself or it does not. If we take the set of all sets who do *not* contain themselves, to which sort does it belong? Challenge

To avoid this and similar problems, mathematics needs a precise definition of the concept of 'set'. The first such definition was given by the german mathematician Ernst Zermelo (1871, Berlin–1951, Freiburg i.B.) and the german-israeli mathematician Adolf/Abraham Fraenkel (1891, München–1965, Jerusalem); later on, the so-called *axiom of choice* was added, in order to make possible to manipulate a wider class of infinite sets. The result of these efforts is called the ZFC definition.* From this basic definition one can construct all mathematical concepts used in physics. From a practical point of view, it is sufficient to keep in mind that for the whole of physics, the "naive" definition of a set given above is equivalent to the precise ZFC definition, actually even to the simper ZF definition. Subtleties appear only for some special types of infinite sets, but these are not used in physics. In short, from the basic, naive set definition one can construct all concepts used in physics.

But the naive set definition is far from boring. To make two people happy when dividing a cake, one follows the rule: I cut, you choose. What rule is needed for three people? And for four? Ref. 20 Challenge

Apart from defining sets, every child and every brain creates links between the different aspects of experience. For example, when it hears a voice, it automatically makes the connection that a human is present. Connections of this type are called *relations* in formal language. Relations connect and differentiate elements along other lines than sets: the two form a complementing couple. Defining a set unifies many objects and at the same time divides them into two: those belonging to the set and those which do not; defining a (binary) relation unifies elements two by two and divides them into many, namely into the many couples it defines.

Sets and relations are closely interrelated concepts. Indeed one can define (mathematical) relations with the help of sets. A *(binary) relation* between two sets X and Y is a subset of the product set, where the *product set* or *cartesian product* $X \times Y$ is the set of all ordered pairs (x,y) with $x \in X$ and $y \in Y$. An *ordered pair* (x,y) can be defined with sets. Can you find out how? For example, in the case of the relation 'is wife of', the set X is the set of all women and the set Y that of all men; the relation is given by the list all the appropriate ordered pairs, which is much smaller than the product set, i.e. the set of all possible woman-man combinations. Challenge

It should be noted that the definition of relation just given is not really complete, since every construction of the concept 'set' already contains certain relations, such as the relation 'is element of.' It does not seem to be possible to reduce either of the concepts 'set' or 'relation' completely to the other one. This situation is reflected in the physical cases of sets

* A global overview of axiomatic set theory is given by Paul J. Cohen, Reuben Hersch, *Non-cantorian set theory*, Scientific American **217**, pp. 104–116, 1967. Those were the times in which Scientific American was a quality magazine.

Ref. 19 Other types of entities, more general than standard sets, obeying other properties, can also be defined, and are also subject of (comparatively little) mathematical research. For an example, see the section on cardinals later on. Such more general entities are called classes whenever they contain at least one set. Can you give an example? In the third part of our escalation we will meet physical concepts which are not described by sets nor by classes, i.e. which do not contain any set at all. That is were the real fun starts. See page 385 Challenge and relations, such as space (as a set of points) and distance, which also seem impossible to separate completely from each other.

Infinity

Mathematicians soon discovered that the concept of 'set' is only useful if one can call also collections such as $\{0,1,2,3...\}$, i.e. of the number 0 and all its successors, a 'set'. To achieve this, one property in the Zermelo-Fraenkel list defining the term 'set' explicitly specifies that this collection can be called a set. (In fact, also the axiom of replacement states that sets may be infinite.) Infinity is thus put into mathematics and into the tools of our thought right at the very beginning, in the definition of the term 'set'. When describing nature, with or without mathematics, one should never forget this fact. In addition, a few other points about infinity should be general knowledge of any expert on motion.

Only *sets* can be infinite. And sets have parts, namely their elements. When a thing or a concept is called 'infinite' one can *always* ask and specify what its parts are; for space the parts are the points, for time the instants, for the set of integers the integers, etc. An indivisible or an only finitely divisible entity cannot be called infinite.*

A set is infinite if there is a function from it into itself that is *injective* (i.e. different elements map to different results) but not *onto* (i.e. some elements do not appear as images of the map); e.g. the map $n \mapsto 2n$ shows that the set of integers is infinite. Infinity can be checked also in another way: a set is infinite if it remains so also after removing one element. Even repeatedly. Of course one needs to remember that the empty set is *finite*.

There are *many types* of infinities, all of different size.** This important result was discovered by the danish-russian-german mathematician Georg Cantor (1845–1918). He showed that from the countable set of natural numbers one can construct other infinite sets which are not countable. He did this by showing that the *power set* $P(\omega)$, namely the set of all subsets, of a countably infinite set is infinite, but *not* countably infinite. Sloppily speaking, the power set is "more infinite" than the original set. The real numbers **R**, to be defined shortly, are an example of an uncountably infinite set; there are many more of them than there are natural numbers. However, *any* type of infinite set contains at least one subset which is countably infinite.

Even for an infinite set one can define size as the number of its elements. Cantor called this the *cardinality* of a set. The cardinality of a finite set is simply given by the number of its elements. The cardinality of a power set is 2 exponentiated by the cardinality of the set. The cardinality of the set of integers is called $\aleph_0$, pronounced 'aleph zero', after the first letter of the hebrew alphabet. The smallest *uncountable* cardinal is called $\aleph_1$. The next cardinal is called $\aleph_2$ etc. A whole branch of mathematics is concerned with the manipulation of these infinite "numbers"; addition, multiplication, exponentiation are easily defined. For some of them, even logarithms and other functions make sense.***

See page 719

* Therefore, most gods, being concepts and thus sets, are either finite, or, in case they are infinite, they are divisible.

** In fact, there such a huge number of types of infinities, that none of these infinities itself actually describes this number. Technically speaking, there are as many infinities as there are ordinals.

*** Many results are summarized in the excellent and delightful paperback by Rudy RUCKER, *Infinity and the mind – the science and philosophy of the infinite*, Bantam, Toronto, 1983.

The cardinals defined in this way, including $\aleph_n$, $\aleph_\omega$, $\aleph_{\aleph_\aleph}$ are called accessible, because in the meantime, people have defined even larger types of infinities, called *inaccessible*. These numbers (inaccessible cardinals, measurable cardinals, supercompact cardinals, etc.) need additional set axioms, extending the ZFC system.

The real numbers have the cardinality of the powerset of the integers, namely $2^{\aleph_0}$. Can you show this? One thus has the famous question: Is $\aleph_1 = 2^{\aleph_0}$ or not? The statement that this be so is called the *continuum hypothesis* and was unanswered for several generations. Only in 1963 came the surprising answer: the definition of the concept of set is not specific enough to fix the answer. By specifying the concept of set in more detail, with additional axioms – remember that axioms are defining properties – one can make the continuum hypothesis come out either right or wrong, as one prefers. Challenge

Another result of research into transfinites is important: for every definition of a type of infinite cardinal, it seems to be possible to find a larger one. In everyday life, the idea of infinity is often used to stop discussions about size: "My big brother is stronger than yours." "But mine is infinitely stronger than yours!" Mathematics has shown that questions on size continue also afterwards: "The strength of my brother is the powerset of that of yours!" Rucker reports that mathematicians think there is no possible nor any conceivable end to these discussions. Ref. 21

For our escalation, a question appears directly: do infinite quantities exist in nature? Or better, is it necessary to use infinite quantities to describe nature? You might want to specify your own opinion on the issue. It will be settled during the rest of our escalation. Challenge

Functions and structures

Which relations are useful to describe patterns in nature? A typical example is "larger stones are heavier". Such a relation is of a specific type: it relates one specific value of an observable 'volume' to one specific value of the observable 'weight'. Such a one-to-one relation is called a *(mathematical) function* or *mapping*. Functions are the most specific types of relations; thus they convey a maximum of information. In the same way as the use of numbers for observables, functions allow easy and precise communication of relations between observations. All physical rules and "laws" are therefore expressed with help of functions, and since physical "laws" are about measurements, functions of numbers are their main building blocks.

A *function* f, or *mapping*, is a thus binary relation, i.e. a set $\{(x,y)\}$ of ordered pairs, where for every value of the first element x, called the *argument*, there is only *one* pair (x,y). The second element y is called the *value* of the function at the argument x. The set X of all arguments x is called the *domain of definition* and the set Y of all second arguments y is called the *range* of the function. One writes

$$f : X \to Y \quad \text{and} \quad f : x \mapsto y \quad \text{or} \quad y = f(x) \tag{377}$$

where the type of arrow shows whether one is speaking about sets or about elements.

We note that it is also possible to use the couple 'set' and 'mapping' to define all mathematical concepts; in this case a relation is defined with help of mappings. A modern school of mathematical thought formalized this approach by the use of (mathematical) *categories*, a concept which includes both sets and mappings on an equal footing in its definition. *

To think and talk more clearly about nature, one needs to define more specialized concepts than sets, relations, and functions, because these basic terms are too general. The most important concepts derived from them are operations, algebraic structures, and numbers.

An *binary operation* is a function that maps the cartesian product of two copies of a set X into itself. In other words, an operation w takes an ordered couple of arguments $x \in X$ and assigns to it a value $y \in X$:

$$w : X \times X \to X \quad \text{and} \quad w : (x,x) \mapsto y \quad . \tag{378}$$

Challenge Is division of numbers an operation following this definition? An *algebraic structure*, also called an *algebraic system*, is (in the most restricted sense) a set together with certain operations. The most important algebraic structures appearing in physics are groups, vector spaces, and algebras.

In addition to algebraic structures, mathematics is based on *order structures* and on *topological structures*. Order structures are building blocks of numbers and necessary to define comparisons of any sort. Topological structures are built, via subsets, on the concept of neighbourhood. They are necessary to define continuity, limits, dimensionality, topological spaces, and manifolds.

Obviously, most mathematical structures are combinations of various examples of these three basic structure types. For example, the system of real numbers is given by the set of real numbers with the operations of addition and multiplication, the order relation 'is larger than', and a continuity property. They are thus built by combining an algebraic structure, an
Ref. 22 order structure and a topological structure.

The mathematical systems of importance in physics are presented partly in the following and partly in appendix D.

Numbers

Which numbers are multiplied by six when their
Challenge last digit is taken away and transferred to the front?

Numbers are the oldest mathematical concept and are found in all cultures. The notion of number, in greek αριθμος, has often been changed with the aim to include more and more general objects, but always retaining the general idea that numbers are entities which can be added, subtracted, multiplied and divided.

The modern way to write numbers, as e.g. in $12\,345\,679 \cdot 45 = 666\,666\,666$, is essential for science.** It can be argued that the lack of a good system for writing down and for

∗ A *category* is defined as a collection of objects and a collection of 'morphisms' or mappings. Morphisms are composable, the composition is associative, and there is an identity morphism. The strange world of category theory, sometimes called the abstraction of all abstractions, is presented in ...

Note that every category contains a set; since it is unclear whether nature contains sets, as we will discuss on page 413, it is questionable whether categories will be useful in the unification of physics, despite their abstract charm.

∗∗ However, there is no need for written numbers for doing mathematics, as shown by Marcia Ascher, *Ethnomathematics – a multicultural view of mathematical ideas*, Brooks/Cole, 1991.

calculating with numbers has kept the progress of science back for several centuries. (By the way, the same can be said for the affordable mass reproduction of written texts.)

The simplest numbers, $1,2,3,4,\ldots$, are usually seen as being taken directly from experience. However, they can also be constructed from the notions of 'relation' and 'set'. One of the many possible ways to do this (can you fin one?) is by identifying a natural number with the set of its predecessors. With the relation S, 'successor of', this definition can be written as Challenge

$$0 := \emptyset \quad , \quad 1 := S0 = \{0\}\{\emptyset\} \quad ,$$
$$2 := S1 = \{0,1\} = \{\emptyset,\{\emptyset\}\} \quad \text{and} \quad n+1 := Sn = \{0,\ldots,n\} \qquad (379)$$

This set, together with the binary operations 'addition' and 'multiplication,' constitutes the algebraic system $N = (N,+,\cdot,1)$ of the *natural numbers*.* (Sometimes the number zero is not counted as a natural number.) For all number systems the algebraic system and the set are often sloppily designated by the same symbol.

Table 35 Some large numbers

Number	examples in nature
Around us	
1	number of angels which can be in one place, following Thomas Aquinas Ref. 23
20	number of digits in precision measurements which will probably never be achieved
34, 55, 89	petals of common types of daisy and sunflower Ref. 24
57	faces of a diamond with brilliant cut
2000	stars visible in the night sky
10^5	leaves of a tree (10 m beech)
6 to $7 \cdot 10^9$	humans in the year 2000

* Any system with the same properties as the natural numbers is called a *semi-ring*. A *ring* $(R,+,\cdot)$ is a set R of elements with two binary operations, called *addition* and *multiplication*, usually written $+$ and $\cdot$ (or simply dropped), for which the following properties hold for all elements $a,b,c \in R$:

- R is a commutative group with respect to addition, i.e. one has $a+b \in R,\ a+b = b+a,\ a+0 = a,\ a+(-a) = a-a = 0$ as well as $a+(b+c) = (a+b)+c$
- R is closed under multiplication, i.e. $ab \in R$
- multiplication is associative, i.e. $a(bc) = (ab)c$
- distributivity holds, i.e. $a(b+c) = ab+ac$ and $(b+c)a = ba+ca$.

Defining properties such as these are also called *axioms*. Note that axioms are not basic beliefs, as often, but wrongly stated; axioms are the basic properties used in the definitions of a concept, in this case, that of ring. A *semi-ring* is a set with all the properties of a ring, except that the existence of neutral and negative elements for addition is replaced by the weaker requirement that if $a+c = b+c$ then $a = b$. A *field* K is a ring with

- an identity 1, such that for all elements a one has $1a = a$,
- at least one element different from zero, and most importantly
- a (multiplicative) inverse a^{-1} for every element $a \neq 0$.

A ring or field are said to be *commutative* if the multiplication is commutative. A non-commutative field is also called a *skew field*. Fields can be finite or infinite. All finite fields are commutative. In a field, all equations of the type $cx = b$ and $xc = b$ $(c \neq 0)$ have solutions for x; they are unique if $b \neq 0$. To sum up sloppily by focusing on the most important property, a field is a set of elements for which, together with addition, subtraction and multiplication, a *division* is also defined.

Number		examples in nature
10^{17}		ants in the world
ca. 10^{20}		number of snowflakes falling on the earth per year
ca. 10^{23}		grains of sand in the Sahara desert
10^{22}		stars in the universe
10^{25}		cells on earth
$1.1 \cdot 10^{50}$		atoms making up the earth $(6370^3\,\mathrm{km}^3 \cdot 4 \cdot 3.14/3 \cdot 5500\,\mathrm{kg/m}^3 \cdot 30\,\mathrm{mol/kg} \cdot 6 \cdot 10^{23}\,/\mathrm{mol})$
10^{81}		atoms in the visible universe
10^{90}		photons in the visible universe
10^{169}		number of atoms fitting in the visible universe
10^{244}		number of space-time points inside the visible universe
Information		
51		record number of languages spoken by one person
ca. 5000		words spoken on an average day by a man
ca. 7000		words spoken on an average day by a woman
ca. 350 000		words of the english language (more than any other language, with the possible exception of german)
ca. 2 000 000		number of scientists on earth around the year 2000
$3 \cdot 10^{8}$		words spoken during a lifetime (2/3 time awake, 30 words per minute)
$4 \cdot 10^{9}$		pulses exchanged between both brain halves every second
10^{9}		words heard and read during a lifetime (2/3 time awake, 30 words per minute)
10^{17}		image pixels seen in a lifetime $(3 \cdot 10^9 s \cdot (1/15\,\mathrm{ms}) \cdot 2/3$ (awake) $\cdot 10^6$ (nerves to the brain) Ref. 25
10^{19}		bits of information processed in a lifetime (the above times 32)
ca. $5 \cdot 10^{12}$		printed words available in (different) books around the world (ca. $100 \cdot 10^6$ books consisting of 50000 words)
$2^{10} \cdot 3^7 \cdot 8! \cdot 12!$ $4.3 \cdot 10^{19}$	=	possible positions of the $3 \times 3 \times 3$ Rubik's cube Ref. 26
$5.8 \cdot 10^{78}$		possible positions of the $4 \times 4 \times 4$ Rubik-like cube
$5.6 \cdot 10^{117}$		possible positions of the $5 \times 5 \times 5$ Rubik-like cube
ca. 10^{200}		possible games of chess
ca. 10^{800}		possible games of go
ca. 10^{10^7}		possible states in a personal computer
Parts of us		
300 000		hairs on the head
900 000		neurons in the brain of a grasshopper
$127 \cdot 10^{6}$		light sensitive cells per retina
10^{10} to 10^{11}		neurons in the human brain
$> 10^{16}$		memory bits in the human brain
600		numbers of muscles in the human body, of which about half are in the face
10^{13} to 10^{14}		cells in the human body
10^{14}		bacteria carried in the human body
$500 \cdot 10^{6}$		blinks of the eye during a lifetime (about once every four seconds when awake)
$300 \cdot 10^{6}$		breaths taken during human life
$3 \cdot 10^{9}$		heart beats during a human life
$3 \cdot 10^{9}$		letters (base pairs) in haploid human DNA

Number	examples in nature
$6.1 \cdot 10^9$	bits in a compact disk

The system of *integers* $Z = (..., -2, -1, 0, 1, 2, ..., +, \cdot, 0, 1)$ is the minimal ring which is an extension of the natural numbers. The system of *rational numbers* $Q = (Q, +, \cdot, 0, 1)$ is the minimal field which is an extension of the ring of the integers. The system of the *real numbers* $R = (R, +, \cdot, 0, 1, >)$ is the minimal extension of the rationals which is continuous and totally ordered. (For the definition of continuity, see page 761.) Equivalently, it is the minimal extension of the rationals which is a complete, totally strictly-archimedean ordered field. But the construction, i.e. the definition, of integer, rational and real numbers from the natural numbers is not only possible in the way just mentioned. Perhaps the most beautiful definition of all these types of numbers is the one discovered in 1969 by John Conway, and popularized by him, Donald Knuth, and Martin Kruskal. Ref. 27

• A number is a sequence of bits. They are usually called ups and downs, and examples are shown in figure 145.

• The empty sequence is zero.

• A finite sequence of n ups is the integer number n, and a finite sequence of n downs is the integer $-n$. Finite sequences of mixed ups and downs give the *dyadic rational numbers* are the numbers made of a finite sequence of ups and downs. Examples are 1, 2, 3, -7, 19/4, $37/256$ etc. They all have denominators with a power of 2. The other *rational numbers* are those which end in an infinitely repeating string of ups and downs, such as $2/3$. Simply countably infinite series give the *reals*, the *infinitesimals*, and simple infinite numbers. Longer countably infinite series give even more crazy numbers. The complete class is called the *surreal* numbers.*

There are two ways to write surreal numbers. The first is the just mentioned sequence of bits. But to define addition and multiplication, one usually uses another notation, deduced from figure 145. A surreal s is defined as the earliest number of all those between two series of earlier surreals, the left and the right series:

$$\alpha = \{a, b, c, ...|A, B, C, ...\} \quad \text{with} \quad a, b, c, < \alpha < A, B, C \tag{380}$$

For example, one has

$$\{0|\} = 1 \quad , \quad \{0,1|\} = 2 \quad , \quad \{|0\} = -1 \quad , \quad \{|-1,0\} = -2 \quad , \quad \{0|1\} = 1/2 \quad ,$$
$$\{0|1/2, 1/4\} = 1 \quad , \quad \{0, 1, 3/2, 25/16 | 41/16, 13/8, 7/4, 2\} = 1 + 37/64$$

showing that the finite surreals are the *dyadic numbers* $m/2^n$. Given two surreals $\alpha = \{..., a, ...|..., A, ...\}$ with $a < \alpha < A$ and $\beta = \{..., b, ...|..., B, ...\}$ with $b < \beta < B$, addition is defined recursively, using earlier, already defined numbers, as

$$\alpha + \beta = \{..., a + \beta, ..., \alpha + b, ...|..., A + \beta, ..., \alpha + B, ...\} \quad . \tag{381}$$

This definition is used for the simple reason that it gives the same results as usual addition for integers and reals. Can you confirm this? By the way, addition is not always commutative. Are you able to find the exceptions, and to find the definition for subtraction? Challenge

* The surreal numbers do *not* form a set because they contain all *ordinal numbers*, which themselves do not form a set, even though they of course *contain* sets. In short, ordinals and surreals are classes which are larger than sets.

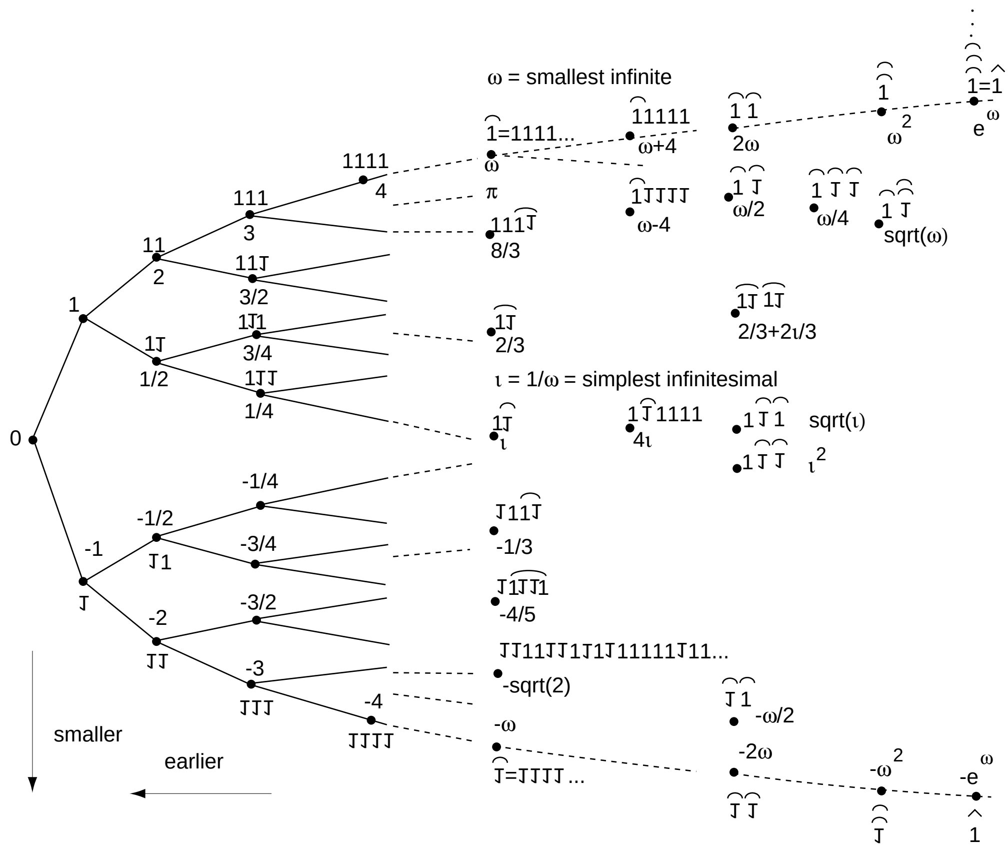


Figure 145 The surreal numbers in conventional and in bit notation

Multiplication is also defined recursively, by the expression

$$\alpha\beta = \{\ldots, a\beta+\alpha b - ab, \ldots, A\beta + \alpha B - AB, \ldots | \ldots, a\beta + \alpha B - aB, \ldots, A\beta + \alpha b - Ab, \ldots\} \quad . \qquad (382)$$

These definitions allow to write $\iota = 1/\omega$, and to talk about numbers such as $\sqrt{\omega}$, the square root of infinity, about $\omega + 4$, $\omega - 1$, 2ω, e^{ω} and about other strange numbers shown in figure 145. However, the surreal numbers are not commonly used. More common is one of their subsets. Ref. 27

The *real numbers* are all those surreals whose length is not larger than infinity and who do not have periodic endings with a period of length 1. In other words, the surreals distinguish the number $0.999999\overline{9}$ form the number 1, whereas the reals do not. In fact, between the two, there are infinitely many surreal numbers. Can you name a few? Challenge

Reals are more useful to describe nature than surreals, because first of all they form a set, which the surreals do not, and secondly because they allow the definition of integration. Other numbers defined with the help of reals, e.g. the complex numbers C and the quaternions H, are also presented in appendix D. A few more elaborate number systems are also presented there. See appendix D

To conclude, in physics it is usual to call *numbers* the elements of any set which is a semi-ring (e.g N), a ring (e.g. Z) or a field (Q, R, C, H). Since numbers allow to compare magnitudes, all play important roles in the description of observations.

When a series of equal balls is packed in such a way that the area of necessary wrapping paper is minimal, for a small number of balls the linear package, with all balls in one row, is the most efficient. For which number of balls is the linear package not a minimum any more? Ref. 28 Challenge

Why use maths?

Die Forderung der Möglichkeit der einfachen Zeichen ist die Forderung der Bestimmtheit des Sinnes.*
Ludwig Wittgenstein, *Tractatus*, 3.23

Several well-known physicists have asked this question repeatedly. For example, Niels Bohr is quoted to have said: "We do not know why the language of mathematics has been so effective in formulating those laws in their most succinct form." Eugene Wigner wrote an often cited paper entitled *The unreasonable effectiveness of mathematics.* At the start of science, many centuries earlier, Pythagoras and his contemporaries were so overwhelmed by the usefulness of numbers in describing nature, that Pythagoras was able to organize a sect based on this connection. The members of the inner circle of this sect were called 'learned people,' in greek "*mathematicians,* " from the greek μάθημα 'teaching'. This sect title then became the name of the profession. Ref. 29 Ref. 30

All these men forgot that numbers, as well as a large part of math, are concepts developed precisely with the aim to describe nature. And most of all, these concepts were developed right from the start to provide a description as succinct as possible. That is one aspect of the fact that mathematics is the science of symbolic necessities.

But perhaps this answer is too dismissive. Perhaps these thinkers mainly wanted to express their feeling of wonder when experiencing that language works, that thinking and our brain works, and that life and nature are so beautiful. This would put the title question nearer to the well-known statement by Albert Einstein: 'The most incomprehensible fact about the universe is that it is comprehensible.' Comprehension is another word for description, i.e. for classification. Obviously, any separable system is comprehensible, and there is nothing strange about it. But is the universe separable? As long as is it described as being made of particles and vacuum, this is the case. But whether this actually is correct will be revealed only in the third part of this walk.

Die Physik ist für Physiker viel zu schwer.**
David Hilbert (1862–1943), mathematician.

* The requirement that simple signs be possible is the requirement that sense be determinate.
** Physics is much too difficult for physicists.

Is mathematics a language?

Die Sätze der Mathematik sind Gleichungen, also Scheinsätze.
Der Satz der Mathematik drückt keinen Gedanken aus.*
Ludwig Wittgenstein, *Tractatus*, 6.2, 6.21

Surely, mathematics is a *vocabulary* which helps to talk with precision. Mathematics can be seen as the exploration of *all* possible concepts which can be constructed from the two fundamental bricks 'set' and 'relation' (or some alternative pair). Therefore, *mathematics* is the science of symbolic necessities. Rephrased again, mathematics is the exploration of all possible types of classifications. This explains its usefulness in all situations where complex, yet precise classifications of observations are necessary, such as in physics.

However, mathematics cannot express everything humans want to communicate, e.g. an idea one had or the fun of swimming. Mathematics is the science of symbolic necessities; thus mathematics is not a language, nor does it contain one. The basic reason for this limitation is that mathematical concepts, being based on *abstract* sets and relations, do not pertain to nature. Mathematics does not allow to talk about nature nor about its basic property which we mentioned at the start of our escalation: the observation of motion and of change.

In his famous 1900 lecture in Paris, the german mathematician David Hilbert** had given a list of 23 great challenges facing mathematics. The sixth of Hilbert's problems was to find a mathematical treatment of the axioms of physics. Our adventure so far has shown that physics started with *circular* definitions which are not yet eliminated after 2500 years of investigations; most important is the definition of space-time with help of objects, and See page 98 the definition of objects with help of space and time. Physics is thus *not* modeled after mathematics, even if many physicists and mathematicians, such as Hilbert, would like it to be so. Physicists have to live with logical problems, and have to walk on unsure ground in order to achieve progress.

If physics were an axiomatic system, it would not contain contradictions, it would cease to be a language, and it would cease to describe nature. We will settle this topic in the third See page 624 part of this escalation.

In short, mathematics is *not* a language, the main reason being that one cannot use it to express the existence or the observation of motion. However, we can and indeed will use mathematical concepts in the description of nature wherever possible.

Physical concepts and patterns of nature

Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt.***

∗ The propositions of mathematics are equations, and therefore pseudo-propositions. A proposition of mathematics does not express a thought.

∗∗ David Hilbert (1862, Königsberg–1943, Göttingen), professor of mathematics in Göttingen, greatest mathematician of his time. He was central to many parts of mathematics, and also played an important role in the birth both of general relativity and of quantum theory. His books are still in print. His famous motto was: "Wir müssen wissen, wir werden wissen." The famous lecture is published e.g. in *Die Hilbertschen Probleme*, Akademische Verlagsgesellschsaft Geest & Portig, 1983. The lecture galvanized all of mathematics. (Despite efforts and promises of similar fame, *nobody* in the world had a similar overview of mathematics which allowed him or her to repeat the feat in the year 2000.) In his last decade he suffered the persecution of the Nazi regime, which eliminated Göttingen from the list of important science universities up to this day.

∗ ∗ ∗ *The limits of my language* are the limits of my world.

Ludwig Wittgenstein, *Tractatus*, 5.6

Der Satz ist ein Bild der Wirklichkeit.
Der Satz ist ein Modell der Wirklichkeit, so wie wir sie uns denken.*
Ludwig Wittgenstein, *Tractatus*, 4.01

In contrast to mathematics, physics does aim at being a language. Through the description of motion it aims to express *everything* observed, and in particular, all examples and possibilities of change.** Like any language, physics consists of concepts and sentences. In order to be able to express everything, it must aim to make few words about a lot of facts.*** Physicists are essentially *lazy* people: they try to minimize the effort in everything they are doing. The concepts in use today have been optimized by the combined effort of many people to be as practical, i.e. as powerful as possible. A concept is called *powerful* when it allows to express in a compact way a large amount of information, meaning that it can convey rapidly a large number of details about observations.

General statements about many examples of motion are called *rules* or *patterns*. In the past, one often said "laws govern nature", using an old and inappropriate ideology. A physical "law" is a way to talk about as much as possible with as few words as possible. Indeed, laws essentially are precise descriptions. Why precise? Because of their laziness, people want to say as much as possible in as few words as possible. When saying "laws govern nature" one actually means to say "being lazy, we describe observations with patterns." Laws are the epitome of laziness. *Making* laws is pure sloth. In fact, the correct expression is *patterns describe nature*.

Physicists have defined the laziness necessary for their field in much detail. In order to become a master of laziness, one needs to distinguish lazy patterns from those which are not, such as lies, beliefs, statements which are not about observations, and statements which are not about motion. We do this shortly.

The principle of extreme laziness is the origin, among others, of the use of numbers in physics. Observables are often best described with the help of numbers, because numbers allow easy and precise communication and classification. Length, velocity, angles, temperature, voltage, field strength, etc. are of this type. The notion of 'number', used in every measurement, is constructed, often unconsciously, from the notions of 'set' and 'relation',

∗ A proposition is a picture of reality. A proposition is a model of reality as we imagine it.

∗∗ All observations are about change or variation. The various types of change are studied by the various sciences; they are usually grouped in the three categories of *human sciences*, *formal sciences* and *natural sciences*. Among the latter, the oldest are astronomy and metallurgy. Then, with the increase of curiosity in early antiquity, came the natural science concerned with the topic of motion: *physics*. In the course of our walk it will become clear that this seemingly restrictive definition indeed covers the whole set of topics studied in physics. In particular it includes the more common definition of physics as the study of matter, its properties, its components, and their interactions.

∗∗∗ A particular, specific observation, i.e. a specific example of input shared by others, is called a *fact*, or in other contexts, an *event*. A striking and regularly observed fact is called a *phenomenon*, and a general observation made in many different situations is called a *(physical) principle*. (Often, when a concept is introduced which is used with other meaning in other fields, in this walk it is preceded by the qualifier 'physical' or 'mathematical' in between brackets.) Actions performed towards the aim of collecting observations are called *experiments*. The concept of experiment became established in the sixteenth century; in the evolution of a child, it is best be compared to that activity which has the same aim of collecting experiences: *play*.

as shown above. Apart from the notion of number, other concepts are regularly defined to allow fast and compact communication of the "laws" of nature; all are "abbreviation tools." In this sense, the statement "the level of the Kac-Moody algebra of the lagrangian of the heterotic superstring model is equal to one" contains precise information, explainable to everybody but which would take dozens of pages if one would express it only using the terms 'set' and 'relation.' In short, the *precision* common in physics results from its *quest for laziness*.

Are physical concepts discovered or created?

Das logische Bild der Tatsachen ist der Gedanke.*
Ludwig Wittgenstein, *Tractatus*, 3

The question is often rephrased as: are physical concepts free of beliefs, tastes, or of choices? The question has been discussed so much that in the meantime it even appears in Hollywood movies. A short summary.

Creation, in contrast to discovery, implies free choice between many alternative possibilities. The chosen alternative would then be due to the beliefs or tastes implied in any created concept. In physics (and in obvious contrast to other, more ideological fields), one knows that different physical descriptions of observations are either equivalent, or, in the opposite case, partly imprecise or even wrong. A description of observations is thus essentially unique: choices are only apparent. There is no freedom in the definition of physical concepts, except for equivalent reformulations, in strong contrast to the case of any creative activity.

If two different concepts could be used to describe the same aspect of observations, they must be equivalent, even if the relation that leads to the equivalence is not clear immediately. (By the way, one knows no physical concept which can be called "created" instead of discovered.) In fact, the requirement that people with different standpoints observing the same event form equivalent descriptions lies at the very basis of physics. It forms the symmetry requirements of nature: examples are the principle of relativity and the principle of gauge invariance. In short, the requirement of viewpoint independence makes the free choice of concepts a logical impossibility.

The conclusion that concepts describing observations are discovered is also reached independently in the field of linguistics by the mentioned research on semantic primitives,** in the field of psychology by the observations on the formation of the concepts in the development of young children, and in the field of ethology by the observations of animal development, especially in the case of mammals. All three fields have observed in detail how the interactions between an individuum and its environment lead to concepts, of which in particular the most basic ones, such as space, time, object, interaction etc., are common across the sexes, cultures, races, and across many animal species populating the world. Curiosity and the way nature works leads to the same concepts for all people and even the animals; the world offers only one possibility, without room for imagination. Thinking the opposite is a belief – often a useful exercise, but never successful.

* A logical picture of facts is a thought.
** Anna Wierzbicka concludes that her research clearly indicates that semantic primitives are *discovered*, in
Ref. 18 particular that they are deduced from the fundamentals of human experience, and not invented.

Physical concepts are classifications of observations. The activity of classification itself follows the patterns of nature; it is a mechanical process which also machines can perform. This means that any distinction, i.e. any statement that A is different from B, is a theory free statement. No belief system is necessary to distinguish different entities in nature. Physicists *can* be replaced by machines. The end of our escalation will confirm this point.

As mentioned already, physical concepts are made up in a way to describe observations as succinctly and as accurately as possible. They are formed with the aim to have the largest possible amount of understanding with the smallest possible amount of effort.

In summary, we found that physical concepts are the same for everybody and are free of beliefs: they are first of all *boring*. Moreover, as they could stem from machines instead of people, they are *born of laziness*. Evidently they are *not* discovered. Having handled the case of physical concepts, let us turn to physical statements. The situation is somewhat similar: physical statements must be lazy, arrogant and boring. Let us see why.

Wo der Glaube anfängt, hört die Wissenschaft auf.*
Ernst Haeckel, *Natürliche Schöpfungsgeschichte*, 1879.

How do we find physical patterns and rules?

Grau, treuer Freund, ist alle Theorie,
Und grün des Lebens goldner Baum.**
Goethe (1749–1832), *Faust*.

Physics is usually presented as an objective science, but I notice that physics changes and the world stays the same, so there must be something subjective about physics.
Richard Bandler

Progressing through the study of motion and through this escalation reflects a young child's way of life, which follows the simple program on the left:

Normal description	Lobbyist description
Curiosity	**Scientific method**
1. look around a lot	interact with the world
2. don't believe anything told	forget authority
3. choose something particularly interesting and explore it yourself	observe
4. make up your own mind, and try to describe precisely what you saw	use reason, build hypotheses
5. check if you can describe also other, similar situations in the same way	analyze hypothesis
6. increase the precision of observation until the checks either fail or are complete	perform experiments until hypothesis is falsified or established
7. depending on the case, continue with step 4 or 1 of a new round.	ask for more money

* Where belief starts, science ends.
** Grey, dear friend, is all theory, and green the golden tree of life.

Adult scientists do not have much more to add, except the more fashionable terms on the right, plus several specialized professions to make money from them. The experts of step 7 are variously called lobbyists or professors; instead of calling this program 'curiosity', they call it the 'scientific method.' They mostly talk. Physics being the talk about motion,* and motion being a vast topic, many people specialize in this step.

The experts of step 6 are called *experimental physicists* or simply *experimentalists*, a term derived from the latin 'experiri', meaning 'to try out'. Most of them are part of the category of 'graduate students'. The experts of steps 5 and 4 are called *theoretical physicists* or simply *theoreticians*. This is a rather modern term; for example, the first professors of theoretical physics were appointed only around the start of the twentieth century. The term is derived from the greek θεωρία meaning 'observation, contemplation'. Finally, those people focussed on steps 1 to 3, who get others to work on steps 4 to 6, are called *geniuses*.

But obviously the most important point is hidden in step 6: how do all these people know whether their checks fail? How do they recognize truth? In other words,

All professions are conspiracies against laymen.
George Bernard Shaw

What is a lie?

Get your facts straight, and then you can distort them at you leisure.
Mark Twain (1835–1910)

The pure truth is always a lie.
Bert Hellinger

Lies are useful statements, as everybody learns during youth. One reason they are useful is because one can draw any imaginable conclusion from them. A well-known discussion between two Cambridge professors early in the twentieth century makes the point. McTaggart asked: "If $2+2=5$, how can you prove that I am the pope?" Hardy: "If $2+2=5$, then $4=5$; subtract 3; then $1=2$; but McTaggart and the pope are two; therefore McTaggart and the pope are one." As already noted a long time ago, ex falso quodlibet. From what is wrong, anything imaginable can be deduced. It is true that in our escalation we need to build on previously deduced results and that our trip could not be completed if we had a false statement somewhere in our chain of arguments. But lying is such an important activity that one should learn to perform it properly.

There are various stages in the art of lying. Animals have been shown to deceive their
Ref. 4 kin. Children start just before their third birthday, by hiding experiences. Adults cheat on taxes. And some intellectuals may claim that truth does not exist.

∗ Several sciences have the term 'talk' as part of their name, namely all those whose name finishes in '-logy', such as e.g. biology. The ending stems from ancient greek and is deduced from ληγηιν meaning 'to say, to talk'. Physics as science of motion could thus be called 'kinesiology' from κίνησις, meaning 'motion'; but for historical reasons this term has a different meaning, namely the study of human muscular activity. The term 'physics' is either derived from the greek φύσικη (τέχνη is understood) meaning 'the art of nature', or from the title of Aristoteles' works τά φυσικά meaning 'natural things'. Both expressions are derived from φύσις, meaning 'nature'.

However, in most countries, everybody must know what "truth" is, since in court for example, telling the opposite can lead to a prison sentence. And courts are full of experts in lie detection. So if one lies in court, one better does it properly. For a court, a lie is a statement in contrast with observations.* The truth of a statement is thus checked by observation. The check itself is sometimes called the *proof* of the statement. For courts, as for physics, truth is thus the correspondence with facts. And facts are shared observations. A good lie is thus a lie whose contrast with shared observations is hard to discover.

The first way to lie is to put the emphasis on the sharedness only. Populists and polemics do that regularly. ("Every foreigner is a danger for the values of our country.") Since almost any imaginable opinion, however weird, is held by some group, one can always claim it as true. Unfortunately, it is not a secret that ideas get shared also because they are fashionable, or imposed, or opposed to somebody generally disliked, such as some sibling in the family – remember Cassandra.** For a good lie one thus needs more than sharedness, more than *intersubjectivity*. A good lie should be, like a true statement, really independent of the listener or the observer, and in particular independent of their age, their sex, their education, their civilization, or the group they belong to. For example, it is especially hard – but not impossible – to lie with mathematics. The reason is that the basic concepts of mathematics, be they 'set', 'relation', or 'number' are taken from observation and are intersubjective, so that statements about them are easily checked. Usual lies thus avoid mathematics.

Secondly, a good lie should avoid statements about observations, and use *interpretations* instead. For example, some people like to talk about other universes, which implies talking about one's imagination, not about observations. One has to avoid however, to fall in the opposite extreme, namely to make statements which are meaningless; the most destructive comment one can make about a statement is the one used by the austrian physicist Wolfgang Pauli (1900, Wien–1958, Zürich): that is "not even wrong".

Thirdly, a good lie doesn't care about observations, only about imagination. Only truth needs to be *empirical*, to distinguish it from *speculative* statements. If one wants to lie well even with empirical statements, one needs to distinguish two cases. There are two types of empirical statements: *specific* statements and *universal* statements. For example, "On the 31st of August 1960 I saw a green swan swimming on the northern shore of the lake of Varese" is specific, whereas "All ravens are black" is universal, since it contains the term 'all'. Universal statements are also called *theories*. There is a well-known difference between the two, which is important for lying well: specific statements cannot be falsified, they are only verifiable, and universal statements cannot be verified, they are only falsifiable. Ref. 31

Why is this so? Universal statements such as "the speed of light is constant" cannot be tested for *all* possible cases. (Note that if they could, they would not be universal statements, but just a list of specific ones.) However, they can be reversed by a counterexample. Another

* Statements not yet checked are variously called *speculations*, *conjectures*, *hypotheses*, or – wrongly – simply *theses*. Statements which are in correspondence with observations are called *correct* or *true*, otherwise *wrong*, *false*, or *lies*.

** The implications of birth order on creativity in science and on acceptance of new ideas has been studied in the fascinating book by Frank J. Sulloway, *Born to rebel – birth order, family dynamics, and creative lives*, Panthon Books, 1996. This exceptional book tells the result of a life-long study correlating the personal situation in the family of thousands of people and their openness to about twenty revolutions in the recent history. The book also includes a test in which one can deduce one own propensity to rebel, on a scale from 0 to 100%. Darwin scores 96% on that scale.

example of the universal type is: 'Apples fall upwards'. Since it is falsified by an observation conducted by Newton several centuries ago, or by everyday experience, it qualifies as an (easily detectable) lie. In general therefore, lying by stating the opposite of theories is usually unsuccessful. If somebody insists on doing so, the lie becomes a *superstition*, a *belief*, a *prejudice* or a *doctrine*. Those are the low points in the art of lying. A famous case of insistence on a lie is that of the colleagues of Galileo, who are said to have refused to look through his telescope to be convinced that Jupiter has moons, an observation which would have shaken their statement and belief that everything turns around the earth. Obviously these astronomers were amateurs in the art of lying. A good universal lie is one whose counterexample is not easily spotted.

There should be no insistence on lies in physics. Unfortunately, classical physics is full of them. We try to get rid of the remaining ones during the rest of our walk.

On the other hand, lying with specific statements is much easier. ("I can't remember.") Even a specific statement such as "yesterday the moon was green, cubic, and smelled of cheese" can never be completely falsified: there is no way to show with absolute certainty that this is wrong. The only thing one can do is to check whether the statement is compatible with other observations, such as whether the different shape affected the tides as expected, whether the smell can be found in air collected that day, etc. A good specific lie is thus not in contrast with other observations.*

By the way, universal and specific statements are connected: the *opposite* of a universal statement is always a specific statement, and vice versa. For example, the opposite of the general statement "apples fall upwards" namely "some apples fall downwards" is specific.

In other words, courts and philosophers disagree. Courts have no issue with calling theories true, and specific statements lies. Many philosophers avoid this. For example, the statement 'Ill-tempered gaseous vertebrates do not exist' is a statement of the universal type. If a universal statement is in agreement with observations, and if it is falsifiable, courts call it *true*. The opposite, namely the statement: "ill-tempered gaseous vertebrates do exist" is of the *specific* type, since it means "Person X has observed a ill-tempered gaseous vertebrate in some place Y at some time Z." To verify it, one needs a record of the event. If such records, for example by photographs, witnesses, etc., do not exist, and if the statement *can* be falsified by other observations, courts call the specific statement a *lie*. Even though these are the rules for everyday life and for the law, there is no agreement between philosophers and scientists that this is acceptable. Intellectuals are extremely careful, mainly because many of them have lost their life by exposing various lies too openly.

* It is often difficult or tedious to verify statements from the past, and the difficulty increases with the distance in time. That is why people can insist on the occurrence of events which are supposed to be exceptions to the patterns of nature ("miracles"). Since the advent of rapid means of communication these checks are becoming more and more easy, and there do not seem to be many miracles left. This happened in the miracle place Lourdes in France, where even though the number of visitors is much higher than in the past, no miracles have been seen in decades.

In fact, most modern miracles are kept alive only by consciously eschewing checks, such as the supposed yearly liquefaction of blood in Napoli, the milk supposedly drunk by statues, the supposed healers in television evangelism, etc. Nevertheless, many organizations make money from the difficulty to falsify specific statements. Ref. 32 When the british princess Diana died in a car crash in 1997, even though the events were investigated in extreme detail, the scandal press could go on almost without end about the "mysteries" of the accident.

In short, specific lies, like all specific statements, can never be falsified with certainty. That makes them so popular. Children learn them first. ("I haven't eaten the jam.") General lies, like all general statements, can always be corroborated by examples. That is the reason for the success of ideologies. But the criteria for recognizing lies have become so commonplace that beliefs and lies try all to keep up. It became fashionable to use expressions such as "scientific fact" – there are no non-scientific facts –, or "scientifically proven" – observations cannot be proven otherwise – and similar empty phrases. These are not really good lies, since whenever one encounters sentences starting with "science says ..." or "science and religion do ...", replacing 'science' by 'knowledge' or 'experience' is an efficient way to check whether such statements are to be taken seriously or not.*

An important aspect makes lies more attractive than true statements, be they universal or specific. True statements require the author to stick his neck out to criticism. If one doesn't stick the neck out, it can't be a lie, nor a observation, nor a theory. Lying does make one vulnerable. For this reason, theories are often *arrogant*, *provoking* and at the same time they have to be *vulnerable*. Theories thus resemble a beautiful woman: fragile and haughty at the same time.** On the other hand, specific statements about observations must be boring and rock-solid. They are opposite in character to theories. Reading books which developed daring theories, such as Darwin's *The origin of the species*, one easily feels the stark contrast between the numerous boring and solid facts and the arrogant and vulnerable theory that he deduced.

But public check is not always reliable. For example, collective imagination played a large role when scientists were talking about 'aether', 'UFOs', 'creation science', and 'cold fusion'. Nevertheless, an important aspect of any lie is to make as little *public* statements as possible, so that others can check as little as possible.(For anybody sending corrections of mistakes in this text, the author provides a small reward.) In the heated frenzy of research, it happens to everybody to make statements which are not based on observations. The search of statements without these properties is sometimes called the *scientific method*. But a good lie is always well prepared and told on purpose; accidental lies are frowned upon by experts.

In short, a good *general lie* seems humble and invulnerable, such as "People have free will", and a good *specific lie* is often surprising and shaky, such as "Yesterday I drowned". Feelings can thus be a criterion to judge the quality of lies, if one pays careful attention to the type of statement. A number of common lies are discussed later on in this intermezzo.

To sum up, the central point in the art of lying without being caught is simple: do not tell details. Be *vague*. All methods to get to the bottom of any is to ask for details, for *precision*. For any statement, its degree of precision is the way to gauge the degree that somebody

∗ Just to clarify the vocabulary usage of this text, *religion* is spirituality plus a varying degree of power abuse. The mixture depends on each person's history, background, and environment. *Spirituality* is the open participation in the whole of nature. Most people with a passion for physics are spiritual.

∗∗ A set of not yet falsified patterns of observations on the same topic is called a *(physical) theory*. The term 'theory' will always be used in this sense in this walk, i.e. with the meaning 'set of correct general statements'. This use results from its greek origin: 'theoria' means 'observation'; its original meaning, 'passionate and emphatic contemplation', summarizes all of physics in a single word. ('Theory', like 'theater', is formed from the root θέ, meaning 'the act of contemplating'.) Sometimes however, the term 'theory' is used – confusing it with 'thesis' – with the meaning of 'conjecture', as in "your theory is wrong", sometimes with the meaning of 'model', as in "Chern-Simons" theory and sometimes with the meaning of 'standard procedure', as in "perturbation theory". These incorrect uses are avoided here.

sticks his neck out. The more precision one demands, the more fragile a statement is, and the more likely the fault is found out, if there is one. This is the main reason we chose the increase in precision as guide for our escalation.The same method is used in trials. To find out the truth, investigators typically ask all the people involved a large number of questions, until as many *details* as possible come to light. When one has collected enough details, when the precision has become high enough, the situation becomes clear. Telling good lies is much harder than telling the truth; it requires an excellent imagination.

Truth is an abyss.
Democritos

To teach superstitions as truth is a most terrible thing.
Hypatia of Alexandria (ca. 355–415)

Absolute truth is what scientists claim it to be at the end of their life.
Charles Peirce (1839–1914)

Is this statement true?

Truth is a rhetorical concept.
Paul Feyerabend (1924, Vienna–1994, Zürich)

Not all statements can be divided into true and false. There even are such statements in mathematics, such as the continuum hypothesis. This hypothesis is undecidable because it makes a statement which depends on the precise meaning of the term 'set'; in standard mathematical usage the term is not defined precisely enough that a truth value can be assigned to the continuum hypothesis. In short, statements can be undecidable because the concepts contained in them are not sharply defined.

Statements can also be undecidable for other reasons. Curious phrases such as "This statement is not true" illustrate the situation. The well-known austrian-american logician Kurt Gödel (1906–1978) has even devised a general way to construct such statements in the domain of logic and mathematics. The different variations of these *self-referential* statements, especially popular both in the field of logic and computer science, have captured a large public.* One can construct similarly undecidable statements with terms such as 'calculable', 'provable' and 'deducible'. Ref. 33

In fact, self-referential statements are undecidable because they are meaningless. If the usual definition of 'true', namely correspondence with facts, is substituted into the sentence "This statement is not true", one quickly sees that it has no meaningful content. The most famous meaningless sentence of them all was constructed by the linguist Noam Chomsky:

"*Colorless green ideas sleep furiously.*"

Ref. 10 It is often used as an example for the language processing properties of the brain. But

* A general introduction is given in the beautiful books by Raymond Smullyan, *Satan, Cantor, and Infinity*, 1992, *What is the name of this book? - The riddle of Dracula and other logical puzzles*, 1986, and *The lady or the tiger?*, 1982.

nobody in his right mind elevates it to the status of a paradox and writes philosophical discussions about it. To do that with the title of this section is a similar waste of energy.

The main reason for the popular success of self-reference is the difficulty to perceive the lack of meaning.* A good example is the statement:

This statement is false or you are an angel.

One can actually deduce from it that 'you are an angel.' Can you see how? If you want, you can change the second half and get even more interesting statements. Such examples show that statements referring to themselves have to be treated with great care when they are investigated. Challenge

In physics, in the other natural sciences, and in legal trials these problems do not appear, since self-referential statements are not used. In fact, the work by the logicians confirms, often rather spectacularly, that there is no way to extend the term 'truth' beyond the definition of 'correspondence with facts.'

Ein Satz kann unmöglich von sich selbst
aussagen, daß er wahr ist.**
Ludwig Wittgenstein, *Tractatus*, 4.442

Observations

Knowledge is a sophisticated statement of ignorance.
Attributed to Karl Popper

The collection of a large number of true statements about a type of observations, i.e. of a large number of facts, is called *knowledge*. In case that the domain of observations is sufficiently extended, one speaks of a *science*. A *scientist* is thus somebody who collects knowledge.*** We fond above that an observation is classified input sticking into memory of several people. Since there is a lot of motion around, the description of all these observations is a large piece of work. As for every large task, the use of appropriate tools determines to a large extent the degree of success one can achieve. These tools, in physics and in all other sciences, fall in three groups: tools for the collection of observations, tools to communicate observations, and tools to communicate relations between observations. The latter group has been already discussed in the section on language and on mathematics. We just touch the other two. Ref. 35

* A well-known victim of this difficulty is Paulus of Tarsus. The paradox of the cretan poet *Epimenedes* (6th century B.C.) who said "All cretans lie" is too difficult for the humor impaired Paulus, who in his letter to Titus (chapter 1, verses 12 and 13, in the christian bible) calls Epimenedes a "prophet", adds some racist comments, Ref. 34 and states that this "testimony" is true. But wait; there is a final twist to this story. The sentence is not a paradox at all; a truth value can actually be ascribed to it, because the statement is not really self-referential. Can you Challenge see it? The only *genuine* paradox is "I am lying", to which no truth value can be ascribed indeed.

** It is quite impossible for a proposition to state that it itself is true.

*** The term 'scientist' is a misnomer peculiar to the english language. Properly speaking, a 'scientist' is a follower of *scientism*, a extremist philosophical school which tried to resolve all problems through science. Therefore some sects have the term in their name. Since the english language did not have a shorter term to designate 'scientific persons', as they used to be called before, the term 'scientist' came into use, first in the United States, from the 18th century on. Nowadays the term is used in all english-speaking countries – but not outside them, fortunately.

Have enough observations been recorded?

Every generation is inclined to define 'the end of physics'
as coincident with the end of their scientific contributions.
Julian Schwinger*

Physics is an experimental science; it rests on the collection of observations. To realize this task effectively, all sorts of *instruments*, i.e. tools which facilitate observations, have been developed and built. Microscopes, telescopes, oscilloscopes, as well as thermometers, hygrometers, manometers, pyrometers, spectrometers and many others are familiar examples. The precision of many of these tools is continuously improved even today; their production is a sizable part of modern industrial production, examples being electrical measurement apparatuses and diagnostic tools for medicine, chemistry, and biology. Instruments can be as small as a tip of a few tungsten atoms to produce electron beams with a few volt, and as big as 27 km in circumference, producing electron beams with over 100 GV effective accelerating voltage. People have built instruments which contain the coldest known spot in the universe and instruments which can measure length variations much smaller than a proton diameter for kilometer long distances. Instruments have been put inside the earth, on the moon, on several planets, and sent outside the solar system.

In this walk, instruments are not described; many good textbooks on this topic are available. Most observations collected with them are not mentioned here. The most important results in physics are recorded in standard publications, such as the Landolt-Börnstein and the physics journals (appendix E gives a general overview of information sources).

Ref. 37, 38
Ref. 39

Will there be significant new future observations in the domain of the fundamentals of motion? At present, *in this specific domain*, even though the number of physicists and publications is at an all-time high, the number of new discoveries has diminished for many years and is now rather small; the sophistication and investment necessary for new results has become extremely high; in many cases, measurement instruments have achieved the limits of technology, of budgets, or even those given by nature; the number of new experiments showing no deviation from theoretical predictions is increasing steadily; historical papers trying to enliven boring or stuck fields of enquiry are increasing; claims of new effects which turn out to be false, due to measurement errors, self-deceit or even to fraud have become so frequent that scepticism has become the natural response. Although in many domains of science, including physics, discoveries are still expected, on the fundamentals of motion the arguments just presented seem to give new observations only a remote possibility. The task of collecting observations on motion seems to be *completed* (though not on other topics of physics). And indeed, all observations described here have been completed before the end of the twentieth century. We are not too early with out walk.

Are all observables known?

Scientists have odious manners, except when you prop up their theory;
then you can borrow money from them.
Mark Twain (1835–1910)

* Julian Seymour Schwinger (1918–1994), US-american enfant prodige, famous for his clear thinking and his excellent lectures, developer of quantum electrodynamics, winner of the 1965 Nobel prize in physics together with Tomonaga and Feynman, and thesis advisor to many famous physicists.

Ref. 36

The most practical way to communicate observations has been developed already a long time ago: the measurement. A measurement allows effective communication of an observation to other times and places. This is not always as trivial as it sounds; in the middle ages for example, people were unable to compare precisely the "coldness" of winters of two different years! Only the invention of the thermometer provided a reliable solution to this requirement. A *measurement* is thus the classification of an observation into a standard set of observations; in simple words, a measurement is a *comparison with a standard*. This definition of a measurement is the most precise and practical, and has therefore been universally adopted. For example, when the length of a house is measured, one classifies this aspect of the house into a certain set of standard lengths, namely the set of lengths defined by multiples of a unit. A *unit* is the abstract name of the standard for a certain observable. Numbers and units allow the most precise and most effective communication of measurement results.

For all measurable quantities, practical standard units and measurement methods have been defined; the main ones are listed and defined in appendix B. All units are derived from a few fundamental ones; this is ultimately due to the limited number of our senses: length, time, and mass are related to sight, hearing, and touch.

We call the different measurable aspects of a system its *observables*. Most observables, such as size, speed, position etc. can be described by numbers, and in this case they are *quantities*, i.e. multiples of some standard unit. Observables are usually abbreviated by *(mathematical) symbols*, usually letters from some alphabet. For example, the symbol c commonly specifies the velocity of light. For most observables, standard symbols have been defined by international bodies.* The symbols for those observables describing the state of an object are also called *variables*. Variables on which other observables depend are often called *parameters*. (A parameter is a variable constant.) For example, the speed of light c is a constant, the position x a variable, the temperature T often a parameter, on which e.g. the length of an object can depend. Note that not all observables are quantities; in particular, parities are not multiples of any unit.

Today the task of defining tools for the communication of observations can be considered *complete*. (For quantities, this is surely correct; for parity-type observables there could be a few examples to be discovered.) This is a simple and strong statement. Even the BIPM, the Bureau International des Poids et Mésures, has stopped to add new units.**

As a note, one can rank the greatness of a physicist by the number of observables he has introduced. Even a great scientist like Einstein, who has discovered many "laws" of nature, has introduced only one new observable, namely the metric tensor for the description of gravity. Following this criterion – as well as several others – Maxwell is the most important physicist, having introduced electric and magnetic fields, the vector potential, and several other material dependent observables. For Heisenberg, Dirac and Schrödinger, the wavefunction describing electron motion could be counted as half an observable (in fact it

* All mathematical symbols used in this walk, together with the alphabets from which they are taken, are listed in appendix A on notation. They follow international standards whenever they are defined. The standard symbols of the physical quantities, as defined by the International Standards Organisation (ISO), the International Union of Pure and Applied Physics (IUPAP) and the International Union of Pure and Applied Chemistry (IUPAC), can be found for example in the *bible*, i.e. the *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, 1992.

** The last, the *katal*, was introduced in 1999. All units are explained in appendix B.

is a quantity necessary to calculate measurement results, but not itself an observable). By the way, even introducing *any* word which is taken up by others is the a rare event; ‘gas’, ‘entropy’ and only a few others are such examples. It was always much more difficult to discover an observable than to discover a “law”; usually, observables are developed by many people together. This is shown from a simple aspect of modern science: many “laws” bear people’s names, but almost no observables.

The list of observables necessary to describe nature being complete, does this mean that automatically one knows all the patterns or rules of nature? No; in the history of physics, observables have usually been defined and measured long *before* the precise rules connecting them were found. For example, all observables used in the description of motion itself, such as time, position and its derivatives, momentum, energy, and all the thermodynamic quantities have been defined during or before the nineteenth century, whereas the most precise versions of the patterns or “laws” of nature connecting them, special relativity and non-equilibrium thermodynamics, have been found only in the twentieth century. The same is true for all observables connected to the electromagnetic interaction, and all those connected to the gravitational interaction, except perhaps the metric tensor. The respective patterns of nature, quantum electrodynamics and general relativity, have been discovered long after the corresponding observables. The observables discovered last are the fields of the strong and of the weak nuclear interactions. Also in this case the patterns of nature were formulated much later.*

Do observations take time?

An observation is an interaction with some part of nature leading to the production of a record, such as a memory in the brain, data on a tape, ink on paper, or any other fixed process applied to a support. The irreversible interaction process is often called *writing* the record. Obviously, writing takes a certain amount of time; zero interaction time would give no record at all. Therefore any recording device, also our brain, always records some *time average* of the observation, however short it may be.

What we call a fixed image, be it a mental image or a photograph, always is the time average of a moving situation. Without time averaging, we would not have any fixed memories. On the other hand, the blurring any time averages introduces, hides the details; and in our quest for precision, at a certain moment, these details are bound to become important. The discovery of these details will begin in the second part of the walk, the one centered on quantum theory. In the third part of our escalation we will discover that there is a shortest possible averaging time, and that observations of that short duration show so many details that one cannot even distinguish particles from empty space. All our concepts of everyday life appear only after relatively long time averages. To describe nature without any averaging is one of the challenges of the final part of or escalation.

* Can one talk about observations at all? It is many a philosopher’s hobby to discuss whether there actually is an example for an ‘Elementarsatz’ mentioned by Wittgenstein in his Tractatus. There seems to be at least one which fits: *Differences exist.* It is a simple sentence; at the end of our walk, it will play a central role.

Is induction a problem in physics?

Nur *gesetzmäßige* Zusammenhänge sind *denkbar*.*
Ludwig Wittgenstein, *Tractatus*, 6.361

There is a tradition of opposition between adherents of induction and of deduction. In my view it would be just as sensible for the two ends of a worm to quarrel.
Alfred North Whitehead (1861–1947)

Induction is the usual term used for the act of taking, from a small and finite number of experiments, general conclusions about the outcome of *all* possible experiments performed in other places, or at other times. In a sense, it is the technical term for the sticking out of one's neck that is necessary in every scientific statement. Induction has been a major topic of discussion for science commentators. Frequently one finds the remark that knowledge in general, and physics in particular, relies on induction for its statements. Following some, induction is a type of hidden belief underlying all sciences and at the same time in contrast with it.

To avoid any waste of energy, we make only a few remarks. The first point can be deduced from a simple experiment. Try to convince an induction critic to put his hand into fire. Nobody who calls induction a belief will conclude from a few unfortunate experiences in the past that such an act will also be dangerous in the future... In short, somehow induction works.

A second point is that physical universal statements are always clearly stated; they are never hidden. The refusal to put the hand into fire is a consequence of the invariance of observations under time and space translations. Indeed, all-statements of this type form the very basis of physics. However, no physical statement is a belief only because it is universal; it always remains open to experimental checks. Physical induction is not a hidden method of argumentation, it is an explicit part of experimental statements. In fact, the complete list of “inductive” statements used in physics is given in the table on page 122. These statements are so important that they have been given a special name: they are called *symmetries*. The table lists all known symmetries of nature; in other words, it lists all inductive statements used in physics.

Perhaps the best argument for the use of induction is that there is no way to avoid it when thinking. There is no way to think or to talk without using concepts, i.e. without assuming that most objects or entities have the same properties over time. The only sentences which do not use induction, the sentences of logic, do not have any content (Tractatus, 6.11). Without Ref. 35
induction, one cannot classify observations at all! Evolution has given us memory and a Challenge
brain because induction works. To criticize induction is not to criticize natural sciences, it is to criticize the use of thought in general. One should never take too seriously people who themselves do what they criticize in others; sporadically pointing out the ridicule of this endeavour is just the right amount of attention they deserve.

The topic could be concluded here, were it not for some interesting developments in modern physics which put two more nails in the coffin of arguments against induction. First of

* Only connexions that are *subject to law* are *thinkable*.

all, whenever in physics one makes statements about all experiments, all times, all velocities, etc., such statements are about a *finite number* of cases. We know today that infinities, both in size and in number, do not occur in nature. The infinite number of cases appearing in statements in classical physics and in quantum mechanics are apparent, not real, and due to human simplifications and approximations. Statements that a certain experiment gives the same result 'everywhere' or that a given equation is correct for 'all times', always encompass only a *finite* number of examples. A lot of otherwise often instinctive repulsion to such statements is avoided in this way. In the sciences, as well as in this book, 'all' *never* means an infinite number of cases.

Finally, it is well known that taking conclusions from a few cases to many is false when the few cases are independent of each other. However, it is correct if the cases are interdependent. From the fact that somebody found a penny on the street on two subsequent months, he cannot follow that he will find one the coming month. Induction is only correct if one knows that all cases have similar behaviour, e.g. because they follow from the same origin. For example, if a neighbour with a hole in his pocket carries his salary across that street once a month, and the hole always opens at that point because of the beginning of stairs, then the conclusion would be correct. It turns out that the results of modern physics encountered in the third part of our walk show that all situations in nature are indeed interdependent, and thus prove in detail that what is called 'induction' is in fact a logically correct conclusion.

The quest for precision and its implications

Der Zweck der Philosophie ist die logische Klärung der Gedanken.*
Ludwig Wittgenstein, *Tractatus*, 4.112

To talk well about motion means to talk precisely. Precision requires avoiding three common mistakes in the description of nature:

Concepts should never have a contradiction built into their definition. For example, any phenomenon occurring in nature evidently is a 'natural' phenomenon; therefore, talking either about "supernatural" phenomena or about "unnatural" phenomena is a mistake that nobody interested in motion should let go by unchallenged; the terms contain a logical contradiction. Naturally, *all* observations are natural. By the way, there is a reward of more than a million dollars for anybody showing the opposite. In over twenty years, nobody has yet been able to collect it. Ref. 40

Concepts should not have unclear or constantly changing definitions. Their content and their limits must be kept constant and explicit. This mistake is often encountered when one talks to crackpots or to populist politicians, and distinguishes them from more reliable thinkers. Also physicists fall into the trap; for example, there is of course only a *single* (physical) universe, as even the name says. Talking about more than one universe is a increasingly frequent error of thought. Ref. 41

Concepts should not be used outside their domain of application. Everybody has succumbed to the temptation to transfer results from physics to philosophy without checking the content. An example is the question: "Why do particles follow the laws of nature?" The question is due to a misunderstanding of the term "law of nature" and to a confusion with the laws of the state. Remembering that "law of nature" simply means 'pattern', 'property' or 'description of behaviour', and rephrasing the question correctly as 'Why do particles behave in the way we describe their behaviour?' one recognizes its senselessness.

In the course of our walk, we will often be tempted by these three mistakes. A few such situations follow, together with the way to avoid them.

Consistency is the last refuge of the unimaginative.
Oscar Wilde (1854, Dublin–1900, Paris)

What are interactions?

The whole is always more than the sum of its parts.
Aristotle, *Metaphysica*, 10f–1045a.

In the physical description of nature, the whole is always *more* than the sum of its parts. Actually, the difference between the whole and the sum of its parts is so important that it gets a special name: the *interaction* between the parts. For example, the energy of the whole minus the sum of the energies of its parts is called the energy of interaction. In fact, the study of interactions is the main topic of physics. In other words, physics is concerned *primarily* with the difference between the parts and the whole, contrary to what is often written by bad journalists or other sloppy thinkers.

* Philosophy aims at the logical clarification of thoughts.

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Challenge - Did you find any mistakes?

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

Note that the term 'inter-action' is based on the general observation that anything which affects other things is in turn affected by them; interactions are *reciprocal*. For example, if a body changes the momentum of a second body, then the second changes the momentum of the first by the same (negative) amount. This reciprocity of interactions is the reason that anybody using the term is a heretic for monotheistic religions, since reciprocity implicitly denies the immutability of the deity.

Remembering the definition of interaction also settles the frequently heard question on whether in nature there are "emergent" properties, i.e. properties of systems which cannot be deduced from the properties of their parts *and* of their interactions. The idea of "emergent" properties is a product of minds with a restricted horizon, unable to see the richness of consequences that general principles can produce. See page 141 Ref. 42

The simple definition of interaction just given, so boring it sounds, leads to surprising conclusions. Take the atomic idea of Democritos in its modern form: nature is made of vacuum and of particles.

The first consequence is the *paradox of incomplete description*: Experiments show that there are interactions between vacuum and particles. But interactions are differences between parts and the whole, in this case therefore between vacuum and particles on one hand, and the whole on the other. We thus have deduced that nature is not made of vacuum and particles alone.

The second consequence is the *paradox of overcomplete description*: Experiments also show that interactions happen through exchange of particles. But we have counted particles already as basic building blocks. Does this mean that the description of nature by vacuum and particles is an overdescription, counting things double? Challenge

We will resolve both paradoxes in the third part of the escalation. See page 598

What is existence?

You know what I like most?
Rhetorical questions.

Assume a friend tells you "I have seen a *grampus* today!" You would naturally ask how it looks. What do we expect from the answer? We expect something like "It's an animal with a certain number of heads similar to a X, attached to a body like a Y, with wings like a Z, it make noises like a U and it felt like a V" – the letters denoting some other animal or object. Generally speaking, in the case of an object, this scene from Darwin's voyage to South America shows that in order to talk to each other, one first of all needs certain basic, common concepts ('animal', 'head', 'wing', etc.). In addition, for the definition of a new entity we need a characterization of its parts ('size', 'colour'), of the way these parts relate to each other, and of the way the whole interacts to the outside world ('feel', 'sound'). In other words, for an object to exist, one must be able to give a list of relations with the outside world. In short, an object exists if one can interact with it. (Is observation sufficient to determine existence?) Ref. 43 Challenge

For an abstract concept, such as 'time' or 'superstring', the definition of existence has to be refined only marginally: *(physical) existence is the ability to describe interactions*. This definition applies to trees, time, virtual particles, imaginary numbers, entropy, and many others. It is thus pointless to discuss whether a physical concept "exists" or whether it is "only" an abstraction used as a tool for descriptions of observations. The two possibilities coincide. The point of dispute can only be whether the description provided by a concept is or is not *precise*.

Do things exist?

Wer Wissenschaft und Kunst besitzt,
Hat auch Religion;
Wer jene beiden nicht besitzt,
Der habe Religion.*
J.W. von Goethe, *Zahme Xenien, IX*

Using the above definition of existence, the question becomes either trivial or imprecise. It is trivial in the sense that things necessarily exist if they describe observations, since they were defined that way. But perhaps the questioner meant to ask: Does reality exist independently of the observer?

Using the above, this question can be rephrased: Do the things one observes exist independently of observation? After thousands of years of extensive discussion by professional philosophers, logicians, sophists, amateurs, etc., the result still remains: Yes, because the world did not change after greatgrandmother died. The disappearance of observers does not seem to change the universe. These experimental findings can be corroborated by filling in the definition of 'existence' into the question, which then becomes: Do the things one observes interact with other aspects of nature when they do not interact with people? The answer is evident. Recent popular books on quantum mechanics fantasize about the importance of the "mind" of observers – whatever this term may mean; they provide pretty examples of authors who see themselves as irreplaceable center of the universe, seemingly having lost the ability to do otherwise.

Of course there are other opinions about existence of things. The most famous one is by the irishman George Berkeley (1685–1753) who rightly understood that thoughts based on observation alone, if spreading, would undermine the basis of a religious organization in which he was one of the top managers. To counteract this, in 1710 he published *A treatise concerning the principles of human knowledge*, a book denying the existence of the material world. This reactionary book became widely known in similar circles (it was a time when few books were written) even though it is based on a fundamentally flawed idea: it assumes that the concept of 'existence' and that of 'world' can be defined independently from each
Challenge other. (You may be curious to try the feat.)

Berkeley had two aims when he wrote his book. First, he tried to deny the capacity of people to arrive at judgments on nature or on any other matter *from their own experience*. Secondly, he also tried to deny the *ontological reach* of science, i.e. the conclusions one can take form experience on the questions about human existence. Even though he is despised by many, he actually achieved his main aim: he is the originator of the statement that science and religion do not contradict each other, but *complement* each other. This widely cited belief is still held dearly by many up to this day. However, when searching for the origin of

* He who possesses science and art, also has religion; he who does not possess the two, better have religion.

motion, beliefs such as this one stand in the way. Carrying them means carrying oversized baggage: it prevents from reaching the top of motion mountain.

Does the void exist?

Teacher: "What is there between the electrons and the nucleus?"
Student: "Nothing, only air."

Natura abhorret vacuum.
Antiquity

In philosophical discussions *void* is usually defined as non-existence. It then becomes a game of words to ask whether one has to answer this question by yes or no. The expression "existence of non-existence" is either a contradiction or at least unclearly defined; the topic would not seem of deep interest. However, similar questions do appear in physics, and one should be prepared to see the difference to the previous one. Does the vacuum exist? Does empty space exist? Or is the world 'full' everywhere, as the more conservative biologist Aristoteles maintained? In the past, people used to be killed if they gave the answer not accepted by authorities.

It is not obvious but nevertheless essential that the modern physical concepts of 'vacuum' or 'empty space' are not the same as the philosophical concept of 'void'. 'Vacuum' is not defined as 'non-existence'; on the contrary, it is defined as the absence of matter and radiation, and is an entity with specific observable properties, such as the number of its dimensions, its electromagnetic constants, its curvature, its vanishing mass, its interaction with matter through curvature and through its influence on decay, etc. (A table of the properties of the physical vacuum is given on page 337.) Historically, it took a long time to clarify the distinction between physical vacuum and philosophical void. People confused the two concepts and debated the question in the section title for more than two thousand years; the first to answer it positively, with the courage to try to look through the logical contradiction to the underlying physical reality, were Leucippos and Democritos, the most daring thinkers of antiquity. Their speculations in turn elicited the reactionary response of Aristoteles, rejecting the concept of vacuum. He and his disciples propagated the belief about nature's *horror of the vacuum*.

The discussion changed completely in the 17th century, when the first experimental method to realize a vacuum was discovered by Torricelli.* Using mercury in a glass tube, he produced the first human made vacuum. Can you guess how? Arguments against the existence of the vacuum reappeared around 1900, when it was argued that light needed "aether" for its propagation, using almost the same arguments used two hundred years earlier, just by changing the words. However, experiments failed to detect any supposed property of this unclearly defined concept. Experiments in the field of general relativity showed that the vacuum can move – though in a completely different way than the aether was expected to – that the vacuum can be bent, and that it tends to move back not normal. Then, in the late twentieth century, quantum field theory again argued against the existence of a true Challenge

* Evangelista Torricelli (1608, Faenza–1647), italian physicist, pupil and successor of Galileo. The pressure unit 'torr' is named after him.

vacuum and in favour of a space full of virtual particle-antiparticle pairs, culminating in the
See page 578 discussions around the cosmological constant.

The title question is settled conclusively only in the third part of this walk, in a rather
See page 599 surprising way.

Is nature infinite?

> There is a separation between state and church, but not yet between state and science.
> Paul Feyerabend (1924–1994)

Most of the modern discussions about set theory center on the ways to define the term 'set' for various types of infinite collections. For the description of motion this leads to two questions: Is the universe infinite? And is it a set? We begin with the first one. Illuminating it from various viewpoints, one quickly discovers that it is equally simple and imprecise.

Firstly, does one need infinite quantities to describe nature? In classical and quantum physics one does indeed, e.g. in the case of space-time. Is this necessary? This issue is settled only in the third part of this escalation.

But that does not help now. A second point is that any set can be finite in one aspect and infinite in another. For example, it is possible to walk a finite length in an infinite amount of time. It is also possible to sweep over an infinite length in a finite amount of time, even in
Challenge relativity, as explained there.

For example, these connections make discussions on whether humanity is near the "end of
Ref. 44 science" rather difficult. The amount of knowledge and the time to discover it are unrelated. Depending on the speed with which one advances through it, the end of science can be near of unreachable. In practice, scientists have thus the power to *make* science infinite or not, e.g. by reducing the speed of progress. Since funding is needed for their activity, everybody can guess which stand of the discussion is usually taken.

Thirdly, is it possible at all to say of nature or of one of its aspects that it is *infinite*? Can
Challenge such a statement be compatible with observations? It seems that every statement claiming that in nature something is infinite is a belief, and not taken from observations. We will encounter this issue several times later on.

In short, the universe cannot be said to be infinite. On the other hand, can nature be *finite*? At first sight, this would be the only possibility left. But even though many have tried to described a universe as finite in all its aspects, they were not successful. In order to see the problems it brings, we continue with the other question mentioned above:

Is the universe a set?

There is a simple fact questioning whether the universe is a set. For 2500 years it has been
Ref. 45 said that the universe is made of vacuum and particles. That implies that the universe is made of a certain *number* of particles. Perhaps the only person to have taken this conclusion to the limit was the english astrophysicist Arthur Eddington (1882–1944), who wrote:

> I believe there are 15,747,724,136,275,002,577,605,653,961,181,555,468,044,717,914,527, 116,709,366,231,425,076,185,631,031,296 protons in the universe and the same number of
> Ref. 46 electrons.

Eddington has been ridiculed over and over for this statement and for his beliefs leading to it. His arguments for this result were indeed based on his personal preferences for certain pet numbers. However, one should not laugh too loud. In fact, for 2500 years, almost all scientists have been thinking along the same line, with the only difference that they leave the precise number unspecified! In fact, *any other number* put into the above sentence would be equally ridiculous. Avoiding to name it is only a cowards' way to avoid looking at this unclear side of the particle description of nature.

Is there such a number at all? If you smiled at the sentence by Eddington, or if you shook your head over it, it may mean that you instinctively believe that nature is not a set. Is this so? Whether we define the universe as the totality of events or of observations, or as the totality of all space-time points and of all objects, we imply that space-time points can be distinguished, that objects can be distinguished, and that both can be distinguished from each other. We always assume that nature is separable. But is this correct?

The question is important. The ability to distinguish space-time points and particles from each other is often called *locality*. Thus the universe is a set or separable if and only if our description of it is local.* And in everyday life, locality is observed without exception.

In daily life we also observe that nature is separable and a whole at the same time. It is a 'many that can be thought as one': in daily life nature is a set. Indeed, the basic characteristic of nature is its diversity. In the world around us we observe changes and differences; we observe that nature is separable. Furthermore, all aspects of nature belong together: there are relations between these aspects, usually called "laws," expressing that the different aspects Challenge
of nature form a whole, usually called universe.

In other words, the possibility to describe observations with help of "laws" follows from the separability of nature. The more precisely the separability is specified, the more precisely the "laws" can be formulated. Indeed, if nature were not separable or were not a unity, we could not explain why stones fall downwards. Thus we are led to speculate that we should be able to deduce all "laws" from the fact that nature is separable.

In addition, only the separability allows us to describe nature at all. A description is a classification, i.e. a mapping between certain aspects of nature and certain concepts, i.e. certain combinations of sets and relations. Since the universe is separable, it can be described with help of sets and relations. Both are separable entities with distinguishable parts. A precise description is commonly called an understanding. In short, the universe is comprehensible only because it is separable.

Moreover, only the separability of the universe makes our brain such a good instrument. The brain is built from a large number of connected components, and only the brain's separability allows it to function. In other words, thinking is only possible because nature is separable.

Finally, only the separability of the universe allows to distinguish reference frames, and thus define all symmetries at the basis of our description. And in the same way as separability is thus necessary for *covariant* descriptions, the unity of nature is necessary for *invariant*

* In quantum mechanics also other, less clear definitions of locality are used. We will mention them in the second part of this text. The issue mentioned here is a different, more fundamental one, and not connected with the one of quantum theory.

descriptions. In other words, the so-called “laws” of nature are based on the fact that nature is both separable and unifiable – that it is a set.

These arguments seem overwhelmingly to prove that the universe is a set. However, these arguments only apply to everyday experience, everyday dimensions, and everyday energies. Is nature a set also *outside* the domains of daily life? Are objects different at all energies, i.e. when looking at them with the highest precision possible? Are objects countable at all See page 623 energies? We will discover in the third part of our escalation, that none of this happens to be the case in nature, confirming our suspicion about the number of particles. The consequences will be extensive as well as fascinating. As an example, try to answer the following: Challenge if the universe is not a set, what does that mean for space an time?

Does the universe exist?

> Each progressive spirit is opposed by a thousand men appointed to guard the past.
> Maurice Maeterlink (1862–1949) belgian dramatist

Following the definition above, existence of a concept means its usefulness to describe interactions. Now, there are two common definitions of the universe. The first is to mean the totality of all matter, energy and space-time. But this results in a strange consequence: since nothing can interact with this totality, we cannot claim that the universe exists.

So let us take the more conservative view, namely that the universe is only the totality of all matter and energy. But obviously, also in this case it is impossible to interact with the Challenge universe. Are you able to give a few arguments?

In short, we arrive at the conclusion that the universe does not exist. We will indeed See page 625 confirm this result in more detail later on in our walk. In particular, that means that it does not make sense to even try to answer *why* the universe exists. The best answer might be: Ref. 10 because of furiously sleeping, colourless green ideas.

What is creation?

> (Gigni) De nihilo nihilum,
> in nihilum nil posse reverti.*
> Ref. 47 Persius, *Satira*, III, v. 83-84.

The term is often heard when talking about nature. It is used in various contexts with different meanings:

One talks of creation as characterization of human actions, such as observed in a painting artist or a typing secretary. Obviously, this is a type of change. In the classification of change introduced at the beginning of our walk, such changes are movements of objects, such as the electrons in the brain, the molecules in the muscles, the material of the paint, or the electrons inside the computer. This type of creation is thus a special case of motion.

One also talks of creation in the biological or social sense, such as in ‘the creation of life’, or ‘creation of a business’, or ‘the creation of civilisation’. These events are forms of growth or of selforganisation; again, special cases of motion.

* Nothing (can appear) from nothing, nothing can disappear into nothing.

In physics one often says that a lamp 'creates' light or that a stone falling into a pond 'creates' water ripples. Similarly, one talks of 'pair creation' of matter and antimatter. It was one of the important discoveries of physics that all these processes are special types of motion, namely excitation of fields.

See page 453
See page 505

In popular pieces on cosmology, 'creation' is also a term commonly applied, or better misapplied, to the *big bang*. However, the expansion of the universe is a pure example of motion, and contrary to a frequent misunderstanding, the description of the big bang contains no process which does not fall into one of the previous three cases, as shown in the chapter of general relativity. Relativistic cosmology provides more reasons for which the term 'creation' is not applicable to the big bang. First of all, it turns out that the big bang was not an event. Secondly, it was not a beginning. Thirdly, it did not provide a *choice* from a large set of possibilities. The big bang does not have any properties attributed to the term 'creation'.

See page 260
See page 259

In summary, one concludes that in all cases, *creation is a type of motion*. (The same applies to the notions of 'disappearance' and 'annihilation'.) No other type of creation is observed in nature. In particular, the naive sense of "creation", namely 'appearance from nothing' – *ex nihilo* in latin –, is never observed in nature. All observed types of "creation" require space, time, forces, energy and matter for their realisation.

The opposite of creation is *conservation*. The central statements of physics are conservation theorems: for energy, for mass, for linear momentum, for angular momentum, for charge, for spin, etc. In fact, every conservation "law" is a detailed and accurate rejection of the concept of creation. Already the ancient greek idea of atoms contains this rejection. Atomists stated that there is no creation and no disappearance, only motion of atoms, only transformation of matter. In other words, the idea of atom was a direct consequence of the negation of creation. It took humanity over 2000 years to stop putting people in jail for talking about atoms, as still happened to Galileo.*

However, there is one exception in which the naive concept of creation does apply: it describes what magicians do on stage. When a magician makes a rabbit appear from nowhere, one indeed experiences "creation" from nothing. At its best such magic is a form of entertainment, at its worst, a misuse of gullibility. The idea that the universe results from either of the two does not seem appealing; on second thought though, maybe looking at the universe as the ultimate entertainment could open up a fresh and more productive approach to life.

* To get a clear view of the matters of dispute in the case of Galileo, especially of interest for physicists, the best text is the excellent book by Pietro REDONDI, *Galileo eretico*, Einaudi, 1983, translated into English as *Galileo heretic*, Princeton University Press, 1987, and into many other languages. Redondi, a renowned historical scholar and colleague of Pierre Costabel, tells the story of the dispute between Galileo and the reactionary parts of the catholic church. He recently discovered a document of that time – the anonymous denunciation which started the trial – allowing him to show that the condemnation of Galileo to life imprisonment was organized by his friend the pope to *protect* him from a sure condemnation to death for heresy. But most importantly, the reason for his arrest were not his ideas on astronomy and on motion of the earth, as usually maintained, but his statements on matter. Galileo defended the view that matter is made of 'atoms' or, as he called them, 'piccolissimi quanti' – smallest quanta – which was and of course still is a heresy, since it is not compatible with the change of bread and wine into human flesh and blood, called *transsubstantiation*, which is central belief of the catholic church. And in those days, church tribunals punished heresy, i.e. deviating personal opinions, by the death sentence. Today, the remainders of the catholic church continue to refuse to publish the proceedings and other documents of the trial. Ironically, 'quantum' theory has become the most precise description of nature ever.

Voltaire (1694–1778) popularized an argument against creation often used in the past: we do not know whether it has taken place or not. Today the situation is different: we *do* know that it has *not* taken place, because creation is a type of motion, and, as we will see in the third part of our escalation, motion did not exist near the big bang.

Have you ever heard the expression "creation of the laws of nature"? It is one of the most common examples of disinformation. First of all, this expression confuses the "laws" with nature itself. A description is not the same as the thing; everybody knows that giving to his beloved the description of a rose is different from giving an actual rose. Secondly, the expression implies that nature is the way it is because it is somehow 'forced' to follow the "laws", a rather childish, and moreover incorrect view. And thirdly, the expression assumes that it is possible to "create" descriptions of nature. But a "law" is a description, and a description by definition cannot be created: the expression makes no sense at all. The expression "creation of the laws of nature" is the epitome of confused thinking.

It may well be that calling a great artist 'creative' or 'divine', as became the use during the renaissance, is not a blasphemy, but simply an encouragement to the gods to try to do similarly well. In fact, whenever the term "creation" is used to mean anything else than some form of motion, one is discarding both observations and human reason. It is one of the last pseudo-concepts of our modern time; no expert on motion should forget this. One cannot escalate motion mountain without getting rid of it. We will encounter the next temptation to See page 566 bring it back in during the second part of the escalation.

Every act of creation is first of all an act of destruction.
Pablo Picasso (1881–1973), painter.

Is nature designed?

In the beginning the universe was created. This has made a lot of people very angry and has been widely regarded as a bad move.
Douglas Adams

The tendency to conclude from existence of an object to its creation is widespread. Some jump to this conclusion every time they see a beautiful landscape. This habit stems from the prejudice that a beautiful scene implies a complex description, in turn implying complex building instructions, and therefore *design*.

There are several mistakes in this conclusion. First of all, beauty is not necessarily a consequence of complexity. Usually it is the opposite, as the study of chaos and of selforganisation shows. Ref. 48 These research fields demonstrated how many beautifully complex shapes and patterns can be generated with extremely simple descriptions. True, for most human artifacts, complex descriptions indeed imply complex building processes. A personal computer is a good example. But in nature, this is not the case. We have seen above that the information to construct a human body is about a million times smaller than the information stored in the brain alone. Similar results have been found for plant architecture and for many other examples of patterns in nature. The simple descriptions behind the apparent complexities of nature have been and are still being uncovered by the study of selforganisation, chaos, turbulence, and fractal shapes. In nature, complex structures derive from *simple* processes.

Beware of anybody saying that nature has “infinite complexity:” apart from the fact that complexity is not a measureable entity, despite many attempts, all known complex system are describable by (relatively) few parameters.

The second mistake: complex descriptions for any object do not imply design; they only imply that the object has a long story of evolution behind it. The correct deduction is: something of large complexity, i.e. of low entropy, exists; therefore it has *grown*, i.e. it has been transformed through input of energy over time. This deduction applies to flowers, mountains, stars, life, people, watches, books, personal computers and works of arts; in fact it applies to all objects in the universe.

Third, the idea of “instruction” is often taken to mean that some unknown intelligence is somehow pulling the strings of the world’s stage. But the study of nature has shown in every single case that there is no such hidden intelligence. An instruction is a list of orders to an executioner. But there are no orders in nature, and no executioners. There are no “laws” of nature, only descriptions of processes. Nobody is building a tree; the tree is an outcome of the motion of molecules making it up. No molecule is given any instructions. The idea of design is an ideology, born from an analogy with monarchy or even tyranny, and a typical anthropomorphism.

In fact there is not a single example of observation in nature which implies or requires either design or creation. However, that is not a reason to deny that the phenomena of nature often inspires us with awe. The wild beauty of nature often show us how *small* a part of nature we actually are, both in space and in time. Remaining open to the power and to the details of this experience is of central importance for the rest of this escalation.

What is a description?

In theory, there is no difference between theory and practice.
In practice, there is.

Following standard vocabulary usage, a description of an observation is a list of the details. The above example of the grampus showed this clearly. In other words, a *description* of an observation is the act of categorizing it, i.e. of comparing, by identifying or distinguishing, the observation with all the other observations already made. A description is a classification. In short, *to describe means to see as an element of a larger set*. See page 410

A description is like the ‘you are here’ sign on a road map in a foreign city. It shows, out of a set of possible positions, the particular one one wants to describe. For example, the formula $a = GM/r^2$ is a description of the observations relating motion to gravity because it classifies the observed accelerations a according to distance to the central body r and to its mass M, and sees each specific case as an example of a general pattern. The habit of generalizing is one reason for the often disturbing dismissiveness of scientists: when they observe something, their professional deformation makes them usually see it as a special case of a known phenomenon and thus keeps them from being taken aback or from being enthusiastic about it.

A description is thus the opposite of a *metaphor*; the latter is an analogy relating different *special* cases only, in contrast to a precise relation between *general* cases, namely a physical *theory*.

Reason, purpose, and explanation

> Der ganzen modernen Weltanschauung liegt die Täuschung zugrunde, daß die sogenannten Naturgesetze die Erklärungen der Naturerscheinungen seien.*
> Ludwig Wittgenstein, *Tractatus*, 6.371

• Why are the leaves of most trees green? Because they absorb red and blue light. Why do they absorb those colours? Because they contain chlorophyll. Why is chlorophyll green? Because all chlorophyll types contain magnesium between four pyrrole groups, and this chemical combination gives green colour, as a result of its quantum mechanical energy levels. Why do plants contain chlorophyll? Because that is what land plants can synthesize. Why only that? Because all land plants originally evolved from the green algaes, who are able to synthesize only this compound, and not the compounds found in the blue or in the red algaes, which are also found in the sea.
• Why do children climb trees, and why do some people climb mountains? Because of the sensations they experience during their activity; the feelings of achievement, the symbolic act to go upwards, the wish to get a wider view of the world are part of this type of adventure.

The 'why'-questions in the two preceding paragraphs show the difference between reasons and purposes (although these two terms are not defined the same way by everybody). A *purpose* or *intention* is a classification applied to actions of humans or animals; strictly said, it specifies the quest for a feeling, namely for some type of satisfaction felt after completion of the action. On the contrary, a *reason* is a specific relation of a fact with the rest of the universe, usually its past. What we call a reason always rests outside the observation itself, whereas a purpose always is internal to it.

Reasons and purposes are the two possibilities of explanations, i.e. the two possible answers to questions starting with 'why'. Usually, physics is not concerned with purpose or with feelings of people, mainly because its original aim, to talk about motion with precision, does not seem to be achievable in this domain. Therefore, *physical* explanations of facts are never purposes, but are always reasons. A *physical explanation* of an observation is always
Ref. 49 the description of its relation with the rest of nature.**

This means that – contrary to an often heard opinion – any question starting with 'why' is accessible to physical investigation, as long as it asks for a *reason*, not a purpose. In particular, questions such as "why do stones fall downwards and not upwards?" or "why do electrons have that value of mass, and why do they have mass at all?" or "why does space have three dimensions and not thirtysix?" fall under this class, as these ask for the connection between specific observations and more general ones. Of course, not all demands for explanation have been answered yet, and there still are problems to be solved. Our present trail only leads along a few answers to some of the questions about motion.

The most general quest for an explanation derives from asking: why is the universe the way it is? Part of the topic is covered in our escalation, using the usual approach, namely:

* The whole modern conception of the world is founded on the illusion that the so-called laws of nature are the explanations of natural phenomena.
** It is important to note that purposes are *not* put aside because they pertain to the future, but because they are inadmissible anthropomorphisms. In fact, for deterministic systems, one can equally say that the future is actually a *reason* for the present and the past, a fact often forgotten.

Unification and demarcation

Studying the properties of motion, paying incessant attention to increase the accuracy of description, one finds that explanations are mostly of two types:*

• "It is like all such cases; also this one is described by ..." The situation is recognized as a *special case* of a general behaviour.

• "If it was different, one would have ..., which is in contrast with the observation that ..." The situation is recognized as the *only possible case*.**

In other terms, the first approach is to formulate rules or "laws" which describe larger and larger numbers of observations, and compare the observation with them. This endeavour is called the *unification* of physics – by those who like it; those who don't like it, call it "reductionism". For example, one finds that the same rule describes the flight of a tennis ball, the motion of the tides at the sea shore, the timing of ice ages, and the time at which the planet Venus ceases to be the evening star and starts to be the morning star. These processes are all consequences of universal gravitation. Similarly, it is not evident that the same rule describes the origin of the colour of the eyes, the formation of lightning, the digestion of food and the working of the brain. These processes are described by quantum electrodynamics.

Unification has its most impressive successes when it predicts an observation which was not made before. A famous example is the existence of antimatter, predicted by Dirac when he investigated the solutions of an equation that describes the precise behaviour of matter.

The second procedure in the search for explanations is the elimination of all other imaginable alternatives in favour of the actually correct one. This endeavour has no commonly accepted name: it could be called the *demarcation* of the "laws" of physics – by those who like it; the others often call it "anthropocentrism", or simply "arrogance".

When one discovers that light travels in such a way to take the shortest possible time to its target, or when one describes motion by a principle of least action, or when one discovers that trees are branched in such a way that they achieve the largest effect with the smallest effort, one is using a demarcation viewpoint.

In summary, unification, answering 'why' questions, and demarcation, answering 'why not' questions, are typical for the progress throughout the history of physics. One can say that the dual aspects of unification and demarcation form the the composing and the opposing traits of physics. They stand for the desire to *know everything*.

However, neither demarcation nor unification can explain the universe. Can you see why? Challenge
In fact, apart from unification and demarcation, there is a third possibility which merges the two and does allow to say more about the universe. Can you find it? Our walk will automatically lead to it later on.

Challenge * Are these the only possible ones?

** These two cases have not to be confused with similar sentences which *seem* explanations, but which aren't:

• "It is like the case of ..." A similarity with another *single* case is *not* an explanation.

• "If it were different, it would contradict the idea that ..." A contradiction with an *idea* or with a theory is *not* an explanation.

Pigs, apes, and the anthropic principle

Das wichtigste Instrument des Wissenschaftlers ist der Papierkorb.*

The wish to achieve demarcation of the patterns of nature is most interesting when one follows the consequences of different rules of nature until one finds them in contradiction with the most striking observation: human existence itself. In this special case the program of demarcation is often called the *anthropic principle* – from the greek ἄνθρωπος, meaning 'man'.

For example, if the gravitational constant were different from the actual one, the resulting temperature change on the earth would have made impossible the emergence of life, which needs liquid water. Similarly, our brain would not work if the moon did not circle the earth. Only because the moon revolves around our planet, the earth's magnetic field becomes big enough to protect the earth by deviating most of the cosmic radiation that would otherwise make all life on earth impossible, but leave enough of it to induce the mutations necessary for evolution. It is also well-known that fewer large planets in the solar system would have made the evolution of humans impossible. They divert large number of comets from hitting the earth. The spectacular collision of comet Shoemaker-Levy-9 with Jupiter, the astronomical event of July 1994, was an example of this mechanism in action.**

Also the anthropic principle has its most impressive successes when it predicts unknown observations. The most famous example stems from the study of stars. Carbon atoms, like all other atoms except hydrogen, helium and lithium, are formed in stars through fusion. While studying the mechanisms of fusion in 1953, the british astrophysicist Fred Hoyle*** found that carbon nuclei could not be formed from the alpha particles present inside stars at reasonable temperatures, except if they had an excited state with an increased cross section. From the fact of our existence, which is based on carbon, Hoyle thus predicted the existence of a previously unknown excited state of the carbon nucleus. Indeed, the excited state was found a few months later by Willy Fowler.**** Ref. 50

In its *serious* form, the anthropic principle is therefore the quest to deduce the description of nature from the experimental fact of our own existence. Ref. 51 In the popular literature however, the anthropic principle is often changed, from a simple experimental method to deduce the patterns of nature, to its *perverted* form, a melting pot of absurd metaphysical ideas in which everybody mixes up his favourite beliefs. Most frequently, the experimental observation of our own existence has been perverted to reintroduce the idea of "design", i.e. that the universe has been constructed with the aim to produce humans; often it is even suggested that the anthropic principle is an *explanation* – a gross example of disinformation.

How can one distinguish between the serious and the perverted form? One gets exactly the same rules and patterns of nature if one would use as starting point the existence of pigs or of monkeys. In other words, if one would get *different* conclusions by using the *porcine principle* or the *simian principle*, one is using the perverted form, otherwise one is using the serious form. (The carbon-12 story is thus an example of the serious form.) This test is

* The most important instrument of a scientist is the waste basket.

** For a collection of pictures about this event, see e.g. the http://garbo.uwasa.fi/pc/gifslevy.html web site.

*** Fred Hoyle (1915, Bingley, Yorkshire–) english astrophysicist.

**** William A. Fowler (1911–), Nobel Prize winner in physics for this and related discoveries.

effective because there is no known pattern or “law” of nature which is particular to humans but which is unnecessary for apes or for pigs.*

Does one need cause and effect in explanations?

In nature there are neither rewards nor punishments
– there are consequences.
Ivan Illich (1926, Vienna –)

No matter how cruel and nasty and evil you may be,
every time you take a breath you make a flower happy.
Mort Sahl

Historically, the two terms have played an important role for philosophical discussions in the time when the “laws of nature” have been formulated the first time with high precision, e.g. during the birth of modern mechanics. In those times, it was important to point out that every effect has a cause, in order to distinguish precise thought from thought based on beliefs such as “evolution from nothing”, “miracles”, or “divine surprises”. It was also essential to stress that effects are different from causes, to avoid pseudo-explanations such as the famous example by Molière where the doctor explains to his patient in elaborate terms that sleeping pills work because they contain a dormitive virtue. Ref. 52

But in physics, the two concepts of cause and effect are not used at all. That miracles do not appear is expressed every time one uses symmetries and conservation theorems, and that cause and effect differ from each other is inherent in any evolution equation. Moreover, the concepts of cause and effect are not clearly defined; for example, it is especially difficult to define what is meant by one cause as opposed to several of them, and the same for one or several effects. Both terms are impossible to quantify and to measure. In other words, useful as “cause” and “effect” may be in personal life for distinction between events which regularly succeed each other, they are not necessary in physics. In physical explanations, they play no special roles.

Is consciousness required?

Variatio delectat.**
Cicero

A lot of mediocre discussions are going on about this topic, and we will avoid them here. Ref. 53 What is consciousness? Most simply and concretely, *consciousness* means to possess a small part of oneself watching what the rest of oneself is perceiving, feeling, thinking, and doing. In short, consciousness is the ability to observe oneself, and especially one’s inner mechanisms and motivations. For this reason, consciousness is *not* a prerequisite for studying motion. Indeed, animals, plants, machines are also able to observe motion. For the same reason, consciousness is obviously not necessary to observe quantum mechanical motion.

* Apes though do not seem to be good physicists, as described in the text by D.J. Povinelli, *Folk physics for apes: the chimpanzee´s theory of how the world works*, Oxford University Press 2000.
** Change pleases.

On the other hand, both the study of motion and that of oneself have a lot in common: the need to observe carefully, to overcome preconceptions, to overcome fear, and the fun of doing so.

For the time being, we have put enough emphasis on the precision of concepts. Talking about motion is also something to be deeply enjoyed. Let us see why.

Precision and clarity obey the uncertainty relation:
their product is constant.

Curiosity

Precision is the child of curiosity.

Like in the history of every person, also in the history of mankind a long struggle took place to avoid the pitfalls of accepting as truth the statements of authorities, without checking the facts. Indeed, whenever curiosity leads somebody to formulate a question, there are always two general ways to proceed. One is to check the facts personally, the other is to ask somebody. But the last way is dangerous: it means to give up a part of oneself. Healthy people, children whose curiosity is still alive, as well as scientists, choose the first way. After all, science is adult curiosity.

Curiosity, also called the *exploratory drive*, plays strange games with people. Starting with the original experience of the world as a big "soup" of interacting parts, curiosity can drive to find *all* the parts and *all* the interactions, as in this walk. And it drives not only people. It has been observed that when rats show curious behaviour, certain brain cells in the hypothalamus get active and secrete hormones which produce positive feelings and emotions. If a rat gets the possibility, via some implanted electrodes, to excite these same cells by pressing a switch, it does so voluntarily: rats get *addicted* to the feelings connected
Ref. 54 with curiosity. Like rats, humans are curious because they enjoy it. And they do so in at least four ways: because they are artists, because they are fond of pleasure, because they are adventurers, and because they are dreamers. Let us see how.

At the origin, curiosity stems from the desire to interact in a positive way with the environment. Young children provide good examples: curiosity is a natural ingredient of their life, in the same way that it is for other mammals and a few bird species; incidentally, the same taxonomic distribution one finds for play behaviour. In short, all animals who play are curious, and vice versa. Curiosity provides the basis for learning, for creativity, and thus e.g. for art. The artist and art theoretician Joseph Beuys (1920–1986) had as his own guid-
Ref. 55 ing principle that *every* creative act is a from of art. Humans, and especially children, enjoy curiosity because they feel its importance for creativity, and for growth in general.

Curiosity regularly leads one to exclaim: "oh!", an experience that leads to the second reason to be curious: relishing feelings of wonder and surprise. Epicuros (341–271 BCE) maintained that this feeling, φα υμάξειν, is the origin of philosophy. These feelings, which today are variously called religious, spiritual, numinous, etc., are the same to which rats can get addicted. Among them, Rudolf Otto has introduced the now classical distinction into the fascinating and the frightening. He named the corresponding experiences 'mysterium fascinans' and 'mysterium tremendum.'* In this division, physicists, the other scientists, children, and other connoisseurs take a clear stand: they choose the fascinans as starting point for their actions and for their approach to the world. Such feelings of fascination induce some of the children who look at the night sky to dream about becoming astronomers, some of those who look through the microscope to become biologists or physicists, and so forth. (It could also be that genetics plays a role in this pleasure of novelty seeking.) Ref. 56

Perhaps the most beautiful moments in the study of physics are those appearing after new observations have shaken one's previously held thinking habits, have forced to give up a previously hold conviction, and have engendered the feeling of being lost. When, in this moment of crisis, one finally discovers the more adequate, more precise description of the observations providing a better insight into the world around or inside oneself, one is pervaded of a feeling usually called illumination. Whoever has kept alive the memory and the taste for these magic moments knows that in those situations one is pervaded by a feeling of union between oneself and the world.** The pleasure of these moments, the adventures of the change of thought structures connected with them, and the joy of insight following them provides the drive for many scientists. Little talking and lots of pleasure is their common denominator. In this spirit the austrian born physicist Viktor Weisskopf likes to say jokingly: "There are two things that make life worth living: Mozart and quantum mechanics."

The choice away from the tremendum towards the fascinans stems from an innate desire, most obvious in children, to reduce uncertainty and fear. This drive is the father of all adventures. It has a well-known parallel in ancient Greece, where the first men studying observations, such as Epicuros, stated explicitly that their aim was to free people from unnecessary fear, and to deepen knowledge with the aim to transform people from frightened passive victims into fascinated, active and responsible beings. Those ancient thinkers started to popularize the idea that like the common events in our life, also the more rare events follow rules. For example, Epicuros underlines that lightning is a natural phenomenon due to interactions between clouds, and stressed that it is a natural process, i.e. a process following rules, in the same way as does the falling of a stone or any more familiar process of everyday life.

Investigating the phenomena around them, philosophers, and later on scientists, succeeded to free humans from most of their fear due to uncertainty and to the lack of knowledge about nature. This liberation played an important role in the history of human culture and still does so in the personal history of many scientists. The aim to arrive at stable, rock-bottom truths has both inspired (but also hindered) many of them; Albert Einstein is a well-known example for both, discovering relativity, helping to start up but then denying quantum mechanics.

* This distinction is the basis of Rudolf OTTO, *Das Heilige – Über das Irrationale in der Idee des Göttlichen und sein Verhältnis zum Rationalen*, Beck, München, 1991. This is a new edition of the epoch-making work originally published at the beginning of the twentieth century. Rudolf Otto (1869–1937) was one of the most important theologians of his time.

** Several researchers have studied the situations leading to these magic moments in more detail, notably the prussian physician and physicist Hermann von Helmholtz (1821–1894) and the french mathematician Henri Poincaré (1854–1912). They distinguish four stages in the conception of an idea at the basis of such a magic
Ref. 57 moment: saturation, incubation, illumination, and verification.

In the experience and in the development of every human being, curiosity, and therefore the sciences, come *before* the two domains of magic and superstition. The former needs deceit to be effective, and the latter needs indoctrination; curiosity doesn't need either. Conflicts with superstitions, ideologies, authorities, or the rest of society are preprogrammed.

Curiosity is the exploration of limits. There are two possibilities: the limit can turn out to be real or apparent. If the limit is real, the most productive attitude is that of acceptance. Approaching the limit then gives strength. If the limit is only apparent and in fact non-existing, the best attitude is that of reevaluating the mistaken view, extracting the positive role it performed, and then to cross it. Distinguishing between the two is only possible when the limit is investigated with great care. That is the quest for precision. Distinguishing between the two also requires openness and unintentionality. The lack of the latter is often a hindrance for progress.

Das gelüftete Geheimnis rächt sich.*
Bert Hellinger (1925–)

Courage

It is dangerous to be right in matters on which
the established authorities are wrong.
Voltaire (1694–1778)

Manche suchen Sicherheit, wo Mut gefragt ist,
und suchen Freiheit, wo das Richtige keine Wahl läßt.**
Bert Hellinger (1925–)

In the adventure to get to the top of motion mountain, most of the material in this intermezzo is necessary. But one needs more. Like any enterprise, also curiosity requires courage, and complete curiosity, as aimed for in this escalation, requires complete courage. In fact, it is easy to get discouraged from this trip. The walk is often dismissed as useless, uninteresting, confusing, damaging, destructive, or evil. For example, between the death of Socrates in 399 BCE and Paul Thierry, Baron d'Holbach, in the 18th century, there are no books with the statement 'gods do not exist', because of the life threats suffered by those who dared to make it.

Ref. 63

But not only the external repression can be scary. Many tried to make the trip with some hidden intention, usually of ideological nature, and then got tangled up in it before reaching the top. Others were not prepared for the openness required, which can shatter deeply held beliefs. And the lack of a sufficient dose of humility when facing nature also makes the escalation impossible.

Through the constant elimination of uncertainty, both curiosity and scientific activity are implicitly opposed to any idea, person or organization which tries to avoid the comparison of

* The unveiled secret takes avenge.
** 'Some look for security where courage is required and look for freedom where the right way doesn't leave any choice.' This is from the beautiful booklet by Bert Hellinger, *Verdichtetes*, Carl-Auer Systeme Verlag, 1996.

statements with observations. As mentioned above, a statement which one refuses to check with facts is a superstition or a belief. Through the refusal inherent in them, superstitions and beliefs produce fear; a fear on which all unjust authorities are based. Curiosity and science are thus fundamentally opposed to unjust authority, a connection that has made life difficult for people such as Anaxagoras (500–428 BCE) in ancient Greece, Hypatia in the christian roman empire, Galileo Galilei in the church state, Antoine Lavoisier in France, Albert Einstein in Germany; in the second half of the twentieth century victims were Robert Oppenheimer and Chandler Davis in the United States, and Andrei Sakharov in the Soviet Union. Each of them have a horrible but instructive story to tell, as have, more recently, Fang Lizhi, Xu Liangying, Liu Gang and Wang Juntao in China, Kim Song-Man in South Korea, Otanazar Aripov in Uzbekistan, Ramadan al-Hadi al-Hush in Libya, Bo Bo Htun in Burma, as well as many hundreds of others; in many authoritarian societies the antagonism between natural sciences and injustice has hindered or even suppressed completely the development of physics and other sciences as a productive social and cultural movement.

Therefore, when one embarks on this walk, one has to know what one is doing. On the other hand, the risk are worth it: by working towards a satisfaction of curiosity, against disinformation, one helps achieving freedom from beliefs, from fear, and from violence, both inside and outside of oneself.

After this look to the basics, we continue our hike. The trail towards the top of motion mountain opens towards on a new adventure: discovering the origin of sizes and shapes in nature.

And the gods said to man:
"Take what you want,
and pay the price."
(Popular saying)

It is difficult to make a man miserable
while he feels he is worthy of himself.
Abraham Lincoln (1809–1865) US President

References

1 See for example the text by ... *Evolution*, ...See also Stephen J. Gould, 1980, *The Panda's thumb*, W.W. Norton & Co., New York. One of the several interesting and informative books on evolutionary biology by the best writer in the field. The epochal work by Charles Darwin, *On the origin of species*, can be found on the web nowadays, e.g. on the http://entisoft.earthlink.net/origspec.htm web site. Cited on page 370.

2 A simple description is *Woher kommt der Mensch*, Belser Verlag, ... Cited on page 370.

3 There is disagreement among experts about the precise timing of this experience. Some say that only birth itself is that moment. However, there are several standard methods to recall memories of early life, even of the time before birth; one is described and used by J.K. Stettbacher, *Making sense of suffering*, Penguin, New York, 1991, translated from the german original. Even more impressive examples are given by N.J. MAYER, *Der Kainkomplex – neue Wege der systemischen Familientherapie*, Integral Verlag, 1998. Cited on page 370.

4 Sanjida O'CONNELL, *Mindreading – how we learn to love and lie*, Arrow, 1998. This interesting book describes the importance of lying in the development of a human being, and explains the troubles those people have who cannot read other minds and thus cannot lie, such as autistics. Cited on page 371, 397.

5 The approach to describe observations as related parts is called structuralism; the starting point for this movement was de Saussure's *Cours de linguistique générale*, (see the footnote on page 378). A number of thinkers have tried to use the same approach in philosophy, mythology, and literature theory, with little success, however. An overview of the (modest) success of structuralism in linguistics and its failure in other fields is given by L. JACKSON, *The poverty of structuralism: literature and structuralist theory*, Longman, 1991. The author argues that when one reduces systems to interacting parts, one neglects the specific content and properties of the elements of the system, and this approach prevents a full understanding of the system under discussion. Cited on page 372.

6 For a view of the mental abilities different from that of Piaget (described on page 372), a presently much discussed author is the soviet experimental psychologist Lev Vigotsky, whose pathbreaking ideas and complicated life are described e.g. in Lev VIGOTSKY, *Mind in Society*, Harvard University Press, 1978, or in Rene Van der VEER & Jaan VALSINER, *Understanding Vygotsky: a quest for synthesis*, Blackwell Publishers, 1994. More extensive material can be found in the extensive work by René Van der VEER & Jaan VALSINGER, *The Vigotsky Reader*, Blackwell, 1994. Cited on page 372.

7 A somewhat unconventional source for more details is the beautiful text by Bruno BETTELHEIM, *The uses of enchantment: the meaning and importance of fairy tales*, Knopf, 1976. Cited on page 372.

8 Quoted in V. HARLEN, R. RAPPMANN & P. SCHATA, *Soziale Plastik – Materialien zu Joseph Beuys*, Achberger Verlag, 1984, on p. 61. Cited on page 373.

9 The problems one has when one loses the ability to classify or to memorize are told in the beautiful book by the neurologist Oliver SACKS, *The man who mistook his wife for a hat*, Picador, 1985, which collects many case studies he encountered in his work. More astonishing cases are collected in his equally impressive text *An anthroplogist on Mars*, Picador, 1995.

See also the beautiful text Donald D. HOFFMAN, *Visual intelligence – how we create what we see*, W.W. Norton & Co., 1998, and its associated ari.ss.uci.edu/cogsci/personnel/hoffman.html website. Cited on page 373, 375.

10 For a passionate introduction into the connections between language and the brain from a Chomskian perspective, see the bestselling book by Steven PINKER, *The language instinct – How the mind creates language*, Harper Perennial, 1994. The green idea sentence is discussed in a chapter of the book. Cited on page 373, 401, 415.

11 An overview of the status of the research into the origin of bipedalism is given by Bernard WOOD, *Four legs good, two legs better*, Nature **363**, pp. 587–588, 17 June 1983. Cited on page 373.

12 More about the connection between entropy and computers can be found in the classic paper by R. LANDAUER, .., IBM Journal of Research and Development **5**, p. 183, 1961, and in C.H. BENNETT & R. LANDAUER, .., Scientific American **235**, pp. 48–56, 1985. Cited on page 376.

13 W.H. ZUREK, *Thermodynamic cost of computation, algorithmic complexity and the information metric*, Nature **341**, pp. 119–124, 14 August 1989. Cited on page 376.

14 J.J. Hopfield/ Nature **376**, p. 33, 1995. Cited on page 377.

15 The details of the properties of the firing patterns of neurons are nicely described in the article by M. MAHOWALD & R. DOUGLAS, *A silicon neuron*, Nature **354**, pp. 515–518, 19/26 December 1991, in which they show how one can simulate a neuron's electrical behaviour using a silicon circuit. Cited on page 377.

16 The discussion whether the brain is or is not superior to a computer is nicely summarized by G. VOLLMER, *Algorithmen, Gehirne, Computer – Was sie können und was sie nicht können, Teil I und Teil II*, Naturwissenschaften **78**, p. 481, 1991, **78**, pp. 533–542, 1991. Cited on page 378.

17 However, the language with the largest available dictionary is dutch, with the 40 volumes of i the *Wordenboek der Nederlandsche Taal*, which appeared between 1864 and 1998. It has almost 400 000 entries. Cited on page 379.

18 The list and the remark on discovery on concepts is due to a personal communication from Anna Wierzbicka. An older list is published in her book *Semantics, primes, and universals*, Oxford University Press, 1996. Cited on page 380, 395.

19 W.S. HATCHER, *Foundations of mathematics*, W.B. Saunders Co., Philadelphia, 1968. There is also the article by Paul J. COHEN & Reuben HERSCH, *Non-cantorian set theory*, Scientific American **217**, pp. 104–116, 1967. Cohen was the mathematician who in 1963 proved that the negation of the continuum hypothesis could be added to the axioms of set theory and still produce a consistent theory; he calls such sets *non-cantorian*. Cited on page 384.

20 See the beautiful article by Ian STEWART, *Fair shares for all*, New Scientist pp. 42–46, 17 June 1995. Cited on page 384.

21 Rudy RUCKER, *Infinity and the mind – the science and philosophy of the infinite*, Bantam, Toronto, 1983. Cited on page 386.

22 This general division of mathematics is nicely explained in the text by Pierre BASIEUX, *Die Architektur der Mathematik – Denken in Strukturen*, Rororo, 2000. Cited on page 387.

23 The issue is treated in his *Summa Theologica*, in question 52 of the first part. The complete text, several thousand pages, can be found on the http://www.knight.org/advent web site. We come back to it in the part on quantum theory, in the section on the Pauli exclusion principle. It seems that the whole question goes back to the *Liber Sententiarum*, by Peter the Lombard. Cited on page 388.

See pages 495, 463

24 Ian STEWART, *Daisy, daisy, give me your answer, do*, Scientific American pp. 76–79, January 1995. This pedagogical article explains how the growth of plants usually leads to flowers whose number of petals is from the Fibonacci series 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, etc. Deviations from this ideal case are also explained. The original work are two articles by S. DOUADY & Y. COUDER, *La physique des spirales végétales*, La Recherche **24**, pp. 26–36, 1993, and *Phyllotaxis as a self-organized growth process*, in *Growth patterns in physical sciences and biology*, edited by J.M. GARCIA-RUIZ & al., Plenum Press, 1993. Cited on page 388.

25 H. DAVSON, *The eye*, Academic press, 1962. Cited on page 389.

26 See the http://akbar.marlboro.edu/ mahoney/cube/NxN.txt web site. Cited on page 389.

27 An introduction to the surreals is given by the article by Polly SHULMAN, *Infinity plus one, and other surreal numbers*, Discover pp. 99–105, December 1995. There is also the text by D. KNUTH, *Surreal numbers: how two ex-students turned on to pure mathematics and found total happiness*, Addison Wesley, 1974, also on the http://www-cs-faculty.stanford.edu/ knuth/sn.html web site. The usually quoted references on the topic also include John H. CONWAY, *On numbers and games*, Academic Press, 1976, E.R. BERLEKAMP, J.H. CONWAY & R.K. GUY, *Winning ways for your mathematical plays, volume I: games in general*, Academic Press, 1972, and H. GONSHOR, *An introduction to surreal numbers*, Cambridge University Press, 1986. Cited on page 390, 391.

28 This beautiful little problem is discussed by Ian STEWART, *A bundling fool beats the wrap*, Scientific American, pp. 109–111, June 1993. The answer, in japanese, is go-ju-nana. The four-dimensional answer lies somewhere between 50 000 and 100 000, whereas the five-dimensional answer is conjectured to be 'never'. Cited on page 392.

29 A. PAIS, *Niels Bohr's times: in physics, philosophy, and polity*, Ofxord University Press, 1991, page 176. Cited on page 392.

30 Eugene WIGNER, *Symmetries and reflections*, Indiana University Press, 1962. Cited on page 392.

31 About the different aspects of falsifiability of general statements it is commonplace to cite the work by the epitemologist Karl Popper (1902–1994), especially his long and boring book *Logik der Forschung*, first published in 1934. The reason for this boredom is that Popper's work is just Pierre Duhem's warmed up. Cited on page 398.

32 About a good way of making blood that liquefies, see L. GARLASCHELLI, F. RAMACCINI & S. DELLA SCALA, *Working bloody miracles*, Nature **353**, p. 507, 1991. The milk drinking indian statues were common around the world in 1994-1995. About healers, see James RANDI, *Flim-flam!*, Prometheus Books, Buffalo, New York, 1987, and the exquisite book by H.C. ZANDER, *Warum ich Jesus nicht leiden kann*, Rowohlt, 1994. Cited on page 399.

33 For an opinion completely contrary to the one described here, see the book by Gregory J. CHAITIN, *The limits of mathematics*, Springer Verlag, 1997, also to be found at the http://www.cs.auckland.ac.nz/CDMTCS/chaitin/lm.html web site, as well as his other works. Chaitin has devoted most of his life to the questions discussed in the section, especially on computability. Cited on page 401.

34 See the book by J. BARWISE & J. ETCHEMENDY, *The liar*, Oxford University Press, New York, 1987. Cited on page 402.

35 This definition (statement 4.11) and many other statements about science are in the beautiful and rightly famous text by Ludwig WITTGENSTEIN, *Tractatus logico-philosophicus*, Edition Suhrkamp, 1963. It gives a condensed summary of the basis of science, thought and language in a collection of highly structured and numbered sentences. Cited on page 402, 406.

36 Cited by Max DRESDEN, *The Klopsteg memorial lecture*, American Journal of Physics **66**, pp. 468–482, 1998. Cited on page 403.

37 Well-known books are e.g. Friedrich KOHLRAUSCH, *Praktische Physik*, Teubner, 24. Auflage, 1996. Cited on page 403.

38 Results are described e.g. in L. BERGMANN & C. SCHÄFER, *Lehrbuch der Experimentalphysik*, Band I, II, III und IV, W. de Gruyter. Cited on page 403.

39 LANDOLT-BÖRNSTEIN, edited by K.-H. HELLWEGE & O. MADELUNG, *Zahlenwerte und Funktionen aus Naturwissenschaften und Technik*, Neue Serie, Springer-Verlag, Berlin, 1984. This series of more than one hundred volumes contains all important observations in the domain of physics. Cited on page 403.

40 The strange and sometimes dangerous consequences of beliefs can be found e.g. in Martin GARDNER, *Fads and Fallacies*, Dover, 1957, and in James RANDI, *Faith healers*, Prometheus Books, 1989. The million dollar prize for showing any paranormal or supernormal effect is available from his http://www.randi.org web site. Cited on page 408.

41 See the nice collection of cranks on the www.crank.net web site. Cited on page 408.

42 As an example, the opposite view on the emergence of properties is strongly defended by ...in Physics Today 1998. Cited on page 410.

43 See page 133 of the bibliography by John BOWLBY, *Charles Darwin*, Pimlico, 1991. Cited on page 410.

44 John HORGAN, *The end of science – Facing the limits of knowledge in the twilight of scientific age*, Addison Wesley, 1996. Cited on page 413.

45 A beautiful introduction into greek philosophy is Eduard ZELLER, *Outlines of the history of greek philosophy*, Dover, 1980, a reprint of a book published in 1928. Among others, it gives a clear exposition of the philosophy of Democritos and the other presocratics. Cited on page 413.

46 The famous quote is found at the beginning of one of the chapters in Arthur EDDINGTON, *The Philosophy of Physical Science*, Cambridge, 1939. Cited on page 413.

47 Giuseppe FUMAGALLI, *Chi l'ha detto?*, Hoepli, 1983. Cited on page 415.

48 For a beautiful text on fractals, see the footnote on page 45. Cited on page 417.

49 As has been pointed out by René Descartes. Cited on page 419.

50 The famous carbon ^{12}C resonance was found by Willy Fowler, as described in E.M. BURBRIDGE, G.R. BURBRIDGE, W.A. FOWLER & F. HOYLE, .., Reviews of Modern Physics **29**, pp. 547–560, 1957. Cited on page 421.

51 An extensive overview of the topic is given by John D. BARROW & Frank J. TIPLER, *The anthropic cosmological principle*, Oxford University Press, 1986.The term itself is due to Brandon Carter, who coined it in 1973 and presented it on a symposium devoted to the 500th anniversary of Nicolaus Copernicus. For more literature, see Yuri I. BALASHOV, *Resource Letter AP-1: the anthropic principle*, American Journal of Physics **59**, pp. 1069–1076, 1991. Cited on page 421.

52 VOLTAIRE, *Candide ou l'optimisme*, 1759. See the footnote on page 103. The book is so good that it was still seized by the US customs in 1930, and the US post office refused to transport it as late as 1944. For more details, search for “banned books online” on the world-wide web. Cited on page 422.

53 The number of books on consciousness is large and the contents not always interesting, such as John ECCLES & ... ..., ..., Cited on page 422.

54 See e.g. the *Encyclopedia Britannica*, Macropaedia, in the entry on animal behaviour. Cited on page 423.

55 A straight and informative introduction to the work and ideas of Joseph Beuys (in german) is by Renate GEORGI, *Joseph Beuys*, RAAbits Kunst, Raabe Fachverlag, September 1996. Cited on page 423.

56 Two studies, one by R.P. EPSTEIN, in Nature Genetics **...**, January 1996, and another by J. BENJAMIN, found that people with a special form of the D4 dopamine receptor gene, or D4DR, are more prone to novelty seeking than people with the usual form. The D4DR gene regulates the formation of dopamine receptors, a chemical messenger in the brain which has been a candidate for some time for a substance involved in the propensity for novelty seeking. Cited on page 424.

57 More can be found in Jacques HADAMARD, *The mathematician's mind – The psychology of invention in the mathematical field*, Princeton Science Library, 1996. Cited on page 424.

58 For an exploration of “inner” motions, see the beautiful text by Richard SCHWARTZ, *Internal Family Systems Therapy*, The Guilford Press, 1995. Cited on page .

59 A clear overview of philosophy of science, often called epistemology, without unnecessary detail, is given by ... BLANCHÉ, ..., Presses Universitaires de France, 19▷▷. Cited on page .

60 An introduction to neurology ... Cited on page .

61 An introduction to computer science ... Cited on page .

62 Alfred TARSKI, *Introduction to modern logic*, Dover, 19▷▷. See also Lewis CARROLL, *Alice in wonderland*, ... Cited on page .

63 For example, one nees the courage to face envy. About this topic see the classic text by Helmut SCHOECK, *Der Neid*, 19▷▷, the standard work on the field. Cited on page 425.

Typeset in April 2001

Chapter VII Permutation of Particles

Why are we able to distinguish twins from each other? Why can we distinguish what looks alike, such as copies from originals? Interestingly, all methods one can imagine get into difficulties for pointlike particles, especially in collisions. The important consequences form the topic of this leg of our walk. (Challenge)

18. Are particles like condoms?

Some usually forgotten properties of objects are highlighted by studying condoms. To make this clear, we look at one of the prettier combinatorics puzzles, the *condom problem*. It asks:

How many condoms are necessary if w women want to encounter each of m men in a hygienical way, i.e. in such a way that nobody can in fact get in contact with the body fluids of any other? *

The problem is not as restricted as it seems. For example, it also applies to doctors, patients and surgical gloves, or to computers, interfaces, and computer viruses. Nevertheless, the way it is formulated here is the most common one. And obviously, the optimal number of condoms is *not* the product *wm*. In fact, the problem has three subcases. (Ref. 1)

- The case $m = w = 2$ already provides the most important ideas one needs. Are you able to find the optimal solution and procedure? (Challenge)
- In the cases $w = 1, m =$ odd or $m = 1$, $w =$ odd, the solution is $(m+1)/2$ condoms. This is the optimal solution, as one can easily check oneself. (Challenge)
- A solution for all other cases is given by $\lceil 2w/3 + m/2 \rceil$ condoms, where $\lceil x \rceil$ means the smallest integer greater than or equal to x. For example, for two men and three women this gives only three condoms. (But this formula does not always give the optimal solution; better values exist in certain subcases, as you might want to check.) (Ref. 2; Challenge)

After solving the puzzle, one can ponder which properties of condoms determine the solution. Firstly, they have two sides, the interior and the exterior. Secondly, they can be distinguished from each other. Do these two properties also apply to particles? We discuss the first and deeper issue, especially the question whether particles can be turned inside out, in the third part of the escalation. In the present chapter we concentrate on the second aspect, namely whether particles can always be distinguished. We will find that *elementary* (See page 573)

* This is the conventional formulation; you might want to modify or expand it in various ways depending on your personal preferences.

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Did you find any mistakes? Challenge

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

particles do not behave like condoms – a conclusion which might not come unexpectedly – but in a much more surprising manner.

In everyday life, distinction is possible in two ways. We are able to distinguish objects or people from each other because they differ in their *intrinsic properties*, such as their colour, size, shape, etc. In addition, we are able to distinguish objects even if they have the *same* structure and the *same* intrinsic properties, as any game of billiard teaches us. By following the path of a ball, we are sure to distinguish it from the others. In short, this second method is base on the state of the objects. In the case of billiard balls, this is possible because the measurement error for the position of the ball is much smaller than the size of the ball itself.

However, in the microscopic domain this is not correct. Obviously, microscopic particles have the same intrinsic properties. In collisions, one would thus be required to keep track of microscopic particles. In the nineteenth century it was shown experimentally that even nature is not able to do this, in the case of atoms of the same type. This was discovered studying systems which incorporate a large number of collisions, namely *gases*.

See page 134 When one calculates the entropy of a simple gas, made of N simple particles of mass m in a volume V, one gets

$$S = k\ln\left[V(\frac{mkT}{2\pi\hbar^2})^{3/2}\right]^N + \frac{3}{2}kN + k\ln\alpha \tag{420}$$

where k is the Boltzmann constant and T the temperature. In this formula, α is equal to 1 if the particles are distinguishable, and equal to $1/N!$ if not. It turns out that only the second case describes nature. This can be seen with a simple thought experiment: the entropy of two volumes of identical gas would not add. Indeed, the value observed by experiment, is given by the so-called Sakur-Tetrode formula

Challenge

Challenge

$$S = kN\ln\left[\frac{V}{N}(\frac{mkT}{2\pi\hbar^2})^{3/2}\right] + \frac{5}{2}kN \tag{421}$$

The result, often called *Gibbs' paradox*,* thus proves that the microscopic components of matter are *indistinguishable*: in a set of particles, there is no way to say which particle is which. Ref. 3

In fact, the properties of matter would be completely different without this indistinguishability; for example, as we will see later on, knifes and swords would not cut. In addition, the soil would not carry us; we would fall right through it.** To illuminate the issue from another side, let us continue with a related question.

Can particles be counted?

If particles cannot be distinguished, one needs to explain carefully how to count them. The first step is the definition of what is meant by a situation without any particle. This seems obvious, but later on we will encounter situations where this is impossible. Technically,

* Josiah Willard Gibbs (1839–1903), US-american physicist who was, with Maxwell, one of the founders of statistical mechanics and thermodynamics; introduced the concepts of *ensemble*, and of *phase*.

** When radioactivity was discovered, people thought that it contradicted the indistinguishability, as it seems to single out certain atoms for decay, instead of others. Only quantum theory resolved the issue.

this step is also called the *specification of the vacuum*. All counting methods require that situations with particles be clearly different from situations without particles.

The second step is the definition of an observable useful for determining particle number. The easiest way is to take one of those quantum numbers which add up under composition, such as electric charge.*

This method has several advantages. First of all, it is not important whether particles are distinguishable or not; it works in both cases. Secondly, virtual particles are not counted. A welcome state of affairs, as we will see, because for virtual particles, i.e. for particle for which $E^2 \neq p^2c^2 + m^2c^4$, there is *no way* to define a particle number anyway.

The other side is that antiparticles count negatively. In other words, any way of counting particles with the above method can produce an error due to this effect. In everyday life this plays no role, as there is no antimatter. The issue plays a role only at high energies. However, it turns out that a better way to count particles does not exist.

With this restriction we thus have a way to count objects which works also in the microscopic domain. In particular, it allows to count particles even when they cannot be distinguished.

What is permutation symmetry?

Since particles are indistinguishable but countable, there exists a symmetry of nature for systems composed of identical particles: *permutation symmetry* or *exchange symmetry* is the property of nature that observations are unchanged under exchange of identical particles. Together with space-time symmetry, gauge symmetry, and the not yet encountered renormalization symmetry, permutation symmetry forms one of the four pillars of quantum mechanics. Permutation symmetry is a property of *composed* systems, i.e. systems made of many (identical) subsystems. Only for such systems one can talk of indistinguishability.

In other words, 'indistinguishable' is not the same as 'identical'. Two particles are not the *same*; they are more like *copies* of each other. But since everyday life experience shows us that copies can be distinguished under close inspection, the term is not appropriate either. In the microscopic domain, particles are countable but completely indistinguishable.**

In addition, we will discover shortly that permutation is partial rotation. Permutation symmetry thus is a symmetry under partial rotations. Can you confirm this? Challenge

Indistinguishability and symmetry

The indistinguishability of particles leads to important conclusions about the description of their state of motion. For example, it is impossible to formulate a description that includes indistinguishability right from the start. Are you able to see why? Challenge

Therefore one describes a n-particle state with a wavefunction $\Psi_{1...i...j...n}$ which assumes that distinction is possible, as expressed by the ordered indices in the notation, and then introduces indistinguishability afterwards. Indistinguishability means that the exchange of

* In everyday life, weighing a composite of indistinguishable particles, i.e. measuring its mass, even though it does not fulfill the counting procedure, is usually a sufficient method to count them. But the method obviously does not work in the quantum domain. Can you give at least two reasons, one from special relativity, and one from general relativity? Challenge

** The word 'indistinguishable' is so long that many physicists sloppily speak of 'identical' particles.

any two particles leads to the same result. We therefore have the same situation as seen already several times: *an overspecification of the mathematical description*, here the explicit ordering of the indices, *implies a symmetry of this description*, which in our case is a symmetry under exchange of indices, i.e. pairs of particles. Now, two quantum states have the same physical implications if they differ at most by a phase factor; indistinguishability thus requires

$$\Psi_{1...i...j...n} = e^{i\alpha}\,\Psi_{1...j...i...n} \tag{422}$$

for some unknown angle α. Applying this expression two times by exchanging the same couple of indices again allows to conclude that $e^{2i\alpha} = 1$, which means that

$$\Psi_{1...i...j...n} = \pm\Psi_{1...j...i...n} \tag{423}$$

i.e., a wavefunction is either *symmetric* or *antisymmetric* under exchange of indices. In short, quantum theory predicts that particles are indistinguishable in one of two distinct ways.* Particles corresponding to symmetric wavefunctions are called *bosons*, those corresponding to antisymmetric wavefunctions are called *fermions*.**

Observation now shows that the behaviour depends on the *type* of particle. Electrons, protons and neutrons are found to be fermions. Also about half of the atoms are found to behave as bosons. In fact, any composite of an *even* number of fermions turns out to be a boson, and one composed of an *odd* number of fermions is a fermion. For example, almost all of the known molecules are bosons. Fermionic molecules are rather special and have even a special name in chemistry; they are called *radicals* and are known for their eagerness to react. Inside the human body, too many radicals can have averse effects on health; it is well known that vitamin C is important because effective in reducing the number of radicals.

Challenge Which class do mountains, trees, people, and all other macroscopic objects belong to?

The energy dependence of permutation symmetry

If experiments force us to conclude that nobody, not even nature, can distinguish any two particles of the same type, one concludes that they do not form two separate entities, but that they form some sort of unity. Our naive, classical sense of particle as a separate entity from the rest of the world is thus an incorrect description of the phenomenon of 'particle'. Indeed, no experiment can track particles with identical intrinsic properties in such a way that they can be distinguished with certainty. This has been checked experimentally with most elementary particles, with nuclei, with atoms, and with numerous molecules.

* This result is for three-dimensional space only. In two dimensions there are more possibilities.

** The first name is derived from the name of the italian physicist and Nobel Prize winner Enrico Fermi (1901, Roma–1954, Chicago) famous for his all-encompassing genius in theoretical and experimental physics. He published on nuclear and elementary particle physics, on spin and statistics, and for his experimental work was called "quantum engineer". Nevertheless, his highly deserved Nobel Prize was one of the few cases in which the prize was given for a discovery which turned out to be wrong. Bosons are named after the indian physicist Satyenra Nath Bose (1894, Calcutta–1974, Calcutta) who first described the statistical properties of photons, later expanding the work in collaboration with Albert Einstein.

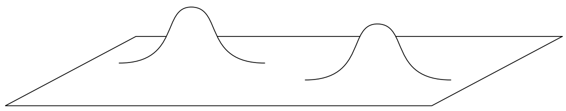


Figure 152 Particles as localized excitations

How does this fit with everyday life, i.e. with classical physics? We can distinguish electrons by the wire in which they flow, and we can distinguish our fridge from that of our neighbour. Well, the simplest way is to imagine a microscopic particle, especially an elementary one, as a bulge, i.e. as a localized excitation of the vacuum. Figure 152 shows two such bulges representing two particles. It is evident that if particles are too near to each other, it makes no sense to distinguish them; one cannot say anymore which is which.

One deduces that either for large distances or for high potential walls separating them, distinction of identical particles does become possible. In such situations, there *are* measurements allowing to track them independently. In other words, the energy at which permutation symmetry of objects or particles separated by a distance d becomes important is given by Challenge

$$E = \frac{c\,\hbar}{d} \quad . \tag{424}$$

Are you able to confirm it? For example, at everyday temperatures one *can* distinguish atoms in a solid from each other, since the energy so calculated is much higher than the thermal energy of atoms. You might also want to calculate at what energy two truly identical twins become indistinguishable. Challenge One can even estimate at what energies the statistical character of trees, or fridges will become apparent. Challenge

Therefore the bulge image also purveys the fact that in everyday life objects are distinguishable even though this is not the case in the microscopic domain. In summary, in daily life we are able to distinguish objects and thus e.g. also to recognize people because they are made of *many* parts, and because we live in a *low energy* environment.

But the energy issue brings an additional twist into the discussion. How does one describe fermions and bosons in the presence of virtual particles or of antiparticles?

Indistinguishability in quantum field theory

Quantum field theory, as we will see shortly, simply puts the bulge idea of figure 152 into mathematical language. A situation with no bulge is called vacuum state. Quantum field theory describes all particles of a given type as *excitations* of a single fundamental field. Particles are indistinguishable because they are all excitations with the same properties of the same basic substrate. A situation with one particle is then described by a vacuum state acted upon by a *creation operator*. Adding a particle is described by adding a creation operator, and subtracting a particle by adding a *annihilation operator*; the latter turns out to be the adjunct of the former.

Quantum field theory then studies how these operators behave if they are supposed to describe observations.* It arrives at the following conclusions:

• Fields with half-integer spin are fermions and imply local anticommutation.

* Whenever one has the relation

$$[b, b\dagger] = bb\dagger - b\dagger b = 1 \tag{425}$$

• Fields with integer spin are bosons and imply local commutation.

• For all fields at spacelike separations, the commutator respectively anticommutator vanishes.

• In addition, antiparticles of fermions are fermions, and antiparticles of bosons are bosons.

• Virtual particles behave like their real counterparts.

These connections are at the basis of quantum field theory.

Why are all electrons identical? Quantum field theory describes electrons as identical excitations of the vacuum, and as such as identical by construction. Of course, this answer is only partially satisfying. We will find a better one only in the third part of our escalation.

How precisely is permutation symmetry verified?

A simple but effective experiment that tests whether electrons are fermions or not was car-
Ref. 4
ried out by Ramberg and Snow. They sent an electric current of 30 A through a copper wire for one month, and since they did not observe any X-ray emission, they concluded that electrons are always in an antisymmetric state, with a symmetric component of less than $2 \cdot 10^{-26}$. The reasoning behind this elegant experiment is the following. If electrons would not always be fermions, every now and then an electron could fall into the lowest level of a copper atom, emitting X rays. The lack of such X-rays implies that electrons are fermions to a very high accuracy, and that the Pauli exclusion principle is indeed an accurate description of nature.

See page 495
As mentioned above, the Pauli exclusion principle, which states that only one electron can be in one give state, is the modern answer to the question about how many angels can dance on the top of a pin. Experiments such as this one in a sense confirm the answer already
Ref. 5
given by Thomas Aquinas in the middle ages: only one.

19. Spin, rotations, and statistics

We saw above that spin is the observation that matter beams, like light beams, can be polarized. Spin thus describes how particles behave under rotations, and it proves that particles
Ref. 6
are not purely point-like. The general background for the appearance of spin was elucidated by Eugene Wigner in 1939.* He started from the fact that any quantum mechanical particle, if elementary, must behave like an irreducible representation of the symmetry group of flat space-time, the so-called *inhomogeneous Lorentz group*. We have seen in the chapter
See page 115
on classical mechanics how this connection comes about. To be of physical relevance for quantum theory, the representations also have to be *unitary*; the full list of such unitary ir-

between the creation operator $b\dagger$ and the annihilation operator b, they describe a *boson*. If the operators for particle creation and annihilation anticommute, i.e. if one has

$$\{d, d\dagger\} = dd\dagger + d\dagger d = 1 \qquad (426)$$

one has a *fermion*. The so defined bracket is called the *anticommutator bracket*.

* Eugene Wigner (1902, Budapest–1995), hungarian-american theoretical physicist, Nobel prize for physics in 1993; wrote over 500 papers, many about symmetry in physics; famous also for being the most polite physicist in the world.

reducible representations thus provides the range of possibilities for any particle that wants to be *elementary*.

Cataloguing the possibilities, one fins first of all that any elementary particle is described by four-momentum – no news so far – and by an internal angular momentum, namely the spin. The four-momentum results from translation symmetry of nature, and the spin from its rotation symmetry. In fact, the types of spins result from the various possibilities an object can behave under rotations in three dimensions.* As is well known, the magnitude of four-momentum is an invariant property, and is given by the mass, whereas its orientation in spacetime is free. Similarly, the magnitude of spin is an invariant and its orientation has various possibilities; for spin, massive and massless particles thus behave differently.

For *massive* particles, one finds from the inhomogeneous Lorentz group that there is an invariant value $\sqrt{J(J+1)}\hbar$, or J for short, of this internal angular momentum. It specifies the representation under rotations of a given particle type. The value of J can be any multiple of $1/2$, i.e. it can take the values $0, 1/2, 1, 3/2, 2, 5/2$, etc. For example, electrons have spin $1/2$, helium atoms spin 0. The representation is $2J+1$ dimensional, meaning that the orientation has $2J+1$ possible values with respect to the axis. For electrons there are thus two possibilities; often they are called 'up' and 'down'. Spin thus only takes *discrete* values. This is in contrast with linear momentum, whose representations are infinite dimensional, and whose possible values form a *continuous* range.

Also *massless* particles are characterized by the value of their spin, which can take the same values as in the massive case. For example, photons have spin 1. But its representations are one-dimensional, so that massless particles are completely described by their *helicity*, defined as the projection of the spin onto the direction of motion. There is no other freedom for the orientation of spin in these cases.

To finish the list of all possible symmetries, one needs to include motion inversion parity, spatial parity, and charge inversion. Since these symmetries are parities, each elementary particle has to be described by three additional numbers, called T, C, an P, which each can take values either of $+1$ or -1.

A list of the values observed in nature is given in appendix C. Spin and parities together are called *quantum numbers*. As we will see later, additional interaction symmetries will add a few more. But let us return to spin. See page 739

The new result is that spin 1/2 is a possibility in nature, even though it does not appear in everyday life. Why? The answer is that only a rotation of 720 degrees is equivalent to one of 0 degrees, while one of 360 degrees is not. The mathematician Hermann Weyl used a simple image explaining this connection.

Take two cones with their tips touching, and touching each other along a line. Hold one cone and roll the other around it. When the rolling cone has come back to the original position, it has rotated by some angle. If the cones are wide, the rotation angle is small. If the cones are very thin, almost like needles, the moving cone has rotated by almost 720 degrees. A rotation of 720 degrees is thus similar to one by 0 degrees. This list of possible representations thus shows that rotations *require* the existence of spin. But why then do experiments show that all fermions have half-integer spin, and all bosons have integer spin?

* The group of physical rotations is also called SO(3), since it is equivalent to the group of **S**pecial **O**rthonormal **3** by 3 matrices.

Spin	system unchanged after rotation by	massive examples elementary	massive examples composite	massless examples all elementary
0	any angle	none[a,b]	mesons, nuclei, atoms	none[b]
1/2	2 turns	e, μ, τ, q	nuclei, atoms, molecules	ν_e, ν_μ, ν_τ
1	1 turn	W, Z	mesons, nuclei, atoms, molecules, toasters	g, γ
3/2	2/3 turn	none[b]	baryons, nuclei, atoms	none[b]
2	1/2 turn	none	nuclei	"graviton"[c]
5/2	2/5 turn	none	nuclei	none
3	1/3 turn	none	nuclei[d]	none
etc.[d]				

a. Whether the Higgs particle is elementary or not is still unknown.
b. Supersymmetry predicts particles in these and other boxes.
c. The graviton has not yet been observed.
d. Nuclei exist with spins values up to at least $101/2$ and 51. Ref. 8

Table 36 Irreducible representations of the rotation group

Why do electrons obey the Pauli exclusion principle? At first, it is not clear what the spin has to do with the statistical properties of a particle.

In fact, there are several ways to show that rotations and statistics are connected. Historically, the first proof used the details of quantum field theory, and was so complicated that its essential ingredients are hidden. It took quite some years to convince everybody that a simple observation about belts was the central part of the proof. Ref. 7

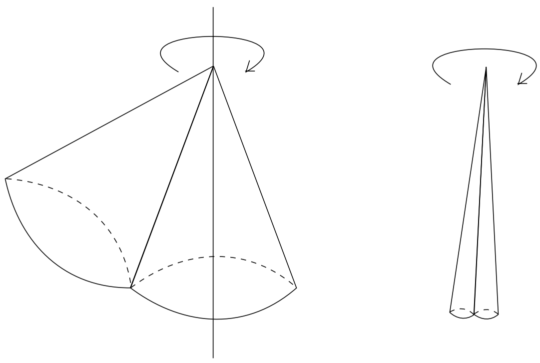


Figure 153 An argument showing why rotations by 4π are equivalent to no rotation

The belt trick

The well-known *belt trick* was often used by Dirac to explain the features of spin 1/2. Taking figure 152, which models particles as indistinguishable excitations, it is not difficult to imagine a sort of sheet connecting them, similar to a belt connecting the two parts of the buckle, as shown in figure

154. If one end of the belt is rotated by 2π along any axis, a twist is inserted into the belt. If the end is rotated for another 2π, bringing the total to 4π, the ensuing double twist can easily be undone without moving or rotating the ends. You need to experience this yourself in order to believe it. Challenge

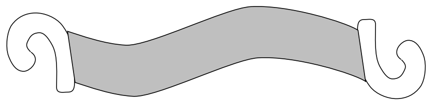


Figure 154 A belt visualizing spin 1/2

Now, if you take the two ends and simply *swap* positions, a twist is introduced into the belt. Again, a second swap will undo the twist.

In other words, if one takes each end to represent a particle, and a twist to mean a factor -1, the belt exactly describes the phase behaviour of spin $1/2$ wavefunctions under exchange and under rotations. In particular, we see that spin and exchange behaviour are related.

The human body has such a belt built in: the *arm*. Just take your hand, put an object on it for clarity, and turn the hand and object by 2π by twisting the arm. After a second rotation the whole system will be untangled again.

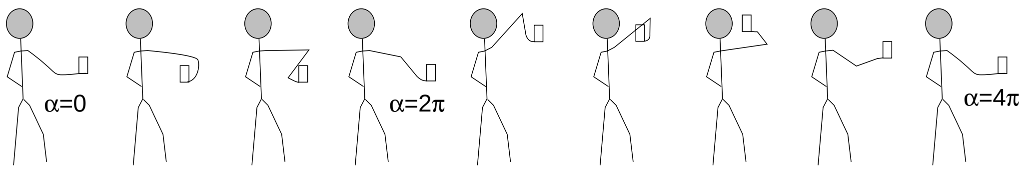


Figure 155 The human arm as spin 1/2 model

The trick is even more impressive when many tails are used. In fact, there are two ways. One is to take two buckles, like in figure 155, and connect them with *many* bands or threads. Either rotation by 2π of one end or exchange of both ends produces quite a tangle, and in both cases a second operation leads back to the original situation.

But there is an even more interesting way to show the connection between rotation and exchange. Just glue any number of threads or bands, say half a meter long, to two asymmetric objects. Like the arm, the bands are supposed to go to infinity and be attached there. If any of the objects, which represent the particles, is rotated by 2π, twists appear in its strings. If the object is rotated by an additional turn, to a total of 4π, all twists and tangles can be made to disappear, without moving or turning the object. One really has to experience this in order to believe it. And the trick really works with *any* number of bands glued to the object.

Even more astonishing is the other half of the experiment. If one exchanges the positions of two such spin 1/2 particles, always keeping the ends at infinity fixed, a tangled mess is created. But incredibly, if one exchanges the objects a second time, everything untangles neatly, independently of the number of attached strings. You might want to test that the behaviour is still the same with sets of three or more particles. Challenge

All these observations together form the spin statistics theorem for spin 1/2 particles: *spin and exchange behaviour are related*. Indeed, these almost 'experimental' arguments can be put into exact mathematical language, by studying the behavior of the configuration space of particles. These investigations result in the following statements: Ref. 9

▷ *Objects of spin 1/2 are fermions.* *

▷ *Exchange and rotation of spin 1/2 particles are similar processes.*

Note that all these arguments only work in three dimensions, because there are no tangles or knots in more or fewer dimensions. And indeed, spin exists only in three spatial dimensions.

One could be led to the conclusion that permutation properties and spin properties, so well described by the belt model, are perhaps really consequence of such belt-like connections between particles and the outside world. Maybe for some reason, we only observe the belt buckles, not the belts themselves, thus allowing us to localize particles. In the third part of this walk we will discover how correct this idea is.

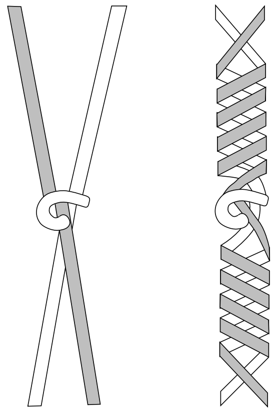


Figure 156 The extended belt trick, modeling a spin 1/2 particle: two situations which can be transformed into each other either by rotating the central object by 4π or by keeping the central object fixed and moving the bands around it

Integer spin

Integer spin particles behave differently under rotations. They do not show the strange sign changes under rotations by 2π. In the belt imagery, integer spin particles need no attached strings. The spin 0 particle obviously corresponds to a sphere. Models for other spin values are shown in figure 157. Including their properties in the same way as above, one arrives at the so-called *spin-statistics theorem*:

▷ *Exchange and rotation of objects are similar processes.*

▷ *Objects of half-integer spin are fermions. They obey the Pauli exclusion principle.*

▷ *Objects of integer spin are bosons.*

Challenge You might prove by yourself that this suffices to show that

▷ *Composites of bosons, as well as composites of an even number of of bosons, are bosons; composites of an uneven number of fermions are fermions.* **

These connections express basic characteristics of the three-dimensional world in which we live.

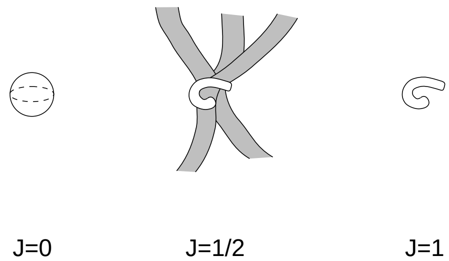


Figure 157 Some visualizations of spin representations

Is spin a rotation about an axis?

The spin of a particle behaves experimentally like an intrinsic angular momentum, adds up like angular momentum, is conserved together with angular momentum, is described like an

Challenge * A mathematical observable behaving like a spin 1/2 particle is neither a vector nor a tensor, as you may want to check. An additional concept is necessary; such an observable is called a *spinor*.

Challenge ** This sentence implies that spin 1 and higher can also be achieved *with* tails; can you find such a representation?

angular momentum, and has a name synonymous of angular momentum. Despite all this, a myth is spread in physics classes and textbooks around the world, namely that spin is *not* a rotation about an axis. The reason given is that any rotating object must have integer spin. How can one reconcile these two contradicting observations?

The way out is to remember what rotation is in detail. The belt trick for spin 1/2 as well as the integer spin case remind us of the answer: a *rotation* of one body around another is a fraction or a multiple of an exchange. What we call a rotating body in everyday life is a continuous exchange of positions between parts of the object.

In addition, we just saw that spin is exchange behaviour. Since rotation is exchange and spin is exchange, it follows that spin *is* rotation. As an additional argument, the belt model of a spin 1/2 particle tells us that such a particle can rotate continuously without any hindrance. Challenge In short, we can indeed maintain that spin is rotation about an axis, without any contradiction to observations, even for spin 1/2, if we keep the belt model in mind, never forgetting that only the buckles, not the belts themselves can be observed. Ref. 10

Rotation requires antiparticles

The connection between rotation and antiparticles may be the most astonishing conclusion from the experiments showing the existence of spin. So far, we have seen that rotation requires the existence of spin, that spin appears when relativity is introduced into quantum theory, and that relativity introduces antimatter already at the classical level. In a sense, the See page 169 conclusion of the title is thus not surprising. But there is a simple direct argument making the same point, without any help of quantum theory, when the belt model is extended from space alone, as done above, to *space-time*.

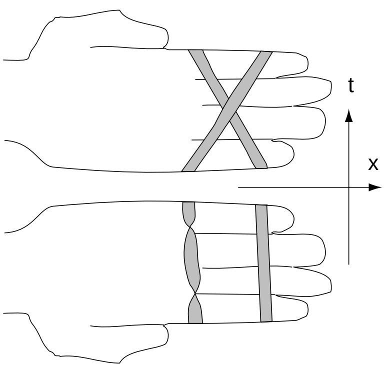


Figure 158 Equivalence of exchange and rotation in space-time

A particle, say of spin 1, i.e. one looking like a detached belt buckle in three dimensions, when moving in a 2+1 dimensional spacetime becomes a ribbon. Challenge Playing around with ribbons in spacetime, instead of belts in space, tells many interesting stories. For example, figure 158 shows that wrapping a rubber ribbon around the fingers can show that a rotation of a body by 2π in presence of a second one is the same as exchanging the positions of the two bodies.*

Now we can use this result for or original aim. When a particle in space-time is a ribbon, the process pictured in figure 159 shows the steps allowing to identify a rotation of a particle with an exchange – which we know to be correct already –, provided an particle-antiparticle pair is used. Without antiparticles, this equivalence could not be extended to spacetime.

* Obviously, the next step would be to check the full spin 1/2 model of figure 156 in four-dimensional space-time. But this is not an easy task; there is no generally accepted solution yet. Challenge

Why can't one fence with laser beams?

When a sword is approaching dangerously, one can stop it with a second sword. Many old movies use such scenes. When a laser beam is approaching, it is impossible to fend it off with a second beam, despite all science fiction movies showing so. Banging two laser beams against each other is impossible.

The above discussion shows why. The electrons in the swords are fermions and obey the Pauli exclusion principle. That makes matter impenetrable. On the other hand, photons are bosons. Bosons can be in the same state; they allow penetration. The distinction between fermions and bosons thus explains why objects can be touched while images cannot. In the first part of our escalation we started by noting this difference; now we know its origin. See page 54

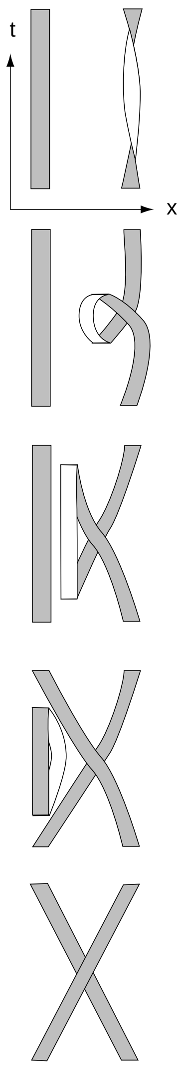


Figure 159 Belts in space-time

Limits and open questions

The topic of statistics is an important research field in theoretical and experimental physics. In particular, people are looking for generalizations of the exchange behaviour of particles.

In two spatial dimensions, exchange behaviour is not described by a sign, but by a continuous phase of the wavefunction. Such objects, called *anyons* because they can have 'any' spin, have experimental importance, since in many experiments in solid state physics the set-up is effectively two-dimensional. The fractional quantum Hall effect, perhaps the most interesting discovery of modern solid state physics, has pushed anyons onto the stage of modern research.

Other theorists generalized the concepts of fermions in other directions, introducing to parafermions, parabosons, plektons, and others. Ref. 11 O.W. Greenberg has spent most of his professional life on the issue. So far, his conclusion is that in $3+1$ dimensions, only fermions and bosons exist. (Can you show that this implies that scottish *ghosts* do not exist?) Challenge

On the other hand, if one looks at the above belt model, one is lead to study the behaviour of braids and knots. (In mathematics, a *braid* is a knot extending to infinity.) This fascinating part of mathematical physics has become important with the advent of string theory, where particles are not pointlike, but extended entities.

Another generalization of statistical behaviour is the concept of quantum group, which we will encounter later on.

In most cases, the quest is to understand what happens to permutation symmetry in a unified theory of nature. This is an issue for the third part of our escalation.

References

1 See the paper by Martin GARDNER, *Science fiction puzzle tales*, Clarkson Potter, **67**, pp. 104–105, 1981, or his book *aha! Insight*, Scientific American & W.H. Freeman, New York, 1978. The rabbit story is from A. HAJNAL & P. LOVÁSZ, *An algorithm to prevent the propagation of certain diseases at minimum cost*, in *Interfaces between computer science and operations research*, edited by J.K. LENSTRA, A.H.G. RINNOOY KAN & P. Van EMDE BOAS, Mathematisch Centrum, Amsterdam 1978, whereas the computer euphemism is used by A. ORLITZKY & L. SHEPP, *On curbing virus propagation*, Technical memorandum, Bell Labs 1989. Cited on page 457.

2 A complete discussion of the problem can be found in chapter 10 of Ilan VARDI, *Computational Recreations in Mathematica*, Addison Wesley, Redwood City, California, 1991. Cited on page 457.

3 On Gibbs' paradox, see your favourite text on thermodynamics or statistical mechanics. See also W.H. ZUREK, *Algorithmic randomness and physical entropy*, Physical Review A **40**, pp. 4731–4751, 1989. Zurek shows that the Sakur-Tetrode formula can be derived from algorithmic entropy considerations. Cited on page 459.

4 The experiment is described in E. RAMBERG & G.A. SNOW, Physics Letters B **238**, pp. 438–441, 1990. Other experimental tests are reviewed in O.W. GREENBERG, *Particles with small violations of Fermi or Bose statistics*, Physical Review D **43**, pp. 4111–4120, 1991. Cited on page 463.

5 The issue is treated in his *Summa Theologica*, in question 52 of the first part. The complete text, several thousand pages, can be found on the http://www.knight.org/advent web site. Cited on page 463.

6 E. WIGNER, *On unitary representations of the inhomogeneous Lorentz group*, Annals of Mathematics **40**, pp. 149–204, 1939. This famous paper summarizes the work which later brought Eugene Wigner the Nobel Prize for physics. Cited on page 463.

7 Wolfgang PAULI, ..., ... Cited on page 465.

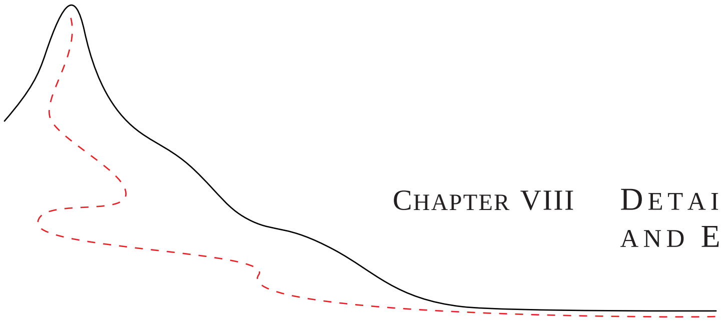
8 For a full list of isotopes, see R.B. FIRESTONE, *Table of Isotopes, Eighth Edition, 1999 Update*, with CDROM, John Wiley & Sons, 1999. Cited on page 465.

9 M.V. BERRY & J.M. ROBBINS, *Indistinguishability for quantum particles: spin, statistics, and the geometric phase*, Proceedings of the Royal Society in London A **453**, pp. 1771–1790, 1997. See also the comments to this result by J. TWAMLEY, *Statistics given a spin*, Nature **389**, pp. 127–128, 11 September 1997. Cited on page 466.

10 The point that spin can be seen as a rotation is made by F.J. BELINFANTE, Physica **6**, p. 887, 1939, and taken up again by Hans C. OHANIAN, *What is spin?*, American Journal of Physics **54**, pp. 500–505, 1986. See also E. DURAN & A. ERSCHOW, Physikalische Zeitschrift der Sowietunion **12**, p. 466, 1937. Cited on page 468.

11 Generalizations of bosons and fermions are reviewed in O.W. GREENBERG, D.M. GREENBERGER & T.V. GREENBERGEST, *(Para)bosons, (para)fermions, quons and other beasts in the menagerie of particle statistics*, hep-th/9306225. A newer summary is O.W. GREENBERG, *Theories of violation of statistics*, hep-th/0007054. Cited on page 469.

Typeset in
April 2001



Chapter VIII Details on Quantum Theory and Electromagnetism

20. Superpositions and probabilities in quantum mechanics

> Niels Bohr brainwashed a whole generation of physicists into believing that the problem [of the interpretation of quantum mechanics] had been solved fifty years ago.
> Murray Gell-Mann, Nobel price acceptance speech.

Why is this famous problem arousing such strong emotions? In particular, who is brainwashed, Gell-Mann, the discoverer of the quarks, or most of the other physicists working on quantum theory, following Niels Bohr's* opinion?

In the twentieth century, quantum mechanics has thrown many in disarray. Indeed, it radically changed the two most basic concepts of classical physics: state and system. The *state* is not described any more by the specific values taken by position and momentum, but by the specific wavefunction "taken" by the position and momentum operators.** In addition, in classical physics a *system* was described as a set of permanent aspects of nature; permanence was defined as negligible interaction with the environment. Quantum mechanics shows that this definition has to be modified as well.

In order to clarify the issues, we take a short walk around the strangest aspects of quantum theory. The section is essential if we want to avoid getting lost on our way to the top of motion mountain, as happened to quite a number of people in the twentieth century.

Why are people either dear or alive?

The evolution equation of quantum mechanics is linear in the wavefunction; thus one can imagine and try to construct systems where the state ψ is a superposition of two very distinct situations, such as those of a dead and of a living cat. This famous fictional animal is called *Schrödinger's cat* after the originator of the example. Is it possible to produce it? How would it evolve in time? Similarly, one can ask for the evolution of the superposition of a state where a car is inside a closed garage with a state where it is outside the closed garage.

∗ Niels Bohr (1885, Copenhagen–1962) made his university, Copenhagen, into one of the centers of quantum theory, overshadowing Göttingen. He developed the description of the atom with quantum theory, for which he received the 1922 Nobel prize in physics. He had to fly Denmark in 1943 after the german invasion, because of his jewish background, but returned there after the war.

∗∗ It is equivalent, but maybe conceptually clearer, to say that the state is described by a complete set of commuting operators. In fact, the discussion is somewhat simplified in the Heisenberg picture. However, here we study the issue in the Schrödinger picture, using wavefunctions.

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Did you find any mistakes? Challenge

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

All these situations are not usually observed in everyday life. What can be said about them? The answer to these questions is an important aspect of what is often called the "interpretation" of quantum mechanics. In principle, such strange situations are indeed possible, and the superposition of macroscopically distinct states has actually been observed in a few cases, though not for cats, people, or cars. To get an idea of the constraints, let us specify the situation in more detail.* The object of discussion are linear superpositions of the type $\psi = a\psi_a + b\psi_b$, where ψ_a and ψ_b are macroscopically distinct states of the system under discussion, and where a and b are some complex coefficients. States are called *macroscopically distinct* when each state corresponds to a different macroscopic situation, i.e. when the two states can be distinguished using the concepts or measurement methods of classical physics. In particular, this means that the action S necessary to transform one state into the other must be much larger than $\hbar$. For example, two different positions of any body composed of a large number of molecules are macroscopically distinct.

Figure 160 Artist's impression of a macroscopic superposition

Let us work out the essence of macroscopic superpositions more clearly. Given two macroscopically distinct states ψ_a and ψ_b, a superposition of the type $\psi = a\psi_a + b\psi_b$, is called a *pure state*. Since the states ψ_a and ψ_b can interfere, one also talks of a *(phase) coherent superposition*. In the case of a superposition of macroscopically distinct states, the scalar product $\psi_a^\dagger \psi_b$ is obviously vanishing. In case of a coherent superposition, the coefficient product $a^* b$ is different from zero. This fact can also be expressed with help of the *density matrix* ρ of the system, defined as $\rho = \psi \otimes \psi^\dagger$. In the present case it is given by

$$\begin{aligned} \rho_{\text{pure}} = \psi \otimes \psi^\dagger &= |a|^2 \psi_a \otimes \psi_a^\dagger + |b|^2 \psi_b \otimes \psi_b^\dagger + a\,b^* \psi_a \otimes \psi_b^\dagger + a^*\,b\,\psi_b \otimes \psi_a^\dagger \\ &= (\psi_a, \psi_b) \begin{pmatrix} |a|^2 & a\,b^* \\ a^*\,b & |b|^2 \end{pmatrix} \begin{pmatrix} \psi_a^\dagger \\ \psi_b^\dagger \end{pmatrix} \quad . \end{aligned} \tag{427}$$

One can then say that whenever the system is in a pure state, its density matrix, or *density functional*, contains off-diagonal terms of the same order of magnitude as the diagonal ones.** Such a density matrix corresponds to the above-mentioned situations so contrasting with daily life experience.

We now have a look at the opposite situation. In contrast to the case just mentioned, a density matrix for macroscopic distinct states with *vanishing* off-diagonal elements, such as

* Most what can be said about this topic has been said by two people: John von Neumann, who in the nine-
Ref. 24 teen thirties stressed the differences between evolution and decoherence, and by Hans Dieter Zeh, who in the
Ref. 25 nineteen seventies stressed the importance of baths in this process.

** Using the density matrix, we can rewrite the evolution equation of a quantum system:

$$\dot{\psi} = -iH\psi \quad \text{becomes} \quad \frac{d\rho}{dt} = -\frac{i}{\hbar}[H, \rho] \quad . \tag{428}$$

Both are completely equivalent. (The new expression is sometimes also called the *von Neumann equation*.) We won't actually do any calculations here. The expressions are given so that you recognize them when you encounter them elsewhere.

the two state example

$$\begin{aligned}\rho = & |a|^2 \psi_a \otimes \psi_a^\dagger + |b|^2 \psi_b \otimes \psi_b^\dagger \\ = & (\psi_a, \psi_b) \begin{pmatrix} |a|^2 & 0 \\ 0 & |b|^2 \end{pmatrix} \begin{pmatrix} \psi_a^\dagger \\ \psi_b^\dagger \end{pmatrix}\end{aligned} \quad (429)$$

describes a system which possesses *no* phase coherence at all. Such a diagonal density matrix cannot be that of a pure state; it describes a system which is in the state ψ_a with probability $|a|^2$ and which is in the state ψ_b with probability $|b|^2$. Such a system is said to be in a *mixed state*, because one does not *know* its state, or equivalently, it is said to be in a *(phase) incoherent superposition*, because one cannot observe interference effects in such a situation. A system described by a mixed state is always *either* in the state ψ_a *or* in the state ψ_b. In other words, a diagonal density matrix for macroscopically distinct states is not in contrast, but in agreement with everyday experience. In the picture of density matrices, the non-diagonal elements contain the difference between normal, i.e. incoherent, and unusual, i.e. coherent, superpositions.

The experimental situation is clear: for macroscopically distinct states, only diagonal density matrices are observed. Any system in a coherent macroscopic superposition somehow loses its off-diagonal matrix elements. How does this process of *decoherence* take place? The density matrix itself shows the way.

Indeed, the density matrix for a large system is used, in thermodynamics, for the
Ref. 26 definition of its entropy and of all its other thermodynamic quantities. These studies show
Challenge that

$$S = -k \operatorname{tr}(\rho \ln \rho) \quad (430)$$

where tr denotes the *trace*, i.e. the sum of all diagonal elements. We also remind ourselves that a system with a large and constant entropy is called a *bath*. In simple physical terms, a bath is thus a system to which we can ascribe a temperature. More precisely, a *(physical) bath*, or *reservoir*, is any large system for which the concept of *equilibrium* can be defined. Experiments show that in practice, this is equivalent to the condition that a bath consists of many interacting subsystems. For this reason, all macroscopic quantities describing the state of a bath show small, irregular *fluctuations*, a fact that will be of central importance shortly.

It is easy to see from the definition (430) of entropy, that the loss of off-diagonal ele-
Challenge ments corresponds to an increase in entropy. And it is known that increases in entropy of a reversible system, such as the quantum mechanical system in question, are due to interactions with a bath.

Where is the bath interacting with the system? It obviously must be in its outside the system one is talking about, i.e. in its *environment*. Indeed, we know experimentally that any environment is large and is characterized by a temperature; examples are listed in table 37. Any environment therefore contains a bath. We can even go further: for every experimental situation, there is a bath *interacting* with the system. Indeed, every system which can be observed is not isolated, as it obviously interacts at least with the observer; and every observer contains a bath, as we will show in more detail shortly. Usually however, the most important baths one has to take into consideration are the atmosphere around a system, the radiation attaining the system or, if the system itself is large enough to have a temperature, those degrees of freedom of the system which are not involved in the superposition one is investigating.

At first sight, this direction of thought is not convincing. The interactions of a system with its environment can be made very small by using clever experimental setups. That wold imply that the time for decoherence can be made arbitrary large. We need to check how much time a superposition of states needs to decohere. It turns out that there are two standard ways to estimate the *decoherence time*: either modelling the bath as large number of hitting particles, or by modelling it as a continuous field.

Table 37 Some common and less common baths with their main properties

Bath type	temperature T	wavelength λ_{eff}	particle flux φ	hit time $t_{hit} = 1/\sigma\varphi$ for atom[a]	object[a]
matter baths					
solid, liquid	300 K	10 pm	10^{31} /m^{2}s	10^{-12} s	10^{-25} s
air	300 K	10 pm	10^{28} /m^{2}s	10^{-9} s	10^{-22} s
laboratory vacuum	50 mK	10 µm	10^{18} /m^{2}s	10 s	10^{-12} s
photon baths					
sunlight	5800 K	900 nm	10^{23} /m^{2}s	10^{-4} s	10^{-17} s
“darkness”	300 K	20 µm	10^{21} /m^{2}s	10^{-2} s	10^{-15} s
cosmic microwaves	2.7 K	2 mm	10^{17} /m^{2}s	10^{2} s	10^{-11} s
terrestrial radio waves	300 K				
Casimir effect	.. K				
Unruh radiation of earth	.. K				
nuclear radiation baths					
radioactivity		10 pm		$10^{..}$ s	$10^{..}$ s
cosmic radiation	>1000 K	10 pm		$10^{..}$ s	$10^{..}$ s
solar neutrinos	≈ 10 MK	10 pm	10^{15} /m^{2}s	$10^{..}$ s	$10^{..}$ s
cosmic neutrinos	2.0 K	3 mm	10^{17} /m^{2}s	$10^{..}$ s	$10^{..}$ s
gravitational baths					
gravitational radiation		unknown		>$10^{..}$ s	>$10^{..}$ s

a. The cross section σ in the case of matter and photon baths was assumed to be 10^{-19} m^2 for atoms; for the macroscopic object a size of 1 mm was used as example. For neutrino baths, ...

If the bath is described as a set of particles randomly hitting the microscopic system, it is characterized by a characteristic wavelength λ_{eff} of the particles, and by the average interval t_{hit} between two hits. A straightforward calculation shows that the decoherence time t_d is in any case smaller than this time interval, so that Challenge

$$t_d \leqslant t_{hit} = \frac{1}{\varphi\sigma} \quad , \tag{431}$$

where φ is the flux of particles and σ is the cross section for the hit.* Typical values are given in table 37. One easily notes that for macroscopic objects, decoherence times are extremely short. Scattering leads to fast decoherence. However, for atoms or smaller systems, the situation is different, as expected.

A second method to estimate the decoherence time is also common. Any interaction of a system with a bath is described by a relaxation time t_r. The term *relaxation* designates any process which leads to the return to the equilibrium state. The terms *damping* and *friction* are also used. In the present case, the relaxation time describes the return to equilibrium of the combination bath and system. Relaxation is an example of an irreversible evolution. A process is called *irreversible* if the reversed process, in which every component moves in opposite direction, is of very low probability.** For example, it is usual that a glass of wine poured into a bowl of water colours the whole water; it is very rarely observed that the wine and the water separate again, since the probability of all water and wine molecules to change directions together at the same time is rather low, a state of affairs making the happiness of wine producers and the despair of wine consumers.

Now let us simplify the description of the bath. We approximate it by a single, unspecified, scalar field which interacts with the quantum system. Due to the continuity of space, such a field has an infinity of degrees of freedom. They are taken to model the many degrees of freedom of the bath. The field is assumed to be in an initial state where its degrees of freedom are excited in a way described by a temperature T. The interaction of the system with the bath, which is at the origin of the relaxation process, can be described by the repeated transfer of small amounts of energy E_{hit} until the relaxation process is completed.

The objects of interest in this discussion, like the mentioned cat, or person or car, are described by a mass m. Their main characteristic is the maximum energy E_r which can be transferred from the system to the environment. This energy describes the interactions between system and environment. The superpositions of macroscopic states we are interested in are solutions of the hamiltonian evolution of these systems.

Again one finds that the initial coherence of the superposition, so disturbingly in contrast with our everyday experience, disappears exponentially within a *decoherence time* t_d given
Ref. 29 by***

* The decoherence time is derived by studying the evolution of the density matrix $\rho(x,x')$ of objects localized at two points x and x'. One finds that the off-diagonal elements follow $\rho(x,x',t) = \rho(x,x',0)e^{-\Lambda t(x-x')^2}$, where the localisation rate Λ is given by

$$\Lambda = k^2 \varphi \sigma_{\text{eff}} \tag{432}$$

where k is the wave number, φ the flux, and σ_{eff} the cross section of the collisions, i.e. usually the size of the
Ref. 27 macroscopic object.

One also finds the surprising result that a system hit by a particle of energy E_{hit} collapses the density matrix
Ref. 28 roughly down to the de Broglie (or thermal de Broglie) wavelength of the hitting particle. Both results together give the formula above.

** Beware of other definitions which try to make something deeper out of the concept of irreversibility, such as claims that 'irreversible' means that the reversed process is *not at all* possible. Many so-called "contradictions" between the irreversibility of processes and the reversibility of evolution equations are due to this mistaken interpretation of the term 'irreversible'.

$$t_d = t_r \frac{E_{\text{hit}}}{E_r} \frac{e^{E_{\text{hit}}/kT} - 1}{e^{E_{\text{hit}}/kT} + 1} \qquad (435)$$

where k is the *Boltzmann constant* and like above, E_r is the maximum energy which can be transferred from the system to the environment. Note that one always has $t_d \leqslant t_r$. After a time interval of length t_d is elapsed, the system has evolved from the coherent to the incoherent superposition of states, or, in other words, the density matrix has lost its off-diagonal terms. One also says that the phase coherence of this system has been destroyed. Thus, after a time t_d, the system is found either in the state ψ_a or in the state ψ_b, respectively with the probability $|a|^2$ or $|b|^2$, and not anymore in a coherent superposition which is so much in contradiction with our daily experience. Which final state is selected depends on the precise state of the bath, whose details were eliminated from the calculation by taking an *average* over the states of its microscopic constituents.

The important result is that for all macroscopic objects, the decoherence time t_d is very small. In order to see this more clearly, one can study a special simplified case. A macroscopic object of mass m, like the mentioned cat or car, is assumed to be at the same time in two locations separated by a distance l, i.e. in a superposition of the two corresponding states. One assumes further that the superposition is due to the object moving as a quantum mechanical oscillator with frequency ω between the two locations; this is the simplest possible system that shows superpositions of an object located in two different positions. The energy of the object is then given by $E_r = m\omega^2 l^2$, and the smallest transfer energy $E_{\text{hit}} = \hbar\omega$ is the difference between the oscillator levels. In a macroscopic situation, this last energy is much smaller than kT, so that from the preceding expression on gets Ref. 31

$$t_d = t_r \frac{E_{\text{hit}}^2}{2E_r kT} = t_r \frac{\hbar^2}{2mkTl^2} = t_r \frac{\lambda_T^2}{l^2} \qquad (436)$$

in which the frequency ω has disappeared. The quantity $\lambda_T = \hbar/\sqrt{2mkT}$ is called the *thermal de Broglie wavelength* of a particle.

It is straightforward to see that for practically all macroscopic objects the typical decoherence time t_d is very short. For example, setting $m = 1\,\text{g}$, $l = 1\,\text{mm}$ and $T = 300\,\text{K}$ one gets $t_d/t_r = 1.3 \cdot 10^{-39}$. Even if the interaction between the system and the environment would be so weak that the system would have as relaxation time the age of the universe, which is about $4 \cdot 10^{17}\,\text{s}$, the time t_d would still be shorter than $5 \cdot 10^{-22}\,\text{s}$, which is over a million times faster than the oscillation time of a beam of light (about 2 fs for green light).

$*$ $*$ $*$ This result is derived as in the above case. A system interacting with a bath always has an evolution given by the general form Ref. 30

$$\frac{d\rho}{dt} = -\frac{i}{\hbar}[H,\rho] - \frac{1}{2t_o}\sum_j [V_j\rho, V_j^\dagger] + [V_j, \rho V_j^\dagger] \quad . \qquad (433)$$

Challenge Are you able to see why? Solving this equation, one finds for the elements far from the diagonal $\rho(t) = \rho_o e^{-t/t_o}$. In other words, they disappear with a characteristic time t_o. In most situations one has a relation of the form

$$t_o = t_r \frac{E_{\text{hit}}}{E_r} = t_{\text{hit}} \qquad (434)$$

or some variations of it, as in the example above.

For Schrödinger's cat, the decoherence time would be even shorter. These times are so short that one cannot even hope to *prepare* the initial coherent superposition, let alone to observe its decay or to measure its lifetime.

For microscopic systems however, the situation is different. For example, for an electron in a solid cooled to liquid helium temperature one has $m = 9.1 \cdot 10^{-31}$ kg, and typically $l = 1$ nm and $T = 4$ K; one then gets $t_d \approx t_r$ and therefore the system can stay in a coherent superposition until it is relaxed, which confirms that for this case coherent effects can indeed be observed if the system is kept isolated. A typical example is the behaviour of electrons in superconducting materials. We'll mention a few more below. Ref. 32

Conclusions on decoherence, life, and death

In summary, both estimates of decoherence times tell us that for most macroscopic objects, in contrast to microscopic ones, both the preparation and the survival of superpositions of macroscopically different states is made practically impossible by the interaction with any bath found in their environment, even if the usual measure of this interaction, given by the friction of the motion of the system, is very small. Even if a macroscopic system is subject to an extremely low friction, leading to a very long relaxation time, its decoherence time is still vanishingly short.

Our everyday environment if full of baths. Therefore, *coherent superpositions of macroscopic objects or macroscopically distinct states never appear in nature.* In short, we cannot be dead and alive at the same time.

We also take a second conclusion: *decoherence results from coupling to a bath in the environment.* Decoherence is a thermodynamic, statistical effect. We will return to this issue below.

What is a system? What is an object?

In classical physics, a system is a part of nature which can be isolated from its environment. However, quantum mechanics tells us that isolated systems do not exist, since interactions cannot be made vanishingly small. The results above allow us to define the concept of system with more accuracy. A *system* is any part of nature which interacts *incoherently* with its environment. In other words, an *object* is a part of nature interacting with its environment only through baths.

In particular, a system is called a *microscopic system* or a *quantum mechanical system* and can described by a wavefunction ψ whenever

- it is almost isolated, with $t_{\mathrm{evol}} = \hbar/\Delta E < t_{\mathrm{r}}$, and
- it is in *incoherent* interaction with its environment. Ref. 33

In short, a microscopic system interacts incoherently and weakly with its environment.

In contrast, a bath is never isolated in the sense just given, because its evolution time is always much larger than its relaxation time. Since all macroscopic bodies are in contact with baths – or even contain one – they cannot be described by a wavefunction. In particular, one cannot describe any measuring apparatus with help of a wavefunction.

We thus conclude that a *macroscopic system* is a system with a decoherence time much shorter than any other evolution time of its constituents. Obviously, macroscopic systems also interact incoherently with their environment. Thus cats, cars, and television news speakers are all macroscopic systems.

A third possibility is left over by the two definitions: what happens in the situation in which the interactions with the environment are *coherent*? We will encounter some examples shortly. Following this definition, such situations are *not* systems, and cannot be described by a wavefunction. For example, it can happen that a particle forms neither a macroscopic nor a microscopic system!

Nature is composed of many parts. Matter is composed of particles. Can parts be defined precisely? Can they be isolated from each other and pinned down unambiguously? In quantum theory, nature is not found to be made of isolated entities, but is still made of *separable* entities. The criterion of separability is the incoherence of interaction. Any system whose parts interact coherently is not separable. So the discovery of coherent superpositions includes the surprising consequence that there are systems which, even though they look separable, are not. In nature, systems exist which are *not* divisible. Quantum mechanics thus also stresses the *interdependence* of the parts of nature. By the way, in the third part of the walk we will encounter much stronger types of interdependence.

All surprising properties of quantum mechanics, such as Schrödinger's cat and all the other examples, are consequences of the classical prejudice that a system made of two or more parts must necessarily be divisible into two subsystems. Whenever one tries to divide indivisible systems, one gets strange or incorrect conclusions, such as apparent faster-than-light propagation, or, as one says today, non-local behaviour. Let us have a look at a few typical examples.

Is quantum theory non-local? – A bit about EPR

Mr. Duffy lived a short distance away from his body.
James Joyce

We asked this same question about general relativity. What is the situation in quantum mechanics? See page 276

Let us first look at the wavefunction collapse for an electron hitting a screen after passing a slit. Following the description just deduced, the process looks roughly as depicted in figure 161. A movie of the same process can be seen in the lower right corners on the pages of the present, second part of our escalation. The situation is surprising: a wavefunction collapse gives the impression to involve faster than light propagation, because the maximum of the function changes position at extremely high speed, due to the short decoherence times. Does this happen faster than light?

Yes, it does. But is it a problem? A situation is called *acausal* or *nonlocal* if energy is transported faster than light. Using figure 161 you can calculate the energy velocity involved, using the results on signal propagation. One finds a value smaller than c. A wavefunction maximum moving faster than light does *not* imply energy motion faster than light.* Challenge See page 333

* In classical electrodynamics, the same happens with the scalar and the vector potential, if the Coulomb gauge is used.

Another often cited Gedankenexperiment was proposed by Bohm* in the discussion around the so-called Einstein-Podolsky-Rosen paradox. In the famous EPR paper the three authors try to find a contradiction between quantum mechanics and common sense. Bohm translated their rather confused paper into a clear thought experiment. When two particles in a spin 0 state move apart, measuring one particle's spin orientation implies an *immediate* collapse also of the other particle's spin, namely in the exactly opposite direction. This happens instantaneously over the whole separation distance; no speed limit is obeyed.

Ref. 34, 35

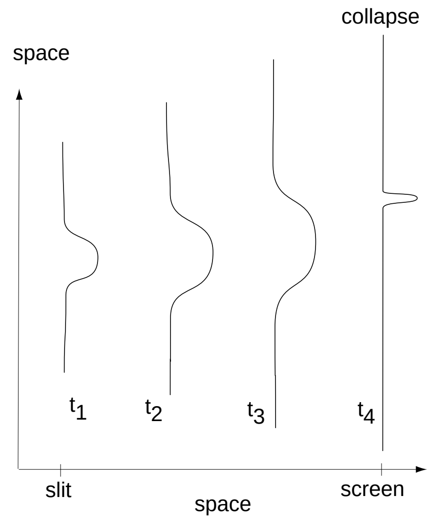


Figure 161 Quantum mechanical motion: an electron wave function (actually its module squared) from the moment it passes a slit until it hits a screen

One notes again that no energy is transported faster than light. No nonlocality is present, against numerous claims of the contrary in older literature. The two electrons belong to one system: assuming that they are separate only because the wavefunction has two distant maxima is a conceptual mistake. In fact, no signal can be transmitted with this method; it is a case of prediction which looks like a signal, as we already discussed in the section on special relativity.

See page 335

This experiment has been actually been performed. The first and most famous is the one performed with photons instead of electrons by Alain Aspect. Also the latter versions have confirmed quantum mechanics completely.

Ref. 36

In fact, this experiment just confirms that it is not allowed to treat either of the two particles as a system, and to ascribe them any property by themselves, such as spin. The Heisenberg picture would express this even more clearly.

These first two examples can be dismissed with the remark that since obviously no energy flux faster than light is involved, no problems with causality appear. Therefore the following example is more interesting. Take two identical atoms, one in an excited state, one in the

* David Bohm (19.. –1997/8) british physicist, codiscovered the Aharonov-Bohm effect; he also investigated the connections between physics and philosophy.

ground state, and call l the distance that separates them. Common sense tells that if the first atom returns to its ground state emitting a photon, the second atom can be excited only after a time $t = l/c$ has been elapsed, i.e. after the photon has traveled to the second atom.

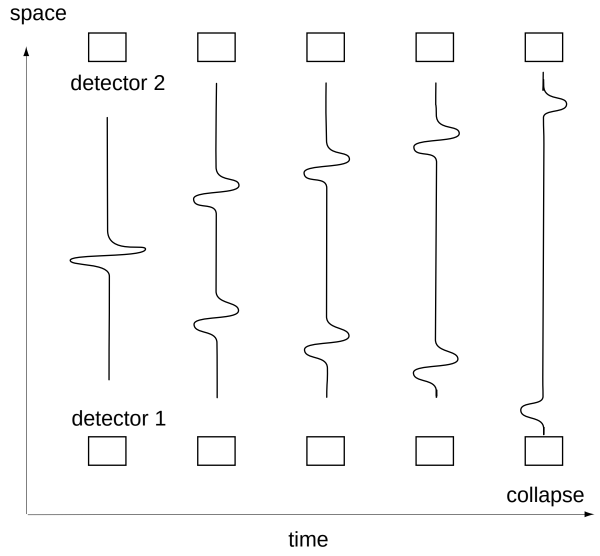


Figure 162 Bohm's Gedankenexperiment

Surprisingly, this conclusion is wrong. The atom in its ground state has a nonzero probability to be excited directly at the same moment in which the first is deexcited. This has been shown most simply by Hegerfeldt. The result has even been confirmed experimentally. Ref. 37

In fact, the result depends on the type of superposition of the two atoms at the beginning: coherent or incoherent. For incoherent superpositions, the intuitive result is correct; the surprising result appears only for coherent superpositions. That is pretty, isn't it?

Curiosities

• In a few rare cases, the superposition of different macroscopic states can actually be observed by lowering the temperature to sufficiently small values and by carefully choosing suitably small masses or distances. Two well-known examples of coherent superpositions are those observed in gravitational wave detectors and in Josephson junctions. In the first case, one observes a mass as heavy as 1000 kg in a superposition of states located at different points in space: the distance between them is of the order of 10^{-17} m. In the second case, in superconducting rings, superpositions of a state in which a macroscopic current of the order of 1 pA flows in clockwise direction with one where it flows in counterclockwise direction have been produced. Ref. 31 Ref. 38

• Obviously, superpositions of magnetization in up and down direction have also be observed. Ref. 39

• Since the 1990s, the sport of finding and playing with new systems in coherent superpositions has taken off world-wide. Its challenges lie in the clean experiments necessary. Experiments with single atoms in superpositions of states are among the most popular ones. Ref. 40

• In 1997, coherent atom waves were extracted from a cloud of sodium atoms. Ref. 41

• Macroscopic objects thus usually are in incoherent states. This is the same situation as for light. The world is full of "macroscopic", i.e. incoherent light: daylight, and all light from lamps, from fire, and from glow-worms is incoherent. Only very special and carefully constructed sources, such as lasers or small point sources, emit coherent light. Only these allow to study interference effects. In fact, the terms 'coherent' and 'incoherent' originated in optics, since for light the difference between the two, namely the capacity to interfere, had been observed centuries before the case of matter.

Coherence and incoherence of light and of matter manifests themselves differently, since matter can stay at rest but light cannot, and because light is made of bosons, but matter See page 468 is made of fermions. Coherence can be observed easily in systems composed of bosons, such as light, sound in solids, or electron pairs in superconductors. Coherence is less easily observed in systems of fermions, such a s systems of atoms. However, in both cases a decoherence time can be defined. In both cases coherence in many particle systems is best observed if all particles are in the same state (superconductivity, laser light), and in both cases the transition from coherent to incoherent is due to the interaction with a bath. In everyday life, the rareness of observation of coherent matter superpositions has the same origin as the rareness of observation of coherent light.

• We will discuss the relation between the environment and the decay of unstable systems See page 500 later on. The phenomenon is completely described by the concepts given here.

• Another conclusion deserves to be mentioned: *teleportation contradicts correlations*. Challenge Can you confirm it?

What is all the fuzz about measurements in quantum theory?

Measurements in quantum mechanics are disturbing. They lead to statements in which *probabilities* appear. That is puzzling. For example, one speaks about the probability of finding an electron at a certain distance from the nucleus of an atom. Statements like this belong to the general type "when the observable A is measured, the probability to find the outcome a is p." In the following we will show that the probabilities in such statements are inevitable for any measurement, because, as we will show, any measurement and any observation is a special case of decoherence process. (Historically however, the process of measurement was studied before the more general process of decoherence. That explains in part why the topic is so confused in many peoples' minds.)

See page 375 What is a measurement? As already mentioned in the intermezzo a measurement is any interaction which produces a record or a memory. Measurements can be performed by machines; when they are performed by people, they are called observations. In quantum theory, the action of measurement is not as straightforward as in classical physics. This is seen most strikingly when a quantum system, such as a single electron, is first made to pass a diffraction slit, or better – in order to make its wave aspect become apparent – a double slit, and then is made to hit a photographic plate, in order to make its particle aspect appear. One observes the well known fact that the blackened dot, the spot where the electron has hit the screen, cannot be determined in advance. (The same is true for photons or any other particle.) However, for large numbers of electrons, the spatial distribution of the black dots, the so-called *diffraction pattern*, can be calculated in advance with high precision.

The outcome of experiments on microscopic systems thus forces one to use probabilities for the description of microsystems. One finds that the probability distribution $p(\mathbf{x})$ of the spots on the photographic plate can be calculated from the wavefunction ψ of the electron at the screen surface and is given by $p(\mathbf{x}) = |\psi^\dagger(\mathbf{x})\psi(\mathbf{x})|^2$. This is in fact a special case of the general *first property of quantum measurements*: the measurement of an observable A for a system in a state ψ gives as result one of the eigenvalues a_n, and the probability P_n to get the result a_n is given by

$$P_n = |\varphi_n^\dagger \psi|^2 \quad , \tag{437}$$

where φ_n is the eigenfunction of the operator A corresponding to the eigenvalue a_n. See page 757

Experiments also show a *second property of quantum measurements:* after the measurement, the observed quantum system is in the state φ_n corresponding to the measured eigenvalue a_n. One also says that during the measurement, the wavefunction has *collapsed* from ψ to φ_n. By the way, both properties can also be generalised to the more general cases with degenerate and continuous eigenvalues.

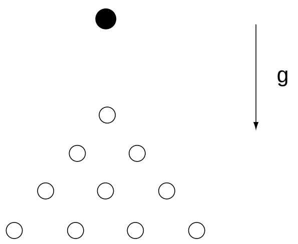


Figure 163 A system showing probabilistic behaviour

The sort of probabilities encountered in quantum theory is different, at first sight from the probabilities one encounters in everyday life. Roulette, dice, pachinko machines, the direction in which a pencil on its tip falls, have been measured experimentally to be random (assuming no cheating) to a high degree of accuracy. These systems do not puzzle us. We unconsciously assume that the random outcome is due to the small, but incontrollable variations of the starting conditions every time one repeats the experiment.*

But microscopic systems seem to be different. The two measurement properties just mentioned express what physicists observe in every experiment, even if the initial conditions are taken to be *exactly* the same every time. But why then is the position for a single electron, or most other observables of quantum systems, not predictable? In other words, what happens during the collapse of the wavefunction? How long does it take? In the beginning of quantum theory, there was the perception that the observed unpredictability is due to the lack of information about the state of the particle. This lead many to search for so-called 'hidden variables'; all these attempts were doomed to fail, however. It took some time for the scientific community to realize that the unpredictability is *not* due to the lack of information about the state of the particle, which is indeed described *completely* by the state vector ψ.

In order to uncover the origin of probabilities, let us recall the nature of a measurement, or better, of a general observation. *Any observation is the production of a record.* The record can be a visual or auditive memory in our brain, or a written record on paper, or a tape recording, or any such type of object. As explained in the intermezzo, an object is a record See page 375 if it cannot have arisen or disappeared by chance. To avoid the influence of chance, all records have to be protected as much as possible from the outer world; e.g. one typically puts archives in earthquake safe buildings with fire protection, keeps documents in a safe, avoids brain injury as much as possible, etc.

On top of this, records have to be protected from their internal fluctuations. These internal fluctuations are due to the fact that a record, being an object, consists of many components. If the fluctuations were too large, they would make it impossible to distinguish between the possible contents of a memory. Now, fluctuations decrease with increasing size of a system,

* To get a feeling for the limitations of these unconscious assumptions, you may want to read the story of those physicists who build a machine who could predict the outcome of a roulette ball from the initial velocity imparted by the croupier. The story is told by ... ...,

typically with the square root of the size. For example, if a hand writing is too small, it is difficult to read if the paper gets brittle; if the magnetic tracks on tapes are too small, they demagnetize and loose the stored information. In other words, a record is rendered stable against internal fluctuations by making it of sufficient size. Every record thus consists of many components and shows small fluctuations.

Therefore, every system with memory, i.e. every system capable of producing a record, contains a *bath*. In summary, the statement that any observation is the production of a record can be expressed more precisely as: *Any observation of a system is the result of an interaction between that system and a bath in the recording apparatus.*

But one can say more. Obviously, any observation measuring a physical quantity uses an interaction *depending* on that same quantity. With these seemingly trivial remarks, one can describe in more detail the process of observation, or as it is usually called in the quantum theory, the measurement process.

Any measurement apparatus, or *detector*, is characterized by two main aspects: the interaction it has with the microscopic system, and the bath it contains to produce the record. Any description of the measurement process thus is the description of the evolution of the microscopic system *and* the detector; therefore one needs the hamiltonian for the particle, the interaction hamiltonian, and the bath properties, such as the relaxation time. The interaction specifies what is measured, and the bath realizes the memory.

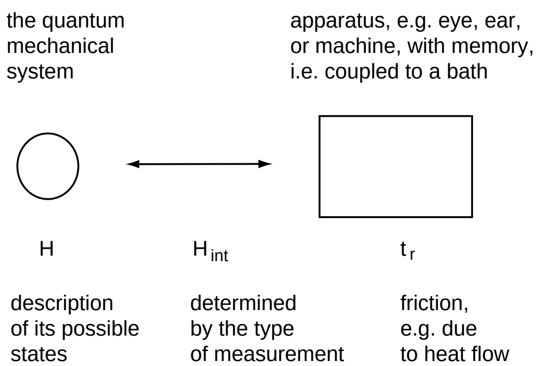


Figure 164 The concepts used in the description of measurements

We know that only classical thermodynamic systems can be irreversible; quantum systems are not. We therefore conclude: a measurement system *must* be described classically: otherwise it has no memory and is not measurement system: it produces no record! Nevertheless, let us see what happens if one describes the measurement system quantum mechanically. Let us call A the observable which is measured in the experiment and its eigenfunctions φ_n. We describe the quantum mechanical system under observation – often a particle – by a state ψ. This state can always be written as $\psi = \psi_p \psi_{\text{other}} = \sum_n c_n \varphi_n \psi_{\text{other}}$, where ψ_{other} represents the other degrees of freedom of the particle, i.e. those not described – *spanned*, in mathematical language – by the operator A corresponding to the observable we want to measure. The numbers $c_n = |\varphi_n^\dagger \psi_p|$ give the expansion of the state ψ_p, which is taken to be normalized, in terms of the basis φ_n. For example, in a typical position measurement, the functions φ_n would be the position eigenfunctions and ψ_{other} would contain the information about the momentum, the spin, and all other properties of the particle.

How does the system-detector interaction look like? Let us call the state of the apparatus before the measurement χ_{start}; the measurement apparatus itself, by definition, is a device which, when it is hit by a particle in the state $\varphi_n \psi_{\text{other}}$, changes from the state χ_{start} to the state χ_n. One then says that the apparatus has *measured* the eigenvalue a_n corresponding to the eigenfunction φ_n of the operator A. The index n is thus the record of the measurement;

it is called the *pointer* index or pointer variable. This index tells us in which state the microscopic system was before the interaction. The important point, taken from our previous discussion, is that the states χ_n, being records, are macroscopically distinct, precisely in the sense of the previous section. Otherwise they would not be records, and the interaction with the detector would not be a measurement.

Of course, during measurement, the apparatus sensitive to φ_n changes the part ψ_{other} of the particle state to some other situation $\psi_{\text{other},n}$, which depends on the measurement and on the apparatus; we do not need to specify it in the following discussion.* Let us have an intermediate check of or reasoning. Do apparatuses as described here exist? Yes, they do. For example, any photographic plate is a detector for the position of ionizing particles. A plate, and in general any apparatus measuring position, does this by changing its momentum in a way depending on the measured position: the electron on a photographic plate is stopped. In this case, χ_{start} is a white plate, φ_n would be a particle localised at spot n, χ_n is the function describing a plate blackened at spot n and $\psi_{\text{other,n}}$ describes the momentum and spin of the particle after it has hit the photographic plate at the spot n.

Now we are ready to look at the measurement process itself. For the moment, let us disregard the bath in the detector. Before the interaction between the particle and the detector, the combined system was in the initial state ψ_i given simply by

$$\psi_i = \psi_p \chi_{\text{start}} = \sum_n c_n \varphi_n \psi_{\text{other}} \chi_{\text{start}} \quad . \tag{440}$$

After the interaction, using the just mentioned characteristics of the apparatus, the combined state ψ_a is

$$\psi_a = \sum_n c_n \varphi_n \psi_{\text{other},n} \chi_n \quad . \tag{441}$$

This evolution from ψ_i to ψ_a follows from the evolution equation applied to the particle detector combination. The state ψ_a however, is a superposition of macroscopically distinct states, since it corresponds to a superposition of distinct states of the detector, which is a macroscopic apparatus following the definition given above. In our example ψ_a could correspond to a superposition of a state where a spot on the left upper corner is blackened on an otherwise white plate with one where a spot on the right lower corner of the otherwise white plate is blackened. Such a situation is never observed. This is due to the fact that the density matrix ρ_a of this situation, given by

$$\rho_a = \psi_a \otimes \psi_a^\dagger = \sum_{n,m} c_n c_m^* (\varphi_n \psi_{\text{other},n} \chi_n) \otimes (\varphi_m \psi_{\text{other},m} \chi_m)^\dagger \quad , \tag{442}$$

* How does the interaction look like mathematically? From the description we just gave, we specified the final state for every initial state. Since the two density matrices are related by

$$\rho_{\text{f}} = T \rho_{\text{i}} T^\dagger \tag{438}$$

Challenge — we can deduce the hamiltonian from the matrix T. Are you able to see how?

By the way, one can say in general that an apparatus measuring an observable A has a system interaction hamiltonian depending on the pointer variable A, and for which one has

$$[H + H_{int}, A] = 0 \quad . \tag{439}$$

contains non-diagonal terms, i.e. terms for $n \neq m$, whose numerical coefficients are different from zero.

Now let's take the bath back in. From the previous section we know the effect of a bath on such a macroscopic superposition. We found that a density matrix such as ρ_a decoheres extremely rapidly. We assume here that the decoherence time is negligibly small, in practice thus instantaneous,* so that the off-diagonal terms vanish, and only the the final, diagonal density matrix ρ_f, given by

$$\rho_f = \sum_n |c_n|^2 (\varphi_n \psi_{\text{other},n} \chi_n) \otimes (\varphi_n \psi_{\text{other},n} \chi_n)^\dagger \tag{443}$$

has experimental relevance. As explained above, such a density matrix describes a mixed state, and the numbers $P_n = |c_n|^2 = |\varphi_n^\dagger \psi_p|^2$ give the probability of measuring the value a_n and of finding the particle in the state $\varphi_n \psi_{\text{other,n}}$ as well as the detector in the state χ_n. But this is precisely what the two properties of quantum measurements state.

One finds therefore that describing a measurement as an evolution of a quantum system interacting with a macroscopic detector, itself containing a bath, one can *deduce* the two properties of quantum measurements, and thus the collapse of the wave function, from the quantum mechanical evolution equation. The decoherence time of the previous section becomes the time of collapse in the case of a measurement:

$$t_{\text{collapse}} = t_{\text{d}} < t_r \tag{444}$$

We thus have a formula for the time the wavefunction takes to collapse. The first experi-
Ref. 42 mental measurements of the time of collapse are appearing, and confirm these results.

Hidden variables

Obviously a large number of people is not satisfied with the train of thought just presented. They long for more mystery in quantum theory. The most famous approach is the idea that the probabilities are due to some hidden aspect of nature which is still unknown to humans. But the beautiful thing about quantum mechanics is that it allows both conceptual and experimental tests on whether such *hidden variables* exist without the need of knowing them.

• The first argument against hidden variables was given by John von Neumann.**

– CS – to be written – CS –

Ref. 43 • An additional no-go theorem for hidden variables was published by Kochen and Specker in 1967, (and independently by Bell in 1969). It states that noncontextual hidden variables are impossible, if the Hilbert space has a dimension equal or larger than three.

* Note however, that an *exactly* vanishing decoherence time, which would mean a *strictly* infinite number of degrees of freedom of the environment, is in contradiction with the evolution equation, and in particular with unitarity, locality and causality. It is essential in the whole argument not to confuse the logical consequences of a very small decoherence time with those of an exactly vanishing decoherence time.

** John von Neumann (1903, Budapest–1957, Washington DC) mathematician, father of the modern computer.

The theorem is about noncontextual variables, i.e. about hidden variables *inside* the quantum mechanical system. The Kochen-Specker theorem thus states that there is no noncontextual hidden variables model, because mathematics forbids it. This result essentially eliminates all possibilities, because usual quantum mechanical systems have dimensions much larger than three.

But also common sense eliminates hidden variables, without any recourse to mathematics, with an argument often overlooked. If a quantum mechanical system had internal hidden variables, the measurement apparatus would have zillions of them.* And that would mean that it could not work as a measurement system.

Of course, one cannot avoid noting that about *contextual* hidden variables, i.e. variables in the environment, there are no restricting theorems; indeed, their necessity was shown earlier in this section.

• Obviously, despite these results, people have also looked for experimental tests on hidden variables. Most tests are based on the famed *Bell's equation*, a beautifully simple relation published by John Bell** in the 1960s.

– CS – to be written – CS –

So far, all experimental checks of Bell's equation have confirmed standard quantum mechanics. No evidence for hidden variables has been found. This is not really surprising, since the search for such variables is based on a misunderstanding of quantum mechanics or on personal desires on how the world should be, instead of relying on experimental evidence.

Conclusions on probabilities and determinism

From the argument presented here, we draw a number of conclusions which we need for the rest of our escalation. Note that these conclusions are not shared by many physicists! The whole topic is touchy.

• Probabilities appear in measurements because the details of the state of the bath are unknown, not because the state of the quantum system is unknown. *Quantum mechanical probabilities are of statistical origin and are due to baths*. The probabilities are due to the large number of degrees of freedom contained in baths. These degrees of freedom make the outcome of experiments unpredictable. If the state of the bath were known, the outcome of an experiment could be predicted. The probabilities of quantum theory are "thermodynamic" in origin.

In other words, there are *no* fundamental probabilities in nature. All probabilities in nature are due to statistics of many particles. Modifying well-known words by Albert Einstein, "nature really does not play dice." We therefore called ψ the *wave function* instead of "probability amplitude", as is often done. 'State function' would be an even better name.

• Any observation in everyday life is a special case of decoherence. What is usually called the collapse of the wavefunction is a process due to the interaction with the bath present in any measuring apparatus. Because humans are warm-blooded and have memory, humans

* Which leads to the definition: one zillion is 10^{23}.

** John Stewart Bell (1928–1990), theoretical physicist.

themselves are thus measurement apparatuses. The fact that our body temperature is 37 °C is thus the reason that we see only a single world, and no superpositions.*

• A measurement is complete when the microscopic system has interacted with the bath in the measuring apparatus. Quantum theory as a description of nature does not require detectors; the evolution equation describes all examples of motion. However, *measurements* do require the existence of detectors; and detectors have to include a bath, i.e. have to be classical, macroscopic objects. In this context one speaks also of a *classical apparatus*. This necessity of the measurement apparatus to be classical had been already stressed in the very early stages of quantum theory.

• All measurements, being decoherence processes, are irreversible processes and increase entropy.

• A measurement is a special case of quantum mechanical evolution, namely the evolution for the combination of a quantum system, a macroscopic detector and the environment. Since the evolution equation is relativistically invariant, no causality problems appear in measurements, no locality problems and no logical problems.

• Since the evolution equation does not involve quantities other than space-time, hamiltonians and wave-functions, no other quantity plays a role in measurement. In particular, no observer nor any consciousness are involved or necessary. Every measurement is complete when the microscopic system has interacted with the bath in the apparatus. The decoherence inherent in every measurement takes place even if "nobody is looking." This trivial consequence is in agreement with the observations of everyday life, for example with the fact that the moon is orbiting the earth even if nobody looks at it.** Similarly, a tree falling in the middle of a forest makes noise even if nobody listens. Decoherence is independent of human observation, of the human mind, and of human existence.

• In every measurement the quantum system interacts with the detector. Since there is a minimum value for the magnitude of action, one cannot avoid the fact that *observation influences objects*. Therefore every measurement *disturbs* the quantum system. Any precise description of observations must also include the the description of this disturbance. In this section the disturbance was modeled by the change of the state of the system from ψ_{other} to $\psi_{\mathrm{other,n}}$. Without such a change of state, without a disturbance of the quantum system, a measurement is impossible.

• Since the complete measurement is described by Dirac's equation, unitarity is and remains the basic property of evolution in quantum mechanics. There are no non-unitary processes in quantum mechanics.

• The argument in this section for the description of the collapse of the wavefunction is an explanation exactly in the sense in which the term 'explanation' was defined in the intermezzo; it describes the relation between an observation and all the other aspects of reality, in this case the bath in the detector. The collapse of the wavefunction has been *explained*, it is not a question of "interpretation", i.e. of opinion, as unfortunately often is suggested.***

See page 419

Challenge

* Actually, there are more reasons; can you name a few?

** The opposite view is sometimes falsely attributed to Niels Bohr; the moon is obviously in contact with many radiation baths. Can you list a few?

Challenge

*** This implies that the so-called "many worlds" interpretation is wishful thinking. One also reaches this conclusion also when studying the details of this religious approach.

Ref. 44

• It is not useful to speculate whether the evolution for a *single* quantum measurement could be determined, if the state of the environment around the system were known. Measurements need baths. But baths cannot be described by wavefunctions.* Quantum mechanics is deterministic. Baths are probabilistic.

• In summary, there is *no* irrationality in quantum theory. Whoever uses quantum theory as argument for irrational behaviour, for ideologies, or for superstitions is guilty of disinformation. A famous example is the following quote.

Nobody understands quantum mechanics.
Richard Feynman

What is the difference between space and time?

This question can be answered by making it a little more specific: Why are objects localised in space but not in time?

Most bath-system interactions are mediated by a potential. All potentials are by definition position dependent. Therefore, every potential, being a function of the position **x**, commutes with the position observable (and thus with the interaction hamiltonian). The decoherence induced by baths – except if special care is taken – thus first of all destroys the non-diagonal elements for every superposition of states centered at different locations. In short, *objects are localized because they interact with baths via potentials.*

For the same reason, objects also have only one spatial orientation at a time. If the system bath interaction is spin-dependent, the bath leads to 'localisation' in the spin variable. This happens for all microscopic systems interacting with magnets. For this reason, one practically never observes macroscopic superpositions of magnetization. Since electrons, protons and neutrons have a magnetic moment and a spin, this conclusion can even be extended: everyday objects are never seen in superpositions of different rotation states, because of spin-dependent interactions with baths.

As a counterexample, most systems are not localized in time, but on the contrary exist for very long times, because practically all system-bath interaction do *not* commute with time. This is in fact the way a bath is defined to begin with. In short, *objects are permanent because they interact with baths.*

Are you able to find an interaction which is momentum dependent? What is the consequence for macroscopic systems? Challenge

In other words, in contrast to general relativity, quantum theory produces a distinction between space and time. In fact, one can *define* position as what commutes with interaction hamiltonians. This distinction between space and time is due to the properties of matter and its interactions; we could not have found this result in general relativity. Later, in the third part of our escalation, this result will become of central importance.

* This very strong type of determinism will be very much softened in the last part of this text, in which it will be shown that time is not a fundamental concept, and therefore that the debate around determinism looses most of its interest.

Are we good observers?

Are human classical apparatuses? Yes, they are. Even though several prominent physicists claim that free will and probabilities are related, a detailed investigation shows that this in not the case. Human brains are classical machines in the sense described above. Ref. 45

In addition, we have stressed several times that observing entities need memory, which means they need to incorporate a bath. That means that observers have to be made of matter; one cannot make an observer from radiation. Our description of nature is thus severely biased: we describe it from the standpoint of matter. That is a little like describing the stars by putting the earth at the centre of the universe. Can we eliminate this most basic anthropomorphism? We will discover this question in the third part of our escalation.

23. Some curiosities of quantum electrodynamics

Motion is an interesting topic, and when a curious person asks a question abut it, most of the time quantum electrodynamics is needed for the answer. Together with gravity, quantum electrodynamics explains almost all of our everyday experience, including numerous surprises. A few are touched upon here. Typeset in April 2001

• Quantum electrodynamics explains why there are only a *finite* number of different atom types. In fact, it takes only two lines to prove that pair production of electron-antielectron pairs make it impossible that a nucleus has more than about 137 protons. Can you show this? The effect at the basis of this limit, the polarisation of the vacuum, also plays a role in much larger systems, such a charged black holes, as we will see shortly. Ref. 19 Challenge See page 522

• Is there a critical magnetic field as there is a critical electric field? Challenge

• In classical physics, the field energy, and hence the mass, of a pointlike charged particle was predicted to be infinite. But QED effectively *smears out* the charge of the electron, so that in the end the field energy contributes only a small correction to its total mass. See page 346

• Microscopic evolution can be pretty slow. Light is always emitted from some metastable state. Usually, the decay times are much smaller than a microsecond. But there are metastable atomic states with a lifetime of ten years: for example, an ytterbium ion in the $^2F_{7/2}$ state achieves this value, because the emission of light requires an octupole transition, in which the angular momentum changes by $3\hbar$, an extremely unlikely process. Ref. 74

• Microscopic evolution can be pretty fast. Can you imagine how to deduce or to measure the speed of electrons inside atoms? And inside metals? Challenge

• Take a horseshoe. The distance between the two ends is not fixed, since otherwise their position and velocity would be known at the same time, contradicting the uncertainty relation. Of course, this reasoning is also valid for any other solid object. This inherent fuzziness of position and velocity can be seen as the main difference between the classical and the quantum domain. In short, also quantum mechanics, like special relativity, shows that rigid bodies do not exist, albeit for different reason.

• As Albert Einstein showed, every lamp whose brightness is turned up high enough will show special behaviour, after a certain intensity threshold is passed. Nowadays such a special lamp is called a *laser*. People have become pretty good at building lasers. They are used to cut metals, to operate people, to drill holes in teeth, to measure distances, and for many other applications. Even a laser made of a single atom (and two mirrors) has been built; in this example, only eleven photons were moving between the two mirrors on average. Quite a small lamp. Ref. 73

• Have you ever observed a quartz crystal? The beautiful shape and atomic arrangement formed spontaneously, as a result of the motion of atoms under high temperature and pressure, during the time that the material was deep under the earth's surface. The details of crystal formation are quite complex.

For example, are regular crystal lattices *energetically* optimal? This simple question leads to a wealth of problems. For example, one might start with the much simpler question whether a regular dense packaging of spheres is the most *dense* possible. Its density is $\pi/\sqrt{18}$, i.e. a bit over 74%. Even though this was conjectured to be the maximum possible value already in 1609, by Johannes Kepler, the statement was proven only in 1998, by Tom Challenge

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Did you find any mistakes? Challenge

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

Hales. The proof is difficult because in small volumes it is possible to pack spheres up to almost 78%. To show that over large volumes the lower value is correct is a tricky business. Ref. 75

Next, does a regular crystal of solid spheres, in which the spheres do *not* touch, have the lowest possible entropy? This simple problem has been the subject of research only in the 1990s. Interestingly, *for low energies*, regular sphere arrangements indeed show the largest possible entropy. Can you imagine why? Challenge

This result, and many similar ones deduced from the research into these so-called *entropic forces* show that the transition from solid to liquid is – at least in part – simply a geometrical effect. It is a beautiful example of how classical thinking can explain certain material properties, without any quantum theory. For the same reason, one gets the surprising result that even slightly repulsing spheres (or atoms) can form crystals, and melt at higher temperatures. Ref. 76

But the energetic side of crystal formation remains interesting. Quantum theory shows that it is possible that two atoms repel each other, while *three* attract each other. This beautiful effect was discovered and explained by Hans-Werner Fink in 1984. He studied rhenium atoms on tungsten surfaces and showed, as observed, that they cannot form dimers, but readily form trimers. This is an example contradicting classical physics; the effect is impossible if one pictures atoms as immutable spheres, but becomes possible when one remembers that the electron clouds around the atoms rearrange depending on their environment. Ref. 77

For an exact study of crystal energy, the interactions between all atoms have to be included. The simplemost question is to determine whether a regular array of alternatively charged spheres has lower energy than some irregular collection. Already such simple questions are still topic of research.

Then there is the topic of face formation in crystals. Can you confirm that the faces are those planes with the *slowest* growth speed, because all fast growing planes are eliminated? (Just use a paper drawing.) But the finer details of the process form a complete research field. Challenge Ref. 78

Finally, there remains the question of symmetry: why are crystals often symmetric, such as the famous snow-flakes, instead of asymmetric? This is a topic of selforganisation, as we mentioned already in the section of classical physics. See page 139

• A similar breadth of physical and mathematical problems are encountered in the study of liquids and of polymers. Ref. 79

• The ways people handle single atoms with electromagnetic fields is a beautiful example of modern applied technologies. Nowadays it is possible to levitate, to trap, to excite, to photograph, to deexcite, and to move single atoms just by shining light onto them. In 1997, the Nobel Prize has been awarded to the originators of the field.

• In 1997, a czech group built a quantum version of the Foucault pendulum, using the superfluidity of Helium. In this beautiful piece of research, they cooled a small ring of fluid helium below the temperature of 0.28 K, below which the helium moves without friction. In such situations it thus behaves like a Foucault pendulum. With a clever arrangement, it is possible to measure the rotation of the helium in the ring using phonon signals. Ref. 80

• An example of modern research is the study of hollow atoms, i.e. atoms missing a number of inner electrons. They have been discovered in 1990 by J.P. Briand and his group when a completely ionized atom, i.e. one without any electrons, is brought in contact with a metal. They can also be formed by intense laser irradiation. Ref. 81

• There are other differences between classical and quantum mechanics. In the past, the description of motion with formulas was taken rather seriously. Before computers appeared, only those examples of motions were studied which could be described with simple formulas. It turned out that galilean mechanics cannot solve the three-body problem, special relativity cannot solve the two-body problem, general relativity the one-body problem, and quantum field theory the zero-body problem. It took some time to the community of physicists to appreciate that understanding motion does not depend on the description by formulas, but on the description by clear equations based on space and time.

• Two observables can commute for two different reasons: either they are very *similar*, such as the coordinate x and x^2, or they are very *different*, such as the coordinate x and the
Challenge momentum p_y. Can you give an explanation?

Challenge • Can you explain why mud is not clear?

Challenge • Can the universe ever have been smaller than its Compton wavelength?

In fact, quantum electrodynamics, or QED, provides a vast number of curiosities, and every year there is at least one new discovery. We conclude the theme with a more general approach.

How can one move on perfect ice? – The ultimate physics test

In our escalation so far, we have encountered motion of many sorts. Therefore, the following test is the ultimate physics test; it allows to check your understanding and to compare it with that of others.

Imagine that you are on a perfectly frictionless surface, and that you want to move to its border. How many methods can you find to achieve this?

Classical physics provided quite a number of methods. We saw that for rotating ourselves, we just need to turn our arm above the head. For translation motion, throwing a shoe or inhaling vertically and exhaling horizontally are the simplest possibilities. Can you list at least six additional methods, maybe some using the fact that the surface is located the surface of the earth? What would you do in space?

Electrodynamics and thermodynamics taught us that in vacuum, heating one side of the body more than the other will work as motor; the imbalance of heat radiation will push you, albeit rather slowly. Are you able to find at least four other methods from these two
Challenge domains?

General relativity showed that turning one arm will emit gravitational radiation unsym-
Challenge metrically, leading to motion as well. Can you find at least two better methods?

Quantum theory offers a wealth of methods. Of course, quantum mechanics shows that one actually is always moving, since the uncertainty relation makes rest an impossibility. However, the average motion can be zero even if the spread increases with time. Are you
Challenge able to find at least four methods of moving on perfect ice due to quantum effects?

Material science, geophysics, atmospheric physics, and astrophysics also provide ways
Challenge to move, such as cosmic rays or solar neutrinos. Can you find four more methods?

Selforganisation, chaos theory, and biophysics also provide ways to move, when the inner
Challenge workings of the human body are taken into account. Can you find at least two methods?

Assuming that you read already the section following the present one, on the effects of semiclassical *quantum gravity*, here is an additional puzzle: Is it possible to move by accelerating a pocket mirror, using the emitted Unruh radiation? Can you find at least two other methods using quantum gravity effects to move yourself? Challenge

The marking of the exam is simple: for students, every working method gives one point. Eight points is ok, twelve points is good, sixteen points is very good, and twenty points or more is excellent.* For graduated physicists, the point is given only when a back-of-the-envelope estimate for the ensuing momentum or acceleration is provided.

Summary of quantum electrodynamics

The shortest possible summary of quantum electrodynamics is: everything, matter and images, is made of interacting particles. Charged particles interact through photon exchange, as described by figure 166.

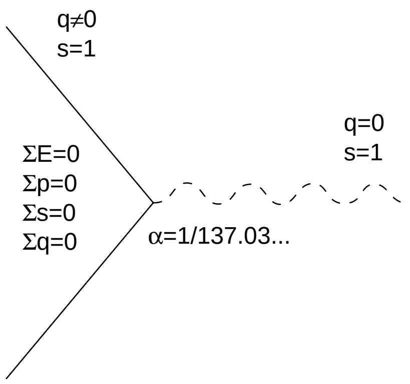


Figure 166 QED as perturbation theory in space-time

In a bit more detail, quantum electrodynamics starts with elementary *particles*, characterized by their mass, their spin, and their charge, and with the *vacuum*, essentially a sea of virtual particle-antiparticle pairs. The *interactions* between charged particles are described as the exchange of virtual photons, and *decay* is described as the interaction with the vacuum.

All physical results of QED can be calculated by using the diagram of figure 166, which describes the first order effects, and its composites, which describe the higher order effects. QED is a perturbative theory.

Quantum theory describes three properties of matter and radiation: the *divisibility* down to the smallest constituents, the *isolability* from the environment, and *impenetrability* of matter in contrast with the *penetrability* of radiation. These properties are due to electromagnetic interactions of constituents. Matter is divisible because the interactions are of finite strength, matter is separable because the interactions are of finite range, and matter is impenetrable because interactions among the constituents increase in intensity when they approach each other. Radiation divisible into photons, and is penetrable because of the lack of first order photon-photon interactions.

Elementary matter particles are indivisible, isolable, indistinguishable, and impenetrable, and distinct from antimatter. Photons are indivisible, isolable, indistinguishable, and penetrable.

Quantum electrodynamics with all its effects is necessary to describe the motion of small objects in all situations where the characteristic dimensions d are of the order of the Comp-

∗ The author keeps track of the answers on the motion mountain web site.

ton wavelength

$$d \approx \lambda_C = \frac{h}{m\,c} \tag{464}$$

In situations where the dimensions are of the order of the de Broglie wavelength, or equivalently, where the action is of the order of the Planck value, simple quantum mechanics is sufficient:

$$d \approx \lambda_{dB} = \frac{h}{m\,v} \tag{465}$$

For larger dimensions, classical physics will do.

Together with gravity, quantum electrodynamics explains almost all observations of motion on earth; QED unifies the description of matter and radiation in daily life. All objects made of atoms, and all images are described by it: their properties, their shape, their transformation, an their changes. Even selforganisation, chemical and biological evolution falls into this field. In other words, we now have a full grasp of the effects and the variety of motion due to electromagnetism.

Open questions in QED

Even though QED describes motion without any discrepancy from experiment, that does not mean that we understand every detail of every example of electric motion. For example, nobody has described the motion of an animal with QED yet.* In fact, there is beautiful and fascinating work going on in many branches of electromagnetism.

Atmospheric physics still provides many puzzles, and regularly delivers new, previously unknown phenomena. For example, the detailed mechanisms at the origin of auroras are still controversial; and the recent unexplained discoveries of discharges *above* clouds should not make one forget that even the precise mechanism of charge separation *inside* clouds, which leads to lightning, is not completely clarified. Ref. 83 Ref. 84

In fact all examples of electrification, such as the charging of amber through rubbing, the experiment which gave electricity its name, are still badly understood.

Material science in all its breadth, including the study of solids, fluids, and plasmas, as well as biology and medicine, still provide many topics of interest. The 21st century will undoubtedly be the century of the life sciences.

The research into the interaction of atoms with intense light is an example of present research in atomic physics. Strong lasers can strip atoms of many of their electrons; for such phenomena, there are not yet precise descriptions, since they do not comply to the weak field approximations usually assumed in physical experiments. In strong fields, new effects take place, such as the so-called Coulomb explosions. Ref. 86

But also the skies have their mysteries. In the topic of cosmic rays, it is still not clear how rays with energies of 10^{22} eV are produced outside the galaxy. Researchers are intensely trying to locate the electromagnetic fields necessary for their acceleration, and to understand their origin and working. Ref. 85

∗ On the other hand, there is beautiful work going on how humans move their limbs; it seems that humans move by combining a small set of fundamental motions. Ref. 82

In the theory of quantum electrodynamics, discoveries are expected by all those who study it in sufficient detail. For example, Dirk Kreimer has found that the complex interaction diagrams contain relations to the theory of knots. Ref. 87 This research topic will provide even more interesting results in the near future.

Results such as the relations to knot theory appear because QED is a *perturbative* description, with the vast richness of its nonperturbative effects still hidden. Studies of QED at high energies, where perturbation is *not* a good approximation, where particle numbers are not conserved, a wealth of new insights is expected. We will return to the topic later on.

High energy provides many more questions. So far, the description of motion was based on the idea that measurable quantities can be multiplied and added. In technical jargon, observables form an algebra. The idea behind this approach is that when two things happen at one spot, the result depends only on what happens at that spot. Every time one multiplies observables, one performs the operation at a given point in space or time. Thus the structure of an algebra contains, implies, and follows from the idea that local properties *lead* to local properties, and inversely, that local properties *derive* from local properties. We will discover later on that this basic assumption is wrong at high energies.

Many people are studying other topics not mentioned so far; in fact, by far the largest numbers of physicists get paid for some form of applied QED. But here we continue with our quest for the description of the *fundaments* of microscopic motion. So far, we have not achieved it. We need to understand motion in the realm of atomic nuclei. That is the topic of the next chapter. But before that, we take a first glimpse of the strange issues appearing when gravity and quantum theory meet.

24. Quantum mechanics with gravitation

Gravitation is a weak effect. Every seaman knows it: storms are the worst part of his life, not gravity. Nevertheless, including gravity into quantum mechanics yields quite a list of important issues.

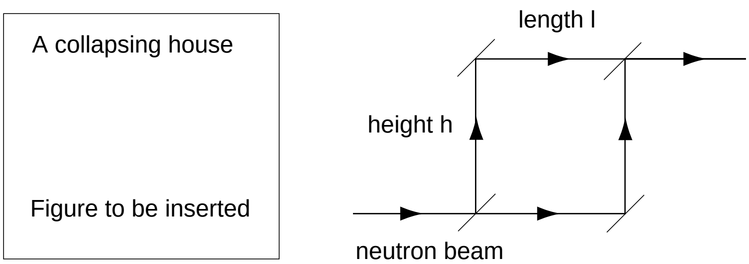


Figure 167 The weakness of gravitation

In the chapter on general relativity we already mentioned the change of light frequency with height. But gravity also changes the *phase* of wavefunctions of matter. Can you imagine how? Challenge The effect was first confirmed by the use of neutron interferometers, where neutron beams are brought to interference after having climbed a height at different locations. The experiment is shown schematically in figure 167; it confirmed the predicted phase difference phase change

$$\delta\varphi = \frac{mghl}{\hbar v} \tag{466}$$

where h is the height, l is the distance of the two climbs, and v is the speed of the neutrons. These beautifully simple experiments have confirmed the formula within experimental errors.*

Ref. 88

Ref. 90

Nowadays, similar experiments have even been performed with complete atoms. These modern set-ups allow to build interferometers so sensitive that local gravity g can be measured with a precision of more than eight significant digits.

Uncertainty relations

The uncertainty relation for momentum and position leads to an expression for the uncertainty of the metric tensor g in a region of size l, given by

$$\Delta g \sim \frac{{l_{\mathrm{Pl}}}^2}{l^2} \tag{467}$$

Challenge

Can you deduce it? Quantum theory thus shows that the metric tensor is a *fuzzy* quantity, like the momentum or the position of a particle.

But gravity leads to several much more spectacular quantum consequences.

Limits to disorder

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

See page 269

Ref. 91

We encountered this famous statement already. Interestingly, for over hundred years nobody asked whether there is a theoretical maximum for entropy, until Jakob Bekenstein answered the question in 1973, when he investigated the consequences gravity has for quantum physics. He found that the entropy of an object of energy E and size R is bound by

$$S \leqslant \frac{ER}{\hbar} \tag{468}$$

Challenge

for all physical systems. In particular, he deduced that (nonrotating) black holes saturate the bound, and have an entropy given by

$$S = \frac{kc^3}{G\hbar}\frac{A}{4} = \frac{kG}{\hbar c}4\pi M^2 \tag{469}$$

Challenge

where A is now the area of the *horizon* of the black hole given by $A = 4\pi R^2 = 4\pi(2GM/c^2)^2$. The result shows that every black hole has an entropy. As 't Hooft explains, this expression implies that the number of degrees of freedom of a black hole is about (but not exactly) one per Planck area.

For a rotating black hole,

$$S = \ldots \tag{470}$$

Ref. 89

* Due to the influence of gravity on phases of wavefunctions, some people who do not believe in bath induced decoherence, have even studied their influence on the decoherence process. Predictably, the results have not convinced.

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

The charge of a black hole enters only though the electrical field energy, in the mass value. Obviously, the black hole entropy is due to gravitation, to quantum physics, and to the horizon. So what are the different microstates leading to this macroscopic entropy?

It took many years to convince physicists that the microstates have to do with the various possible states of the horizon itself, and that they are due to the diffeomorphism invariance at this boundary. Ref. 92

As a note, the maximum entropy also gives a memory limit for memory chips. Can you imagine why? Challenge

If black holes have entropy, they must have a temperature. What does it mean? In fact, nobody believed this conclusion until two unrelated developments confirmed it within a few months. All these results were waiting to be discovered since the 1930s, but incredibly, nobody had thought about them for over 40 years.

How to measure acceleration with a thermometer – Davies-Unruh radiation

Also in 1973, Paul Davies and William Unruh independently made a theoretical discovery: Ref. 93 if an inertial observer observes that he is surrounded by vacuum, a second observer *accelerated* with respect to the first doesn't: he observes blackbody radiation, with a spectrum corresponding to the temperature

$$T = a \frac{\hbar}{2\pi kc} \qquad (471)$$

That also means that there is no vacuum on earth, because an observer on its surface can maintain that he is accelerated with $g = 9.8\,\mathrm{m/s^2} \neq 0$, leading to $T = 40\,\mathrm{zK}$! One can thus measure gravity, at least in principle, using a thermometer. However, the values are so small that it is questionable whether the effect will ever be confirmed. Ref. 94

– CS – more to be added – CS –

When this effect was discovered, people studied it from all sides. For example, it was found that the acceleration of a *mirror* leads to radiation emission. Mirrors are thus harder to accelerate than other bodies of the same mass.

Obviously, after the previous result one directly wants to know what is measured by an observer in rotational motion. In a few words, the result (471) also applies to him.

No particles

Gravity has another important consequence. It turns out that particles do not exist when the curvature radius of spacetime, instead of being larger than the Compton wavelength, becomes comparable to it. In such highly curved space-times, particles cannot be defined. Can you confirm the conclusion?

A rephrased large number hypothesis

The weakness of gravitation surfaces in unexpected places. For example, a number of curious coincidences can be found when quantum mechanics and gravitation are combined.

They are usually grouped under the heading 'large number hypotheses' because they usually involve large dimensionless numbers. For example, a less well known version connects
Ref. 107 the Planck length, the cosmic horizon, and the number of baryons:

$$(N_b)^3 \approx \left(\frac{R_o}{l_{Pl}}\right)^4 = \left(\frac{t_o}{t_{Pl}}\right)^4 \approx 10^{244} \qquad (472)$$

when using $N_b = 10^{81}$ and $t_o = 1.2 \cdot 10^{10}$ a. This is nothing new however; the approximate
Ref. 108 equality can be deduced from $Gn_b m_p = 1/t_o^2$, which, as Weinberg explains, is required by several cosmological models,* using the definition of the baryon density $n_b = N_b/R_o^3$, in
Ref. 108 combination with Dirac's large number hypothesis, substituting protons for pions, which conjectures that $m_p^3 = \hbar^2/(Gct_o)$.

The approximate equality (472) can be taken to suggest that certain microscopic properties, namely the mass of the proton, is connected to some general properties of the universe as a whole. This has lead to numerous discussions, especially since the time dependence of the two sides of (472) differs. Some people even speculate whether equation (472) expresses
Ref. 109 a relation between local and global topological properties. Up to this day, the only correct statement seems to be that it is a *coincidence* connected to the time at which we happen to live.

Black holes aren't black

In 1974, Stephen Hawking surprised the world of general relativity with a fundamental theoretical discovery. He found that if in the vacuum near the horizon a virtual particle-antiparticle pair appeared, there was the chance that one particle escapes and the antiparticle is captured by the black hole. From far away this looks like the emission of a particle. Hawking found that black holes radiate as black bodies.

The so-called Hawking temperature of a black hole of mass M turns out to be

$$T = \frac{\hbar c^3}{8\pi kGM} = \frac{\hbar}{2\pi kc}\, g_{surf} \quad \text{with} \quad g_{surf} = \frac{c^4}{4GM} \quad . \qquad (474)$$

For example, a black hole with the mass of the sun would have the rather small temperature of 62 nK, whereas a smaller black hole with the mass of a mountain, say 10^{12} kg, would have a temperature of 123 GK. That would make quite a good oven. However, in the general case the radiation is extremely small, as the emitted wavelength is of the order of the black hole
Challenge radius.

Challenge Anyway, this academic effect leads to a luminosity

$$L \sim \frac{1}{M^3} or L = nA\sigma T^4 = \frac{n\pi^3 k^4}{15c^2\hbar^3} T^4 \qquad (475)$$

where n is the number of particle degrees of freedom that can be radiated; if only photons are radiated, one has $n = 2$. (Actually, massless neutrinos are emitted in more frequently
Ref. 95 than photons.) *Black holes thus shine*, and the more the smaller they are. This is a genuine

See page 620 * Note that this relation can be rewritten simply as

$$m_o^2/R_o^2 \approx m_{Pl}^2/R_{Pl}^2 = c^4/G^2 \quad . \qquad (473)$$

quantum effect, since classically, black holes cannot emit any light. The radiation of black holes is often also called *Bekenstein-Hawking radiation*. Due to the emitted radiation, black holes gradually lose mass. Therefore their theoretical lifetime is *finite*. A short calculation shows that it is given by Challenge

$$t = M^3 \frac{20480\,\pi\, G^2}{\hbar c^4} \approx M^3\; 3.4 \cdot 10^{-16} \mathrm{s/kg}^3 \tag{476}$$

as function of their initial mass M. For example, a black hole with mass of 1 gram would have a lifetime of $3.4 \cdot 10^{-25}$ s, whereas a black hole of the mass of the sun, $2.0 \cdot 10^{30}$ kg, would have a lifetime of about 10^{68} years. Obviously, these numbers are purely academic. Anyway, *black holes evaporate*. However, this extremely slow process determines the lifetime only if no other, faster process comes into play. We will present a few shortly. In particular, Hawking radiation is the weakest of all known effects, and applies only for non-rotating, electrically neutral black holes with no matter falling into them from the surroundings. In practice, this situation never appears.

When this became clear, a beautiful Gedankenexperiment (thought experiment) was published by Unruh and Wald, showing that the whole result could have been deduced already 50 years earlier! Ref. 96

– CS – to be completed – CS –

So far, none of these formulas has been confirmed experimentally, as the values are much too small to be detected. But the deduction of a Hawking temperature is confirmed most beautifully by a theoretical discovery of Unruh, who found that there are configurations of fluids in which sound waves cannot escape, so-called '*silent holes*'. Consequently, these silent holes radiate sound waves with a temperature satisfying the same formula as real black holes. Another type of analog system, namely *optical black holes*, are presently being studied. Ref. 97 Ref. 98

Also in 1975, a much more dramatic radiation effect was predicted for *charged* black holes by Damour and Ruffini. Charged black holes have a much shorter lifetime than just presented, because during their formation, in the surrounding region in which the electric field is larger than the so-called vacuum polarization value, electron-positron pairs are prduced, which then almost all annihilate. This process reduces the charge of the black hole to a value for which the field is below critical everywhere, and reduces the mass by up to 30% in an extremely short time, much shorter than 10^{68} years. This process thus produces an extremely intense *gamma ray burst*. Ref. 99

Such gamma ray bursts had been discovered in the late 1960s by military satellites which were trying to spot nuclear exposions around the world through their gamma ray emission. The satellites found about two such bursts every day, coming from all over the sky. Another satellite, the Compton satellite, confirmed that they were extragalactical in origin, that their duration varied between a sixtieth of a second and about a thousand seconds. In 1996, the italian-dutch BeppoSAX satellite started mapping and measuring gamma ray bursts systematically. It discovered that they were followed by an *afterglow* of many hours, sometimes of days, both in the X-ray and in 1997 also in the optical domain. The satellite also allowed to find the corresponding X-ray, optical, and radio signals and sources for each burst. These Ref. 100

measurements in turn allowed to determine the distance of the burst sources; red shifts between 0.0085 and 4.5 were measured. In 1999 it was even possible to find the optical burst corresponding to the gamma ray one.* (Ref. 101)

All this data together shows that the bursts have energies in the range 10^{40} W up to $3 \cdot 10^{47}$ W; that is (almost) the same brightness as that of all stars of the whole visible universe taken together! Or put differently, that is the same energy released when converting several solar masses into radiation within a few seconds. In fact, the measured luminosity is near the theoretical maximum luminosity a body can have, namely

$$L < L_{\text{Pl}} = \frac{c^5}{2G} = 1.8 \cdot 10^{52}\,\text{W} \tag{477}$$

as you might want to check yourself. (Challenge) In short, the sources of gamma ray bursts are the biggest bombs found in the universe. (Ref. 101) In fact, more detailed investigations of experimental data confirm that gamma ray bursts are the primal screams of black holes in formation.

With all this new data, Ruffini took up his 1974 model again in 1997, and with his collaborators showed that the gamma ray bursts generated by the annihilation of electron-positrons pairs created by vacuum polarization, in the region they called the *dyadosphere*, have a luminosity and a duration exactly as measured, if a black hole of about a few up to 30 solar masses is assumed. Charged black holes therefore reduce their charge and mass through the vacuum polarization and electron positron pair creation process. (The process reduces the mass because it is one of the few processes which is *reversible*; in contrast, most other attempts to reduce charge on a black hole, e.g. by throwing in a particle with the opposite charge, increase the mass of the black hole, and are thus irreversible.) The left over remnant then can lose energy in various ways, and also turns out to be responsible for the after-glow discovered by the BeppoSAX satellite. Among others, Ruffini´s team speculates that the remnants are the sources for the high energy cosmic rays, whose origin had not been localized so far. All these exciting studies are still ongoing. (Ref. 102)

Other radiation processes for black holes appear when matter falls into the black hole and heats up, or when matter is ejected from rotating black holes through the Penrose process, (See page 284) or when charged particles fall into the black hole. These mechanisms are at the origin of *quasars*, the extremely bright quasi-stellar sources found all over the sky, might be related to black holes. The details of what happens there, the enormous voltages (up to 10^{20} V) and magnetic fields generated, as well as their effects on the surrounding matter are still objects of research.

Black hole material properties

Once the concept of entropy of a black hole was established, people started to think about black holes like about any other material. In this sense, black holes have a matter density, which can be defined by relating their mass to a fictitious volume defined by $4\pi R^3/3$. This density, given by

$$\rho = \frac{1}{M^2}\,\frac{3c^6}{32\pi G^3} \quad , \tag{478}$$

* For more about this fascinating topic, see the http://www.aip.de/˜ jcg/grb.html web site by Jochen Greiner.

can be quite low for large black holes. For the highest black holes known, with 1000 million solar masses or more, the density is of the order of the density of air. By the way, the gravitational acceleration at the horizon is still appreciable, as it is given by

$$g_{\mathrm{h}} = \frac{1}{M}\frac{c^4}{4G} = \frac{c^2}{2R} \qquad (479)$$

which is still $15\,\mathrm{km/s^2}$ for the air density black hole. Challenge

Obviously, the black hole temperature is related to the entropy as usual, namely by its definition

$$\frac{1}{T} = \left.\frac{\partial S}{\partial E}\right|_{\rho} = \left.\frac{\partial S}{\partial (Mc^2)}\right|_{\rho} \qquad (480)$$

All other thermal properties can be deduced by the standard relations from thermostatics.

For example, it turns out that black holes have a *negative* heat capacity, meaning that they cannot achieve equilibrium with a bath. However, this is not a surprise, since *any* gravitationally bound material system has negative specific heat.

Indeed, it takes only a bit of thinking to see that any gas or matter system collapsing under gravity follows $dE/dR > 0$ and $dS/dR > 0$. That means that while collapsing, the energy and the entropy of the system shrink. (Can you find out where they go?) Since temperature is defined as $1/T = dS/dE$, temperature is always positive; from the temperature increase $dT/dR < 0$ during collapse one deduces that the specific heat dE/dT is negative. Challenge Challenge Ref. 104

Black holes have no magnetic field, as was established already early on by russian physicists. We only mention that black holes have something akin to a finite electrical conductivity, and to a finite viscosity. These properties can be understood if the horizon is described as a membrane, even though this idea is not always applicable. In any case, one can treat, study and describe macroscopic black holes like any other macroscopic material body. Ref. 95 Ref. 103

How do black holes evaporate?

When a nonrotating and uncharged black hole loses mass by radiating Hawking radiation, eventually its mass reaches values approaching the Planck mass, namely a few micrograms. What happens in that final region, usually called *evaporation*? The expression (476) for the lifetime, applied to a black hole of Planck mass, yields a value of over sixty thousand Planck times. A surprising large value.

Now, a black hole approaching the Planck mass starts to be smaller than its own Compton wavelength; that means that it behaves like an elementary particle, and in particular, that quantum effects have to be taken into account. It is still unknown how evaporation takes place, whether the mass continues to diminish smoothly or in steps (e.g. with mass values decreasing as $\sqrt{n}$ when n approaches zero), how its internal structure changes, whether a stationary black hole starts to rotate (as the author predicts) or not, how the emitted radiation deviates from black body radiation, etc. There is enough to study for the future. However, one important issue has been settled.

The information paradox of black holes

When the thermal radiation of black holes was discovered, one question was hotly debated for many years. The matter forming a black hole can contain lots of information; e.g., imagine the black hole formed by a large number of books collapsing onto each other. On the other hand, a black hole radiates thermally until it evaporates. Since thermal radiation carries no information, it seems that information somehow disappears.

An incredible number of papers have been written about this problem, some even claiming that this example shows that physics as we know it is incorrect and needs to be changed. As usual, to settle the issue, one needs to look at it with precision, laying all prejudice aside. Three intermediate questions can help finding the answer.

- What happens when a book is thrown into the sun? When and how is the information radiated away?

- How precise is the sentence that black hole radiate *thermal* radiation? Could there be a slight deviation?

- Could the deviation be measured? In what way would black holes radiate information?

Challenge You might want to make up your own mind before reading on.

Let us walk through a short summary. When a book or any other highly complex object is thrown into the sun, the information contained is radiated away, and is contained in some slight deviations from black hole radiation, namely in slight correlations between the emitted radiation emitted over the burning time of the sun. A short calculation shows, like the comparison between the entropy of a room temperature book and the information contained in it, that these effects are extremely small, and difficult to measure.

Ref. 105 A clear exposition of the topic is given by Don Page. To make his point, he calculates what information one would measure in the radiation if the the system of black hole and radiation *together* would be in a *pure* state, i.e. a state containing specific information. He shows that even if a system is large (meaning consisting of many degrees of freedom) and in pure state, nevertheless any smaller subsystem looks almost perfectly thermal. More specifically, if a total system has a Hilbert space dimension $N = nm$, where n and $m \leqslant n$ are the dimensions of two subsystems, and if the total system is in a pure state, the subsystem m would have an

Challenge entropy S_m given by

$$S_m = \frac{1-m}{2n} + \sum_{k=n+1}^{mn} \frac{1}{k} \qquad (481)$$

which is approximately given by

$$S_m = \ln m - \frac{m}{2n} \quad \text{for} \quad m \gg 1 \quad . \qquad (482)$$

To discuss the result, let us think of n and m as counting degrees of freedom, instead of Hilbert space dimensions. The first term in equation (482) is the usual entropy of a mixed state. The second term is a small deviation and describes the amount of specific information contained in the original pure state; inserting numbers, one finds that it is extremely small compared to the first. In other words, the subsystem m is almost indistinguishable from a mixed or thermal one.

It turns out that the second, small term on the right of equation (482) is indeed sufficient to radiate away, during the lifetime of the black hole, any information contained in it. Page then goes on to show that the second term is so small that it not only lost in measurements, but that it is even lost in the usual, perturbative calculations about physical systems. The question whether any radiated information could be measured can now be answered directly. Don Page shows that even measuring half of the system only gives about 1/2 bit of that information. It is necessary to measure the complete system to measure all the contained information. Ref. 106

In summary, at every moment the radiated amount of information is negligible when compared with the total black hole radiation, and practically impossible to detect by measurements or even by usual calculations.

More paradoxes

A black hole is a macroscopic object, similar to a star. Like all objects, it can interact with its environment. Black holes have the special property to swallow everything that falls into them. This immediately leads one to ask if one can use this property to cheat around the usual everyday “laws” of nature. Some of them have been studied in the section on general relativity; here are a few more. See page 285

- The second law of thermodynamics was discussed already above. One can also look for methods to cheat around conservation of energy, of angular momentum, of charge. Challenge
- All these Gedankenexperiments come to the same conclusion: the maximum number of degrees of freedom in a region is proportional to the surface area of the region, and *not* to its volume. This intriguing result will keep us busy later on.
- If a twin falls into a black hole, he himself feels nothing. However, in his brother’s view he burned on the horizon. And as explained before, he burns while spreading all over the horizon. Adding quantum theory to the study of black holes thus shows that in contrast to what general relativity says, there is no way to cross the horizon unharmed. Challenge
- A black hole transforms matter into antimatter with a certain efficiency. Thus one might look for departures from particle number conservation. Are you able to find an example? Challenge

Quantum mechanics of gravitation

There is a conceptual step at this stage. We just looked at quantum theory *with* gravitation; now we have a glimpse at quantum theory *of* gravitation.

If one insists on a similarity between the gravitational “field” and the electromagnetic field, one can try to find its quantum description. Despite attack by many brilliant minds for almost a century, this approach was not successful.* Let us see why.

The gravitational Bohr atom

A short calculation shows that an electron circling a proton due to gravity alone, without electrostatic attraction, would do so at a distance of Challenge

$$r_{\text{gr.B.}} = \frac{\hbar^2}{G m_e^2 m_p} = 1.1 \cdot 10^{29}\,\text{m} \qquad (483)$$

* Modern approaches take another direction, as explained in the third part of the escalation.

which is about a thousand times the distance to the cosmic horizon. In fact, even in the normal hydrogen atom there is not a *single way* to measure gravitational effects. (Are you able to confirm this?) But why is gravity so weak? Or equivalently, why are the universe and normal atoms so much smaller than a gravitational Bohr atom? At the present point of our escalation these questions cannot be answered. Worse, the weakness of gravity even means that with high probability, future experiments will not provide *additional* data helping to decide among competing answers. Only careful thinking can help.

Challenge

Space-time foam

Quantum theory is based on the principle that actions below $\hbar/2$ cannot be observed. This implies that the observable values for the metric g in a region of size L are bound by

$$g \geqslant \frac{2\hbar G}{c^3}\frac{1}{L} \quad . \tag{484}$$

Can you confirm this? On other words, it is impossible to say what happens about the shape of space-time at extremely small dimensions. John Wheeler introduced the term *space-time foam* to describe this situation, and to make clear that space-time is not continuous nor a manifold in those domains. This issue will form the start of the third part of the escalation.

See page 581

Decoherence of space-time

If the gravitational field evolves like a quantum system, one encounters all issues found in other quantum systems. In particular, one can ask why no superpositions of different macroscopic space-times are observed.

Ref. 110

The calculation is easily performed for the simplest case of all, namely the superposition, in a cube of length l, of a homogenous gravitational field with value g and one with value g'. As in the case of superpositions of macroscopic distinct wavefunctions, one one finds a decay time for such a superposition. In particular, such a superposition decays when particles cross the volume, and the decay time is given by

See page 475

Challenge

$$t_d = \left(\frac{2kT}{\pi m}\right)^{3/2}\frac{nl^4}{(g-g')^2} \tag{485}$$

where n is the particle number density, kT their kinetic energy, and m their mass. Inserting typical numbers, one finds that the variations in gravitational field strength are *extremely* small. In fact, the numbers are so small that one can deduce that the gravitational field is the *first* variable which behaves classically in the history of the universe. This provides another reason that quantum gravity effects will probably never be detected. In short, matter not only tells space-time how to curve, it also tells it to behave with class.

Challenge

Do gravitons exist?

If the gravitational field is be treated quantum mechanically, like the electromagnetic field, its waves should be quantized. What kind of particles are associated with it? Most properties can be derived in a straightforward way.

The $1/r^2$ dependence of universal gravity, like that of electricity, means that the particles have *vanishing* mass and move at light speed.

The observation that gravity is always attractive, never repulsive, means that field quanta have *integer* and *even* spin. However, vanishing spin is ruled out, since it implies no coupling to energy. To comply with the fact that "all energy has gravity", one needs $S = 2$. In fact, one Ref. 111 can show that *only* the exchange of a massless spin 2 particle leads, in the classical limit, to general relativity.

The independence of gravity from electromagnetic effects implies a *vanishing* electric charge.

The coupling strength of gravity, corresponding to the fine structure constant of electromagnetism, is given either by

$$\alpha_{G1} = \frac{G}{\hbar c} = \mathrm{kg}^{-2} \quad \text{or by} \quad \alpha_{G2} = \frac{Gmm}{\hbar c} = \left(\frac{m}{m_{\mathrm{Pl}}}\right)^2 = \left(\frac{E}{E_{\mathrm{Pl}}}\right)^2 \qquad (486)$$

However, the first expression is not a pure number; the second expression is, but it depends on the mass one inserts. These difficulties reflect the fact that gravity is not properly speaking an interaction, as was explained in the section on general relativity. One often argues that m should be taken as the value corresponding to the energy of the system in question. For everyday life, typical energies are 1 eV, leading to a value $\alpha_{G2} \approx 1/10^{56}$. Gravity is weak compared to electromagnetism, for which $\alpha_{em} = 1/137.04$.

Up to this day, the so-called *graviton* has not yet been detected, and there is in fact little hope that it ever will. In fact, the problems with the coupling constant probably make it impossible to construct a *renormalizable* theory of gravity, and the lack of renormalization means the impossibility to define a perturbation expansion, and thus to define particles in the usual way.

No cheating any more

Already this short excursion into the theory of quantum gravitation shows that a lot of trouble is waiting there. The reason is that up to now, we deluded ourselves. In fact, we cheated. We hid a simple fact: quantum theory and general relativity *contradict* each other. That was the reason that we stepped back to special relativity before we started exploring quantum theory. In this way we avoided all problems, as quantum theory does not contradict *special* relativity, but does contradict *general* relativity. In fact the contradictions are staggering. The issues are so dramatic, changing everything from the basis of classical physics onwards, that we devote the start of the third part to their exploration. The situation will have incredible consequences on the nature of mass, charge, spin, and motion. But before we arrive there, we complete the theme of the the present, second part of the escalation, namely the essence of matter and interactions.

References

1 F. RIEKE & D.A. BAYLOR, *Single-photon detection by rod cells of the retina*, Reviews of Modern Physics **70**, pp. 1027–1036, 1998. They also mention that the eye usually works at photon fluxes between 10^8 /µm^{2}s (sunlight) and 10^{-2} /µm^{2}s (starlight). The cones detect, in color, light intensities in the uppermost 7 or 8 decades, whereas the rods detect, in black and white, the lower light intensities. Cited on page 434, 434.

2 For a detailed account, See Jeffrey J. PRENTIS, *Poincaré's proof of the quantum discontinuity of nature*, American Journal of Physics **63**, pp. 339–350, 1995. The original papers are Henri POINCARÉ, *Sur la théorie des quanta*, Comptes Rendus de l'Académie des Sciences (Paris) **153**, pp. 1103–1108, 1911, as well as Henri POINCARÉ, *Sur la théorie des quanta*, Journal de Physique (Paris) **2**, pp. 5–34, 1912. Cited on page 436.

3 J. JACOBSON, G. BJÖRK, I. CHANG & Y. YAMAMOTO, *Photonic de Broglie waves*, Physical Review Letters **74**, pp. 4835–4838, 1995. The first measurement was published by E.J.S. FONSECA, C.H. MONTEN & S. DE PÁDUA, *Measurement of the de Broglie wavelength of a multiphoton wave packet*, Physical Review Letters **82**, pp. 2868–2671, 1995. Cited on page 437.

4 Wolfgang TITTEL, J. BRENDEL, H. ZBINDEN & N. GISIN, *Violation of bell inequalities by photons more than 10 km apart*, Physical Review Letters **81**, pp. 3563–3566, 26 October 1998. Cited on page 438.

5 Giuseppe FUMAGALLI, *Chi l'ha detto?*, Hoepli, 1983. Cited on page 440.

6 Louis de Broglie ... Cited on page 443.

7 M. ARNDT, O. NAIRZ, J. VOS-ANDREAE, C. KELLER, G. VAN DER ZOUW & A. ZEILINGER, *Wave-particle duality of C60 molecules*, Nature **401**, pp. 680–682, 14 October 1999. Cited on page 443.

8 See for example the papers by Jan HILGEVOORD, *The uncertainty principle for energy and time*, American Journal of Physics **64**, pp. 1451–1456, 1996, and by Paul BUSCH, *On the time-energy uncertainty reaction, part 1 & 2*, Foundations of Physics **20**, pp. 1–43, 1990. A classic is the paper by Eugene P. WIGNER, *On the time-energy uncertainty reaction*, in Abdus SALAM & Eugene P. WIGNER, editors, *Aspects of quantum theory*, Cambridge University Press, 1972. Cited on page 444.

9 P. CARRUTHERS & M.M. NIETO, *Phase and angle variables in quantum mechanics*, Review of Modern Physics **40**, pp. 411–440, 1968. Cited on page 444.

10 The uncertainty principle for rotational motion is well explained by ... Cited on page 444.

11 E.I. BUTIKOV, *The rigid pendulum – an antique but evergreen physical model*, European Journal of Physics **20**, pp. 429–441, 1999. Cited on page 444.

12 The whole bunch of atoms behaves as one single molecule; one speaks of a Bose-Einstein condensate. This was observed e.g. by M.H. ANDERSON & al., *Observation of Bose-Einstein condensation in a dilute atomic vapour*, Science **269**, p. 198›201, 1995. Cited on page 449.

13 W. GERLACH & O. STERN, *Der experimentelle Nachweis des magnetischen Moments des Silberatoms*, Zeitschrift für Physik **8**, p. 110, 1921. See also the pedagogical explanation by M. HANNOUT, S. HOYT, A. KRYOWONOS & A. WIDOM, *Quantum measurement and the Stern-Gerlach experiment*, American Journal of Physics **66**, pp. 377–379, 1995. Cited on page 450.

14 R.H. DICKE & J.P. WITTKE, *Introduction to quantum theory*, Addison-Wesley, Reading, Massachusetts, 1960. Cited on page 450.

15 Y. AHARONOV & D. BOHM, ..., Physical Review **115**, p. 485, 1959. Cited on page .

16 K. von MEYENN & E. SCHUCKING, *Wolfgang Pauli*, Physics Today pp. 43–48, February 2001. Cited on page 451.

17 L.L. FOLDY & S.A. WOUTHUYSEN, *On the Dirac theory of spin 12 particles and its non-relativistic limit*, Physical Review **78**, pp. 29–36, 1950. Cited on page 451.

18 John P. COSTELLA & Bruce H.J. MC KELLAR, *The Foldy-Wouthuysen transformation*, American Journal of Physics **63**, pp. 1119–1121, 1995. Cited on page 452.

19 W. Heisenberg, ... Euler/ ...1935. See also J. SCHWINGER, ...1951. Cited on page 453, 512.

20 D.L. BURKE & al., *Positron production in multiphoton light-by-light scattering*, Physical Review Letters **79**, pp. 1626–1629, 1997. A simple summary is given in Bertram SCHWARZSCHILD, *Gamma rays create matter just by plowing into laser light*, Physics Today **51**, pp. 17–18, February 1998. Cited on page 454.

21 About tachyons, see ... Cited on page 455.

22 The g-factors for composite nuclei are explained by ...See also Hans DEHMELT, *Is the electron a composite particle?*, Hyperfine Interactions **81**, pp. 1–3, 1993. Cited on page 455.

23 The nearest to an image of a hydrogen atom is found in A. YAZDANI, *Watching an atom tunnel*, Nature **409**, pp. 471–472, 2001. Cited on page 456.

24 See e.g. the reprints of his papers in the standard collection by J.A. WHEELER & W.H. ZUREK, *Quantum theory and measurement*, Princeton University Press, 1983. Cited on page 473.

25 H.D. ZEH, *On the interpretation of measurement in quantum theory*, Foundations of Physics **1**, pp. 69–76, 1970. Cited on page 473.

26 Linda REICHL, *A modern course in statistical physics*, Wiley, 2nd edition, 1998. An excellent introduction into thermodynamics. Cited on page 474.

27 E. JOOS & H.D. ZEH, *The emergence of classical properties through interactions with the environment*, Zeitschrift für Physik B **59**, pp. 223–2443, 1985. Cited on page 476.

28 M. TEGMARK, *Apparent wave function collapse*, Foundation of Physics Letters, **6**, pp. 571–590, 1993. Cited on page 476.

29 The decoherence time is bound from above by the relaxation time. See A.O. CALDEIRA & A.J. LEGGETT, *Influence of damping on quantum interference: an exactly soluble model*, Physical Review A **31**, 1985, pp. 1059–1066. This is the main reference about effects of decoherence for a harmonic oscillator. The general approach to relate decoherence to the influence of the environment is due to Niels Bohr, and has been pursued in detail by Dieter Zeh. Cited on page 476.

30 G. LINDBLAD, *On the generators of quantum dynamical subgroups*, Communications on Mathematical Physics **48**, pp. 119–130, 1976. Cited on page 477.

31 W.H. ZUREK, *Decoherence and the transition from quantum to classical*, Physics Today pp. 36–44, October 1991. An easy but somewhat confusing article. His reply to the numerous letters of response in Physics Today, April 1993, pp. 13–15, and pp. 81–90, exposes his ideas in a clearer way, and gives a taste of the heated discussions on this topic. Cited on page 477, 481.

32 John BARDEEN, his talk in Japan. Cited on page 478.

33 K. BAUMANN, *Quantenmechanik und Objektivierbarkeit*, Zeitschrift für Naturforschung **25a**, pp. 1954–1956, 1970. Cited on page 478.

34 David BOHM, *Quantum Theory*, Prentice-Hall, 1951, pp. 614–622. Cited on page 480.

35 Albert EINSTEIN, Boris PODOLSKY & Nathan ROSEN, *Can quantum-mechanical description of reality be considered complete?*, Physical Review **48**, pp. 696–702, 1935. Cited on page 480.

36 Alain ASPECT, J. DALIBARD & G. ROGER, *Experimental tests of Bell's inequalities using time-varying analysers*, Physical Review Letters **49**, pp. 1804–1807, 1982, Cited on page 480.

37 John MADDOX, ... Cited on page 481.

38 The first experimental measurement of superpositions of left and right flowing currents is ... In the year 2000, superpositions of 1 μA clockwise and anticlockwise have been detected; for more details, see J.R. FRIEDMAN & al., *Quantum superposition of distinct macroscopic states*, Nature **406**, p. 43, 2000. Cited on page 481.

39 ... ..., On the superposition of magnetization in up and down directions. Cited on page 481.

40 S. HAROCHE, *Entanglement, decoherence, and the quantum-classical transition*, Physics Today **51**, pp. 36–42, July 1998. Cited on page 481.

41 M.R. ANDRES & al., *Observations of interference between two Bose condensates*, Science **275**, pp. 637–641, 31 January 1997. See also the www.aip.org/physnews/special.htm web site. Cited on page 481.

42 Collapse times have been measured ... Cited on page 486.

43 S. KOCHEN & E.P. SPECKER, *The problem of hidden variables in quantum mechanics*, **17**, pp. 59–87, 1967. Cited on page 486.

44 BRYCE DE WITT & al., *The many-worlds interpretation of quantum mechanics*, Princeton University Press, 1973. This interpretation talks about entities which cannot be observed, namely the many worlds, and often assumes that the wavefunction of the universe exists. Both habits are beliefs, and in contrast with facts. Cited on page 488.

45 Max TEGMARK, *The importance of quantum decoherence in brain processes*, Physical Review D **61**, pp. 4194–4206, 2000, or also http://arxiv.org/quant-ph/9907009. Cited on page 490.

46 The most famous reference on the wave function collapse is chapter IV of the book by K. GOTTFRIED, *Quantum mechanics*, Benjamin, New York, 1966, It is the favorite reference by Victor Weisskopf, and cited by him on every occasion he talks about the topic. Cited on page .

47 E. FERMI, *Quantum theory of radiation*, Review of Modern Physics **4**, p. 87, 1932. Cited on page .

48 G. HEGERFELDT, *Causality problems for Fermi's two-atom system*, Physical Review Letters **72**, pp. 596–599, 1994, Cited on page .

49 Recent experiments on the subject of quantum tunneling in magnetic systems are described in D.D. AWSCHALOM, J.F. SMITH, G. GRINSTEIN, D.P. DIVICENZO & D. LOSS, *Macroscopic quantum tunneling in magnetic proteins*, Physical Review Letters **88**, pp. 3092–3095, 1992, and in C. PAULSEN & al., *Macroscopic quantum tunneling effects of Bloch walls in small ferromagnetic particles*, Europhysics Letters **19**, pp. 643–648, 1992. Cited on page .

50 The prediction that quantum tunneling could be observable when the dissipative interaction with the rest of the world is small enough was made by Leggett; the topic is reviewed in A.J. LEGGETT, S. CHAHRAVARTY, A.T. DORSEY, M.P.A. FISHER, A. GARG & W. ZWERGER, *Dynamics of dissipative 2-state systems*, Review of Modern Physics **59**, pp. 1–85, 1987. Cited on page .

51 It was verified in Josephson junctions by R.F. VOSS & R.A. WEBB, *Macroscopic quantum tunneling in 1 mm Nb Josephson junctions*, Physical Review Letters **47**, pp. 265–268, 1981, Cited on page .

52 The beautiful analogy between optics and two-dimensional electron systems, in GaAs at helium temperature, has been studied in detail by C.W.J. BEENAKKER & H. VAN HOUTEN, for example in their review in *Solid State Physics*, **44**, pp. 1–228, 1991, Cited on page .

53 The use of radioactivity for breeding is described by ... Cited on page 492.

54 Reviews on the results and ideas on this topic are summarized by Barbara GOSS LEVI, *Measured steps advance the understanding of molecular motors*, Physics Today pp. 17–19, April 1995, as well as ... Cited on page 493, 493.

55 R. BARTUSSEK, P.HÄNGGI, *Brownsche Motoren*, Physikalische Blätter **51**, pp. 506–507, 1995. Cited on page 493.

56 For good figures of atomic orbitals, see for example the chemistry text by ... Cited on page 496.

57 For pictures of the orbitals of dangling bonds, see for example F. GIESSIBL & al., Science **28**, p. 422, 2000. Cited on page 496.

58 The group of John Wearden in Manchester has shown by experiments with persons that the accuracy of a few percent is possible for any action of duration between a tenth of a second and a few minutes. See John MCCRONE, *When a second lasts forever*, New Scientist pp. 53–56, 1 November 1997. Cited on page 497.

59 E.J. ZIMMERMAN, American Journal of Physics **30**, 1962, pp. 97–105. Cited on page 498.

60 Peter D. PEŞ I´C, *The smallest clock*, European Journal of Physics **14**, pp. 90–92, 1993. Cited on page 499.

61 A pretty example of a quantum mechanical system showing exponential behaviour at all times is given by H. NAKAZATO, M. NAMIKI & S. PASCAZIO, *Exponential behaviour of a quantum system in a macroscopic medium*, Physical Review Letters **73**, pp. 1063–1066, 1994. Cited on page 500.

62 See the delightful book about the topic by Paolo FACCHI & Saverio PASCAZIO, *La regola d'oro di Fermi*, Bibliopolis, 1999. An experiment observing deviations at short times is S.R. WILKINSON, C.F. BHARUCHA, M.C. FISCHER, K.W. MADISON, P.R. MORROW, Q. NIU, B. SUNDARAM & M.G. RAIZEN, Nature **387**, p. 575, 1997. Cited on page 500.

63 H.B.G. CASIMIR, *On the attraction between two perfectly conducting bodies*, Proc. Kon. Ned. Acad. Wetensch. **51**, p. 793?795, 1948. Cited on page .

64 H.B.G. CASIMIR & D. POLDER, *The influence of retardation on the London-Van der Waals forces*, Physical Review **73**, pp. 360–372, 1948. Cited on page .

65 S.L. BOERSMA, *A maritime analogy of the Casimir effect*, American Journal of Physics **64**, pp. 539–541, 1996. Cited on page 507.

66 A. LARRAZA & B. DENARDO, *An acoustic Casimir effect*, Physics Letters A **248**, p. 151? 155, 1998. Cited on page 508.

67 M.J. SPARNAAY, Nature **180**, p. 334, 1957. M.J. SPARNAAY, *Measurements of attractive forces between flat plates*, Physica (Utrecht) **24**, pp. 751–764, 1958. Cited on page 508, 508.

68 S.K. LAMOREAUX, *Demonstration of the Casimir force in the 0.6 to 6 µm range*, Physical Review Letters **78**, pp. 5–8, 1997. U. MOHIDEEN & A. ROY, *Precision Measurement of the Casimir Force from 0.1 to 0.9 µm*, Physical Review Letters **81**, pp. 4549–4552, 1998. Cited on page 508.

69 M. KARDAR, R. GOLESTANIAN, *The "friction" of the vacuum, and other fluctuation-induced forces*, Reviews of Modern Physics **71**, pp. 1233–1245, 1999. Cited on page 509.

70 G. GOUR & L. SRIRAMKUMAR, *Will small particles exhibit Brownian motion in the quantum vacuum?*, http://www.lanl.gov/quant/phys/9808032 Cited on page 509.

71 P.A.M. DIRAC, ..., 1985. This article transcribes a speech by Paul Dirac just before his death. Disturbed by the infinities of quantum electrodynamics, he rejects as wrong the theory which he himself found, even though it correctly describes all experiments. The humility of a great man who is so dismissive of his main and great achievement was impressive when hearing him, and still is impressive when reading the speech. Cited on page 510.

72 N. BOHR & L. ROSENFELD, *Field and charge measurements in quantum electrodynamics*, Physical Review **78**, pp. 794–798, 1950. Cited on page .

73 K. An, J.J. Childs, R.R. Desari & M.S. Field, *Microlaser: a laser with one atom in an optical resonator*, Physical Review Letters **73**, pp. 3375–3378, 19 December 1994. Cited on page 512.

74 M. Roberts & al., *Observation of an electric octupole transition in a single ion*, Physical Review Letters **78**, pp. 1876–1879, 1997. A lifetime of 3700 days was determined. Cited on page 512.

75 See the web site by Tom Hales on www.math.lsa.umich.edu/~ hales/countdown. Cited on page 514.

76 For a short overview, also covering binary systems, see C. Bechinger, H.-H. von Grünberg & P. Leiderer, *Entropische Kräfte - warum sich repulsiv wechselwirkende Teilchen anziehen können*, Physikalische Blätter **55**, p. 53'56, 1999. Cited on page 514.

77 Hans-Werner Fink, *Direct observation of three-body interactions in adsorbed layers: Re on W(111)*, Physical Review Letters **52**, pp. 1532–1534, 1984. Cited on page 514.

78 See for example the textbook by A. Pimpinelli & J. Villain, *Physics of crystal growth*, Cambridge University Press, 1998. Cited on page 514.

79 P.G. de Gennes, *Soft matter*, Reviews of Modern Physics **63**, p. 645, 1992. Cited on page 514.

80 The fluid helium based quantum gyroscope 1: czech group, 2: berkley group, Nature 10. April 1997 Cited on page 514.

81 J.P. Briand & al., *Production of hollow atoms by the excitation of highly charged ionas in interaction with a metallic surface*, Physical Review Letters **65**, pp. 159–162, 1990. Cited on page 515.

See also C. Rhodes, *Hollow atoms*, ..., and ...Physikalische Blätter Oktober 1996.

82 The present knowledge on how humans move is summarizedby ... Cited on page 517.

83 See e.g. D.A. Bryant, *Electron acceleration in the aurora*, Contemporary Physics **35**, pp. 165–179, 1994. Cited on page 517.

84 The discharges above clouds are described e.g. by ... Cited on page 517.

85 See e.g. L.O'C. Drury, *Acceleration of cosmic rays*, Contemporary Physics **35**, pp. 231–242, 1994. Cited on page 517.

86 About Coulomb explosions, see ... Cited on page 517.

87 See the text by Dirk Kreimer, ..., 2000. Cited on page 518.

88 R. Colella, A.W. Overhauser & S.A. Werner, *Observation of gravitationally induced quantum mechanics*, Physical Review Letters **34**, pp. 1472–1475, 1975. This experiment is usually called the COW experiment. Cited on page 519.

89 G.C. Ghirardi, A. Rimini & T. Weber, *Unified dynamics for microscopic and macroscopic systems*, Physical Review D **34**, pp. 470–491, 1986. Cited on page 519.

90 Ch.J. Bordé & C. Lämmerzahl, *Atominterferometrie und Gravitation*, Physikalische Blätter **52**, pp. 238–240, 1996. See also A. Peters, K.Y. Chung & S. Chu, *Observation of gravitational acceleration by dropping atoms*, Nature **400**, pp. 849–852, 26 August 1999. Cited on page 519.

91 Jakob Bekenstein, *Black holes and entropy*, Physical Review D p. 2333, 1973. Cited on page 519.

92 This is clearly argued by S. Carlip, *Black hole entropy from horizon conformal field theory*, Nuclear Physics B Proceedings Supplement **88**, pp. 10–16, 2000. Cited on page 520.

93 Paul Davies, ..., William G. Unruh, *Notes on black hole evaporation*, Physical Review D **14**, p. 870'892, 1976. Cited on page 520.

94 About the possibility to measure Davies-Unruh radiation, see for example the paper ..., http://arxiv.org:gr-qc/9605032. Cited on page 520.

95 For a delightful popular account from a black hole expert, see Igor NOVIKOV, *Black holes and the universe*, Cambridge University Press 1990. Cited on page 521, 524.

96 W.G. UNRUH, R.M. WALD, ..., Physical Review D **25**, pp. 942–958 , 1982. Cited on page 522.

97 The original paper is W.G. UNRUH, *Experimental black hole evaporation?*, Physical Review Letters **46**, pp. 1351–1353, 1981. A good explanation with good literature overview is the one by Matt VISSER, *Acoustical black holes: horizons, ergospheres and Hawking radiation*, http://xxx.lanl.gov/abs/gr-qc/9712010. Cited on page 522.

98 Optical black holes ... Cited on page 522.

99 T. DAMOUR & R. RUFFINI, Physical Review Letters **35**, p. 463, 1975. Cited on page 522.

100 These were the Vela satellites; their existence and results were announced officially only in 1974, even though they were working already for many years. Cited on page 522.

101 The best general introduction into the topic of gamma ray bursts is S. KLOSE, J. GREINER & D. HARTMANN, *Kosmische Gammastrahlenausbrüche – Beobachtungen und Modelle*, Teil I und II, Sterne und Weltraum March and April 2001. Cited on page 523, 523.

102 When the gamma ray burst encounters the matter around the black hole, it is broadened. The larger the amount of matter, the broader the pulse is. See G. PREPARATA, R. RUFFINI & S.-S. XUE, A & A **338**, p. *L*87, 1998, and R. RUFFINI, J.D. SALMONSON, J.R. WILSON & S.-S. XUE, A & A **350**, pp. 334–343, 1999. For a very personal account by Ruffini on his involvement in gamma ray bursts, see his paper *Black hole formation and gamma ray bursts*, http://arxiv.org:astro-ph/0001425. Cited on page 523.

103 About the membrane despription of black holes, see ... Cited on page 524.

104 Negative heat capacity has also been found in atom clusters and in nuclei. See e.g. XX: SCHMIDT & al., Physical Review Letters 12 February 2001. Cited on page 524.

105 He wrote a series of papers on the topic; a beautiful summary is Don N. PAGE, *How fast does a black hole radiate information?*, International Journal of Modern Physics **3**, pp. 93–106, 1994, based on his earlier papers, such as *Information in black hole radiation*, Physical Review Letters **71**, pp. 3743–3746, 1993, and *Average entropy of a subsystem*, Physical Review Letters **71**, pp. 1291–1294, 1993. The entropy formula of the latter paper, used above, was proven by S.K. FOONG & S. KANNO, *Proof of Page's conjecture on the average entropy of a subsystem*, Physical Review Letters **72**, pp. 1148–1151, 1994. (See also the computer preprint http://xxx.lanl.gov/abs/hep-th/9305040.) Cited on page 525.

106 See Don N. PAGE, ..., Cited on page 526.

107 The author speculates that this version of the coincidences could be original; he has not found it in the literature. Cited on page 521.

108 Steven WEINBERG, *Gravitation and Cosmology*, (Wiley, 1972). See equation 16.4.3 on page 619 and also page 620. Cited on page 521, 521.

109 It could be that knot theory provides a relation between local knot invariant and a global one. Cited on page 521.

110 E. JOOS, *Why do we observe a classical space-time?*, Physics Letters A **116**, pp. 6–8, 1986. Cited on page 527.

111 This point was made repeatedly by Steven WEINBERG, namely in ..., Physics Letters **9**, p. 357, 1964, in .., Physical Review B **135**, p. 1049, 1964, and in .., Physical Review B **138**, p. 988, 1965. Cited on page 528.

—

112 An overview on squeezed light is given in ... Cited on page .

113 An introduction into the history of the concept of precision is given by N.W. WISE, *The values of precision*, Princeton University Press, 1994. Cited on page .

114 E.H. KENNARD, *Zur Quantenmechanik einfacher Bewegungstypen*, Zeitschrift für Physik **44**, pp. 326–352, 1927. The priority of the result is often wrongly attributed to the much later paper by H.P. ROBERTSON, *The uncertainty principle*, Physical Review **34**, pp. 163–164, 1929. The complete result is by Erwin SCHRÖDINGER, Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse Berlin, p. 296, 1930. Cited on page .

115 W. HEISENBERG, *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, Zeitschrift für Physik **43**, pp. 172–198, 1927. Cited on page .

116 M.G. RAYMER, *Uncertainty principle for joint measurement of noncommuting variables*, American Journal of Physics **62**, pp. 986–993, 1994. Cited on page .

117 The question about the origin of blue colour in flowers has been discussed already in the beginning of the twentieth century. A short summary is given by K. YOSHIDA & al., *Cause of blue petal colour*, Nature **373**, p. 291, 26 January 1995. Cited on page .

118 The main reference on nondemolition measurements is the text by C.M. CAVES & P.D. DRUMOND, *Quantum limits on bosonic communication rates*, Reviews of Modern Physics **66**, pp. 481–537, 1994. Cited on page .

119 Mark P. SILVERMAN, *And yet it moves: strange systems and subtle questions in physics*, Cambridge University Press, 1993. It is a treasure chest for anybody interested in the details of physics. Cited on page .

The two other books by Mark P. SILVERMAN, *More than one mystery*, Springer 1995, talking about the Aharonov-Bohm effect, about the EPR paradox, and about the Hanbury-Brown-Twiss experiments, and *Waves and grains: reflections on light and learning*, Princeton University Press, 1998, are also collections of interesting thoughts about physics.

120 N. ASHCROFT & D. MERMIN, *Solid state physics*, Saunders, 1976, is probably the best as well as the most original and entertaining introduction to the topic available in english language. Cited on page .

121 The famous quantum theory book is P.A.M. DIRAC, *Principles of quantum mechanics*, Oxford University Press, 1958. Cited on page .

Typeset in April 2001

Chapter X Quantum Physics in a Nutshell

Achievements in precision

Compared to the classical description of motion, quantum theory is remarkably more complex. Observables form non-commutative algebras; this noncommutativity then describes the strange observations made in the microscopic domain, such as wave behaviour of particles, tunneling, uncertainty relations, quantization of angular momentum, pair creation, decay, etc. Was this part of the walk worth the effort?

It was. Quantum theory improved the accuracy of predictions from the few – if any – digits common in classical mechanics to the full number of digits – sometimes fourteen – one is able to measure today. The limit is *not* given by the theory, it is given by the measurement accuracy. In other words, the agreement is only limited the amount of money the experimenter is willing to spend, as the table shows.

Table 39 A few comparisons between quantum theory and experiment

Observable	Classical prediction	Prediction of quantum theory[a]	Measurement	Cost[b] (est.)
Simple motion of bodies				
Uncertainty	0	$\Delta x \Delta p \geqslant \hbar/2$	$(1 \pm 10^{-2})\,\hbar/2$	10 k$
Wavelength of matter beams	none	$\lambda p = 2\pi\hbar$	$(1 \pm 10^{-2})\,\hbar$	10 k$
Tunneling rate in alpha decay	0	$\tau = \ldots$	$(1 \pm 10^{-2})\,\tau$	0.5 M$
Compton wavelength	none	$\lambda_c = h/m_e c$	$(1 \pm 10^{-3})\,\lambda$	20 k$
Pair creation rate	0			20 M$
Radiative decay time in hydrogen	none	$\tau \sim 1/n^3$		5 k$
Smallest angular momentum	0	$\hbar/2$	$(1 \pm \ldots)\,\hbar/2$	1 k$
Smallest action	0	$\hbar$	$(1 \pm \ldots)\,\hbar$	10 k$
Casimir effect	0	$ma/A = (\pi^2 \hbar c)/(240 r^4)$	$(1 \pm 10^{-3})\,ma$	30 k$
Colours of objects				
Lamb shift	none	$\Delta\lambda = 1057.86(1)$ MHz	$(1 \pm 10^{-6})\,\Delta\lambda$	50 k$
Rydberg constant	none	$R_\infty = m_e c \alpha^2/2h$	$(1 \pm 10^{-9})\,R_\infty$	50 k$

Observable	Classical prediction	Prediction of quantum theory[a]	Measurement	Cost[b] (est.)
Stephan-Boltzmann constant	none	$\sigma = \pi^2 k^4 / 60\hbar^3 c^2$	$(1 \pm 3 \cdot 10^{-8})\,\sigma$	50 k$
Wien displacement constant	none	$b = \lambda_{max} T$	$(1 \pm 10^{-5})\,b$	100 k$
Refractive index of ...	none			
Photon-photon scattering	0			
Particle and interaction properties				
Electron gyromagnetic ratio	1 or 2	2.002 319 304 3(1)	2.002 319 304 3737(82)	30 M$
Z boson mass	none	$m_Z^2 = m_W^2 (1 + \sin\theta_W^2)$	$(1 \pm 10^{-3})\,m_Z$	100 M$
proton mass	none	$(1 \pm 5\%)\,m_p$	$m_p = 1.67$ yg	1 M$
Composite matter properties				
Atom lifetime	$\approx 1\,\mu$s	∞	$> 10^{20}$ a	10 k$
Molecular size	none	from QED	within 10^{-3}	20 k$
Von Klitzing constant	∞	$h/e^2 = \mu_0 c / 2\alpha$	$(1 \pm 10^{-7})\,h/e^2$	1 M$
AC Josephson constant	0	$2e/h$	$(1 \pm 10^{-6})\,2e/h$	5 M$
Heat capacity of metals at 0 K		25 J/K 0	$< 10^{-3}$ J/K	10 k$
Water density	none	...	$1000\,\mathrm{kg/m^3}$	10 k$
Electr. conductivity of ...	none	...	...	3 k$
Proton lifetime	$\approx 1\,\mu$s	∞	$> 10^{35}$ a	100 M$

See page 563
See page 723
a. These predictions are calculated from the values of table 40. For more details about the predicted values, see appendix B.

Challenge
b. Sometimes the cost for the calculation of the prediction is higher than that for the measurement. (Can you spot the examples?) The sum of the two is given.

See page 563
One notices that the calculated values are not noticeably different from the measured ones. If we remember that classical physics does not allow to calculate any of these values we get an idea of the progress quantum physics has allowed. But despite this impressive agreement, there still are *unexplained* observations. In fact, the unexplained observations provide the input for the calculations just cited; we list them in detail below, in table 40.

In summary, in the microscopic domain we are left with the impression that quantum theory is in *perfect* correspondence with nature; despite prospects of fame and riches, despite the largest number of researchers ever, no contradiction with observation has been found yet.

Physical results

All of quantum theory can be resumed in two sentences.

▷ *In nature, actions smaller than* $\frac{\hbar}{2} = 0.53 \cdot 10^{-34}\,\mathrm{Js}$ *are not observed.*

▷ *All intrinsic properties in nature, with the exception of mass, such as electric charge, spin, parities, etc., appear as integer numbers; in composed systems they either add or multiply.*

These sentences directly lead to the main lesson we learned about motion from quantum mechanics: *if it moves, it is made of particles*. This statement applies to everything, to objects and to images, i.e. to matter and to radiation. Moving stuff is made of *quanta*. Stones, water waves, light, sound waves, earthquakes, jell-o, and everything else one can interact with is made of particles. The second part of our escalation was under the title: what is matter and what are interactions? Now we know: composites of elementary particles.

See page 739

To be clear, an *elementary particle* is a countable entity, smaller than its own Compton wavelength, described by energy, momentum, and the following *complete* list of intrinsic properties: mass, spin, electric charge, parity, charge parity, colour, isospin, strangeness, charm, topness, beauty, lepton number, baryon number, and *R*-parity.

To see how deep this result is, one can apply it to all those moving entities for which it is usually forgotten, such as ghosts, spirits, angels, nymphs, daemons, devils, gods, goddesses, and souls. You can check yourself what happens when their particle nature is taken into account.

Challenge

In addition, quantum theory makes quite a number of statements about particle motion:

• There is no rest for microscopic particles. All objects obey the uncertainty principle, which states that

$$\Delta x \Delta p \geqslant \hbar/2 \quad \text{with} \quad \hbar = 1.1 \cdot 10^{-34}\,\text{Js} \tag{507}$$

making rest an impossibility. The state of particles is defined by the same observables as in classical physics, with the difference that observables do not commute. Classical physics appears in the limit that the Planck constant $\hbar$ can effectively be set to zero.

• Quantum theory introduces a probabilistic element into motion. These effects result from the interactions with baths in the environment of the system.

• Large number of identical particles with the same momentum behave like waves. The so-called de Broglie wavelength is given by the momentum of the single particles through

$$\lambda = \frac{h}{p} = \frac{2\pi\hbar}{p} \tag{508}$$

both in the case of matter and of radiation. This relation is the origin of the wave behaviour of light. The light particles are called photons; their observation is now standard practice. Waves interfere, refract, and diffract. But all waves being made of particles, all waves can be seen, touched and moved. Light for example, can be 'seen' in photon-photon scattering, can be 'touched' using the Compton effect, and it can be 'moved' by gravitational bending. Matter particles, such as molecules or atoms, can be seen, e.g. in electron microscopes, as well as touched and moves, e.g. with atomic force microscopes.

• Particles cannot be enclosed. Even though matter is impenetrable, quantum theory shows that tight boxes or insurmountable obstacles do not exist. Waiting long enough always allows to overcome such boundaries, since there is a finite probability to overcome any obstacle; this process is called tunneling when seen from the spatial point of view and is called decay when seen from the temporal point of view. This behaviour explains the working of television tubes as well as radioactive decay.

• Identical particles are indistinguishable. Radiation is made of bosons, matter of fermions. Under exchange, fermions commute at space-like separations, whereas bosons anticommute. All other properties of quantum particles are the same as for classical particles, namely countability, interaction, mass, charge, angular momentum, energy, momentum, position, as well as impenetrability for matter and penetrability for radiation.

• Particles are described by an angular momentum called spin, giving their behaviour under rotations. Bosons have integer spin, fermions have have integer spin. An even number of bound fermions or any number of bound bosons yield a composite boson; an odd number of bound fermions or an infinite number of interacting bosons yield a fermion.

• In collisions, particles interact locally, through the exchange of other particles. When matter particles collide, they interact through the exchange of virtual bosons, i.e. off-shell bosons. Motion change is thus due to particle exchange. Exchange bosons of even spin mediate only attractive interactions. Exchange bosons of odd spin mediate repulsive interactions as well.

• Quantum theory defines elementary particles as particles smaller than their own Compton wavelength. Experiments so far filed to detected a non-vanishing size for any elementary particle.

• The properties of collisions imply the existence of antiparticles, as regularly observed in experiments. Elementary fermions, in contrast to many elementary bosons, differ from their antiparticles, and can be created and annihilated only in pairs. Apart from neutrinos, elementary fermions have non-vanishing mass and move slower than light.

• Quantum electrodynamics is the quantum field description of electromagnetism. Like all the other interactions, its lagrangian is determined by the gauge group, plus the requirements of Poincaré symmetry and renormalizability. The latter requirement follows from the continuity of space-time. Through the effects of virtual particles, QED describes decay, pair creation, vacuum energy, Unruh radiation for accelerating observers, the Casimir effect, i.e. the attraction of neutral conducting bodies, and the limit for the localization of particles. In fact, all objects can be localized only within intervals of the Compton wavelength

$$\lambda_C = \frac{h}{mc} = \frac{2\pi\hbar}{mc} \quad . \tag{509}$$

At the latest at these distances one must abandon the classical description and use quantum field theory. Quantum field theory introduces corrections to classical electrodynamics; among others, the nonlinearities thus appearing produce departures from the superposition principle for electromagnetic fields, resulting in photon-photon scattering.

• Composite matter is separable because of the finite interaction energies of the constituents. Atoms are made of a nucleus made of quarks, and of electrons. They provide an effective minimal length scale to all everyday matter. Solids are impenetrable because of the fermion character of its electrons in the atoms.

• Quantum theory implies, through the appearance of Planck's constant $\hbar$, that length scales exist in nature. Through the fundamental jitter quantum theory introduces in every example of motion, the infinitely small is eliminated. Thus lower limits to structural dimensions and to many other measurable quantities appear. In particular, quantum theory shows that it is impossible that on the electrons of an atom there live small creatures in the same way that people live on the earth circling the sun.

• Clocks and meter bars have finite precision, due to their interactions with baths. And all measurement apparatuses contain baths, since otherwise they would not be able to record results.

• Elementary particles have the same properties as either objects or images, except divisibility. The elementary fermions (objects) are: the six leptons electron, muon, tau, each with its corresponding neutrino, and the six quarks. The elementary bosons (images) are the See page 739 photon, the eight gluons, and the two weak interaction bosons.

• Quantum chromodynamics, the field theory of the strong interactions, explains the masses of mesons and baryons, through its descriptions as bound quark states. At femtometer scales, the strong interaction effectively acts through the exchange of spin 0 pions, and is thus strongly attractive. At fundamental scales, the strong interaction is mediated by the elementary gluons.

• The theory of electroweak interactions describes the unification of electromagnetism and weak interactions through the Higgs mechanism and the mixing matrix.

• Since matter is composed of particles, quantum theory provides a complete list of the See page 739 intrinsic properties which make up what is called an 'object' in everyday life, namely the same which characterize particles. All other properties of objects, such as shape, temperature, (everyday) colour, elasticity, density, magnetism, etc., are merely combinations of the properties from the particle properties. In particular, quantum theory specifies an object, like every system, as a part of nature interacting weakly and incoherently with its environment.

• Since quantum theory explains the origin of material properties, it also explains the origin of the properties of life. Quantum theory, especially the study of the electroweak and the strong forces, has allowed to give a common basis of concepts and descriptions to material science, nuclear physics, chemistry, biology, medicine, and to most of astronomy.

For example, the same concepts allow to answer questions such as why water is liquid at room temperature, why copper is red, why the rainbow is coloured, why the sun and the stars continue to shine, why there are about 110 elements, where a tree takes the material to make its wood, and why we are able to move our right hand at our own will.

• Matter objects are permanent because, in contrast to radiation, matter particles can only disappear when their antiparticles are present. It turns out that in our environment antimatter is almost completely absent, except for the cases of radioactivity and cosmic rays, where it appears in tiny amounts.

• Images, made of radiation, are described by the same properties as matter. Images can only be localized with a precision of the wavelength λ of the radiation producing it.

Ref. 1 • Quantum physics leaves no room for cold fusion, astrology, teleportation, telekinesis, supernatural phenomena, multiple universes, or faster than light phenomena, the EPR paradox notwithstanding.

• The particle description of nature, e.g. particle number conservation, follows from the possibility to describe interactions perturbatively. This is possible only at low and medium energies. At extremely high energies the situation changes, and non-perturbative effects come into play.

Is physics magic?

Studying nature is like experiencing magic. Nature often looks different from what it is. During magic one is fooled only if one forgets one's own limitations. Once one starts to see oneself to be part of the game, one starts to understand the tricks. That is the fun of it.

- The world looks irreversible, even though it isn't. We never remember the future. We are fooled because we are macroscopic.

- The world looks decoherent, even though it isn't. We are fooled again because we are macroscopic.

- Motion seems to disappear, even though it is eternal. We are fooled again, because our senses cannot experience the microscopic domain.

- The world seems dependent on the choice of the frame of reference, even though it is not. We are fooled because we are used to live on the surface of the earth.

- Objects seem distinguishable, even though they are not. We are fooled because we live at low energies.

- Matter looks continuous, even though it isn't. We are fooled because of the limitations of our senses.

In short, quantum theory answers the title question is affirmatively; that is its main attraction.

The dangers of buying a can of beans

The ultimate product warning, which should be printed on its package according to certain well-informed lawyers, gives another summary of our walk so far. Ref. 2

Warning: care should be taken when **looking** at this product:

– It emits heat radiation.

– Bright light has the effect to compress this product.

Warning: care should be taken when **not looking** at this product:

– It could cool down and lift itself into the air.

– This product can disappear from its present location and disappear at any random place in the universe, including your neighbor's garage.

Warning: care should be taken when **touching** this product:

– Part of it could heat up while another part cools down, causing severe burns.

Warning: care should be taken when **handling** this product:

– This product consists of at least 99, 999 999 999 999 % empty space.

– This product contains particles moving with speeds higher than one million kilometers per hour.

– Every kilogram of this product contains the same amount of energy as liberated by about one hundred nuclear bombs.*

– In case this product is brought in contact with antimatter, a catastrophic explosion will occur.

– In case this product is rotated, it will emit gravitational radiation.

* A standard nuclear warhead has an explosive power of about 0.2 megatons of TNT, i.e. of trinitrotoluene; a megaton is defined as $4.2 \cdot 10^{15}$ J, the energy content of about 47 g of matter. That is less than a handful for most solids or liquids.

Warning: care should be taken when **transporting** this product:

– The force needed depends on its velocity, since its kinetic energy increases its weight, which is thus higher than when at rest.
– It will emit additional radiation when accelerated.
– This product attracts, with a force that increases with decreasing distance, every other object around, including its purchaser's kids.

Warning: care should be taken when **storing** this product:

– It is impossible to keep this product in a specific place and at rest at the same time.
– Except when stored underground at a depth of several kilometers, over time cosmic radiation will render this product radioactive.
– This product may disintegrate in the next 10^{35} years.
– Parts of this product are hidden in other dimensions.
– This product warps space and time in its vicinity, including the storage container.
– Even if stored in a closed container, this product is influenced and influences all other objects in the universe, including your parents in law.
– Inside a closed container this product may start to float.

Warning: care should be taken when **traveling away from** this product:

– It will arrive at the expiration date before the purchaser does so.

Warning: care should be taken when **using** this product:

– Any use whatsoever will increase the entropy of the universe.
– The constituents of this product are exactly the same as those of any other object in the universe, including those of rotten fish.
– The use could be disturbed by the (possibly) forthcoming collapse of the universe.

The impression of a certain paranoid side to physics is purely coincidental.

What is unexplained by quantum mechanics and general relativity?

The material gathered in this second part of our escalation, together with the earlier summary of general relativity, allows us to give a complete answer to this question. Even though with the available concepts and theories all observed phenomena of motion are described, there remain some *unexplained* properties of nature. Whenever one asks 'why?' and continues doing so after each answer, one arrives at one of the points in table 40.

Table 40 *Everything* quantum field theory and general relativity do *not* explain; in other words, a list of *the only* experimental data and criteria available for tests of the unified description of motion.

Observed value	Property unexplained so far
Local quantities, from quantum theory	
α_{em}	the low energy value of the electromagnetic coupling constant
α_{w}	the low energy value of the weak coupling constant
α_{s}	the low energy value of the strong coupling constant
m_{q}	the values of the 6 quark masses

Observed value	Property unexplained so far
m_l	the values of 3 lepton masses
m_W	the values of the independent mass of the W vector boson
θ_W	the value of the Weinberg angle
$\beta_1, \beta_2, \beta_3$	three mixing angles
θ_{CP}	the value of the CP parameter
θ_{st}	the value of the strong topological angle
3	the number of particle generations
$0.5\,nJ/m^3$	the value of the observed vacuum mass density or cosmological constant
$3+1$	the number of space and time dimensions
Global quantities, from general relativity	
$1.2(1)\cdot 10^{26}\,m$?	the distance of the horizon, i.e. the “size” of the universe (if it makes sense)
10^{82} ?	the number of baryons in the universe, i.e. the average matter density in the universe (if it makes sense)
$> 10^{92}$?	the initial conditions for more than 10^{92} particle fields in the universe, including those at the origin of galaxies, of stars, etc. (if they make sense)
Local structures, from quantum theory	
$S(n)$	the origin of particle identity, i.e. of permutation symmetry
Ren. group	the renormalisation properties, i.e. the existence of point particles
SO(3,1)	the origin of Lorentz (or Poincaré) symmetry (i.e. of spin, position, energy, momentum)
C^*	the origin of the algebra of observables
Gauge group	the origin of gauge symmetry (and thus of charge, strangeness, beauty, etc.)
in particular, for the standard model:	
U(1)	the origin of the electromagnetic gauge group (i.e. of the quantization of electric charge, as well as the vanishing of magnetic charge)
SU(2)	the origin of weak interaction gauge group
SU(3)	the origin of strong interaction gauge group
Global structures, from general relativity	
maybe $R \times S^3$?	the unknown topology of the universe (if it makes sense)

The table has several notable aspects.* First of all, neither quantum mechanics nor general relativity explain any property unexplained in the other field. They do not help each other; the unexplained parts of both fields simply add up. Secondly, both in quantum theory and in general relativity, motion still remains the change of position with time. In short, in the first two parts of this walk, we did not achieve our goal: we still do not understand motion. The basic questions remain: What is time and space? What is mass? What is charge and what are the other properties of objects? What are fields? Why are all the electrons the same?

We also note that the table lists a lot of extremely *different* concepts. That means that at this point of our walk there is *a lot* we do not understand. Answering will require involved and abstract descriptions.

* If neutrinos have masses, these three values plus four additional mixing angles have to be added to the list. Note also that other, equivalent choices of parameter combinations are possible.

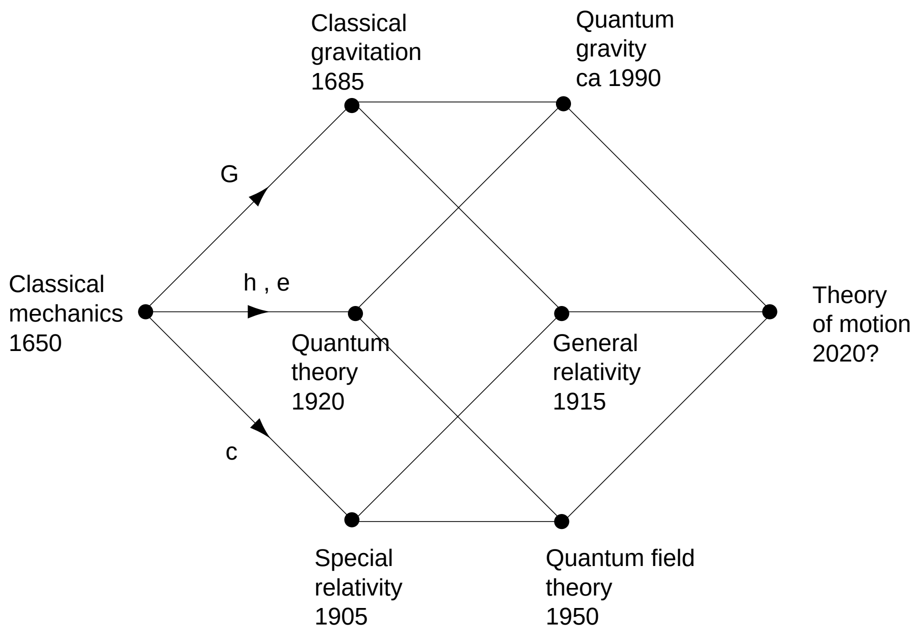


Figure 170 How the description of motion proceeded in the history of physics

On the other hand, the list is also *short*. The description of nature our escalation has produced is concise and precise. No discrepancies with experiments are known. In other words we have a good description of motion *in practice*. Going further is not necessary for the improvement of measurement precision. Simplifying the above list is only important from the *conceptual* point of view. For this reason, the study of physics at university often stops at this point. However, even though we have no known discrepancies with experiments, we are *not* at the top of motion mountain, as the table 40 and figure 170 show; the last leg forms the third part of our walk.

How to delude oneself to have reached the top of motion mountain

Nowadays is deemed chic to pretend that the escalation is over at this stage.* The reasoning is as follows. If in the previous table on unexplained features of nature one changes the values of the constants only ever so slightly, the world would look completely different from what it is.**

Ref. 3 * Actually this is not new. Only he arguments have changed. Maybe the greatest physicist ever once wrote: "..., that, in a few years, all great physical constants will have been approximately estimated, and that the only occupation which will be left to men of science will be to carry these measurements to another place of decimals." This old excuse for giving up the rest of the escalation has now been superseded by newer, more fashionable ones.

** Most of the material below is from the mighty book by John D. Barrow & Frank J. Tipler, *The anthropic cosmological principle*, Oxford University Press, 1986.

Table 41 A tiny selection of consequences of changing the properties of nature

Observed value	Change	Result
Local quantities, from quantum theory		
α_{em}	smaller:	only short lived, smaller, hotter stars; no sun
	larger:	darker sun, animals die of electromagnetic radiation, too much proton decay, no planets, no stellar explosions, no star formation, no galaxy formation
	$+60\%$:	quarks decay into leptons
	$+200\%$:	proton-proton repulsion makes nuclei impossible
α_{w}	-50%:	carbon nucleus unstable
	very weak:	no hydrogen, no p-p cycle in stars, no C-N-O cycle
	$+2\%$:	no protons from quarks
	$G_F m_e^2 \not\approx \sqrt{G m_e^2}$:	either no or only helium in the universe
	much larger:	no stellar explosions, faster stellar burning
α_{s}	-9%:	no deuteron, stars much less bright
	-1%:	no C resonance, no life
	$+3.4\%$:	diproton stable, faster star burning
	much larger:	carbon unstable, heavy nuclei unstable, widespread leukemia
θ_{W}	different:	
θ_{CP}	different:	
m_{q} changes:		
n-p mass difference	larger:	neutron decays in proton inside nuclei; no elements
	smaller:	free neutron not unstable, all protons into neutrons during big bang; no elements
	smaller than m_e:	protons would capture electrons, no hydrogen atoms, star life much shorter
m_{l} changes:		
e-p mass ratio	much different:	no molecules
	much smaller:	no solids
m_{W} different:		
3 generations	6-8:	only helium in nature
	>8:	no asymptotic freedom & confinement
Global quantities, from general relativity		
horizon size	much smaller:	no people
baryon number	very different:	no smoothness
Initial condition changes:		
moon mass	smaller:	small earth magnetic field; too much cosmic radiation; widespread child skin cancer
moon mass	larger:	large earth magnetic field; too little cosmic radiation; no evolution into humans
jupiter mass	smaller:	too many comet impacts on earth; extinction of animal life
jupiter mass	larger:	too little comet impacts on earth; no moon; no dinosaur extinction

Observed value	Change	Result
Oort cloud object number	smaller:	no comets; no irregular asteroids; no moon; still dinosaurs
galaxy centre distance	smaller:	irregular planet motion; supernova dangers
initial cosmic speed	+0.1%:	1000 times faster universe expansion
	−0.0001%:	universe recollapses after 10 000 years
vacuum energy density	change by 10^{-55}:	no flatness
$3+1$ dimensions	different:	no atoms, no planetary systems
Local structures, from quantum theory		
permutation symmetry	none:	no matter
Lorentz symmetry	none:	no communication possible
U(1)	different:	no Huygens principle, no way to *see* anything
SU(2)	different:	no radioactivity, no sun, no life
SU(3)	different:	no stable quarks and nuclei
Global structures, from general relativity		
topology	other:	unknown; possibly correlated gamma ray bursts or star images at the antipodes

The table is overwhelming. Some have summed it up in a simple sentence: if any parameter is changed, the universe would either have too many or too few black holes. Obviously, even the tiniest changes in the properties of nature are incompatible with our existence. What does this mean? Answering this question too rapidly is dangerous. Many fall into a common trap, namely to refuse admitting that the numbers and the other properties need to be explained, i.e. deduced from more general principles. It is easier to throw in some irrational belief; the most fashionable ones are that the universe is *created* or *designed*, or that the universe is *designed for people*, or that the values are random since our universe happens to be *one of many others*.

All these beliefs have in common that they have no factual basis, that they discourage further search, and that they sell many books. Physicists call the issue of the first belief *fine tuning*, and usually, but not always, steer clear form the logical errors contained in the so common belief in "creation" discussed earlier on. However, many physicists subscribe to the second belief, namely that the universe is designed for people, calling it the *anthropic principle*, even though we saw that it is indistinguishable both from the simian principle and from the request that statements be based on observations. The third belief, namely *multiple universes*, is a minority view, but also sells well.

See page 415

See page 421

However, stopping the escalation at the present point with a belief is not different from doing so directly at the beginning. This used to be the case in societies which lacked the passion for rational investigation, and is still the case in circles which discourage the use of reason among their members. Looking for beliefs instead of looking for answers means to give up the escalation while pretending to have reached the top.

That is a pity. During our escalation, accepting the powerful evidence of table 41 is one of the most awe-inspiring, touching, and motivating moments. All the evidence has only one possible implication based on facts: it implies that we are only a *tiny part* of the universe, *linked* with all other aspects of it. Due to our small size and to all the connections with our environment, any imagined tiny change would make us disappear, like a water droplet engulfed by a large wave. Our walk has repeatedly reminded us of this smallness and dependence, and overwhelmingly does so again at this point.

After having faced this conclusion, everybody has to make up his own mind on whether to proceed or not with the present escalation. Of course, there is no obligation to do so. Challenge

What awaits us?

For those who want to proceed with the escalation of motion mountain, the shortness of the list of unexplained aspects of nature means that *no additional experimental data* is available as check of the final description of nature. Everything we need to arrive at the final description of motion will probably be deduced from the experimental data given in this list, and from nothing else. In other words, future experiments will *not* help us – except if they change something in the list, as supersymmetry might do with the gauge groups.

This lack of new experimental data means that to continue the walk is a *conceptual* adventure only. We have to walk into the clouds and storms of the top of motion mountain, keeping our eyes open, without any other guidance except our reason: not an adventure of action, but an adventure of the mind. And an incredible one, as we will soon find out. To provide a feeling of what awaits us, we rephrase a few of the remaining issues in a more challenging way.

What determines *colours*? In other words, what relation of nature fixes the famous proton-electron mass ratio, i.e. the number of about 1836.2, as well as the fine structure constant? There is still no answer to the question.

What fixes the contents of a *teapot*? It is given by its size to the third power. But why are there only three dimensions? Why is the tea content limited in this way?

Was Democritos *right*? The escalation up to this point has shown so; nature is indeed well described by the concepts of particles and of vacuum. At large scales, relativity has added a horizon, and at small scales, quantum theory added vacuum energy and pair creation. Nevertheless, both theories *assume* the existence of particles and assume the existence of space-time, and neither *predicts* them. Even worse, both theories completely fail to predict the existence *any* of the properties either of space-time – such as its dimensionality – or of particles – such as their masses and other quantum numbers. A lot is missing.

Was Democritos *wrong*? One often reads that the standard model has only about twenty unknown parameters; this common mistake negates the remaining 10^{93} initial conditions. To get an idea of the problem, we simply estimate the number N of possible states of all particles in the universe by

$$N = n\,v\,d\,p\,f \qquad (510)$$

where n is the number of particles, v is the number of variables (position, momentum, spin), d is the number of values each of them can take (limited by the maximum of 61 decimal digits), p is the number of spacetime points (usually taken to be 10^{183}, assuming that all of the universe is visible) and f is a factor expressing how many of all these initial conditions are actually independent of each other. One thus has

$$N = 10^{92} \cdot 8 \cdot 10^{61} \cdot 10^{183} \cdot f = f\, 10^{336} \qquad (511)$$

with the small problem that we know nothing whatsoever about f. Its value could be 0, if all data are interdependent, or its value could be 1, in case nothing is. And in any case one still needs to understand how all the particles get their states assigned from this truly enormous range of possibilities.*

See page 98

Was the escalation up to this point *in vain*? Quite at the beginning of our walk we noted that in classical physics, space and time were defined using matter, and matter was defined using space-time. Hundred years of general relativity and of quantum theory as well as dozens of geniuses have not solved this oldest paradox of all. The issue is still open, as you might want to check by yourself.

Challenge

The answers to these questions define the top of motion mountain. Answering them means to know *everything* about motion. In summary, our quest for the unraveling of the essence of motion gets really interesting only from this point onwards!

References

1 An informative account of the world of psychokinesis and the paranormal is given by James Randi, a professional magician, in *Flim-flam!*, Prometheus Books, Buffalo 1987, as well as in several of his other books. See also the http://www.randi.org web site. Cited on page 561.

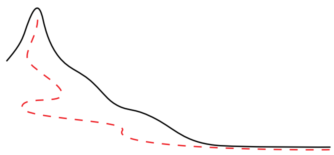
2 This way to look at things goes back to the text by Susan Hewitt & Edward Subitzky, *A call for more scientific truth in product warning labels*, Journal of Irreproducible Results **36**, nr. 1, 1991. Cited on page 562.

3 James Clerk Maxwell, *Scientific papers*, **2**, p. 244, October 1871. Cited on page 565.

* Actually, there are strong arguments for $f = 0$, as mentioned already several times so far. More about this issue in the third part of the escalation.

See page 618

Third Part



Motion Without Motion
What are Space, Time, and Particles?

Where, through the combination of
quantum mechanics and general relativity,
the top of motion mountain is reached,
discovering
that vacuum is indistinguishable from matter,
that space, time and mass are easily confused,
that there is no difference between the very large and the very small,
and that the complete description of motion is possible.
(Ahem – well, wait a few more years for the last line.)

Typeset in April 2001

Chapter XII General Relativity Versus Quantum Mechanics

The contradictions

> Man muß die Denkgewohnheiten durch Denknotwendigkeiten ersetzen.*
> Albert Einstein (Ulm, 1879-Princeton, 1955)

The two stories told in the two parts of the path we followed up to now, namely that on general relativity and that on quantum field theory, are both beautiful and successful. We reached a considerable height in our escalation. The precision we achieved in the description of nature is impressive, and we are able to describe all known examples of motion. There seem to be no exceptions.

However, we have seen in the last chapter that the most important aspects of any type of motion, the masses of the particles and their coupling strengths, are yet unexplained. Also the origin of the universe's particle number, their initial conditions, and the dimensionality of space-time remain in the dark. Obviously, our adventure has not yet reached its completion.

This last part of our hike will be the most demanding. In the escalation of any high mountain, at great heights the head gets dizzy due to lack of oxygen; the finite forces at one's disposal require that one leaves all unnecessary baggage behind, to be able to proceed at all. Like in an actual escalation, we will leave behind everything which hinders us. In order to determine what is unnecessary, we need complete concentration on our aim. We will drop all those concepts which are at the origin of the *contradictions* between general relativity and quantum theory. To pinpoint this useless baggage, we first list the main ones.

• In classical physics, as well as in general relativity, the *vacuum* is a region with no mass, no energy and no momentum. If space-time contains matter or gravitational fields, it is curved. The best way to measure the mass or energy content of the vacuum at large is to measure the average curvature of the universe. Cosmology tells us how we can do this; Ref. 1
measurements yield an average mass density of the vacuum See page 256

$$m/V \approx 10^{-26}\,\mathrm{kg/m^3} \qquad (513)$$

and a corresponding limit for the energy density. This value is of the order of the mass of a proton per cubic meter. However, the textbooks on quantum field theory tell a different Ref. 2
story. Following them, vacuum is a region with zero-point fluctuations. The energy content

* One needs to exchange thinking habits by thinking necessities.

of vacuum then is the sum of the zero point energies of all the fields it contains. Most texts also point out that the Casimir effect “proves” the reality of these zero point energies. This energy density is given, within one order of magnitude by See page 507

$$m/V = \frac{4\pi h}{c^5}\int_0^{\nu_{\rm max}} \nu^3 d\nu = \frac{\pi h}{c^5}\nu_{\rm max}^4 \qquad (514)$$

The approximation is valid for the case that the cutoff frequency $\nu_{\rm max}$ is much larger than the rest mass m of the particles corresponding to the field under consideration. Indeed, particle physicists argue that the cutoff energy has to be at least the energy of grand unification, about 10^{16} GeV= 1.6 MJ. That would give a vacuum mass density of

$$m/V \approx 10^{82}\,\mathrm{kg/m^3} \quad . \qquad (515)$$

about 10^{108} times more than the experimental limit deduced using general relativity. In other words, something is slightly wrong here.

• But let’s have a look at some other contradictions. Gravity is curved space-time. Extensive research has shown that quantum field theory, the description of electrodynamics and of the nuclear forces, fails for situations with strongly curved space-times. In these cases the concept of particle is not uniquely defined; quantum field theory cannot be extended to consistently include gravity, and thus general relativity. Without the concept of particle as a countable entity one also loses possibility to perform perturbation calculations, which are the only ones possible in quantum field theory. In short, quantum theory works only by assuming that gravity does not exist; and indeed, the gravitational constant does not appear in any consistent quantum field theory. Ref. 3

• On the other hand, general relativity neglects the commutation rules between physical quantities discovered by experiments on a microscopic scale. General relativity assumes that position and momentum of material objects can be given the meaning of classical physics and thus ignores Planck’s constant $\hbar$.

• In everyday life, a measurement is a comparison, a counting of a sequence; it is essentially a process of distinction, e.g. on ruler, between ‘mark’ and ‘not mark’. In general relativity, like in classical physics, infinite measurement precision is assumed to be possible. Indeed, by reading off between which finer and finer ruler marks a result lies, one is supposed to get the complete real number corresponding to it. In contrast, in quantum mechanics the measurement precision is limited, via the uncertainty relation, by the mass M of the apparatus. See page 391 See page 497

• Relativity explains that time is what is read from clocks. Quantum theory tells that precise clocks do not exist, especially if the coupling with gravitation is included. What does it mean to wait 10 minutes, if the clock goes into a superposition due to its coupling to space-time geometry?

• Quantum theory associates mass to an inverse length, via the Compton wavelength, whereas general relativity associates mass to length, via the Schwarzschild radius.

• Most dramatically, the contradiction is shown by the failure of general relativity to describe the pair creation of spin 1/2 particles, a typical and essential quantum process. John Wheeler and others have shown that in such a case, the topology of space necessarily has to *change*; in general relativity however, the topology of space is fixed. In short, quantum Ref. 4, 5 Ref. 6, 7 theory says matter is made of fermions; general relativity cannot incorporate fermions.

• General relativity shows that space and time cannot be distinguished, whereas quantum theory tells that matter does so. Quantum theory is a theory of – admittedly weirdly constructed – local observables. General relativity doesn't have any local observables, as Einstein's hole argument shows.

As long as a description is arbitrary or has contradictions, it cannot provide unification nor demarcation from alternatives. This means that it is *not* an explanation, since it does not See page 420 fulfill the criteria to be one. In other words, so far we do not have an explanation for either general relativity or quantum theory. In order to find one, let us take the shortest and fastest path: let us investigate the contradictions in more detail.

30. Does matter differ from vacuum?

Ref. 8 There is a simple, but not well known way to state the origin of all the mentioned contradictions between general relativity and quantum mechanics. Both theories describe motion with objects made of particles and with space-time made of events. Let us see how these two concepts are defined.

A *particle* – and in general any object – is defined as a conserved entity to which a position can be ascribed and which can move. (The etymology of the term 'object' is connected to the latter fact.) In other words, a particle is a small entity with conserved mass, charge etc., which can vary its position with time.

Ref. 9 At the same time, in every physics text *time* is defined with the help of moving objects, usually called 'clocks', or with the help of moving particles, such as those emitted by light sources. Similarly, also the *length* unit is defined with objects, be it a old-fashioned ruler, or nowadays with help of the motion of light, which is a collection of moving particles as well.

The rest of physics has sharpened the definitions of particles and space-time. In quantum mechanics one assumes space-time given (it is included as a symmetry of the hamiltonian), and one studies in all detail the properties and the motion of particles from it, both for matter and radiation. In general relativity and especially in cosmology, the opposite path is taken: one assumes that the properties of matter and radiation are given, e.g. via their equations of state, and one describes in detail the space-time that follows from them, in particular its curvature. But one fact remains unchanged throughout all these advances in standard textbook physics: *the two concepts of particles and of space-time are defined with the help of each other.*

To avoid the contradiction between quantum mechanics and general relativity and to eliminate their incompleteness requires eliminating this circular definition. As argued in the following, this necessitates a radical change in our description of nature, and in particular about the continuity of space-time.

For a long time, the contradictions in the two descriptions of nature were avoided by keeping them separate: one often hears the statement that quantum mechanics is valid at small dimensions, and the other, general relativity, is valid at large dimensions. But this artificial separation is not justified and obviously prevents the solution of the problem. The situation resembles the well-known drawing by M.C. Escher, where two hands, each holding a pencil, seem to draw each other. If one takes one hand as a symbol for space-time, the other as a symbol for particles, and the act of drawing as a symbol for the act of defining, one has a description of standard twentieth century physics. The apparent contradiction is solved when one recognizes that both concepts (both hands) result from a hidden third concept (a third hand) from which the other two originate. In the picture, this third entity is the hand of the painter.

Figure 174 'Tekenen' by M.C. Escher, 1948 – a metaphor for the way in which 'particles' and 'space-time' are usually defined: each with help of the other

In the case of space-time and matter, the search for this underlying concept is presently making renewed progress. The required conceptual changes are so dramatic that they should be of interest to anybody who has an interest in physics. Some of the issues are presented here. The most effective way to study these changes is to focus in detail on that domain where the contradiction between the two standard theories becomes most dramatic, and where both theories are necessary at the same time. That domain is given by the following well-known argument.

Ref. 10, 11

Ref. 12, 13

Planck scales

Both general relativity and quantum mechanics are successful theories for the description of nature. Each of them provides a criterion to determine when classical galilean physics is not applicable any more. (In the following, we use the terms 'vacuum' and 'empty space-time' interchangeably.)

General relativity shows that it is necessary to take into account the curvature of space-time whenever we approach an object of mass m to distances of the order of the Schwarzschild radius r_S, given by

$$r_S = 2Gm/c^2 \quad . \tag{516}$$

Approaching the Schwarzschild radius of an object, the difference between general relativity and the classical $1/r^2$ description of gravity becomes larger and larger. For example, the barely measurable gravitational deflection of light by the sun is due to an approach to $2.4 \cdot 10^5$ times the Schwarzschild radius of the sun. In general however, we are forced to stay away from objects an even larger multiple of the Schwarzschild radius, except in the vicinity of neutron stars, as shown in table 42; for this reason, general relativity is not necessary in everyday life. (An object smaller than its own Schwarzschild radius is called a *black hole*. Following general relativity, no signals from the inside of the Schwarzschild radius can reach the outside world; hence the name 'black hole'.) Ref. 4, 14 Ref. 15

Similarly, quantum mechanics shows that galilean physics must be abandoned and quantum effects must be taken into account whenever one approaches an object to distances of

Object	size: diameter d	mass m	Schwarz-schild radius r_S	ratio d/r_S	Compton wave length λ_C	ratio d/λ_C
galaxy	≈ 1 Zm	$\approx 5 \cdot 10^{40}$ kg	≈ 70 Tm	$\approx 10^{7}$	$\approx 10^{-83}$ m	$\approx 10^{104}$
neutron star	10 km	$2.8 \cdot 10^{30}$ kg	4.2 km	2.4	“$1.3 \cdot 10^{-73}$ m”	$8.0 \cdot 10^{76}$
sun	1.4 Gm	$2.0 \cdot 10^{30}$ kg	3.0 km	$4.8 \cdot 10^{5}$	“$1.0 \cdot 10^{-73}$ m”	$8.0 \cdot 10^{81}$
earth	13 Mm	$6.0 \cdot 10^{24}$ kg	8.9 mm	$1.4 \cdot 10^{9}$	“$5.8 \cdot 10^{-68}$ m”	$2.2 \cdot 10^{74}$
human	1.8 m	75 kg	0.11 ym	$1.6 \cdot 10^{25}$	“$4.7 \cdot 10^{-45}$ m”	$3.8 \cdot 10^{44}$
molecule	10 nm	0.57 zg	“$8.5 \cdot 10^{-52}$ m”	$1.2 \cdot 10^{43}$	$6.2 \cdot 10^{-19}$ m	$1.6 \cdot 10^{10}$
atom (^{12}C)	0.6 nm	20 yg	“$3.0 \cdot 10^{-53}$ m”	$2.0 \cdot 10^{43}$	$1.8 \cdot 10^{-17}$ m	$3.2 \cdot 10^{7}$
proton p	2 fm	1.7 yg	“$2.5 \cdot 10^{-54}$ m”	$8.0 \cdot 10^{38}$	$2.0 \cdot 10^{-16}$ m	9.6
pion π	2 fm	0.24 yg	“$3.6 \cdot 10^{-55}$ m”	$5.6 \cdot 10^{39}$	$1.5 \cdot 10^{-15}$ m	1.4
up-quark u	< 0.1 fm	0.6 yg	“$9.0 \cdot 10^{-55}$ m”	$< 1.1 \cdot 10^{38}$	$5.5 \cdot 10^{-16}$ m	< 0.18
electron e	< 4 am	$9.1 \cdot 10^{-31}$ kg	“$1.4 \cdot 10^{-57}$ m”	$3.0 \cdot 10^{39}$	$3.8 \cdot 10^{-13}$ m	$< 1.0 \cdot 10^{-5}$
neutrino ν_e	< 4 am	$< 3.0 \cdot 10^{-35}$ kg	“$< 4.5 \cdot 10^{-62}$ m”	n.a.	$> 1.1 \cdot 10^{-8}$ m	$< 3.4 \cdot 10^{-10}$

Table 42 The size, Schwarzschild radius, and Compton wavelength of some objects appearing in nature. A short reminder of the new SI prefixes: f: 10^{-15}, a: 10^{-18}, z: 10^{-21}, y: 10^{-24}, P: 10^{15}, E: 10^{18}, Z: 10^{21}, Y: 10^{24}. Note that the lengths between quotes make no physical sense, as explained in this section.

the order of the Compton wavelength λ_C, which is given by

$$\lambda_C = \frac{\hbar}{m\,c} \quad . \tag{517}$$

Of course, this length only plays a role if the object itself is smaller than its own Compton wavelength. At these dimensions we get relativistic quantum effects, such as particle-antiparticle creation and annihilation. Table 42 shows that the approach distance d is near or smaller than the Compton wavelength only in the microscopic world, so that such effects are not observed in everyday life; therefore we do not need quantum field theory to describe common observations.

Taking these two results together, the situations which require the combined concepts of quantum field theory and of general relativity are those in which both conditions are satisfied simultaneously. The necessary approach distance is calculated by setting $r_S = 2\lambda_C$ (the factor 2 is introduced for simplicity). We find that this is the case when lengths or times are of the order of

$$\begin{aligned} l_{Pl} &= \sqrt{\hbar G/c^3} = 1.6 \cdot 10^{-35}\,\text{m, the } \textit{Planck length,} \\ t_{Pl} &= \sqrt{\hbar G/c^5} = 5.4 \cdot 10^{-44}\,\text{s , the } \textit{Planck time.} \end{aligned} \tag{518}$$

Whenever we approach objects at these dimensions, general relativity and quantum mechanics both play a role; at these scales effects of *quantum gravity* appear. The values of the Planck dimensions being extremely small, this level of sophistication is unnecessary in everyday life, in astronomy and even in particle physics.

However, this sophistication is necessary to understand why the universe is the way it is. The questions mentioned at the beginning – why do we live in three dimensions, why is the proton 1836.15 times heavier than the electron – need for their answer a precise and complete description of nature. The contradictions between quantum mechanics and general relativity make the search for these answers impossible. On the other hand, the unified theory, describing quantum gravity, is not yet finished; but a few glimpses on its implications can already taken at the present stage.

Note that the Planck scales are also the *only* domain of nature where quantum mechanics and general relativity come together; therefore they provide the only possible starting point for the following discussion. Planck was interested in the Planck units mainly as *natural units of measurement*, and that is the way he called them. However, their importance in nature is much more pervasive, as will be seen now. Ref. 16

Farewell to instants of time

> Time is composed of time atoms ... which in fact are indivisible.
> Moses Maimonides, 12th century.

Quantum mechanics introduces an additional aspect in the description of nature: the uncertainty relations between momentum and energy and between energy and time, which introduce quantum limits to all measurements. These relations lead to important consequences at Planck dimensions; they appear most clearly when we investigate the properties of clocks and meter bars. Is it possible to construct a clock which is able to measure time intervals shorter than the Planck time? Surprisingly, the answer is no, even though in the time–energy uncertainty relation $\Delta E \Delta t \geqslant \hbar$ it seems that by making ΔE arbitrary large, we can make Δt arbitrary small. Ref. 17, 18

Any clock is a device with some moving parts; such parts can be mechanical wheels, matter particles in motion, changing electrodynamic fields – e.g. flying photons –, decaying radioactive particles, etc. For each moving component of a clock, e.g. the two hands of the dial, the uncertainty principle applies. As has been discussed in many occasions and most clearly by Raymer, the uncertainty relation for two non-commuting variables describes two different, but related situations: it makes a statement about standard *deviations* of *separate* measurements on *many* identical systems, and it describes the measurement *precision* for a *joint* measurement on a *single* system. Throughout this article, only the second viewpoint is used. Ref. 19, 20 Ref. 21

Now, in any clock, we need to know both the time and the energy of each hand, since otherwise it would not be a classical system, i.e. it would not be a recording device. We therefore need the joint knowledge of non-commuting variables for each moving component of the clock; we are interested in the component with the largest time uncertainty, i.e. imprecision of time measurement Δt. It is evident that the smallest time interval δt which can be measured by a clock is always larger than the quantum limit, i.e. larger than the time precision Δt due to the uncertainty relation for its moving components. Thus one has

$$\delta t \geq \Delta t \geq \frac{\hbar}{\Delta E} \tag{519}$$

where ΔE is the energy uncertainty of the moving component. This energy uncertainty ΔE is surely smaller than the total energy $E = mc^2$ of the component itself. (Physically, this condition means that one is sure that there is only *one* clock; the case $\Delta E > E$ would mean that it is impossible to distinguish between a single clock, or a clock plus a pair of clock-anticlock created from the vacuum, or a component plus two such pairs, etc.) Furthermore, any clock provides information; therefore, signals have to be able to leave it. To make this possible, the clock may not be a black hole; its mass m must therefore be smaller than the Schwarzschild mass for its size, i.e. $m \leq c^2 l/G$, where l is the size of the clock (here and in the following we neglect factors of order unity). Finally, the size l of the clock must be smaller than $c\,\delta t$ itself, to allow a sensible measurement of the time interval δt: one has $l \leq c\,\delta t$, since otherwise different parts of the clock could not work together to produce the same time display. (It is amusing to explore how a clock larger than $c\,\delta t$ stops working, due to the loss of rigidity of its components.) Putting all these conditions together one after the other, one gets

$$\delta t \geq \frac{\hbar G}{c^5 \delta t} \quad , \tag{520}$$

or

$$\delta t \geq \sqrt{\frac{\hbar G}{c^5}} = t_{\mathrm{Pl}} \quad . \tag{521}$$

In summary, from three simple properties of every clock, namely that we have only one of them, that we can read its dial, and that it gives sensible readouts, we get the general conclusion that *clocks cannot measure time intervals shorter than the Planck time*.

Note that this argument is independent of the nature of the clock mechanism. Whether the clock is powered by gravitational, electrical, plain mechanical or even nuclear means, the relations still apply. Note also that gravitation is essential in this argument. A well-known study on the limitations of clocks due to their mass and their measuring time has been
Ref. 22 published by Salecker and Wigner and summarized in pedagogical form by Zimmerman;
Ref. 23 the present argument differs in that it includes both quantum mechanics as well as gravity, and therefore yields a different, lower, and much more fundamental limit. Note also that
Ref. 24 the discovery of black hole radiation does not change the argument; black hole radiation notwithstanding, measurement devices cannot be inside black holes.

Ref. 25 The same result can also be found in other ways. For example, any clock small enough to measure small time intervals necessarily, due to the uncertainty relation, has a certain energy uncertainty. At the same time, following general relativity, any energy density induces a deformation of space-time, and signals from that region arrive with a certain delay due to that deformation. The energy uncertainty of the source leads to a uncertainty in deformation
Ref. 14 and thus of the delay. The expression from general relativity for the deformation of the time part of the line element due to a mass m is $\delta t = mG/lc^3$. Using Einstein's mass energy relation we get that an energy spread ΔE produces an uncertainty Δt in the delay

$$\Delta t = \frac{\Delta E\, G}{l\, c^5} \quad . \tag{522}$$

which determines the precision of the clock. Now the energy uncertainty of the clock is bound by the uncertainty relation for time and energy $\Delta E \geqslant \hbar/\Delta t$, with the precision of the clock appearing again. Putting this together, one again finds the relation $\delta t \leqslant t_{\mathrm{Pl}}$. We are forced to conclude that *in nature there is a minimum time interval*. In other words, *at Planck scales the term "instant of time" has no theoretical nor experimental backing.* It therefore makes no sense to use it.

Farewell to points in space

In a similar way we can deduce that it is not possible to make a meter bar or any other length measuring device that can measure lengths shorter than the Planck length. (One can deduce this already from $l_{\mathrm{Pl}} = c\, t_{\mathrm{Pl}}$.) The straightforward way to measure the distance between two points is to put an object at rest at each position. In other words, joint measurements of position and momentum are necessary for every length measurement. Now the minimal length δl that can be measured is surely larger than the position uncertainty of the two objects. From the uncertainty principle it is known that each object's position cannot be determined with a precision Δl smaller than that given by the uncertainty relation $\Delta l\, \Delta p = \hbar$, where Δp is the momentum uncertainty. Requiring to have only one object at each end, i.e. avoiding pair production from the vacuum, means $\Delta p < mc$, which together gives Ref. 26

$$\delta l \geq \Delta l \geq \frac{\hbar}{mc} \quad . \tag{523}$$

Furthermore, the measurement cannot be performed if signals cannot leave the object: they may not be black holes. Their masses must therefore be so small that their Schwarzschild radius $r_{\mathrm{S}} = 2Gm/c^2$ is smaller than the distance δl separating them. Dropping again the factor of 2, we get

$$\delta l \geq \sqrt{\frac{\hbar G}{c^3}} = l_{\mathrm{Pl}} \quad . \tag{524}$$

Another way to deduce this limit reverses the role of general relativity and quantum theory. To measure the distance between two objects, we have to localize the first object with respect to the other within a certain interval Δx. This object thus possesses a momentum uncertainty $\Delta p \geqslant \hbar/\Delta x$ and therefore possesses an energy uncertainty $\Delta E = c(c^2 m^2 + (\Delta p)^2)^{1/2} \geqslant c\hbar/\Delta x$. But general relativity shows that a small volume filled with energy changes the curvature, and thus the metric of the surrounding space. For the resulting distance change Δl, compared to empty space, we find the expression $\Delta l \approx G\Delta E/c^4$. In short, if one localizes a first particle in space with a precision Δx, the distance to a second particle is known only with precision Δl. The minimum length δl that can be measured is obviously larger than each of the quantities; inserting the expression for ΔE, we find again that the minimum measurable length δl is given by the Planck length. Ref. 4, 14 Ref. 26, 27, 28 Ref. 29, 30, 31

As a remark, the Planck length being the shortest possible length, it follows that there can be no observations of quantum mechanical effects for situations in which the corresponding de Broglie or Compton wavelength would be smaller. This is one of the reasons why in everyday, macroscopic situations, e.g. in car-car collisions, one never observes embryo-antiembryo pair production or quantum interference effects. In contrast, in proton-proton collisions one both observes pair production and interference effects.

In summary, from two simple properties common to all length measuring devices, namely that they can be counted and that they can be read out, one arrives at the conclusion that *lengths smaller than the Planck length cannot be found in measurements*. Whatever the method used, whether lengths are measured with a meter bar or by measuring time of flight of particles between the end points: one cannot overcome this fundamental limit. It follows that in its usual sense as entity without size, *the concept of "point in space" has no experimental backing*. In the same way, the term "event", being a combination of "point in space" and "instant of time", also loses its meaningfulness for the description of nature.

Ref. 32 These results are often summarized in the so-called generalized uncertainty principle

$$\Delta p \Delta x \geqslant \hbar/2 + f \frac{G}{c^3} (\Delta p)^2 \qquad (525)$$

or

$$\Delta p \Delta x \geqslant \hbar/2 + f \frac{l_{\mathrm{Pl}}^2}{\hbar} (\Delta p)^2 \qquad (526)$$

where f is a numerical factor of order unity. A similar expression holds for the time-energy uncertainty relation. The first term on the right hand side is the usual quantum mechanical one. The second term, negligible at everyday life energies, plays a role only near Planck energies. It is due to the changes in space-time induced by gravity at these high energies. One easily deduces from (525) that the generalized principle automatically implies that Δx can never be smaller than $f^{1/2} l_{\mathrm{Pl}}$.

The generalized uncertainty principle is derived in exactly the same way in which Heisenberg derived the original uncertainty principle $\Delta p \Delta x \geqslant \hbar/2$ for an object: by using the deflection of light by the object under a microscope. A careful rederivation of the process, not disregarding gravity, yields equation (525). Ref. 32

For this reason all approaches which try to unify quantum mechanics and gravity must yield this relation; indeed it appears in canonical quantum gravity, in superstring theory, and in the quantum group approach. Ref. 33 Ref. 34 Ref. 35, 36

It is instructive to remember that quantum mechanics starts when one realizes that the classical concept of action makes no sense below the value of $\hbar$; similarly, unified theories such as quantum gravity start when one realizes that the classical concepts of time and length make no sense below Planck values. However, the usual description of space-time does contain such small values; the usual description claims the existence of intervals smaller than the smallest measurable one. *Therefore the continuum description of space-time has to be abolished in favor of a more appropriate one.*

Ref. 38 This is clearly expressed in a new uncertainty relation appearing at Planck scales. Inserting $c\Delta p \geqslant \Delta E \geqslant \hbar/\Delta t$ into equation (525) we get

$$\Delta x \Delta t \geqslant \hbar G/c^4 = t_{\mathrm{Pl}} l_{\mathrm{Pl}} \,, \qquad (527)$$

which of course has no counterpart in standard quantum mechanics. A final way to convince oneself that points have no meaning is that a point is an entity with vanishing volume; however, the minimum volume possible in nature is the Planck volume $V_{\mathrm{Pl}} = l_{\mathrm{Pl}}^3$. In other words, the Planck units do not only provide natural units, they also provide – within a factor of order one – the limit values of space and time.

Space-time points are idealizations of events. But this idealization is incorrect. The use of the concept of "point" is similar to the use of the concept of "aether" one century ago: one cannot measure it, and it is useful to describe observations only until one has found the way to describe them without it.

Farewell to the space-time manifold

But the consequences of the Planck limits for time and space measurements can be taken much further. To put the previous results in a different way, points in space and time have size, namely the Planck size. It is commonplace to say that given any two points in space or two instants of time, there is always a third in between. Physicists sloppily call this property continuity, mathematicians call it denseness. However, at Planck dimensions, this property cannot hold since intervals smaller than the Planck time can never be found: thus points and instants are not dense, and *between two points there is not always a third.* But this means that *space and time are not continuous.* Of course, at large scales they are – approximately – continuous, in the same way that a piece of rubber or a liquid seems continuous at everyday dimensions, but is not at small scales. This means that to avoid Zeno's paradoxes resulting from the infinite divisibility of space and time we do not need any more the system of differential calculus; we can now directly dismiss the paradoxes because of their incorrect premises on the nature of space and time.

But let us go on. Special relativity, quantum mechanics and general relativity all rely on the idea that time can be defined for all points of a given reference frame. However, two clocks at a distance l cannot be synchronized with arbitrary precision. Since the distance between two clocks cannot be measured with an error smaller than the Planck length l_{Pl}, and transmission of signals is necessary for synchronization, it is not possible to synchronize two clocks with a better precision than the time $l_{Pl}/c = t_{Pl}$, the Planck time. Due to this impossibility to synchronize clocks precisely, the idea of a single time coordinate for a whole reference frame is only approximate, and cannot be maintained in a precise description of nature.

Moreover, since the difference between events cannot be measured with a precision better than a Planck time, for two events distant in time by this order of magnitude, it is not possible to say with complete certainty which one *precedes* the other. This is an important Ref. 39 result. If events cannot be ordered at Planck scales, the concept of time, which is introduced in physics to describe sequences, cannot be defined at all. In other words, after dropping the idea of a common time coordinate for a complete frame of reference, we are forced to drop the idea of time at a single "point" as well. For example, the concept of 'proper time' loses its sense at Planck scales.

For the case of space, it is straightforward to use the same arguments to show that length measurements do not allow us to speak of continuous space, but only about approximately continuous space. Due to the lack of measurement precision at Planck scales, the concept of spatial order, of translation invariance and isotropy of the vacuum, and of global coordinate systems lack experimental backing at those dimensions.

But there is more to come. The very existence of a minimum length contradicts special relativity, where it is shown that whenever one changes to a moving coordinate system a given length undergoes a Lorentz contraction. A minimum length cannot exist in special relativity; therefore, *at Planck dimensions, space-time is neither Lorentz invariant, nor diffeomorphism invariant, nor dilatation invariant.* All the symmetries at the basis of special and general relativity are thus only approximately valid at Planck scales.

Due to the imprecision of measurement, most familiar concepts used to describe spatial relations become useless. For example, the concept of *metric* also loses its usefulness at Planck scales. Since distances cannot be measured with precision, the metric cannot be determined. We deduce that it is impossible to say precisely whether space is flat or curved. In other words, *the impossibility to measure lengths exactly is equivalent to fluctuations of the curvature and thus of gravitation*. Ref. 26, 40

In addition, even the number of space dimensions makes no sense at Planck scales. Let us remind ourselves how one determines this number experimentally. One possible way is to determine how many points one can choose in space such that all their distances are equal. If we can find at most n such points, the space has $n - 1$ dimensions. One recognizes directly that without reliable length measurements there is no way to determine reliably the number of dimensions of a space at Planck scales with this method.

Another simple way to check for three dimensions is to make a knot in a shoe string and glue the ends together: since it stays knotted we know that space has three dimensions, because it is a mathematical theorem that in spaces with more or less than three dimensions, knots do not exist. Again, at Planck dimensions the measurement errors do not allow to say whether a string is knotted or not, because measurement limits at crossings make it impossible to say which strand lies above the other; in short, we cannot deduce whether space has three dimensions or not at Planck dimensions.

There are many other methods to determine the dimensionality of space. For example, one can determine the dimension using only the topological properties of space. If one draws a covering of a topological space with open sets, there are always points which are elements of several sets of the covering. Call p the maximal number of sets of which a point can be an element in a given covering. Determine this number for all coverings. The minimum value of p, minus one, gives the dimension of the space. (Note that if physical space is not a manifold, the various methods could give different answers for the dimensionality. For linear spaces without norm, the number of dimensions cannot even be defined at all.)

But all these methods use the fact that the concept of dimensionality is based on a precise definition of the concept of neighborhood. But at Planck scales, as just mentioned, length measurements do not allow us to say with certainty whether a given point is inside or outside a given volume. In short, whatever method one uses, the lack of reliable length measurements means that *at Planck scales, the dimensionality of physical space is not defined.* It should therefore not come as a surprise that when we *approach* those scales, we could get a scale-dependent answer, different from three.

The reason for the troubles with space-time become most evident when one remembers the well-known definition by Euclid: “A point is that which has no part.” As Euclid clearly understood, a physical point, and here the stress is on *physical*, cannot be defined *without* some measurement method. A physical point is an idealization of position, and as such includes measurement right from the start. In mathematics however, Euclid’s definition is rejected, because mathematical points do not need metrics for their definition. Mathematical points are elements of a set, usually called a space. In mathematics, a measurable or a metric space is a set of points equipped *in addition* with a measure or a metric. Mathematical Ref. 41

points do not need a metric for their definition; they are basic entities. In contrast to the mathematical situation, the case of physical space-time, the concepts of measure and of metric are *more fundamental* than that of a point. The difficulty of distinguishing physical and mathematical space arises from the failure to distinguish a mathematical metric from a physical length measurement.

Perhaps the most beautiful way to make this point clear is the Banach-Tarski theorem or paradox, which shows the limits of the concept of volume. This theorem states that a sphere made of *mathematical points* can be cut into six pieces in such a way that two sets of three pieces can be put together and form two spheres, each of the *same volume* as the original one. However, the necessary cuts are "infinitely" curved and thin. For physical matter such as gold, unfortunately – or fortunately – the existence of a minimum length, namely the atomic distance, makes it impossible to perform such a cut. For vacuum, the puzzle reappears: for example, the energy of its zero-point fluctuations is given by a density times the volume; following the Banach-Tarski theorem, the zero point energy content of a single sphere should be equal to the zero point energy of two similar spheres each of the same volume as the original one. This paradox is solved by the Planck length, which provides a fundamental length scale also for the vacuum. Therefore, *the concept of volume is only well defined at Planck scales if a minimum length is introduced*. Ref. 42

In summary, *physical space-time cannot be a set of mathematical points.* But the surprises are not finished. At Planck dimensions, since both temporal order and spatial order break down, there is no way to say if the distance between two near enough space-time regions is space-like or time-like. Measurement limits make it impossible to distinguish the two cases. *At Planck scales, time and space cannot be distinguished from each other*. In summary, space-time at Planck scales is not continuous, not ordered, not endowed with a metric, not four-dimensional, not made of points. If we compare this with the definition of the term manifold, (a manifold is what locally looks like an euclidean space) not one of its defining properties is fulfilled. We arrive at the conclusion that *the concept of a space-time manifold has no backing at Planck scales*. But this idea is slow to disappear, because *even though both general relativity and quantum mechanics use continuous space-time, the combination of both theories does not.*

Where does the incorrect idea of continuous space-time take its start? In everyday life, as well as in physics, space-time is introduced to describe observations. It is thus from the beginning a bookkeeping device. The properties of space-time are extracted from the properties of observables. From the fact that observables can be added, multiplied by numbers and between each other, one extrapolates that they can take continuous values. From this extrapolation, one deduces that length and time intervals can take continuous values, and in particular arbitrary small values. From this consequence, one gets the possibility to define points and sets of points. An elaborate part of mathematics, topology, shows, how from a set of points with help of neighbourhood relations, separation properties, one can construct a topological space, then with help of a metric a metric space, and finally, with the appropriate compactness and connectedness relations, a manifold, characterized by its dimension, metric and curvature.

> There is nothing in the world but matter in motion,
> and matter in motion cannot move otherwise than in space and time.
> Lenin

Farewell to observables and measurements

To complete this state of affairs, if space and time are not continuous, all quantities defined as derivatives versus space or time are not defined precisely. Velocity, acceleration, momentum, energy, etc., are only defined in the classical approximation of continuous space and time. Concepts such as 'derivative', 'divergence-free', 'source free', etc., lose their meaning at Planck scales. Even the important tool of the evolution equation, based on derivatives, such as the Schrödinger or the Dirac equation, cannot be used any more.

In fact, all physical observables are defined using at least length and time measurements, as is evident from any list of physical units. Any such table shows that all physical units are products of powers of length, time (and mass) units. (Even though in the SI system electrical quantities have a separate base quantity, the ampere, the argument still holds; the ampere is itself defined by measuring a force, which is measured using the three base units length, time, and mass.) Now, since time and length are not continuous, observables themselves are not continuously varying. *This means that at Planck scales, observables (or their components in a basis) are not to be described by real numbers with – potentially – infinite precision.* Similarly, if time and space are not continuous, the usual expression for an observable quantity A, namely $A(t,x)$, does not make sense: we have to find a more appropriate description. *Physical fields cannot be described by continuous functions at Planck scales.*

In quantum mechanics this means that it makes no sense to define multiplication of observables by continuous, i.e. real numbers, but only by discrete steps. Among others, this means that observables do not form a linear algebra. (One recognizes directly that due to measurement errors, one cannot prove that observables do form such an algebra.) This means that *observables are not described by operators at Planck scales.* But quantum mechanics is based on the superposition principle: without it, it all comes crumbling down. Moreover, the most important observables are the gauge potentials. Since they do not form an algebra, *gauge symmetry is not valid at Planck scales*. Even innocuous looking expressions such as $[x_i, x_j] = 0$ for $x_i \neq x_j$, which are at the basis of quantum field theory, become meaningless at Planck scales. Even worse, also the superposition principle cannot be backed up by experiment at those scales. Even the famous Wheeler-DeWitt equation, often assumed to describe quantum gravity, cannot be valid at those scales.

Similarly, permutation symmetry is based on the premise that we can distinguish two points by their coordinates, and then exchange particles at those two locations. As just seen, this is not possible if the distance between the two particles is small; we concludes that *permutation symmetry has no experimental backing at Planck scales*.

Even discrete symmetries, like charge conjugation, space inversion, and time reversal cannot be correct in that domain, because there is no way to verify them exactly by measurement. *CPT symmetry is not valid at Planck scales.*

All these results are consistent: if there are no symmetries at Planck scales, there also are no observables, since physical observables are representations of symmetry groups. In fact, the limits on time and length measurements imply that *the concept of measurement has no significance at Planck scales*.

Can space-time be a lattice? Can it be dual?

Discretization of space-time has been studied already fifty years ago. The idea that space-time is described as a lattice has also been studied in detail, for example by Finkelstein and by 't Hooft. It is generally agreed that in order to get an isotropic and homogeneous situation for large, everyday scales, the lattice cannot be periodic, but must be random. Moreover any fixed lattice violates the result that there are no lengths smaller than the Planck length: due to the Lorentz contraction, any moving observer would find lattice distances smaller than the Planck value. Finally, where would a particle be *during* the jump from one lattice point to the next? *Space-time cannot be a lattice.* In fact, the idea of space-time as a lattice is based on the idea that if a minimum distance exists, then all distances are a multiple of this minimum. However, as we will see, there is no evidence for this at all and actually there evidence for the contrary. Ref. 43 Ref. 44 Ref. 45 Ref. 46

If space-time is not a set of points or events, it must be something else. Three hints already appear at this stage. The first necessary step to improve the description of motion starts with the recognition that to abandon "points" means to abandon the *local* description of physics. Both quantum mechanics and general relativity assume that the phrase 'observable at a point' had a precise meaning. Due to the impossibility of describing space as a manifold, this expression is no longer useful. The unification of general relativity and quantum physics forces a *non-local* description of nature at Planck scales.

The existence of a minimal length implies that there is no way to physically distinguish locations that are spaced by even smaller distances. One is tempted to conclude that therefore *any* pair of locations cannot be distinguished, even if they are one meter apart, since on any path joining two points, any two nearby locations cannot be distinguished. We notice that this situation is similar to the question on the size of a cloud or that of an atom. Measuring water density or electron density, we find non-vanishing values at any distance from the center of the cloud; however, an effective size of the cloud can still be defined, because it is very improbable to see effects of a cloud's or of an atom's presence at distances much larger than this effective size. Similarly, one guesses that two space-time points at macroscopic distance from each other can be distinguished because the probability that they will be confused drops rapidly with increasing distance. In short, we are thus led to a *probabilistic* description of space-time. It becomes a macroscopic observable, the *statistical, or thermodynamic limit* of some microscopic entities.

We note that a *fluctuating* structure for space-time would also avoid the problems of fixed structures with Lorentz invariance. This property is of course compatible with a statistical description. In summary, the experimental observations of special relativity, i.e. Lorentz invariance, isotropy, and homogeneity, together with that of a minimum distance, point towards a fluctuating description of space-time. In the meantime, research efforts in quantum gravity, superstring theory and in quantum groups have confirmed independently from each other that a probabilistic description of space-time, together with a *non-local* description of observables at Planck dimensions, indeed resolves the contradictions between general relativity and quantum theory. To clarify the issue, we have to turn to the concept of particle.

Before that, a few remarks on one of the most important topics in theoretical physics at present: duality. String theory is build around a new symmetry, *space-time duality* and its generalizations. It states that in nature, physical systems of size R are indistinguishable from those of size $l_{\rm Pl}^2/R$. (In natural units this symmetry is often written $R \leftrightarrow 1/R$.) Duality is another way to look at small dimensions. Intervals below the Planck values make no physical sense. If one nevertheless insists on using such small intervals in a mathematical description, as we automatically do if we use real numbers as coordinates, then one finds the strange result that these small intervals are equivalent to intervals of the dual size.

Farewell to particles

But let us take a step backwards and follow another line of arguments, this time about particles. Apart from space and time, in every example of motion, there is some object involved. One of the important discoveries of the natural sciences was that all objects are made of small constituents, called *elementary particles*. Quantum theory shows that all composite, non-elementary objects have a simple property: they have a finite, non-vanishing size. This property allows us to determine whether a particle is elementary or not. If it behaves like a point particle, it is elementary. At present, only the leptons (electron, muon, tau and the neutrinos), the quarks, and the radiation quanta of the electromagnetic, the weak and the strong nuclear interaction (the photon, the W and Z bosons, the gluons) have been found to be elementary. A few more elementary particles are predicted by various refinements of the standard model. Protons, atoms, molecules, cheese, people, galaxies, etc., are all composite, See page 583 as shown in table 42. Elementary particles are characterized by their vanishing size, their spin, and their mass.

Even though the definition of 'elementary' as point particle is all one needs in the following argument, it is not complete, because it seems to leave open the possibility that future experiments show that electrons or quarks are not elementary. This is not so! In fact any particle smaller than its own Compton wavelength is elementary. If it were composite, there would be a lighter component inside it; this lighter particle would have a larger Compton wavelength than the composite particle. This is impossible, since the size of a composite particle must be larger than the Compton wavelength of its components. (The possibility that all components be heavier than the composite, which would avoid this argument, does not lead to satisfying physical properties; e.g. it leads to intrinsically unstable components.)

The *size* of an object, e.g. the one given in table 1, is defined as the length at which one observes differences from point-like behavior. This is the way in which, using alpha particle scattering, the radius of the atomic nucleus was determined for the first time in Rutherford's experiment. In other words, the size d of an object is determined by measuring the interference pattern it creates when it scatters a beam of probe particles. For examples, this is the way we determines sizes of objects when we look at them: we make use of scattered photons. In general, in order to make use of these interference effects, the wavelength λ of the probe must be smaller than the object size d to be determined. The de Broglie wavelength of the probe is given by the mass and the relative velocity v between the probe and the unknown object, i.e. we need $d > \lambda = \hbar/(mv) \geqslant \hbar/(mc)$. On the other hand, in order to make a scattering experiment possible, the object must not be a black hole, since then it would simply swallow the infalling particle. This means that its mass m must be smaller than that of a black hole of its size, i.e. from equation (516), that $m < dc^2/G$; together with the previous condition we get

$$d > \sqrt{\frac{\hbar G}{c^3}} = l_{\mathrm{Pl}} \quad . \tag{528}$$

In other words, there is no way to observe that an object is smaller than the Planck length. *There is thus no way in principle to deduce from observations that a particle is point-like.* In fact, it makes no sense to use the term "point particle" at all. Of course the existence of a minimal length both for empty space and for objects, are related. If the term "point of space" is meaningless, then the term "point particle" is so as well. Note that as in the case of time, the lower limit on length results from the combination of quantum mechanics and general relativity. (Note also that the minimal size of a particle has nothing to do with the impossibility, quantum theory, to localize a particle to within better than its Compton wavelength.)

The size d of any elementary particle, which following the conventional quantum field description is zero, is surely smaller than its own Compton wavelength $\hbar/(mc)$. Moreover, we have seen above that a particle's size is always larger than the Planck length: $d > l_{\mathrm{Pl}}$. Eliminating the size d we get a condition for the mass m of any elementary particle, namely

$$m < \frac{\hbar}{c\, l_{\mathrm{Pl}}} = \sqrt{\frac{\hbar c}{G}} = m_{\mathrm{Pl}} = 2.2 \cdot 10^{-8}\,\mathrm{kg} = 1.2 \cdot 10^{19}\,\mathrm{GeV/c^2} \tag{529}$$

(This limit, the so-called *Planck mass*, corresponds roughly to the mass of a ten days old human embryo, or, equivalently that of a small flea.) In short, *the mass of any elementary particle must be smaller than the Planck mass*. This fact is already mentioned as "well-known" by Sakharov who explains that these hypothetical particles are sometimes called 'maximons'. And indeed, the known elementary particles all have masses well below the Planck mass. (Actually, the question why their masses are so incredibly much smaller than the Planck mass is one of the main questions of high energy physics. But this is another story.) Ref. 47

There are many other ways to arrive at this mass limit. For example, in order to measure mass by scattering – and that is the only way for very small objects – the Compton wavelength of the scatterer must be larger than the Schwarzschild radius; otherwise the probe would be swallowed. Inserting the definition of the two quantities and neglecting the factor 2, we get again the limit $m < m_{\mathrm{Pl}}$. (In fact it is a general property of descriptions of nature that a minimum space-time interval leads to an upper limit for elementary particle masses.) The importance of the Planck mass will become clear shortly. Ref. 48

Another property connected with the size of a particle is its electric dipole moment, which describes the deviation from a spherical shape of its charge distribution. The standard model of elementary particles gives as *upper* limit for the dipole moment of the electron d_e a value of Ref. 49

$$|d_e| < 10^{-39}\,\mathrm{m}\; e \tag{530}$$

where e is the charge of the electron. This value is ten thousand times smaller than $l_{\mathrm{Pl}}\, e$; using the fact that the Planck length is the smallest possible length, this implies either that charge can be distributed in space, or that estimate is wrong, or that the standard model is wrong, or several of these. Only future will tell.

Let us return to some other strange consequences for particles. In quantum field theory, the difference between a virtual and a real particle is that a real particle is on shell, i.e. it obeys $E^2 = m^2c^4 + p^2c^2$, whereas a virtual particle is off shell, i.e. $E^2 \neq m^2c^4 + p^2c^2$. Due to the fundamental limits of measurement precision, *at Planck scales we cannot determine whether a particle is real or virtual.*

But that is not all. Since antimatter can be described as matter moving backwards in time, and since the difference between backwards and forwards cannot be determined at Planck scales, *matter and antimatter cannot be distinguished at Planck scales.*

Particles are also characterized by their spin. Spin describes two properties of a particle: its behavior under rotations (and if the particle is charged, the behavior in magnetic fields) and its behavior under particle exchange. The wave function of spin 1 particles remain invariant under rotation of 2π, whereas that of spin 1/2 particles changes sign. Similarly, the combined wave function of two spin 1 particles does not change sign under exchange of particles, whereas for two spin 1/2 particles it does.

One sees directly that both transformations are impossible to study at Planck scales. Given the limit on position measurements, the position of the axis of a rotation cannot be well defined, and rotations become impossible to distinguish from translations. Similarly, position imprecision makes the determination of precise separate positions for exchange experiments impossible. In short, *spin cannot be defined at Planck scales, and fermions cannot distinguished from bosons, or, differently phrased, matter cannot be distinguished from radiation at Planck scales.* We can thus easily imagine that supersymmetry, a unifying symmetry between bosons and fermions, somehow becomes natural at those dimensions. But let us now move to the main property of elementary particles.

Farewell to mass

The Planck mass divided by the Planck volume, i.e. the Planck density, is given by

$$\rho_{\mathrm{Pl}} = \frac{c^5}{G^2\hbar} = 5.2 \cdot 10^{96}\,\mathrm{kg/m^3} \tag{531}$$

and is a useful concept in the following. If one wants to measure the (gravitational) mass M enclosed in a sphere of size R and thus (roughly) of volume R^3, one way to do this is to put a test particle in orbit around it at that same distance R. The universal "law" of gravity then gives for the mass M the expression $M = Rv^2/G$, where v is the speed of the orbiting test particle. From $v < c$, we thus deduce that $M < c^2R/G$; using the fact that the minimum value for R is the Planck distance, we get (neglecting again factors of order unity) a limit for the mass density, namely

$$\rho < \rho_{\mathrm{Pl}} \quad . \tag{532}$$

In other words, *the Planck density is the maximum possible value for mass density.* In particular, in a volume of Planck dimensions, one cannot have a larger mass than the Planck mass.

Interesting things happen when one starts to determine the error ΔM of the mass measurement in a Planck volume. Let us return to the mass measurement by an orbiting probe. From the relation $GM = rv^2$ one deduces by differentiation $G\Delta M = v^2\Delta r + 2vr\Delta v > 2vr\Delta v = 2GM\Delta v/v$. For the error Δv in the velocity measurement one has the uncertainty relation $\Delta v \geqslant \hbar/(m\Delta r) + \hbar/(MR) \geqslant \hbar/(MR)$. Inserting this in the previous inequality, and forgetting again the factor of 2, we get that the mass measurement error ΔM of a mass M enclosed in a volume of size R follows

$$\Delta M \geqslant \frac{\hbar}{cR} \quad . \tag{533}$$

Note that for everyday situations, this error is extremely small, and other errors, such as the technical limits of the balance, are much larger.

As a check of this result, we take another situation, and use relativistic expressions, in order to show that the result does not depend on the details of the situation or the approximations. Imagine having a mass M in a box of size R and weighing the box. (It is supposed that either the box is massless, or that its mass is subtracted by the scale.) The mass error is given by $\Delta M = \Delta E/c^2$, where ΔE is due to the uncertainty in kinetic energy of the mass inside the box. Using the expression $E^2 = m^2c^4 + p^2c^2$ we get that $\Delta M \geqslant \Delta p/c$, which again reduces to equation (533). Now that we are sure of the result, let us continue.

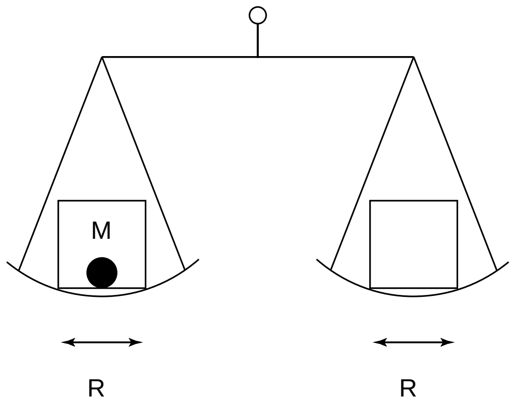


Figure 175 A *Gedankenexperiment* showing that at Planck scales, matter and vacuum cannot be distinguished

From equation (533) we deduce that for a box of Planck dimensions, the mass measurement *error* is given by the Planck mass. But from above we also know that the mass which can be put inside such a box is itself not larger than the Planck mass. In other words, for a box of Planck dimensions, the mass measurement *error* is larger (or at best equal) to the mass contained in it: $\Delta M \geqslant M_{\rm Pl}$. In other words, if one builds a balance with two boxes of Planck size, one empty and the other full, as shown in the figure, nature cannot decide which way the balance should hang! Note that even a repeated or a continuous measurement would not resolve the situation: the balance would then randomly change inclination.

A rephrased argument is as follows. If one has a box of size R, the largest mass one can put in it is a black hole with a Schwarzschild radius of the same value; the smallest mass present in such a box – corresponding to what we call vacuum – is due to the uncertainty relation and is given by that mass whose Compton wavelength matches the size of the box. One therefore has, inside any box of size R, a mass m whose limits are given by:

$$\text{(full box)}\ \frac{c^2R}{G} > m > \frac{\hbar}{cR}\ \text{(empty box)} \tag{534}$$

One sees directly that for sizes R of the order of the Planck scale, the two limits coincide; in other words, one cannot distinguish a full from an empty box.

To be sure of this strange result, we check whether we find the same result if instead of measuring the gravitational mass, as done just now, we measure the inertial mass. The inertial mass for a small object is determined by touching it, i.e. physically speaking, by performing a scattering experiment. To determine the inertial mass inside a region of size R, a probe must have a wavelength smaller that R, and thus a correspondingly high energy. A high energy means that the probe also attracts the particle through gravity. (We thus find the intermediate result that *at Planck scales, inertial and gravitational mass cannot be distinguished.* Even the balance experiment shown in the figure makes this point: at Planck scales, the two effects of mass are always inextricably linked.) Now, in any scattering experiment, e.g. in a Compton-type experiment, the mass measurement is performed by measuring the wavelength change $\delta\lambda$ of the probe before and after the scattering experiment. The mass uncertainty is given by

$$\frac{\Delta M}{M} = \frac{\Delta\delta\lambda}{\delta\lambda} \tag{535}$$

In order to determine the mass in a Planck volume, the probe has to have a wavelength of the Planck length. But we know from above that there always is a minimal wavelength uncertainty given by the Planck length l_{Pl}. In other words, for a Planck volume the mass error is always as large as the Planck mass itself: $\Delta M \geqslant M_{\mathrm{Pl}}$. Again, this limit is a direct consequence of the limit on length and space measurements.

But this result has an astonishing consequence. In these examples, the measurement error is independent of the mass of the scatterer, i.e. independent of whether we start with a situation in which there is a particle in the original volume, or whether there is none. We thus find that in a volume of Planck size, it is impossible to say if there is something or not when weighing it or probing it with a beam!

In short, all arguments lead to the same conclusion: *vacuum, i.e. empty space-time, cannot be distinguished from matter at Planck scales.* Another, often used way to express this is to say that when a particle of Planck energy travels through space it will be scattered by the fluctuations of space-time itself, making it thus impossible to say whether it was scattered by empty space-time or by matter. These surprising results stem from the fact that whatever definition of mass we use, it is always measured via length and time measurements. (This is even the case for normal weight scales: mass is measured by the displacement of some part of the machine.) And the error in these measurements makes *impossible to distinguish vacuum from matter.*

To put it another way, if we measure the mass of a piece of vacuum of size R, the result is always at least $\hbar/cR$; there is no possible way to find a perfect vacuum in an experiment. On the other hand, if we measure the mass of a particle, we finds that the result is size dependent; at Planck dimensions it approaches the Planck mass for every type of particle, be it matter or radiation.

Using another image, when two particles are approached to lengths of the order of the Planck length, the uncertainty in the length measurements makes it impossible to say whether there is something or nothing between the two objects. In short, *matter and vacuum get mixed-up at Planck dimensions.* This is an important result: since both mass and empty space-time can be mixed-up, one has confirmed that they are made of the same "fabric”, confirming the idea presented at the beginning. This idea is now commonplace in all attempts to find a unified description of nature.

This approach is corroborated by the attempts of quantum mechanics in highly curved space-time, where a clear distinction between the vacuum and particles is not possible; the well-known Unruh radiation, namely that any observer either accelerated in vacuum or in a gravitational field in vacuum still detects particles hitting him, is one of the examples which shows that for curved space-time the idea of vacuum as a particle-free space does not work. Since at Planck scales it is impossible to say whether space is flat or not, it follows that it is also impossible to say whether it contains particles or not. Ref. 50

Curiosities

These strange results imply many others; here are a few.

• In summary, we have a new answer to the old question: Why is there anything instead of nothing? Well, there is *no difference* between anything and nothing.

• The Planck energy is rather large. Imagine that we want to impart electrons this amount of energy using a particle accelerator. How large would that accelerator be? Challenge

• We now can honestly say about ourselves: we are made of nothing.

• Quantum mechanics alone gives, via the Heisenberg uncertainty relation, a lower limit on the spread of measurements, but strangely enough not on their precision, i.e. not on the number of significant digits. Jauch gives the example that atomic lattice constants are known much more precisely than the position uncertainty of single atoms inside the crystal. Despite claims to the contrary, can you show why this cannot be the case for space and time? Ref. 51 Ref. 36 Challenge

• Of course, the idea that vacuum is not empty is not new. Already Aristoteles argued for a filled vacuum, even though he used incorrect arguments, seen from today’s perspective. Also in the fourteenth century the discussion on whether empty space was composed of indivisible entities was rather common, but died down again later.

• One way to generalize the results presented here is to assume that at Planck energy, nature is *event symmetric*, i.e. that nature is symmetric under exchange of any two events. This approach, developed by Phil Gibbs, provides an additional insight to the strange behaviour of nature at extreme scales. Ref. 25

• Naked singularities do not appear in nature, and the question becomes uninteresting, thus ending decades of speculations. See page 287

• Since mass density and therefore energy volume density is limited, one knows that the number of degrees of freedom of any object of finite volume is limited. This means among others that perfect baths do not exist. Baths play an important role in thermodynamics (which is thus found to be only an approximation) and also in recording and measuring devices: when a device measures, it switches from a neutral state to a state in which it shows the result of the measurement. In order to avoid that the device returns to the neutral state, it must be coupled to a bath. Without a bath, a reliable measuring device cannot be made. In short, perfect clocks and length measuring devices do not exist because nature puts a limit on their storage ability. More about this shortly. Ref. 52

• We had seen earlier that characterizing nature as made of particles and vacuum creates problems when interactions are included, since interactions on one hand are the difference between the parts and the whole, and on the other hand, as quantum theory says, interactions

See page 410 are exchanges of particles. This connection can be used to show that either vacuum and particles are not everything nature is made of, or that something is counted double. Since matter and space-time are made of the same 'stuff,' both paradoxes are solved.

The baggage left behind

In this rapid walk, we have destroyed all the experimental pillars of quantum theory: the superposition principle, space-time symmetry, gauge symmetry, and the discrete symmetries. We also have destroyed the foundations of general relativity, namely the existence of a space-time manifold, of the field concept, the particle concept, and of the concept of mass. It was even shown that matter and space-time cannot be distinguished. It seems that one has destroyed every concept used for the description of motion, and thus made the description of nature impossible. One naturally asks whether one can save the situation. On first sight it seems that we have to leave everything behind.

In summary, the usual concepts of matter and of radiation are not applicable at Planck dimensions. Usually, it is assumed that matter and radiation are made of interacting elementary particles. The concept of an elementary particle is that of an entity which is countable, point-like, real and not virtual, with a definite mass, a definite spin, distinct from its antiparticle, and most of all, distinct from vacuum, which is assumed to have zero mass. All these properties are found to be incorrect at Planck scales. *At Planck dimensions, it does not make sense to use the concepts of 'mass', 'vacuum', 'elementary particle', 'radiation', and 'matter'.*

This result is encouraging: since matter is not distinguishable for vacuum, and since this is correct for all types of particles, be they matter or radiation, we have a argument showing that the quest for unification in the description of elementary particles is correct and necessary.

Moreover, since the concepts 'mass', 'time', and 'space' cannot be distinguished from each other, we also know that a new, *single* entity is necessary to define both particles and space-time. To find out more about this new entity, three approaches have being pursued at the end of the twentieth century. The first, quantum gravity, especially the one using the Ashtekar's new variables and the loop representation, starts by generalizing *space-time symmetry*. The second, string theory, starts by generalizing *gauge symmetries* and interactions, and the third, the algebraic quantum group approach, looks for generalized *permutation symmetries*. We will describe them in more detail shortly. Ref. 10 Ref. 11 Ref. 12

We also found that there is no argument showing that space and time are continuous or made of points, and that in contrast, the combination of relativity and quantum theory makes this impossible. In order to proceed in our escalation, we need to leave behind us our usual concept of space-time. *At Planck dimensions, the concepts of 'space-time points' or 'mass points' are not applicable to the description of nature.* In other words, we now know that the entity we are looking for is not point-like. How does it look? To get to the top of motion mountain as rapidly as possible, we make use of some explosives to blast away some disturbing obstacles.

Some experimental predictions

There is a race both in experimental and in theoretical physics going on at present: which will be the first experiment that will detect quantum gravity effects, i.e. effects sensitive to the Planck energy?*

A good candidate is the measurement of light speed at different frequencies from far away light flashes. There are flashes in nature, called gamma ray bursts, which have an extremely broad spectrum, from 100 GeV down to visible light of about 1 eV. These flashes often originate at cosmological distances d. From the difference in arrival time Δt for the two frequencies one can construct a characteristic energy given by

$$E_{\text{char}} = \frac{\hbar(\omega_1 - \omega_2)d}{c\Delta t} \quad . \tag{536}$$

This energy value is $8 \cdot 10^{16}$ GeV for the best measurement to date. This value is not far from the Planck value, even more so when the missing factors of order unity are included. It is expected that the Planck scale will be reached in a few years, so that tests will become possible on whether the quantum nature of space-time influences the dispersion of light signals. Planck scale effects should produce a minimum value for this dispersion. This effect would allow to confirm that Lorentz symmetry is not valid at Planck scales. Ref. 53 Ref. 54, 55

Another candidate is the direct detection of distance fluctuations between bodies. Gravitational wave detectors are sensitive to extremely small noise signals in length measurements. One noise signal should be due to the distance fluctuations induced at Planck energies. There is hope that the sensitivity to noise of the detectors will reach the required levels. Ref. 56, 55

A third candidate is the detection of effects signalling the loss of CPT symmetry at high energies. Especially in the case of the decay of certain elementary particles, in particular neutral kaons, the experimental measurement precision is approaching the detection of Planck scale effects. Ref. 55

A fourth candidate is the possibility that quantum gravity effects might change the threshold energy at which certain particle reactions become possible. It might be that extremely high energy photons or cosmic rays allow to deduce that Lorentz invariance is indeed broken near the Planck scale. Ref. 57

A few candidates for quantum gravity effects by the author follow. One way to summarize the situation so far is the following.** Special relativity started with the discovery that observable speeds are limited by the speed of light c. Quantum theory starts with the result that observable actions are limited by $\hbar/2$. Gravitation shows that for every system with length L and mass M, the observable ratio L/M is larger or equal than the constant $4G/c^2$.

Combining these results, we deduced that all physical observables are *bound*, namely by what are usually called the Planck values, though modified by a factor of square root of 2 (or several of them) to compensate the numerical factors from the previous sentence lost over time by physicists. In fact, one needs to exchange $\hbar$ by $\hbar/2$ and G by $4G$ in all the defining expressions. One finds directly that the limit for lengths and times is $\sqrt{2}$ times the Planck value, and that the important limit for energy is the Planck value divided by $\sqrt{8}$.***

∗ As more candidates appear, they will be added to this section.

∗∗ This subsection, in contrast to the ones so far, is speculative; it was added in February 2001.

∗∗∗ The entropy of a black hole is thus given by the ratio between its horizon and *half* the minimal area.

Interestingly, the existence of bounds on all observables allows to deduce several experimentally testable predictions for the unification of quantum theory and general relativity. These predictions do not depend on the detailed final theory.

These prediction are possible because so far, we cheated. The (corrected) Planck values do not seem to be the actual limits to measurements. The actual measurement limits are more strict.

First of all, for any measurement, one needs certain fundamental conditions to be realized. Take the length measurement of an object. One needs to be able to distinguish between matter and radiation, as the object to be measured is made of matter, and radiation is the measurement tool which is used to read of distances on the ruler. For a measurement process, one needs an interaction, and that implies the use of radiation. Note that even the use of particle scattering to determine lengths does not invalidate this general requirement.

Even for the measurement of wavelengths one needs to distinguish matter and radiation, as the matter is necessary to compare two wavelengths. In fact, all length measurements whatsoever require the distinction between matter and radiation.* But this distinction is impossible at the energies of the grand unification, in which the electroweak and the strong nuclear interactions are unififed. At higher energies, particles of matter and of radiation cannot be distinguished from each other. To sum up, no measurement can be performed at energies at or above the GUT unification energy.

In short, the smallest length in nature is $\sqrt{2}$ times the Planck length multiplied by the ratio between the maximal energy, namely $E_{Pl}/\sqrt{8}$, and the unification energy. Following
Ref. 58 present estimates, one has $E_{GUT} = 10^{16}\,\text{GeV}$, which implies that

$$L_{min} = \sqrt{2}\, l_{Pl} \frac{E_{Pl}}{\sqrt{8} E_{GUT}} \approx 10^{-32}\,\text{m} \approx 600\, l_{Pl} \quad . \tag{537}$$

It is unlikely that measurements at these dimensions will ever be possible. Thus the smallest *measurable* length is quite a bit larger than the Planck scale of nature discussed above. The reason is that the Planck scale is that length for which particles and vacuum cannot be distinguished, whereas the minimal measurable length is the distance at which particles of matter and radiation cannot be distinguished. This happens at lower energy.

If above the impression was created that measurements could reach the Planck scale, we thus have to retract: *the minimum measurable length cannot be smaller than* L_{min}.

The experimentally determined factor of about 600 is one of the great riddles of physics. It is the high energy translation of the quest to understand why the electromagnetic coupling constant is about $1/137$, or simpler, why all things have the colours they have. Only the final theory of motion will answer this question.

As a consequence, this result also puts a stronger bound on the electric dipole moment d of elementary particles, i.e. on any particles with no constituents. One gets

$$d_{min} = eL_{min} = 1.5 \cdot 10^{-51}\,\text{Cm} \tag{538}$$

Ref. 59, 49 which seems be in the reach of future experiments. This improved limit might be the

Of course, a detailed investigation also shows that the Planck mass (divided by $\sqrt{8}$) is a limit for elementary particles from *below*, and for black holes from *above*. For usual systems, there is no limit.

∗ To speak in modern high energy concepts, all measurements require broken supersymmetry.

simplest possible measurement of yet unpredicted quantum gravity effects. Measuring the dipole moment could be a way to determine the unification enrgy (the factor 600) independently of high energy physics experiments, and possibly to higher precision.

Finally, the electromagnetic, the weak, and the strong interactions are characterized by coupling constants whose inverse depends linearly on the logarithm of the energy. It is usually assumed that these three lines meet at the already mentioned unification energy. Measurements put the unification coupling value at about 1/26. Ref. 58

Interestingly, the bound on the measurability of observables also puts a bound on the measurement *precision* for each observable. This bound is of no importance at everyday life energy, but it is important at high energy.

What is the precision with which a coupling constant can be measured? It is sufficient to study the electromagnetic constant as an example. This constant α, also called the fine structure constant, is related to the charge by

$$q = \sqrt{4\pi\varepsilon_o \hbar c \alpha} \tag{539}$$

Now, any electrical charge itself is defined and measured by comparing, in an electrical field, the acceleration the charged object is subjected to with the acceleration of some unit charge. In other words,

$$\frac{q}{q_{\text{unit}}} = \frac{ma}{m_{\text{unit}} a_{\text{unit}}} \tag{540}$$

Therefore, any error in mass and acceleration measurements implies errors in the charge and thus in coupling constant measurements.

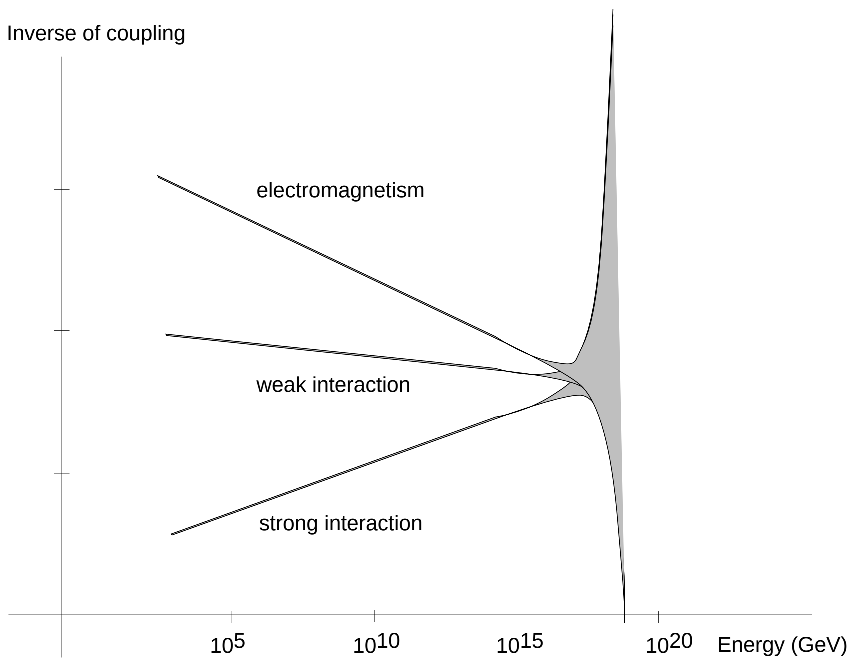


Figure 176 Coupling constants running with energy

In addition, we know from the above discussions that for all measurements of observables at the limit energy, the measurement error is the same as the value to be measured. For measurements at lower energies, the minimum values for length and time measurements imply minimum values for the relative measurement errors.

Therefore, for any energy measurement, the minimal measurement error is given by the ratio between the energy to be measured and the limit energy. Inserting this into the graph of the running coupling constants, one gets the result shown in figure 176.

The search for consequences of this *fan-out effect* is delightful. One way to put the result is to say that coupling constants are by definition affected with an error. For example, all measurement devices, be they clocks, meter bars, scales, or something else, work using electromagnetic effects at energies of around 1 eV. This is about 10^{-25} times the GUT energy. As a consequence, the measurement precision of any observable is limited to about 25 digits. The present precision record is about 15 digits, and for the electromagnetic coupling constant it is about 9 digits. The prediction can thus be tested only in quite some time.*

The fun is thus to find a system in which the variations of the coupling constant appear more clearly in the measurements. It might be that high precision measurements of the *g*-factor of elementary particles or high energy cosmic ray reactions can show some effects of the fan-out. The lifetime of elementary particles could also be affected. Can you find another?

Challenge

In summary, the experimental detection of quantum gravity effects might be possible in the 21st century, despite their weakness. The successfull detection of any such effect will be one of the highlights of physics.

References

1 Steven WEINBERG, *The cosmological constant problem*, Reviews of Modern Physics **61**, pp. 1–23, 1989. Cited on page 578.

2 Steven WEINBERG, *The quantum theory of fields*, Cambridge University Press, volume I, 1995, and II, 1996. Cited on page 578.

3 The difficulties are summarized by B.S. DE WITT, *Quantum field theory in curved space-time*, Physics Reports **19**, pp. 295–357, 1975. Cited on page 580.

4 C.W. MISNER & K.S. THORNE, J.A. WHEELER, *Gravitation*, Freeman, 1973. Cited on page 580, 582, 586.

5 J.A. WHEELER, in *Relativity, Groups, and Topology*, edited by C. DeWitt & B. DeWitt (Gordon and Breach, 1994). See also J.A. WHEELER, *Physics at the Planck length*, International Journal of Modern Physics A **8**, pp. 4013–4018, 1993. Cited on page 580.

6 J.L. FRIEDMAN & R.D. SORKIN, *Spin 1/2 from gravity*, Physical Review Letters **44**, pp. 1100–1103, 1980. Cited on page 580.

7 A.P. BALACHANDRAN, G. BIMONTE, G. MARMO & A. SIMONI, *Topology change and quantum physics*, Nuclear Physics B **446**, pp. 299–314, 1995, also preprint hep-th/9503046. Cited on page 580.

8 Many themes here resulted from an intense exchange with Luca Bombelli, now at the Department of Physics and Astronomy of the University of Mississippi. These pages are a reworked

* It might be that the correct energy of everyday life has to be taken as the electron rest energy; that would change the prediction to only 19 digits.

version of the ones published in french as Christoph SCHILLER, *Le vide diffère-t-il de la matière?*, in E. GUNZIG & S. DINER, editeurs, *Le vide – Univers du tout et du rien – Des physiciens et des philosophes s'interrogent*, Les Éditions de l'Université de Bruxelles, 1998. An older english version is also available as Christoph SCHILLER, *Does matter differ from vacuum?*, http://xxx.lanl.gov/abs/gr-qc/9610066. Cited on page 581.

9 See for example R.P. FEYNMAN, *The Feynman lectures of physics*, Addison Wesley, 1963. Cited on page 581.

10 A. ASHTEKAR, *Recent mathematical developments in quantum general relativity*, preprint gr-qc/9411055. C. ROVELLI, *Ashtekar Formulation of general relativity and loop-space non-perturbative quantum gravity: a report*, Classical and Quantum Gravity **8**, pp. 1613–1675, 1991. Cited on page 582, 599.

11 M.B. GREEN, J.H. SCHWARZ & E. WITTEN, *Superstring theory*, volume 1 and 2 (Cambridge University Press, 1987). G. VENEZIANO, *A stringy nature needs just two constants*, Europhysics Letters **2**, pp. 199–204, 1986. Cited on page 582, 599.

12 S. MAJID, *Introduction to Braided Geometry and q-Minkowski Space*, preprint hep-th/9410241. See also S. MAJID, *Beyond supersymmetry and quantum symmetry (an introduction to braided groups and braided matrices)*, in *Quantum groups, integrable statistical models, and knot theory*, edited by M.-L. GE & H.J. DE VEGA, World Scientific, 1993, pp. 231–282. Cited on page 582, 599.

13 C.J. ISHAM, *Prima facie questions in quantum gravity*, preprint gr-qc/9310031. Cited on page 582.

14 Steven WEINBERG, *Gravitation and Cosmology*, (Wiley, 1972). Cited on page 582, 585, 586.

15 The observations of black holes at the centre of galaxies and elsewhere are summarized by R. BLANDFORD & N. GEHRELS, *Revisiting the black hole*, Physics Today **52**, June 1999. Cited on page 582, 661.

16 M. PLANCK, *Über irreversible Strahlungsvorgänge*, Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften zu Berlin, pp. 440–480, 1899. Today it is commonplace to use Dirac's $\hbar$ instead of Planck's h, which he originally called b. Cited on page 584.

17 The argument is given e.g. in E.P. WIGNER, *Relativistic invariance and quantum phenomena*, Reviews of Modern Physics **29**, pp. 255–258, 1957. Cited on page 584.

18 The starting point for the following arguments is taken from M. SCHÖN, *Operative time definition and principal uncertainty*, preprint hep-th/9304024, and from T. PADMANABHAN, *Limitations on the operational definition of space-time events and quantum gravity*, Classical and Quantum Gravity **4**, pp. $L107$–$L113$, 1987; see also his earlier papers referenced there. Cited on page 584.

19 W. HEISENBERG, *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, Zeitschrift für Physik **43**, pp. 172–198, 1927. Cited on page 584.

20 E.H. KENNARD, *Zur Quantenmechanik einfacher Bewegungstypen*, Zeitschrift für Physik **44**, pp. 326–352, 1927. Cited on page 584.

21 M.G. RAYMER, *Uncertainty principle for joint measurement of noncommuting variables*, American Journal of Physics **62**, pp. 986–993, 1994. Cited on page 584.

22 H. SALECKER & E.P. WIGNER, *Quantum limitations of the measurement of space-time distances*, Physical Review **109**, pp. 571–577, 1958. Cited on page 585.

23 E.J. ZIMMERMAN, *The macroscopic nature of space-time*, American Journal of Physics **30**, pp. 97–105, 1962. Cited on page 585.

24 S.W. HAWKING, *Particle creation by black holes*, Communications in Mathematical Physics **43**, pp. 199–220, 1975; see also S.W. HAWKING, *Black hole thermodynamics*, Physical Review D**13**, pp. 191–197, 1976. Cited on page 585.

25 P. Gibbs, *The small structure of space-time: a bibliographical review*, preprint hep-th/9506171. Cited on page 585, 598.

26 C.A. Mead, *Possible connection between gravitation and fundamental length*, Physical Review B **135**, pp. 849–862, 1964. Cited on page 586, 586, 589, 605.

27 P.K. Townsend, *Small-scale structure of space-time as the origin of the gravitational constant*, Physical Review D **15**, pp. 2795–2801, 1977. Cited on page 586.

28 M.-T. Jaekel & S. Renaud, *Gravitational quantum limit for length measurement*, Physics Letters A **185**, pp. 143–148, 1994. Cited on page 586.

29 D.V. Ahluwalia, *Quantum measurement, gravitation, and locality*, Physics Letters B **339**, pp. 301–303, 1994, also preprint gr-qc/9308007. Cited on page 586.

30 L. Garay, *Quantum Gravity and minimum length*, International Journal of Modern Physics A **10**, pp. 145–165, 1995, also preprint gr-qc/9403008; this paper also includes an extensive bibliography. Cited on page 586.

31 G. Amelino-Camelia, *Limits on the measurability of space-time distances in (the semiclassical approximation of) quantum gravity*, Modern Physics Letters A **9**, pp. 3415–3422, 1994, also preprint gr-qc/9603014. See also Y.J. Ng & H. Van Dam, *Limit to space-time measurement*, Modern Physics Letters A **9**, pp. 335–340, 1994. Cited on page 586.

32 The generalized uncertainty relation is an implicit part of reference 26, especially on page 852, but the issue is explained rather unclearly. Obviously the author found the result too simple to be mentioned explicitly. Cited on page 587, 587.

33 C. Rovelli & L. Smolin, *Discreteness of area and volume in quantum gravity*, Nuclear Physics B **442**, pp. 593–619, 1995. R. Loll, *The volume operator in discretized quantum gravity*, preprint gr-qc/9506014. Cited on page 587.

34 D. Amati, M. Ciafaloni & G. Veneziano, *Superstring collisions at Planckian energies*, Physics Letters B **197**, pp. 81–88, 1987. D.J. Gross & P.F. Mende, *The high energy behavior of string scattering amplitudes*, Physics Letters B **197**, pp. 129–134, 1987. K. Konishi, G. Paffuti & P. Provero, *Minimum physical length and the generalized uncertainty principle*, Physics Letters B **234**, pp. 276–284, 1990. P. Aspinwall, *Minimum Distances in Non-Trivial String Target Spaces*, Nuclear Physics B **431**, pp. 78–96, 1994, also preprint hep-th/9404060. Cited on page 587.

35 M. Maggiore, *A generalized uncertainty principle in quantum mechanics*, Physics Letters B **304**, pp. 65–69, 1993. Cited on page 587.

36 A simple approach is S. Doplicher, K. Fredenhagen & J.E. Roberts, *Space-time quantization induced by classical gravity*, Physics Letters B **331**, pp. 39–44, 1994, Cited on page 587, 598.

37 A. Kempf, *Uncertainty relation in quantum mechanics with quantum group symmetry*, Journal of Mathematical Physics **35**, pp. 4483–4496, 1994. A. Kempf, *Quantum Groups and quantum field theory with Nonzero minimal uncertainties in positions and momenta*, Czechoslovak Journal of Physics **44**, pp. 1041–1048, 1994. Cited on page .

38 E.J. Hellund & K. Tanaka, *Quantized space-time*, Physical Review **94**, pp. 192–195, 1954. Cited on page 587.

39 The impossibility to determine temporal ordering in quantum theory is discussed by J. Oppenheimer, B. Reznik & W.G. Unruh, *Temporal ordering in quantum mechanics*, ... 2001. Cited on page 588.

40 A. Peres & N. Rosen, *Quantum Limitations on the measurement of gravitational fields*, Physical Review **118**, pp. 335–336, 1960. Cited on page 589.

41 It is the first definition in Euclids's *Elements*. For an english translation see T. HEATH, *The thirteen books of the elements*, Dover, 1969. Cited on page 589.

42 A beautiful description of the Banach-Tarski paradox is the one by Ian STEWART, *Paradox of the spheres*, New Scientist, 14 January 1995, pp. 28–31. Cited on page 590.

43 H.S. SNYDER, *Quantized space-time*, Physical Review **71**, pp. 38–41, 1947. H.S. SNYDER, *The electromagnetic field in quantized of space-time*, Physical Review **72**, pp. 68–74, 1947. A. SCHILD, *Discrete space-time and integral Lorentz transformations*, Physical Review **73**, pp. 414–415, 1948. E.L. HILL, *Relativistic theory of discrete momentum space and discrete space-time*, Physical Review **100**, pp. 1780–1783, 1950. H.T. FLINT, *The quantization of space-time*, Physical Review **74**, pp. 209–210, 1948. A. DAS, *Cellular space-time and quantum field theory*, Il Nuovo Cimento **18**, pp. 482–504, 1960. Cited on page 592.

44 D. FINKELSTEIN, *"Superconducting" causal nets*, International Journal of Theoretical Physics **27**, pp. 473–519, 1985. Cited on page 592.

45 N.H. CHRIST & R. FRIEDBERG, T.D. LEE, *Random lattice field theory: general formulation*, Nuclear Physics B **202**, pp. 89–125, 1982. G. 'T HOOFT, *Quantum field theory for elementary particles – is quantum field theory a theory?*, Physics Reports **104**, pp. 129–142, 1984. Cited on page 592.

46 See for example L. BOMBELLI, J. LEE, D. MEYER & R.D. SORKIN, *Space-time as a causal set*, Physical Review Letters **59**, pp. 521–524, 1987. G. BRIGHTWELL & R. GREGORY, *Structure of random space-time*, Physical Review Letters **66**, pp. 260–263, 1991. Cited on page 592.

47 A.D. SAKHAROV, *Vacuum quantum fluctuations in curved space and the theory of gravitation*, Soviet Physics – Doklady, **12**, pp. 1040–1041, 1968. Cited on page 594.

48 C. WOLF, *Upper limit for the mass of an elementary particle due to discrete time quantum mechanics*, Il Nuovo Cimento B **109**, pp. 213–218, 1994. Cited on page 594.

49 W. BERNREUTHER & M. SUZUKI, *The electric dipole moment of the electron*, Reviews of Modern Physics **63**, pp. 313–340, 1991. N.F. RAMSEY & A. WEIS, *Suche nach permanenten elektrischen Dipolmomenten: ein Test der Zeitumkehrinvarianz*, Physikalische Blätter **52**, pp. 859–863, 1996. See also the musings in Hans DEHMELT, *Is the electron a composite particle?*, Hyperfine Interactions **81**, pp. 1–3, 1993. Cited on page 594, 601.

50 W.G. UNRUH, *Notes on black hole evaporation*, Physical Review D **14**, pp. 870–875, 1976. W.G. UNRUH & R.M. WALD, *What happens when an accelerating observer detects a Rindler particle*, Physical Review D **29**, pp. 1047–1056, 1984. Cited on page 598.

51 W. JAUCH, *Heisenberg's uncertainty relation and thermal vibrations in crystals*, American Journal of Physics **61**, pp. 929–932, 1993. Cited on page 598.

52 H.D. ZEH, *On the interpretation of measurement in quantum theory*, Foundations of Physics **1**, pp. 69–76, 1970. Cited on page 598.

53 B.E. SCHAEFER, *Severe limits on variations of the speed of light with frequency*, Physical Review Letters **82**, pp. 4964–4966, 21 June 1999. Cited on page 600.

54 G. AMELINO-CAMELIA, J. ELLIS, N.E. MAVROMATOS, D.V. NANOPOULOS & S. SAKAR, *Potential sensitivity of gamma-ray-burster observations to wave dispersion in vacuo*, Nature **393**, p. 763, 1998, and preprint http://arxiv.org/astro-ph/9712103. Cited on page 600.

55 G. AMELINO-CAMELIA, *Phenomenological description of space-time foam*, preprint http://arxiv.org/gr-qc/0104005, 2001. The paper includes a clearly written overview of present experimental approaches to detect quantum gravity effects. Cited on page 600, 600, 600.

56 G. AMELINO-CAMELIA, *An interferometric gravitational wave detector as a quantum gravity apparatus*, Nature **398**, p. 216, 1999, and preprint http://arxiv.org/gr-qc/9808029. Cited on page 600.

57 G. Amelino-Camelia & T. Piran, *Planck scale-deformation of Lorentz symmetry as a solution to the UHECR and the TeV γ paradoxes*, preprint http://arxiv.org/astro-ph/0008107, 2000. Cited on page 600.

58 See the paper by U. Amaldi, W. de Boer & H. Fürstenau, *Comparison of grand unified theories with elektroweak and strong coupling constants measured at LEP*, Physics Letters **260**, pp. 447–455, 1991. This is the standard reference on the topic. Cited on page 601, 602.

59 M.V. Romalis, W.C. Griffith, J.P. Jacobs & E.N. Fortson, *New limit on the permanent electric dipole moment of* ^{199}Hg, Physical Review Letters **86**, pp. 2505–2508, 19 March 2001, or arxiv.org/hep-ex/0012001. Cited on page 601.

31. Nature at large scales – is the universe something or nothing?

Die Grenze ist der Ort der Erkenntnis.*
Paul Tillich

This strange question is the topic of the present leg of our escalation. We explored he properties of nature in the vicinity of Planck dimensions in the previous section; Ref. 1
it is equally fascinating to explore the other limit, i.e. to study the description of motion at large, cosmological scales. Step by step many incredible results will appear, and at the end we will discover a surprising answer to the title question.

This section is not standard textbook material; a large part is original and thus more speculative and questionable.** Even though it aims at explaining in simple words the ongoing research in the domains of quantum gravity and superstring theory, watch out. You will probably find a physicist disagreeing for each sentence of this section!

Typeset in April 2001

We asked questions about the universe already several times during our escalation. In classical physics we enquired about its initial conditions, and whether it is isolated. In the first intermezzo we asked whether the universe is a set, a concept, and whether it exists. In general relativity we gave the classical definition of the term, namely the sum of all matter and space-time, studied its expansion, and asked about its size and topology. In quantum theory we asked whether the universe has a wavefunction, whether it is born from a quantum fluctuation, and whether it allows to define a particle number.

Here we will settle all these issues by combining general relativity and quantum theory at cosmological scales. That will lead us to some of the strangest results we will encounter in our hike.

Cosmological scales

Hic sunt leones.***

The description of motion requires general relativity whenever the scales d of the situation are in the order of the Schwarzschild radius, i.e. whenever

$$d \approx r_S = 2Gm/c^2 \quad . \tag{541}$$

It is straightforward to confirm that with the usually quoted mass and size of all visible components of the universe, this condition is indeed fulfilled. One does need general relativity and thus curved space-time when talking about the whole of nature. Challenge

Similarly, quantum theory is required for the description of motion of an object whenever one approaches it to distances of the order of the Compton wavelength, i.e. whenever

$$d \approx \lambda_C = \frac{h}{mc} \quad . \tag{542}$$

∗ The frontier is the place of understanding.
∗∗ Written in summer to december 2000.
∗ ∗ ∗ ‘Here are lions.’ Written in ancient maps across unknown and dangerous regions.

Obviously, for the total mass of the universe this condition is not fulfilled. But we are not interested in the motion of the universe itself, we are interested in the motion of its components. Quantum theory is required whenever pair production and annihilation play a role. Especially in the early history of the universe and near the horizon, i.e. for the most distant events we can observe, this is indeed the case. In short, we are obliged to include quantum theory in any precise description of the universe.

Since at cosmological scales we need both quantum theory and general relativity, we start our investigation with the study of time, space, and mass, by asking at large scales the same questions we asked above at Planck scales.

Maximum time

Is it possible to measure time intervals of any imaginable size? General relativity shows that in nature there is a maximum time interval, with a value of about twelve thousand million yearsor 380 Ps, providing an upper limit to time measurements. It is called the "age" of the universe. It is deduced from the limits found in two measurements: the expansion of space-time, and the age of matter.

We all know clocks ticking for long times: the hydrogen atoms in our body. They were formed just after the big bang. One can almost say that their electrons orbit the nuclei since the dawn of time. In fact, inside their protons, the quarks move for even a few hundred thousand years longer. One thus gets the same maximum time limit for any clock made of atoms. Even clocks made of radiation (can you describe one?) yield a similar maximum time. In fact there are no ways to show that any imaginable clock has been ticking *before* this maximum time; none could provide a record of having done so. On the contrary, all known arguments maintain that clocks have not been ticking before. Challenge

In summary, it is *not* possible to measure time intervals larger than the maximum one, neither by using the history of space-time nor by using the history of matter or radiation.* It is thus rightly called the 'age' of the universe. Of course, all this is not a surprise. But looking at the issue in more detail is.

Does the universe have a certain age?

One should never trust a woman who tells one her real age.
A woman who would tell one that,
would tell one anything.
Oscar Wilde

This seems a silly question, since we just talked about it; in addition, the value is found in many books and tables, including that of appendix B, and its precise determination is actually one of the most important quests in modern astrophysics. But is this quest reasonable? See page 733

In order to measure the duration of a movement or the age of a system, one needs a clock. The clock one uses has to be *independent* of that movement and thus has to be *outside* the system. However, there are no clocks outside the universe. And inside it, a clock cannot be independent. In fact we just saw that inside it, no clock can run during *all* the history of the

* This conclusion implies that so-called 'oscillating' universe models, in which it is claimed that "before" the big bang there are other phenomena, have nothing to do with nature or observations.

universe. Indeed, time can be defined only once matter and space-time can be distinguished. And from this distinction onwards, only the two possibilities just discussed remain: one can either talk about the age of space-time, as is done in general relativity, by assuming that matter provides suitable clocks; or one can talk about the age of matter, such as stars, galaxies, etc., by assuming that either space-time extension or some other matter provides the clock. Both possibilities are being explored experimentally by modern astrophysics, and give the same mentioned result of about twelve thousand million years. But for the universe as a *whole*, an age cannot be defined.

The issue of the starting point of time makes this difficulty even more apparent. We might imagine that going back in time, there should be only two possibilities: either the instant $t=0$ is part of time or it is not. (Mathematically, this means that the segment describing time should be either closed or open.) Both cases assume that it is possible to measure arbitrary small times. But we know from the combination of general relativity and of quantum theory that this is *not* the case. In other words, both possibilities are incorrect: the beginning cannot *be* part of time, nor can it *not be* part of it. To this situation there is only one solution: there has not been any beginning at all.

In other words, the situation is consistently muddled. Neither does the age of the universe make sense, nor does its origin. What goes wrong? Or better, *how* do things go wrong? In other words, what happens if instead of jumping at the big bang directly, one *approaches* it as much as possible? The best way to clarify the issue is to ask about the measurement *error* one makes when saying that the universe is twelve thousand million years old. This turns out to be a fascinating topic.

How precisely can ages be measured?

> No woman should ever be quite accurate about her age.
> It looks so calculating.
> Oscar Wilde

The first way to measure the age of the universe* is to look at clocks in the usual sense of the term, namely clocks made of *matter*. As explained in the part on quantum theory, Ref. 3 Salecker and Wigner showed that a clock built to measure a total time T with a precision Δt has a minimum mass m given by

$$m > \frac{\hbar}{c^2}\frac{T}{(\Delta t)^2} \quad . \tag{543}$$

A simple way to include general relativity into this result was suggested by Ng and Van Ref. 4 Dam. Any clock of mass m has a minimum resolution Δt due to the curvature of space it introduces, given by

$$\Delta t > \frac{Gm}{c^3} \quad . \tag{544}$$

* Note that the age t_o is not the same as the Hubble time $T=1/H_o$. The Hubble time is only a computed quantity and (almost) always larger than the age; the relation between the two depends on the value of the cosmological constant, the density, and on other parameters of the universe. For example, for the standard hot big bang Ref. 2 scenario, i.e. for the matter dominated Einstein-de Sitter model, one has the simple relation $T=(3/2)\,t_o$.

Eliminating m, these two results imply that any clock with a precision Δt can only measure times T up to a certain maximum value, namely

$$T < \frac{(\Delta t)^3}{t_{Pl}^2} \quad , \tag{545}$$

where $t_{Pl} = \sqrt{\hbar G/c^5} = 5.4 \cdot 10^{-44}$ s is the already familiar Planck time. (As usual, we have *omitted* factors of order one in this and all the following results of this section.) In other words, the higher the accuracy of a clock, the shorter the time the clock works dependably! The precision of a clock is not (only) limited by the budget spent to build it, but by nature itself. Nevertheless, it does not take much to check that for clocks in daily life, this limit is not even remotely reached. For example, you might want to deduce how precisely your own age can be specified. Challenge

As a consequence of (545), a clock trying to achieve an accuracy of one Planck time can do so for at most one single Planck time. Simply put, *a real clock cannot achieve Planck time accuracy.*

If one tries to go beyond limit (545), fluctuations of space-time hinder the working of the clock and prevent higher precision. With every Planck time passing by, the clock accumulates at least one Planck time of measuring error. At the end, the total measurement error is at least as large as the measurement result itself.

We note in passing that result (545) is also valid for clocks made of radiation, such as background radiation. In short, measuring an age with a clock always involves some errors; whenever one tries to reduce these errors, the clock becomes so imprecise that age measurements become impossible. Challenge

Does time exist?

Time is waste of money.
Oscar Wilde

The limit (545) also tells something more drastic. From the origins of physics onwards, the concept of 'time' has alway been the name for the quantity measured by a clock. Therefore equation (545), expressing the non-existence of perfect clocks, also implies that time is only an approximate concept, and that perfect time does not exist. Thus there is no 'idea' of time, in the sense of Plato. In fact, all discussions of the previous and the present section can be seen as proofs that there are no perfect or 'ideal' examples of any classical concept. See page 38

Despite this conclusion, time is obviously a useful concept in everyday life. A simple explanation appears when one focuses on the importance of *energy*. Any clock, in fact any system of nature, is characterized by a simple number, namely the highest fraction of kinetic energy to rest energy of its components. In daily life, this fraction is about $1\,\text{eV}/10\,\text{GeV} = 10^{-10}$. Such *low energy* systems are well suited to build clocks. The better the motion of the main moving part – the pointer of the clock – can be kept constant and be monitored, the better the precision of the clock. To achieve the highest possible clock precision, the highest possible mass of the pointer is required, since both its position and speed must be measured an since the two measurement errors are related by $\Delta v \Delta x > \hbar/m$. Then one needs Challenge even more mass to screen the pointer from outside influences, thus explaining why more money usually buys better clocks.

But the relation is valid only at everyday energies. Increasing the mass is not possible without bounds, since general relativity changes the right hand side to $\Delta v \Delta x > \hbar/m + G(\Delta v)^2 m/c^3$. The additional term, negligible at everyday scales, is proportional to mass *and* energy fraction. Increasing either of the two by too large an amount limits the achievable precision of the clock. And thus at Planck energies the maximum measurable time interval is given by the Planck time.

Ref. 1

See page 587

Challenge

In summary, time exists as a good approximation only for *low energy* systems. Any increase in precision beyond a certain limit would require an increase of the energy of the components; but this energy increase will then prevent the increase in precision.

What is the measurement error for the age of the universe?

Applying the discussion about time measurements to the age of the universe is now straightforward. Expression (545) yields, for the the best accuracy possible, a value of about 10^{-23} s, or about the time light takes to move across a proton.

Challenge

Among others, the age of the universe yields also a maximal measurement precision. Expression (545) can be written as

$$\frac{\Delta t}{T} > \left(\frac{t_{\mathrm{Pl}}}{T}\right)^{2/3} \tag{546}$$

which shows that no time interval can be measured with more than about 40 decimals.

Another way to clarify the issue is one calculates the measurement error as function of the observation energy. One gets two limits. For *small* energies, the error is given by quantum theory as

$$\frac{\Delta t}{T} \sim \frac{1}{E_{\mathrm{meas}}} \tag{547}$$

and thus goes down with measurement energy. For high energies, the error is given by gravitational effects as

$$\frac{\Delta t}{T} \sim \frac{E_{\mathrm{meas}}}{E_{\mathrm{Pl}}} \tag{548}$$

so that the total result is given in figure 177. In short, high energies do not help to reduce measurement errors.

In particular, any attempt to reduce the measurement error for the age of the universe below 10^{-23} s would require such high energies that the limits of space-time are reached, and the measurement itself becomes impossible.

But maybe this conclusion was due to the fact that the argument used clocks made of particles, either of matter or of radiation. In the following we will find a confirmation as well as more details about this limit by looking at the methods to determine the age of the universe from space-time, through its expansion rate, at all possible extremes of observation methods.

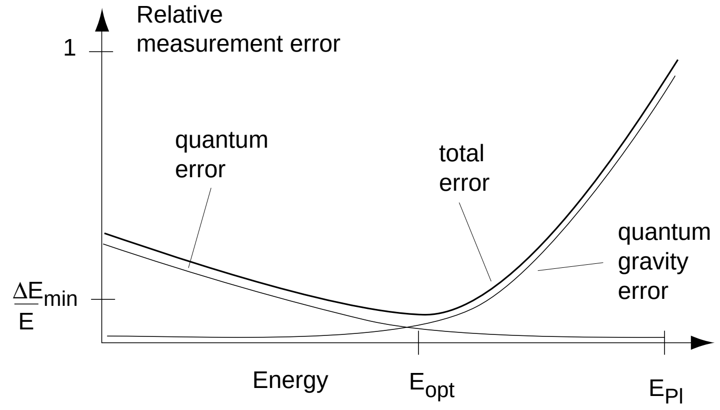


Figure 177 Measurement errors as a function of measurement energy

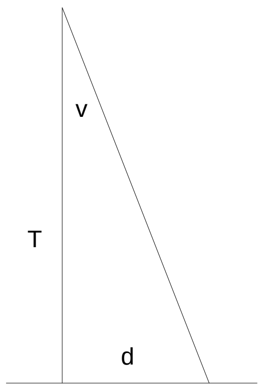


Figure 178 Trees and galaxies

Imagine you see a tree which, due to some wind storm, fell towards another, touching it at the very top. It is possible to determine the height of both trees by measuring the separation and the angles at the base. The height *error* will depend on the measurement errors of the separation and of the angles. Similarly, the age of the universe follows from the distance and the speed of objects, such as galaxies, observed in the night sky. The distance d corresponds to the ground separation of the trees and the speed v to the angle between the two trees. The Hubble time T of the universe – as already mentioned, it is usually assumed to be larger than the age of the universe – then corresponds to the height at which the two trees meet, since the age starts, in a naive sense, when the galaxies "separated". That time is given, within a factor of order one, by

$$T = \frac{d}{v} \quad . \tag{549}$$

This is in a few words the method used to determine the age of the universe from the expansion of space-time, for galaxies with redshifts below unity.* Of interest in the following is the (positive) measurement error ΔT, which becomes

$$\frac{\Delta T}{T} = \frac{\Delta d}{d} + \frac{\Delta v}{v} \quad ; \tag{550}$$

studying it in more detail is worthwhile. For any measurement of T one has to choose the object, i.e. a distance d, as well as an observation time Δt, or equivalently, an observation energy $\Delta E = 2\pi\hbar/\Delta t$. We will now investigate the consequences of these choices for expression (550), always taking into account both quantum theory *and* general relativity.

∗ At higher redshifts, the speed of light as well as the details of the expansion come into play; in the image of inclined trees, one finds that the trees are not straight all the way up to the top, and that they grow on a slope, as shown in figure 179.

At everyday energies, the result of the determination of the age t_0, about $12 \pm 2 \cdot 10^9$ years, is well known. The value is deduced by measuring red shifts, i.e. velocities, and distances, for stars and galaxies in distance ranges from some hundred thousand light years up to a red shift of about 1. Measuring redshifts does not produce large velocity errors. The main source of experimental error is the difficulty to determine galaxy distances.

What is the smallest possible distance error? Obviously, one gets

$$\frac{\Delta d}{T} > \frac{l_{\mathrm{Pl}}^{2/3}}{d^{2/3}} \tag{551}$$

which implies the same age uncertainty for the universe as found above in the case of material clocks. Challenge

One can try to reduce this error in two ways: either choosing objects at small or at large distances. Let us start with the smallest possible distances. In order to get high precision at small distances, one needs high observation energies. It does not take much to note that at observation energies near the Planck value, the value of $\Delta T/T$ approaches unity. In fact, both terms on the right hand side of expression (550) become of order one. At these energies, Δv approaches c and the maximum value for d approaches the Planck length, for the same reason that at Planck energies the maximum measurable time is the Planck time. In short, *at Planck scales it is impossible to say whether the universe is old or young*. Challenge

Let us continue with the other extreme, namely objects extremely far away, say with a redshift of $z >> 1$. Relativistic cosmology requires that the diagram of figure 178 be replaced by the more realistic diagram of figure 179. The "light onion" replaces the familiar light cone of special relativity: light converges near the big bang. Ref. 2

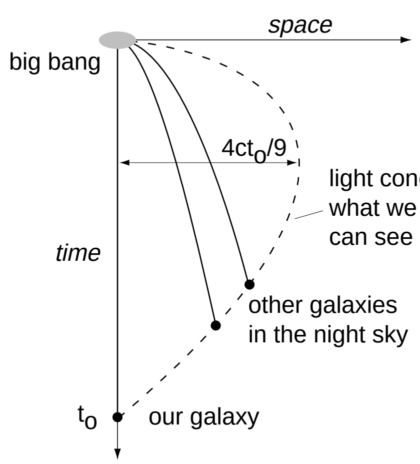


Figure 179 Speed and distance of remote galaxies

Also in this case the measurement error for the age of the universe depends on the distance and velocity errors. At the largest possible distances, the signals an object must send away must be of high energy, because the emitted wavelength must be smaller than the universe itself. One inevitably reaches Planck energies. But we saw that in such high energy situations, the emitted radiation, as well as the object itself, are indistinguishable from the space-time background. In other words, the redshifted signal we would observe today would have a wavelength as large as the size of the universe, with a correspondingly small frequency.

Another way to describe the situation is the following. At Planck energies or near the horizon, the original signal has an error of the same size as the signal itself. At present time, the redshifted signal still has an error of the same size as the signal. In short, also for large distances, the error on the horizon distance becomes as large as the value to be measured.

In short, also using space–time expansion and large scales, the instant of the so-called beginning of the universe cannot be determined with an error smaller than the age of the universe itself, as was found at Planck distances. Whenever one aims for perfect precision, one finds that the universe is 12 ± 12 thousand million years old! In other words, *at both extremal situations it is impossible to say whether the universe has a non-vanishing age*.

We conclude that the anthropomorphic concept of 'age' does not make any sense for the universe as a whole. The usual textbook answer is useful only for domains in time, space and energy for which matter and space-time are clearly distinguished, namely at everyday, *human scale* energy; however, this anthropocentric value has no overall meaning.

By the way, you might like to discuss the issue of the *fate* of the universe using the same arguments. Here however, we continue on the path outlined at the start of this section; the next topic is the measurement of length. Challenge

Maximum length

General relativity shows that in the standard cosmological model, for hyperbolical (open) and parabolic (marginal) universe evolutions, the actual *size* of the universe is infinite. It is only the *horizon distance*, i.e. the distance of objects with infinite redshift, which is finite. The horizon can also be defined as the set of the most distant events which can be observed.* For elliptical evolution, the total size is finite and depends on the curvature; but also in this case the present measurement limits yield a minimum size for the universe many times larger than the horizon distance. At least, this is what general relativity says. See page 256

On the other hand, quantum field theory is based on flat and infinite space-time. Let us us thee what happens when both theories are combined. What can one say about length measurements in this case? For example, would it be possible to construct and use a meter bar to measure lengths larger than the distance to the horizon? It is true that one would have no time to push it up to there, since in the standard Einstein-de Sitter big bang model the horizon moves away from us with more than the speed of light. One should have started installing the meter bar right at the big bang. Ref. 2

For fun, let us assume that we actually managed to do this. How far could we read read off distances? In fact, since the universe was smaller in the past, and since every observation of the sky is an observation of the past, the maximum *spatial distance* an object can be seen away from us is only $(4/9)ct_o$, as shown in figure 179. For space-time intervals, the maximum remains ct_o. Ref. 2

In any case, it turns out to be impossible to measure lengths larger than the horizon distance, even though general relativity predicts such distances. This unsurprising result is in obvious agreement with the existence of a limit for time interval measurements. The surprises come now.

Is the universe really a big place?

Astronomers and Hollywood movies answer by the affirmative. Indeed, the horizon distance of the universe is usually included in tables. Cosmological models specify that the scale Ref. 5 See page 733

* In cosmology, one needs to distinguish between the scale factor R, the Hubble radius $c/H = cR/\dot{R}$, the horizon distance h, and the size d of the universe. The Hubble radius is a computed quantity giving the distance at which objects move away with the speed of light. It is always *smaller* than the horizon distance, at which e.g. in the standard Einstein-de Sitter model objects move away with *two* times the speed of light. However, the horizon itself moves away with *three* times the speed of light. Ref. 2

factor R, which fixes the horizon distance, grows with time; for the case of the usually assumed mass dominated Einstein-de Sitter model, i.e. for vanishing cosmological constant and flat space, one has

$$R(t) = C\,t^{2/3} \tag{552}$$

where the constant C relates the commonly accepted horizon distance to the commonly accepted age. Indeed, observation shows that the universe is large and is getting even larger. But let us investigate what happens if to this result from general relativity we add the limitations of quantum theory. Is it really possible to measure the distance to the horizon?

We first look at the situation at large energies. We saw above that space-time and matter are not distinguishable at Planck scales. Therefore, at Planck energy one cannot state whether objects are *localized* or not. At Planck scales, a basic distinction of our thinking, the one between matter and vacuum, becomes obsolete. Equivalently, it is not possible to claim that space-time is *extended* at Planck scales. Our concept of extension derives from the possibility to measure distances and time intervals, and from observations such as the ability to align several objects, e.g. in one room, behind each other. Such observations are not possible at Planck scales. In fact, all observations from which we deduce in daily life that space is extended, are not possible at Planck scales. *At Planck scales, the basic distinction between vacuum and matter, extension and localization, disappears.* As a consequence, at Planck energies the size of the universe cannot be measured. It cannot even be called larger than a match box.

At cosmological distances, the situation is even easier. All arguments given above on the measurement errors for the age can be repeated for the distance of the horizon. Essentially, at largest distances and at Planck energies, the measurement errors are of the same magnitude as the measured value. All this happened because length measurements become impossible near the limits. This is corroborated by the fact that there is nothing to compare the size of the universe with.

Challenge

Also studying the big bang produces strange results. At Planck energies, whenever one tries to determine the size of the big bang, one cannot claim that it was smaller than the present universe. Somehow, Planck dimensions and the size of the universe get confused.

There are also other confirmations. Let us come back to the example above. If one had a meter bar spanning all the universe, even beyond the horizon, with a zero at the place we live, what measurement *error* would it produce for the horizon? It does not take long to discover that the expansion of space-time from Planck scales to the present also expands an uncertainty of the Planck size into one of the horizon size. The error is as large as the measurement result.

Challenge

Since this also applies when one tries to measure the *diameter* of the universe instead of its radius, it becomes impossible to state whether the antipodes in the sky really are distant form each other!

We summarize the situation by noting that anything said about the size of the universe is as limited as anything said about its age. *The height of the sky depends on the observation energy.* At Planck energies, it cannot distinguished from the Planck length. And if one starts at standard observation energies, the measurement errors increase beyond all bounds if one tries to determine the distance of the horizon to high precision. Therefore, at Planck energies, the volume of the universe is indistinguishable from the Planck volume.

The boundary of space-time – is the sky a surface?

The horizon of the universe, essentially the black part of the night sky, is a fascinating entity. Everybody interested in cosmology wants to know what happens there. In newspapers the horizon is sometimes called the *boundary of space-time*. Some surprising insights, not yet common in newspapers, appear when general relativity and quantum mechanics are combined.

We saw above that the measurement errors for the distance of the horizon are substantial. They imply that we cannot pretend that all points of the sky are equally far away from us. Thus we cannot say that the sky is a surface. That implies that there is even no way to determine the dimensionality of the horizon, nor the dimensionality of spacetime near it.

In addition, the measurement errors imply that no statement can be made about translation symmetry at cosmological scales. Are you able to confirm this? We also conclude that at the horizon it is impossible to distinguish spacelike and timelike distances. Even worse, concepts such as 'mass' or 'momentum' are muddled up at the horizon. This means that like at Planck energy, we are unable to distinguish between objects and background, and between state and intrinsic properties. We will come back to this important point shortly. Challenge

Measurements thus also do not allow to determine whether the boundary is a point, a surface, or a line. It could be an arbitrary complex shape, even knotted. In fact, quantum theory tells us that it must be all of this from time to time, in short, that *the sky fluctuates in height and shape*.

In short, it is impossible to determine the topology of the sky. But that is nothing new. As is well known, general relativity is unable to describe pair creation of spin 1/2 particles. The reason is the change of space-time topology required by the process. On the other hand, the universe is full of such processes, implying that it is impossible to define a topology for the universe and in particular, to talk of the topology of the horizon itself. Are you able to find at least two other arguments to show this? Challenge

Worse, quantum theory shows that space-time is not continuous at a horizon, as is easily deduced by applying the Planck scale arguments from the previous section. Time and space are not defined there. Ref. 1

Finally, there is no way to decide whether the various boundary points are *different* from each other. The distance between two points on the boundary is undefined. In other words, it is unclear what the *diameter* of the horizon is.

In summary, the horizon has no specific distance nor shape. The horizon and thus the universe cannot be shown to be manifolds. This leads to the next question:

Does the universe have initial conditions?

One often reads of the quest for the initial conditions of the universe. But before joining the search, one should ask *whether* and *when* such initial conditions make any sense. Obviously, our everyday description of motion requires them. Initial conditions describe the *state* of a system, i.e. all those aspects which differentiate it from a system with the same intrinsic properties. Initial conditions, like the state of a system, thus are attributed to a system by an *outside* observer. Ref. 6

More specifically, quantum theory told us that initial conditions or the state of a system can only be defined by an outside observer with respect to an environment. It is already a difficult feat to be outside the universe. In addition, indipendently of this issue, even inside the universe a state can only be defined if matter can be distinguished from the vacuum. However, this is impossible at Planck energies, near the big bang, or at the horizon. Thus there is no state for the universe. No state also means *no wavefunction of the universe*.

The limits imposed by the Planck values also confirm this conclusion in other ways. First of all, they show that the big bang was not a singularity with infinite density or temperature, Ref. 1 as infinite large values do not exist in nature. Secondly, since instants of time do not exist, it is impossible to define the state of any system at a given time. Thirdly, as the non-existence of instants of time means that events do not exist, also the big bang was not an event, so that neither an initial state nor an initial wavefunction can be ascribed to the universe also for this more prosaic reason. (Note that this also means that the universe cannot have been created.)

In short, *there are no initial conditions of the universe*. Initial conditions make sense only for subsystems and only far away from Planck scales. That means that two conditions must be fulfilled: the system must be away from the horizon, and it must evolve some time "after" the big bang. Only when these two conditions are fulfilled it becomes possible to state that objects *move* in space. Of course, this is always the case in everyday life.

At this point of our escalation, given that time and length are unclearly defined at cosmological scales, it should come as no big surprise that the concept of mass has similar difficulties.

Does the universe contain particles and stars?

Of course, one says. The number of stars, about $10^{23\pm1}$, is included in every book on cos- See page 733 mology, as it is in the table of appendix B. A subset of this number can be counted on clear nights. If one asks the same question about particles instead of stars, the situation is similar. The commonly quoted baryon number is $10^{81\pm1}$, together with a photon number of $10^{90\pm1}$.

But this does not settle the issue. Neither quantum theory nor general relativity alone make predictions about the number of particles, neither inside nor outside the horizon. What happens if one combines them?

In order to define the number of particles in a region, quantum theory first of all requires See page 459 a vacuum state. Only then the particle number can be defined by comparing the system with the vacuum. Neglecting or leaving out general relativity by assuming flat space-time, this procedure poses no problem. But adding general relativity and thus a curved space-time, especially one with such a strangely behaved horizon as we just found, the answer is simple: there is *no* vacuum state to compare the universe to, for two reasons. First of all nobody can explain what an empty universe would look like; second, and most importantly, there is now way to define a state of the universe at all. The number of particles in the universe thus becomes undefinable. Only at everyday energies and for finite dimensions we are able to speak of an approximate particle number.

The definition of a vacuum state is necessary for a simple reason: to count the number of particles in a volume, one must be able to compare the situation to an empty volume.* But for the universe this feat is impossible also for another reason. It is not possible to remove particles from the universe. Therefore the impossibility to define a vacuum state and thus a particle number is not surprising. It is an interesting exercise to investigate the measurement errors appearing when one tries to determine a particle number despite this fundamental impossibility. Challenge

What about stars? In principle, the same conclusion as for particles applies. However, at everyday energies stars can be counted also *classically*, i.e. without taking them out of the volume they are enclosed in. For example, this is possible by differentiating by their mass, their colour, or any other characteristic, individual property. Only near Planck energy or near the horizon these methods are not applicable. In short, the number of stars is only defined as long as the observation energy is low, i.e. as long as one stays away from Planck energies and from the horizon.

In summary, despite the appearances at human scales, *there is no definite number of particles in the universe*. The universe cannot be distinguished from vacuum by counting particles. Even though particles are necessary for our own existence and functioning, they cannot be counted at all.

This conclusion is so strange that one cannot accept it so easily. Let us try the other method: instead of counting, let us weigh.

Does the universe contain masses and objects?

The average density of the universe, of about $10^{-26}\,\mathrm{kg/m^3}$, is frequently cited in texts. See page 733 Is it different from vacuum? Quantum theory shows that due to the uncertainty relation, even an empty volume of size R has a mass. For a zero energy photon inside it, one has $E/c = \Delta p > \hbar/\Delta x$, so that in a volume of size R, one has a minimum mass of at least $m_{min}(R) = h/cR$. For a spherical volume of radius R there is thus a minimal mass density given roughly by

$$\rho_{\mathrm{min}} \approx m_{min}(R)/R^3 = \frac{\hbar}{cR^4} \quad . \tag{553}$$

For the universe, inserting the standard horizon distance R_0 of twelve thousand million light years, the value becomes about $10^{-142}\,\mathrm{kg/m^3}$. It describes the density of the vacuum. In other words, the universe, with its density of about $10^{-26}\,\mathrm{kg/m^3}$, seems to be clearly different from vacuum. But are we sure?

We just deduced that the radius of the horizon is undefined: depending on the observation energy, it can be as small as the Planck length. That implies that the density of the universe lies somewhere between the the lowest possible value, given by the just mentioned vacuum density, and highest possible one, namely the Planck density.** In short, relation (553) does not really provide a clear statement.

* This requirement effectively translates into the requirement that the particle counter be outside the system. Challenge Can you confirm the connection?

Challenge ** In fact, at everyday energies it lies almost exactly between the two values, yielding the strange relation

$$m_0^2/R_0^2 \approx m_{\mathrm{Pl}}^2/R_{\mathrm{Pl}}^2 = c^4/G^2 \quad . \tag{554}$$

Another way to measure the mass of the universe would be the use of the original definition of mass, as given by Mach and modified by special relativity, and apply it. Thus, let us try to collide a standard kilogram with the universe. It is not hard to see that whatever we do, using either low or high energies for the standard kilogram, the mass of the universe cannot be constrained by this method. One would need to produce or to measure a velocity change Δv after the collision for the rest of the universe. To hit all mass in the universe at the same time, one needs high energy; but then one is hindered by Planck energy effects. In addition, a really well-done collision measurement would in addition require a mass outside the universe, a rather difficult feat.

See page 164

Challenge

Still another way to measure the mass would be to try to determine the gravitational mass of the universe through straightforward weighing. But the lack of balances outside the universe makes this an unpractical solution, to say the least.

A way out might be the most precise definition of mass provided by general relativity, namely the so-called *ADM mass*. However, for its definition a specified behaviour at infinity is required, i.e. a background, which the universe lacks.

One is then left with the other general relativistic method: determining the mass of the universe by measuring its average curvature. Let us take the defining expressions for average curvature κ for a region of size R, namely

See page 233

$$\kappa = \frac{1}{r^2_{\text{curvature}}} = \frac{3}{4\pi}\,\frac{4\pi R^2 - S}{R^4} = \frac{15}{4\pi}\,\frac{4\pi R^3/3 - V}{R^5} \quad . \tag{555}$$

We have to insert the horizon radius R_0 plus either its surface S_0 or its volume V_0. However, given the the error margins on the radius and the volume, especially at Planck energies, we again find no reliable result for the radius of curvature.

Challenge

An equivalent method starts with the usual expression for the scalar curvature uncertainty $\Delta\kappa$ for a region of size R provided by Rosenfeld

Ref. 7

$$\Delta\kappa > \frac{16\pi l_{\text{Pl}}^2}{R^4} \tag{556}$$

showing that the curvature radius error behaves like the horizon distance error.

In summary, *at Planck energy, the average radius of curvature of nature turns out to lie between infinity and the Planck length.* This implies that the matter density lies between the minimum value and the Planck value. There is thus no method to determine the mass of the universe at Planck energy. (Can you find one?) The concept of mass cannot be applied to the universe as a whole. *The universe has no mass.*

Challenge

Symmetries

We have already seen that at the horizon, space-time translation symmetry breaks down. Let us have a quick look at the other symmetries.

What happens to permutation symmetry? Exchange is an operation on objects in space-time. Exchange thus automatically requires a distinction between matter, space, and time.

See page 521

But it is nothing new. The approximate equality can be deduced from the equation 16.4.3 (p. 620) of Steven Weinberg, *Gravitation and Cosmology*, Wiley, 1972, namely $Gn_b m_p = 1/t_0^2$. The relation is required by several cosmological models.

If we cannot distinguish positions, we cannot talk about exchange of particles. But this is exactly what happens at the horizon. In short, general relativity and quantum theory together make it impossible to define permutation symmetry at the horizon.

CPT symmetry suffers the same fate. Due to measurement errors or to limiting maximum or minimum values, it is impossible to distinguish the original and transformed situation. It is therefore impossible to maintain that CPT is a symmetry of nature at horizon scales. In other words, matter and antimatter cannot be distinguished at the horizon.

The same happens with gauge symmetry, as you might want to check in detail yourself. Challenge The concept of gauge field requires a distinction between time, space, and mass for its definition, an impossibility at the horizon. We deduce that at the horizon also concepts such as algebras of observables cannot be used to describe nature.

In fact, the complete vocabulary we use to talk about observations, e.g. magnetic field, electric field, potential, spin, charge, speed, position, etc., cannot be used at the horizon. In summary, *all symmetries of nature break down at the horizon.* And that is not all.

Is there a boundary of the universe?

To answer, one needs to specify more precisely what is meant by this expression. As above, we take 'boundary' and 'horizon' to be synonyms, as they are the same for all practical purposes. The knowledge of mathematics does not help us to clarify the issue; the properties of mathematical boundaries, e.g. that they themselves have no boundary, are not applicable in the case of nature, since space-time is not continuous at Planck scales or at large scales. We need other, physical arguments.

The boundary of the universe is obviously supposed to mean the boundary between *something* and *nothing*. This gives three possibilities:

• 'Nothing' could mean 'no matter'. But we just saw that this distinction cannot be made at Planck scales. As a consequence, the boundary would either not exist at all or encompass both the horizon *as well as* the whole universe.

• 'Nothing' could mean 'no space-time'. We then have to look for those domains where space and time cease to exist. That happens at Planck scales and at the horizon. Again, the boundary would either not exist or encompass the whole universe.

• 'Nothing' could mean 'neither space-time nor matter.' The only possibility is a boundary to domains *beyond* the Planck scale and *beyond* the horizon. But such a boundary would also encompass all of nature.

This result is puzzling. When combining quantum theory and relativity, we do not seem to be able to find a conceptual definition of the horizon which distinguishes it from its interior. In fact, if you find one, publish it! In general relativity a distinction *is* possible, obviously. Challenge In quantum theory as well. But as soon as we combine the two, the boundary becomes indistinguishable from its content; *the interior of the universe cannot be distinguished from its horizon*. There is thus no boundary.

That is definitely interesting; it suggests that nature might be *symmetric* under transformations which exchange interiors and boundaries. Such a connection, nowadays called *holography* because it recalls the working of credit card holograms, is an busy research field in present high energy physics. But for the time being we continue with our original theme, Ref. 8 which directly leads us to ask:

Is the universe a set?

We are used to call the universe the sum of all matter and all space-time. In other words, we implied that the universe is a list of components, all different from each other. This idea was introduced in three situations: it was assumed that matter consists of particles, that space-time consists of events (or points), and that the set of states, such as the Hilbert space, consists of different initial conditions. But our discussion so far shows that the universe is *not* a list of such distinguishable elements. We encountered several proofs: at the horizon, at the big bang, and at Planck scales distinction between events, between particles, between observables, and between space-time and matter becomes impossible. In those domains, distinctions of any kind become impossible. We found that any distinction among two entities, even between a fridge and a mountain, is possible only *approximately*, due to the fact that we live at energies much smaller than the Planck energy. Obviously, we are able to distinguish cars from people and from toothpicks; the approximation is so good that we do not notice the error when performing it. But the discussion of the situation at Planck energies shows that a perfect distinction is impossible in principle. In short, *it is impossible to split the universe into separate entities.*

Another way to this result is the following. Distinction of two entities requires different measurement results, such as different positions, masses, sizes, etc. Whatever quantity we choose, at Planck energies the distinction becomes impossible. Thus it is only approximately possible at everyday energies.

In short, since the universe is not a list of entities, *the universe is not a set*. We envisaged
See page 413
this possibility already in the first intermezzo; now it is confirmed. The concept of 'set' is already too specialized to describe the universe. *The universe must be described by a mathematical concept which does not contain any set.*

This is a powerful result: it means that the universe cannot be described precisely if any of the concepts used for its description presuppose sets. But all concepts we used so far to describe nature, such as space-time, phase space, Hilbert-spaces and their generalizations, Fock spaces, particle spaces, are based on sets. They all must be abandoned at Planck energy, and thus also in any precise description. In fact, *any* concept necessarily is a set!

But also many speculations in the literature do not satisfy the criterion. In particular, all studies about quantum fluctuations, mathematical categories, posets, complex mathematical spaces, computer programs, Turing machines, Gödel's theorem, creation of any sort, space-time lattices, quantum lattices, even Bohm's unbroken wholeness, etc. do not satisfy the requirement. Also almost all speculations about the origin of the universe are put to rest with this result. For example, you might want to check the religious explanations you know
Challenge
against this result. In fact, no approach of theoretical physics so far – in the year 2000 – satisfies the requirement to abandon sets; maybe a future version of string or M theory might do so.

Note that this radical conclusion is deduced from only two statements: the necessity of using quantum theory whenever the dimensions are of the order of the Compton wavelength, and the necessity to use general relativity whenever the dimensions are of the order of the Schwarzschild radius. Together, they mean that any precise description of nature cannot contain sets. We reached this result after a long and interesting, but in a sense unnecessary digression. In any case, this radical result may explain why the unification of the two theories was not successful so far. Not only does unification require that we stop using space, time and mass for the description of nature, as we saw in the discussion of Planck scales; it also requires that all distinctions, of any kind, be only approximate. But all physicists have been educated with exactly the opposite credo!

Note that if it is not a set, *the universe is* not *a physical system.* In particular, it has no state, no intrinsic properties, no wavefunction, no initial conditions, no density, no entropy, no cosmological constant, etc. Neither is it thermodynamically closed or open, nor does it contain any *information*. All thermodynamical quantities, such as entropy, temperature, free energy etc., are defined using *ensembles*. Ensembles are limits of systems, either thermodynamically open or closed ones. The universe being neither of the two, no thermodynamic quantity can be defined for it.* All physical properties are only defined for parts of nature which are approximated or idealized as sets, and thus as physical systems. The universe is neither.

Hilbert's sixth problem settled

In the year 1900, David Hilbert gave a famous lecture in which he listed 23 of the great challenges facing mathematics in the twintieth century. Most problems provided challenges to many mathematicians for decades afterwards. A few are still unsolved, among them the sixth problem. This problem was the challenge to find an *axiomatic* treatment of physics.

See page 393
Ref. 9

Interestingly, since the universe is not even a set, one can deduce that such an axiomatic description of nature is *impossible*. The reasoning is simple; all mathematical systems, be they algebraic systems, order systems, or topological systems, are based on sets. Mathematics does not have axiomatic systems which do not contain sets. The reason is that any mathematical concept so far is based on sets or at least contains sets.

This conclusion is also confirmed by the fact that physics started with a *circular* definition which was not eliminated after 2500 years of investigations: the definition of space-time with help of objects and the definition of objects with help of space and time. Physics *could never* be modeled after mathematics. Physicists have to live with logical problems, because any axiomatic description can only be about an *aspect* of nature, and will thus be sometimes in contradiction with the description of other parts.

See page 98

The situation is similar to the description of the sky by a child as "made of air and clouds". Looking closely, one discovers that clouds are made of water droplets. But there is air inside clouds, and there is also water vapour in air everywhere. Among others, when clouds and air are watched through the microscope, there is no clear boundary between the two. One cannot define the 'cloud' and 'air' without the other. There is no axiomatic definition of 'cloud'.

Also objects and vacuum behave in this way. Virtual particles are found in the vacuum, and vacuum is found inside objects. And at Planck scales, there is no clear boundary between the two; one cannot define them without the other.

In both cases, despite this lack of definition and despite the logical problems following, the description works well at large scales. But then the next question naturally becomes:

* There are people who knew this long before physicists; for example, the belief that the universe is or contains information was ridiculed most thoroughly in the popular science fiction parody Douglas ADAMS, *The hitchhiker's guide to the galaxy*, 1979, and its sequels.

Does the universe make sense?

Drum hab ich mich der Magie ergeben,
[...]
Daß ich erkenne, was die Welt
Im Innersten zusammenhält.*
Goethe, *Faust.*

Is the universe really the sum of matter-energy and space-time? We heard this so often up to now that we might be lulled into forgetting to check the statement. To find out, we do not need magic, as Faust thought; we only need to list what we found in this section, with some help from the section on Planck scales, and from the intermezzo on brain and its language: See page 368

Table 43 Physical properties of the universe

• The universe has no age.	• The universe has no beginning.
• The universe has no size.	• The universe has no volume.
• The universe has no shape.	• The universe has no particle number.
• The universe has no mass.	• The universe has no energy.
• The universe has no density.	• The universe contains no matter.
• The universe has no cosmological constant.	• The universe has no initial conditions.
• The universe has no state.	• The universe has no wave function.
• The universe is not a physical system.	• The universe contains no information.
• The universe is not isolated.	• The universe is not open.
• The universe has no boundaries.	• The universe does not interact.
• The universe cannot be measured.	• The universe cannot be said to exist.
• The universe cannot be distinguished from nothing.	• The universe cannot be distinguished from a single event.
• The universe contains moments.	• The universe is not composite.
• The universe is not a set.	• The universe is not a concept.
• The universe cannot be described.	• There is no plural for ‘universe’.
• The universe cannot be distinguished from vacuum.	• The universe is not created.

Not only are we unable to state that the universe is made of space-time and matter; in fact, we are unable to say anything positive about the universe at all!** It is not even possible to say that it exists, since it is impossible to interact with it. The term “universe” does not allow to make a single sensible statement. (Can you find one?) Challenge We are only able to say which properties it does *not* have. We were unable to find any property the universe *does* have. We cannot even say whether the universe is something or nothing. *The universe isn’t anything in particular.* In other words, the term “universe” is not useful at all for the description of motion.

We get a confirmation for this strange conclusion from the first intermezzo, back in the first part of our escalation. See page 381 There we found that any concept needs a defined content, defined limits, and a defined domain of application. In this section, we just concluded that for the term “universe”, neither of these aspects is defined; there is thus *no* such concept. If somebody asks: ‘why does the universe exist?’ the answer is: not only does the use of

* Thus I have devoted myself to magic, [...] that I understand how the innermost of the world is held together.
** There is also a well-known non-physical concept of which nothing positive can be said. Can you spot it? Challenge

'why' wrongly suggest that something might exist outside the universe, contradicting the definition of the term 'universe' itself; most importantly of all, the universe simply does not exist. The term "universe" only *seems* to express something, but it doesn't. It makes no sense. We will therefore avoid using it from now on.*

This conclusion may be interesting, even strangely beautiful; but does it help us to understand motion more precisely? Interestingly so, it does. Not only does it tell us that a precise description of motion needs to avoid the term "universe", making us wary when we hear expressions like "the number of particles in the universe" or similar phrases; more importantly, these results open up a fresh look at the list of properties so far unexplained.

Extremal scales and open questions of physics

In the chapter *Quantum physics in a nutshell* we had listed all the unexplained properties of nature which remain despite general relativity and quantum theory. The present conclusions point to a new connection among them. See page 563

Indeed, many of the cosmological results of this section sound surprisingly familiar, in particular when we compare them systematically with those of the previous section on Planck scale properties. Both sections covered topics – some in more details than others – from that list of unexplained properties of nature.

Table 44 Properties of nature at maximal, at everyday, and at minimal scales

Physical property of nature	at horizon scale	at everyday scale	at Planck scale
requires quantum theory and relativity	true	wrong	true
intervals can be measured precisely	wrong	true	wrong
length and time intervals are	limited	unlimited	limited
space-time is not continuous	true	wrong	true
points and events cannot be distinguished	true	wrong	true
space-time is not a manifold	true	wrong	true
space is 3 dimensional	wrong	true	wrong
space and time are indistinguishable	true	wrong	true
initial conditions make sense	wrong	true wrong	
space-time fluctuates	true	wrong	true
Lorentz and Poincaré symmetry	disappear	correct	disappear
CPT symmetry	disappears	correct	disappears
permutation symmetry	disappears	correct	disappears
interactions	disappear	exist	disappear
number of particles	undefined	defined	undefined
algebras of observables	disappear	apply	disappear
matter indistinguishable from vacuum	true	wrong	true
boundaries exist	wrong	true	wrong
nature is a set	wrong	true	wrong

* Of course, the term 'universe' makes still sense if it is defined more restrictively, such as 'everything which interacts with a particular human or animal observer.' But such a definition is not useful for our escalation, as it does not allow for a shared description of motion.

First of all, we see that *each* of the unexplained properties makes no sense at *both* limits of nature, the small *and* the large one. All open questions remain open at both extremes.

Secondly and more importantly, at horizon scales nature behaves in exactly the same way as at Planck scales. In fact, we have not found any difference between the two cases. Are you able to discover one?* We are thus lead to the hypothesis that nature does not distinguish between large and small scales; there seems to be a property of *extremal identity*. Challenge

Is extremal identity a principle of nature?

The principle of extremal identity incorporates some rather general points:

- all open questions about nature so far appear at its two limits;
- a description of nature requires both general relativity and quantum theory;
- nature is not a set;
- initial conditions make no sense at nature's limits;
- there is a relation between local and global issues in nature;
- the concept of "universe" has no content.

Extremal identity thus looks like a good candidate tool in the search for a unified description of nature. To be a bit more provocative, it might be the *only* known principle incorporating the idea that the universe is not a set, and thus might be the best candidate so far to help in the quest of unification. Extremal identity is beautiful in its simplicity, in its unexpectedness, and in its richness of consequences. Just explore it a little for yourself. Its consequences are presently studied with great effort in high energy particle physics, even though often under different names. But is extremal identity also helpful for our quest? Challenge Ref. 10

A simple model – in fact too simple to be correct – might be the following. It looks as if extremal identity implies a connection such as

$$r \leftrightarrow \frac{l_{\text{Pl}}^2}{r} \quad \text{or} \quad x_\mu \leftrightarrow \frac{l_{\text{Pl}}^2\, x_\mu}{x_\mu x^\mu} \quad . \tag{557}$$

Could this mapping, called *inversion*, be a symmetry of nature? At every point of space? For example, inserting the horizon distance, equation (557) would imply that lengths smaller than $l_{\text{Pl}}/10^{61} \approx 10^{-96}$ m never appear in physics. Is this the case? What would inversion imply for the big bang? Or is there an underlying mapping of similar properties? Numerous fascinating questions are contained in the simple hypothesis of extremal identity. In fact, *two* main directions of investigation appear. Challenge

The first topic is the search for some stronger arguments for the validity of extremal identity. This quest fills the next leg of our escalation. In fact, we will discover quite a number of simple arguments, all showing that extremal identity is indeed a property of nature. Investigating it more closely produces many beautiful insights.

The second topic is obvious. What is the correct version of equation (557)? That oversimplified expression is in fact neither valuable nor correct. It is not valuable because even if it were true, it would not explain *any* of the issues left open by general relativity and quantum theory. It only *relates* some of them, thus reducing their number, but doesn't *solve* any of them. You might want to check this for yourself. Challenge

* If so, send a message to cs@motionmountain.org; but first of all, publish it!

But worse, inversion is also simply *wrong*. Inversion is not the correct description of extremal identity because it does not realize one of the connections discovered above: it does not connect *states* and *intrinsic properties*, but keeps them distinct. This is a somewhat abstract way to say that inversion does not take into account *interactions*. And most open problems at this point of our escalation are properties of interactions.

References

1 Available at http://www.dse.nl/motionmountain/contents.html. Those pages are a reworked version of the ones published in french as C. Schiller, *Le vide diffère-t-il de la matière?*, in E. Gunzig & S. Diner, editeurs, *Le vide – Univers du tout et du rien – Des physiciens et des philosophes s'interrogent*, Les Éditions de l'Université de Bruxelles, 1998. An older english version is also available as Christoph Schiller, *Does matter differ from vacuum?*, http://xxx.lanl.gov/abs/gr-qc/9610066. Cited on page 608, 613, 618, 619.

2 See the lucid discussion by G.F.R. Ellis & T. Rothman, *Lost horizons*, American Journal of Physics **61**, pp. 883–893, 1993. Cited on page 611, 615, 616, 616, 616.

3 H. Salecker & E.P. Wigner, *Quantum limitations of the measurement of space-time distances*, Physical Review **109**, pp. 571–577, 1958. See also the popular account of this paper by E.J. Zimmerman, *The macroscopic nature of space-time*, American Journal of Physics **30**, pp. 97–105, 1962. Cited on page 611.

4 Y.J. Ng & H. van Dam, *Limit to space-time measurement*, Modern Physics Letters A **9**, pp. 335–340, 1994, Y.J. Ng & H. van Dam, Modern Physics Letters A *Remarks on gravitational sources*, **10**, pp. 2801–2808, 1995. Cited on page 611.

5 See e.g. the movie *Contact*, based on the book by Carl Sagan, *Contact*, 1996. Cited on page 616.

6 For example the international bestseller by S. Hawking, *A brief history of time – From the big bang to black holes*, 1988. Cited on page 618.

7 L. Rosenfeld, in H.-J. Treder, *Entstehung, Entwicklung und Perspektiven der Einsteinschen Gravitationstheorie*, Springer Verlag, 1966. Cited on page 621.

8 Holography is connected with the work by 't Hooft and Susskind. See for example G. 't Hooft, *Dimensional reduction in quantum gravity*, pp. 284–296, in A. Ali, J. Ellis & S. Randjbar-Daemi, *Salaamfeest*, 1993, or also the paper by L. Susskind, Journal of Mathematical Physics **36**, p. 6377, 1995. Cited on page 622.

9 See for example *Die Hilbertschen Probleme*, Akademische Verlagsgesellschsaft Geest & Portig, 1983. Cited on page 624.

10 Large part of the study of dualities in string and M theory can be seen as investigations into the detailed consequences of extremal identity. For a review, see ... Cited on page 627.

A classical version of duality is discussed by M.C.B. Abddalla, A.L. Gadelka & I.V. Vancea, *Duality between coordinates and the Dirac field*, hep-th/0002217.

33. The physics of sex – a summary of the first two and a half parts

Sex is the physics urge sublimated.
Graffito

Maybe you have once met a physicist who has told you, in one of those moments of confidentiality, that studying physics is more beautiful than making love. At this statement, many will simply shake their heads in disbelief, and strongly disapprove. In this section we will argue that it is possible to learn so much about physics while making love that the discussion on their relative beauty can be put aside altogether.

Imagine you are with your partner on a beautiful tropical island, just after sunset, and that you look together at the evening sky. Imagine as well that you know little of what is taught at school nowadays, e.g. that your knowledge is that of the late renaissance, which probably is a good description of the average modern education level anyway. Typeset in April 2001

Most important results of physics can be deduced from the following experimental facts: *

Sex is communication.
Sex is an interaction between moving bodies.
Sex is attractive.
Sex is makes noise.
Sex is for reproduction.
Sex needs memory.
Sex uses the sense of sight.
Sex is motion.
Sex is based on touch.
Sex is fun.
Sex makes one dream.
Sex is tiring.
Sex takes time.
Sex is repulsive.
In sex, size matters.
Sex can hurt.
Sex is greek.
Sex is bestial.
Sex is holy.
Sex uses motion again.
Sex is private.

Let us start.

• Sex is *communication*. Communication is possible first of all because nature looks similar from different standpoints, and secondly because in nature there are no surprises. Without similarity we could not understand each other, and a world of surprises would even make thinking impossible; it would not be possible to form concepts to describe observations. But fortunately, the world is regular; it thus allows to use concepts such as time and space for its description.

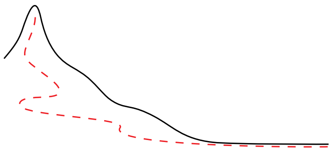
• Sex is an *interaction between moving bodies*. Together with the previous result, this implies that we can and need to describe moving bodies with *mass*, *energy* and *momentum*. That is not a small feat. For example, it allows to deduce that the sun will rise tomorrow if the sea level around the island is the usual one.

• Sex is *attractive*. When feeling attracted to your partner, you may wonder if this attraction is the same which keeps the moon going around the earth. You make a quick calculation,

Ref. 1 * To study the influences of *sex* on physics is mostly a waste of time. Maybe one day we will understand why there do not seem to be any female crackpots proposing pet physical theories. Much more fun is the influence of *sexuality* onto physics. In the following, we bow to the modern habit of saying 'sex' when meaning 'sexuality' instead.

This is a section of the freely downloadable e-textbook

Motion Mountain



Hiking beyond space and time along the concepts of modern physics

available at
www.motionmountain.org

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?

Challenge - Did you find any mistakes?

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

and find that from the expression for universal gravity

$$E_{\text{pot}} = \varphi m = -\frac{GMm}{r} \qquad (608)$$

the involved energy is about as much as the energy added by the leg of a fly on the skin. In short, your partner teaches you that in nature there are other attractive interactions apart from gravity; the average modern education is incomplete.

Nevertheless, this first equation is important: it allows to predict the position of the planets, the time of the tides, the time of eclipses, the return of comets, etc., to a high accuracy for thousands of years in advance.

• Sex is *makes noise*. That is no news. However, one hears noises even after sex, even when everybody and everything is quiet. Even in a completely silent environment we do hear something. The noises one hears are produced within the ear, partly by the blood flowing through the head, partly by the electrical noise generated in the nerves. That is strange. If matter were continuous, there would be no noise even for low signal levels. The existence of noise makes one suspect that matter is made of smallest entities. This suspicion is confirmed in several ways by making love.

In fact, all proofs for the discreteness of matter, of electric current, of energy, or of light are always based on the increase of fluctuations with the smallness of systems under consideration.

• Sex is for *reproduction*. Sex is what we owe our life to, as we all are results of reproduction. But the reproduction of a structure is possible only if it can be constructed, i.e. if the structure can be built from small standard entities. Thus we suspect ourselves to be made of smallest, discrete entities.

Sex is a strange method of reproduction. Mathematics provides a much better one. If matter objects were not made of particles, but were continuous, it would be possible to perform reproduction by cutting and reassembling. A famous mathematical theorem by Banach and Tarski proves that it is possible to take a continuous solid, to cut it into six pieces, and to rearrange the pieces in such a way that one ends up with *two* copies of the same size and volume as the original. In fact, one can even produce volume increases in this way, thus realizing growth without any need for food. However, all these possibilities assume that matter is continuous, without a smallest length scale. The fact that the methods do not work in nature is compatible with the idea that matter is not continuous.

• Sex *needs memory*. If you would not recognize your partner among all possible ones, your love life would be quite complicated. A memory is a device which, in order to store information, must have small internal fluctuations. Obviously, fluctuations in systems get smaller as their number of components increase. Since our memory works so well, we can follow that we are made of a large number of small particles. Ref. 2

In summary, sex shows that we are made of a kind of lego bricks: depending on the level one looks at, these bricks are called molecules, atoms, or elementary particles. One can estimate their size using the sea around the tropical island, as well as a bit of oil. Can you imagine how? Challenge

• Sex *uses the sense of sight*. We are able to see each other: that is only possible because we are cold whereas the sun is hot. If we and our environment all had the same temperature of the sun, we could not see each other; this can be checked experimentally by looking into

Ref. 3 a hot oven: Inside a glowing oven filled with glowing objects it is impossible to discern them against the background.

• Sex *is motion*. Bodies move against each other. Moreover, it is possible to measure speed. Since measurement is a comparison with a standard, there must be a velocity standard in nature, i.e. some characteristic velocity scale. Such a standard must either be the minimum or the maximum possible value. Now, daily life shows that for velocity, a minimum value does not exist. To estimate the value of the maximum, just take your cellular phone and ring home from that island to your family. From the delay in the line and the height of the satellite, you can deduce the telephone speed c.

The existence of a maximum speed c implies that time is different for different observers. Looking into the details, one finds that this effect becomes noticeable at energies

$$E_{\text{different time}} \approx mc^2 \quad . \tag{609}$$

For example, this applies to electrons inside a television tube.

• Sex is based on *touching*. When we touch our partner, sometimes we get small shocks. The energies are larger than than those of fly legs. People are electric. In the dark, one can see that these discharges emit light. Light is thus related to electricity.

Touching also proves that light is a wave: simply observe the dark lines between two fingers near your eye in front of a bright background. The lines are due to interference effects. Light thus does not move with infinite speed. In fact, it moves with the same speed as that of telephone calls.

• Sex is *fun*. To make it more interesting people like to make love in different ways, such as in a dark room. It turns out that rooms get dark when the light is switched off only because we live in a space of odd dimensions. In even dimensions, the light would not turn off directly after the switch is flipped.

Moreover, the fun of sex is related to the fact that we can make knots with our legs, arms, and bodies. Knots are possible only in three dimensions. In short, sex is real fun only because we live in 3 dimensions.

• Sex is *tiring*. That is due to gravity. But was is gravity? Since there is a maximum speed, a little thinking shows that gravity is the change of time with height. More precisely speaking, gravity means that space-time is curved. Curved space also means that *horizons* appear, i.e. a largest possible visible distance. From equations (608) and (609), one deduces that this happens when distances are of the order of

$$R_{\text{horizon}} \approx Gm/c^2 \tag{610}$$

For example, only due to such a horizon, albeit one appearing in a different way, the night sky is dark.

• Sex *takes time*. It is known that men and women have different opinions on durations. It is also said that sex happens between your ears. Biological research indeed showed that we have a clock, due to circulating electrical currents, inside the brain. It provides our normal sense of time. But if this clock in the brain can use such circulating currents, it is because there must be a time standard in nature. Again, such a standard must be a minimum or a maximum time interval. We will discover it later on.

• Sex is *repulsive*. And in sex, *size matters*. Both facts turn out to be the two sides of the same coin. Sex is based on touch, and touch needs repulsion. Repulsion needs a length scale, but neither gravity nor classical electrodynamics provides one. Classical physics cannot explain that the measurement of length, time, or mass is possible.* Classical physics only allows for the measurement of speed. Classically, matter cannot be hard; one should be able to compress it. But sex tells us that this is not the case. Sex tells us that lengths scales do exist in nature, and thus that classical physics is not sufficient for the description of nature.

• Sex can *hurt*. For example, it can lead to injuries. Atoms can get ripped apart. That happens at energies of a few aJ per atom, i.e. when energies are concentrated on small volumes. Investigating such situations more precisely, one finds that at distances r, if energies exceed the value

$$E \approx \frac{\hbar c}{r} \tag{611}$$

strange phenomena appear: energy becomes chunky, motion and things become fuzzy. These phenomena are called quantum phenomena. The new constant $\hbar$ is important: it determines the size of things, because it allows to define distance and time units. In short, objects tear and break because in nature there is a minimum action, given roughly by $\hbar$.

If one tries to concentrate even more energy on small volumes, e.g. if one has energies of the order of mc^2 per particle, one observes transformation of energy into matter, or *pair production*. From equations (609) and (611), we deduce that this happens at distances of

$$r_{\text{pair production}} \approx \frac{\hbar}{m c} \quad . \tag{612}$$

At such small distances one cannot avoid using quantum mechanical description of nature.

• Sex is indeed *greek*. The greek were the first to make theories above love, such as Plato in his *Phaedrus*. But they also described it in another way. Already before Plato, Democritos said that making love is an example of particles moving and interacting in vacuum. If we change ‘vacuum’ to ‘curved 3+1-dimensional space’, and particle to ‘quantum particle’, we do indeed make love in the way Democritos described 2500 years ago.

It seems that physics has not made much progress in the meantime. But take the statement by the british astrophysicist Arthur Eddington:

> I believe there are 15,747,724,136,275,002,577,605,653,961,181,555,468,044,717,914,527, 116,709,366,231,425,076,185,631,031,296 protons in the universe and the same number of electrons. Ref. 4

Compare it with the modern version:

> Baryons in the universe: $10^{81\pm1}$; total charge: near zero.

The second is more honest, but which of the two is more sensible? In any case, both sentences show that there are unexplained facts in the greek description nature, such as the number of involved particles.

* Note that the classical electron radius is not an exception: it contains the elementary charge e, which contains a length scale. See page 275.

• Sex is *bestial*. We have seen that we can learn a lot about nature from the fact that we do have sex. We could be tempted to see this approach of nature as a special case of the so-called *anthropic principle*. But we need to be careful. In fact, we could have learned exactly the same if we had taken as starting point the fact that *apes* have sex. There is no "law" of nature which distinguishes between apes and humans. In fact, there is a simple way to determine whether any 'anthropic' statement makes sense: the reasoning must be equally true for humans, for apes, and for pigs.

For example, while studying stars, the british astrophysicist Fred Hoyle predicted a resonance in the carbon-12 nucleus, because otherwise stars could not have produced the carbon which then was spread out by explosions into interstellar space from which the earth formed. Also pigs and apes could reason this way; therefore Hoyle's statement does make sense.

As an example for the other case, claiming that the universe is made especially for people is not sensible: using the same arguments, pigs would say it is made for pigs. The existence of either requires all "laws" of nature. In summary, the anthropic principle is true only in so far as its consequences are indistinguishable from the porcine or the simian principle. In short, the fact that sex is something bestial tells something about the philosophy of physics.

• Sex is *holy*. Following the famous definition by the theologian Rudolf Otto, holiness
Ref. 5 results from a mixture of a *mysterium tremendum* and a *mysterium fascinans*. Tremendum means that it makes one tremble. Let us see if this is the case. Sex *produces heat*. It thus is a dissipative process. All systems in nature which produce heat have a finite lifetime. That is true for machines, stars, animals, lightning, fire, lamps, and people. Through heat, sex shows us that we are going to die. Physicists call this the second principle of thermodynamics.

But sex also fascinates. Everything which fascinates has a story. Indeed, this is a principle of nature: every dissipative structure, every structure which appears or is sustained through the release of energy, tells us that it has a story. Take atoms, for example. All the protons we are made of formed during the big bang. Most hydrogen we are made of is also that old. The other elements were formed in stars, and then blown into the sky during nova or supernova explosions. They then regrouped during planet formation. We truly are made of stardust.

Why do such stories fascinate? If you only think about how you and your partner have met, you will discover that it is through a chain of incredible coincidences. As in the case of the evolution of the world, if only one of all these coincidences had not taken place, you and your partner would not be together. And of course, we all owe our existence to such a chain of coincidences, since other such coincidences brought our parents together, their parents, etc.

The realization of the importance of coincidences automatically produces two kinds of questions: *why?* and *what if?* Physicists have now produced a list of all the answers to repeated why questions, and many are working at the list of what-if questions. The first list, the why-list, gives all still unexplained facts. One can also call it the complete list of all surprises in nature. (Above, it was said that there are no surprises in nature about *what* happens. However, so far there still are a handful of surprises on *how* all these things happen.)

Table 49 *Everything* quantum field theory and general relativity do *not* explain; in other words, a list of *the only* experimental data and criteria available for tests of the unified description of motion.

Observed value	Property unexplained so far
Local quantities, from quantum theory	
α_{em}	the low energy value of the electromagnetic coupling constant
α_w	the low energy value of the weak coupling constant
α_s	the low energy value of the strong coupling constant
m_q	the values of the 6 quark masses
m_l	the values of 3 lepton masses
m_W	the values of the independent mass of the W vector boson
θ_W	the value of the Weinberg angle
$\beta_1, \beta_2, \beta_3$	three mixing angles
θ_{CP}	the value of the CP parameter
θ_{st}	the value of the strong topological angle
3	the number of particle generations
$0.5\,nJ/m^3$	the value of the observed vacuum mass density or cosmological constant
$3+1$	the number of space and time dimensions
Global quantities, from general relativity	
$1.2(1) \cdot 10^{26}\,m$?	the distance of the horizon, i.e. the “size” of the universe (if it makes sense)
10^{82} ?	the number of baryons in the universe, i.e. the average matter density in the universe (if it makes sense)
$> 10^{92}$?	the initial conditions for more than 10^{92} particle fields in the universe, including those at the origin of galaxies, of stars, etc. (if they make sense)
Local structures, from quantum theory	
$S(n)$	the origin of particle identity, i.e. of permutation symmetry
Ren. group	the renormalisation properties, i.e. the existence of point particles
SO(3,1)	the origin of Lorentz (or Poincaré) symmetry (i.e. of spin, position, energy, momentum)
C^*	the origin of the algebra of observables
Gauge group	the origin of gauge symmetry (and thus of charge, strangeness, beauty, etc.)
in particular, for the standard model:	
U(1)	the origin of the electromagnetic gauge group (i.e. of the quantization of electric charge, as well as the vanishing of magnetic charge)
SU(2)	the origin of weak interaction gauge group
SU(3)	the origin of strong interaction gauge group
Global structures, from general relativity	
maybe $R \times S^3$?	the unknown topology of the universe (if it makes sense)

This why-list fascinates through its shortness, which many researchers are still trying to reduce. But it is equally interesting to study what consequences appear if any of the values from the previous list were only a tiny bit different. It is not a secret that small changes in nature would lead to completely different observations.

Table 50 A tiny selection of the consequences of changing aspect of nature

Observable	Change	Result
Moon size	smaller	small earth magnetic field; too much cosmic radiation; widespread child cancers.
Moon size	larger	large earth magnetic field; too little cosmic radiation; no evolution into humans.
Jupiter	smaller	too many comet impacts on earth; extinction of animal life.
Jupiter	larger	too little comet impacts on earth; no moon; no dinosaur extinction.
Oort belt	smaller	no comets, no irregular asteroids, no moon; still dinosaurs.
Galaxy distance	smaller	irregular planet motion; supernova dangers.
Strong coupling constant	smaller	proton decay; leucemia.

The large number of coincidences of life force our mind onto one main point: we are only a *tiny* part of nature. We are a small droplet shaken around in the ocean of nature. In short, even the tiniest changes would prevent the existence of humans, apes, and pigs. In other words, making love tells us that the universe is much larger than we are, and tells us how much we are connected to the rest of the universe.

• We said above that sex uses *motion*. But that is a remarkable mystery, worth a second look:

- Motion is the change of position with time of some bodies.

- Position is what we measure with a ruler. Time is what we measure with a clock. Both rulers and clocks are bodies.

- A body is en entity distinct from its environment by its shape or its mass. Shape is the extension of a body in space (and time). Mass is measured by measuring speed or acceleration, i.e. by measuring space and time.

This means that we define space-time with bodies – and specify the details in general relativity –, and that we define bodies with space-time – as done in full detail in quantum theory. This circular reasoning shows that making love truly is a mystery. The circular reasoning has not yet been eliminated yet; at present, modern theoretical physicists are busy attempting to do so. The most promising seems to be M-theory, the modern extension of string theory. But any such attempt has to overcome important difficulties; they can also be experienced while making love, as we will see now.

• Sex is *private*. But is it? Privacy assumes that one is able to separate oneself from the rest, without important interactions, at least for a given time, and come back later. One usually assumes that this is possible if one puts enough *empty distance* between oneself and others. In other words, privacy is based on the idea that we are able to distinguish objects from vacuum. Let us check whether this is always possible.

Ref. 6 What is the smallest distance one can measure? This question has been almost, but only almost answered by Max Planck in 1899. The distance δl between two objects of mass m is surely larger than their position uncertainty $\hbar/\Delta p$; and the momentum uncertainty must be smaller that the momentum leading to pair production, i.e. $\Delta p < mc$. This means that

$$\delta l \geqslant \Delta l \geqslant \frac{\hbar}{mc} \quad . \tag{613}$$

In addition, the measurements require that signals leave the objects; the two masses must not be black holes. Their masses must be so small that the Schwarzschild radius is smaller than the distance to be measured. This means that $r_S \approx Gm/c^2 < \delta l$ or that

$$\delta l \geqslant \sqrt{\frac{\hbar G}{c^3}} = l_{Pl} = 1.6 \cdot 10^{-35}\,\mathrm{m} \quad . \tag{614}$$

This expression defines a minimum length in nature, the so-called *Planck length*. All other Gedankenexperiments lead to this characteristic length as well. In fact, this minimum distance (and the corresponding minimum time interval) provides the measurement standard we were looking for at the beginning of our musings about length and time measurements.

A more detailed discussion shows that the smallest *measurable* distance is somewhat larger, a multiple of the Planck length, as measurements require the distinction of matter and radiation. This happens at scales about 600 times larger than the Planck length.

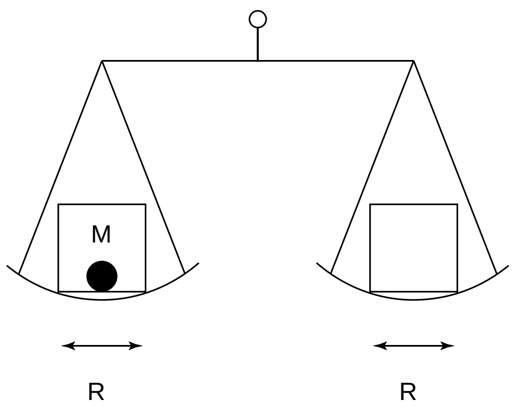


Figure 182 A *Gedankenexperiment* showing that at Planck scales, matter and vacuum cannot be distinguished

In other words, privacy has its limits. In fact, the issue is even more muddled when one explores the consequences for bodies. A body, also a human one, is something one can touch, throw, hit, carry or weigh. Physicists say that a body is something with energy or momentum. Vacuum has none of it; instead, vacuum is unbounded, whereas objects are bounded.

What happens if one tries to weigh objects at Planck scales? If one puts an object of mass M in a box of size R onto a scale quantum theory makes a simple prediction. Equation (611) implies that there is a minimal mass error ΔM given by

$$\Delta M \approx \frac{\hbar}{cR} \quad . \tag{615}$$

If the box has Planck size, the mass error is the Planck mass

$$\Delta M = M_{Pl} = \sqrt{\hbar c/G} \approx 22\,\mu\mathrm{g} \quad . \tag{616}$$

How large is the mass one can put into a box of Planck size? Obviously it is given by the maximum possible mass density. To determine it, imagine a planet and put a satellite in orbit around it, just skimming its surface. The density ρ of the planet with radius r is given by

$$\rho \approx \frac{M}{r^3} = \frac{v^2}{Gr^2} \qquad (617)$$

Using equation (613) we find that the maximum mass density in nature, within a factor of order one, is the so-called *Planck density*, given by

$$\rho_{\rm Pl} = \frac{c^5}{G^2\hbar} = 5.2\cdot 10^{96}\,\mathrm{kg/m^3} \quad . \qquad (618)$$

Therefore the maximum mass in a Planck box is the Planck mass. But that was also the measurement error for that situation. This implies that we cannot say whether the original box we measured was empty or full: vacuum *cannot* be distinguished from matter at Planck scales. This astonishing result is confirmed by all other Gedankenexperiments exploring the issue. Ref. 6

It is straightforward to deduce with similar arguments that objects are not bound in size at Planck scales, i.e. that they are not localized, and that the vacuum is not necessarily extended at those scales. In addition, the concept of particle number cannot be defined at Planck scales. Challenge

So, why is there something instead of nothing? Making love shows that there is no difference between the two options!

• Sex makes one *dream*. When we dream, especially at night, we often look at the sky. How far is it away? How many atoms are enclosed by it? How old is it? These questions have an answer for small distances and for large distances; but for the whole of the sky or the whole of nature they cannot have one, as there is no way to be outside of the sky in order to measure it. In fact, each of the impossibilities to measure nature at smallest distances are found again at the largest scales. There seems to be a fundamental equivalence, or, as physicists say, a duality between the largest and the smallest distances.

The coming years will hopefully show how we can translate these results into an even more precise description of motion and of nature. In particular, this description should allow us to reduce the number of unexplained properties of nature.

In summary, making love is a good physics lesson. Enjoy the rest of your day.

References

1 An attempt to explain the lack of women in physics is made in Margaret WERTHEIM, *Pythagoras' trousers – god, physics, and the gender wars*, Fourth Estate, 1997. Cited on page 696.

2 The consequences of memory loss in this case are already told by VOLTAIRE, *Aventure de la mémoire*, 1775. Cited on page 698.

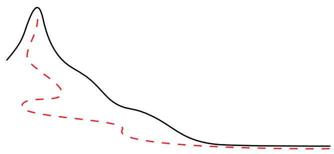
3 A picture of objects in a red hot oven and at room temperature is shown in C.H. BENNETT, *Demons, engines, and the second law*, Scientific American **255**, pp. 108–117, November 1987. Cited on page 699.

4 The famous quote is found at the beginning of one of the chapters in Arthur EDDINGTON, *The Philosophy of Physical Science*, Cambridge, 1939. Cited on page 700.

5 See the first intermezzo for details and references. Cited on page 701.

6 See chapter XII of this text the details and a full list of references. The chapter is a reworked version of the pages published in french as Christoph SCHILLER, *Le vide diffère-t-il de la matière?*, in E. GUNZIG & S. DINER, editeurs, *Le vide – Univers du tout et du rien – Des physiciens et des philosophes s'interrogent*, Les Éditions de l'Université de Bruxelles, 1998. An older english version is also available as Christoph SCHILLER, *Does matter differ from vacuum?*, http://xxx.lanl.gov/abs/gr-qc/9610066. Cited on page 703, 705.

7 See chapter 31. of this text for details of the arguments leading to duality. It also includes suggestions supporting the notion that the universe is not a even a set, and thus indirectly proposes a solution for Hilbert's sixth problem. Cited on page .



APPENDICES

A collection of reference information
useful for mountain escalations
and other adventures.

Typeset in April 2001

Appendix A Notation and Conventions

Newly introduced and defined concepts in this text are indicated by *italic typeface*. New definitions can also be found in the index, referred to with italic page numbers. In all formulas throughout the text SI units are used. They are defined in appendix B. In gravitation, we use the time convention, i.e. a (pseudo) metric g with signature $(+---)$, as used by about 70% of the literature worldwide. We use indices i, j, k for three-vectors, Ref. 1 and indices a, b, c, etc. for four-dimensional indices. Experimental results are cited in the text with limited precision, usually only two digits, since this is usually sufficient for discussion. More precise values of the fundamental constants and other experimental results can be found in appendix B.

The symbols used in the text

> To avoide the tediouse repetition of these woordes: is equalle to: I will sette as I doe often in woorke use, a paire of paralleles, or Gemowe lines of one lengthe, thus: = , bicause noe .2. thynges, can be moare equalle.
>
> Robert Recorde*

The symbols used as abbreviations for physical quantities are always defined in the context where they are used. The symbols designating units, constants and particles are defined in the respective appendices. The following list gives a list of those symbols used in formulas in this text which were not defined there, and their short history. They conform as much as Ref. 2 possible to the ISO standard. Ref. 3

$+, -$	plus, minus; the plus sign is derived from latin 'et' — german mathematicians, end of 15th century
$\sqrt{}$	read as 'square root'; the sign stems from a deformation of the letter 'r', initial of the latin 'radix' — used by K. Rudolff in 1525
$=$	equal to — italian mathematicians, early 16th century, brought to england by R. Recorde 1557
{ }, [ ], ( )	grouping symbols — use starts in the 16th century
$>, <$	larger than, smaller than — T. Harriot 1631
$\times$	multiplied with, times — W. Oughtred 1631

∗ Robert Recorde (ca.1510–1558), english mathematician and physician; he died in prison, though not for his Ref. 2 pretention to be the inventor of the equal sign, which he simply took over from his italian colleagues, but for a smaller crime, namely debth. The quotation is from his *The Whetstone of Witte*, 1557. An image of the quote is found at the http://www.geocities.com/westpasco/witte.jpg web site.

:	divided by — G. Leibniz 1684
·	multiplied with, times — G. Leibniz 1698
a^n	power — R. Descartes 1637
x, y, z	coordinates, unknowns — R. Descartes 1637
$ax + by + c = 0$	constants and equations for unknowns — R. Descartes 1637
d/dx, d^2x, $\int y\,dx$	derivative, differential, integral — G. Leibniz 1675
φx	function of x — J. Bernoulli 1718
fx, $f(x)$	function of x — L. Euler 1734
Δx, $\sum$	difference, sum — L. Euler 1755
$\neq$	is different from — L. Euler 18th century
$\partial/\partial x$	partial derivative, read like 'd/dx'; it was deduced from cursive form of the letter 'dey' of the cyrillic alphabet — A. Legendre 1786
Δ	Laplace operator — R. Murphy 1833
$\lvert x \rvert$	absolute value — K. Weierstrass 1841
∇	read as 'nabla' — introduced by W. Hamilton in 1853, from the shape of an old egyptian musical instrument
$[x]$	the measurement unit of a quantity x — 20th century
∞	infinity — J. Wallis 1655
π	$4 \arctan 1$ — H. Jones 1706
e	$\lim_{n\to\infty}(1+1/n)^n$ — L. Euler 1736
i	$+\sqrt{-1}$ — L. Euler 1777
$\cup, \cap$	set union and intersection — G. Peano 1888
$\in$	element of — G. Peano 1888
$\emptyset$	empty set — André Weil as member of the N. Bourbaki group in the early 20th century

Other signs have more complicated origins. The & sign is a contraction of latin 'et' meaning 'and', as often is more clearly visible in its variations, such as &, the common italic form. Ref. 4

The section sign § dates from the 13th century in northern Italy, as was shown by the german paleographer Paul Lehmann. It was derived from ornamental versions of the capital letter C for 'capitulum', i.e. 'little head' or 'chapter.' The sign appeared first in legal texts, where it is still used today, and then spread also into other domains. Ref. 5

The paragraph ¶ sign was derived from a simpler ancient form looking like the greek letter Γ, a sign which was used in manuscripts from ancient Greece until way into the middle ages to mark the start of a new text paragraph. In the middle ages it took the modern form because probably a letter c for 'caput' was added in front of it.

The punctuation signs used in sentences with modern latin alphabets, such as , . ; : ! ? " " » « – ( ) ... , have their own history. Many are from ancient Greece, but the question mark is from the court of Charlemagne, and exclamation marks appear first in the 16th century.* Ref. 6 The @ or *at-sign* may stem from a medieval abbreviation of latin ad, meaning 'at', in a similar way as the & sign evolved. In recent years, the *smiley* :-) and its variations has become popular; it is in fact a new edition of the 'point of irony' which had been proposed already, without success, by A. de Brahm (1868–1942). Ref. 7

∗ On the parenthesis see the beautiful book by J. LENNARD, *But I disgress*, Oxford University Press, 1991.

Ref. 8 Perhaps the most important sign of all, the white space separating words, was due to keltic and germanic influences when these people started using the latin alphabet. It became commonplace only between the 9th and the 13th century, depending on the language in question.

The latin alphabet

This text is written using the Latin alphabet. By the way, this implies that its pronunciation *cannot* be explained in print, in contrast to that of any other alphabet. The latin alphabet was derived from the etruscan, which itself was a derivation of the greek alphabet. The main forms are

Challenge

from the 6th century BCE onwards, the *ancient* latin alphabet:

A B C D E F Z H I K L M N O P Q R S T V X

from the 2nd century BCE until the 11th century, the *classical* latin alphabet:

A B C D E F G H I K L M N O P Q R S T V X Y Z

The latin alphabet was spread around Europe, Africa, and Asia by the romans during their conquests; due to its simplicity it was adopted by numerous modern languages. The letter G was added in the third century BCE by the first roman to run a fee paying school, Spurius Carvilius Ruga, by adding a horizontal bar to the letter C, and substituting the letter Z, which was not used in latin any more.

In the second century BCE, after the conquest of Greece, the romans included the letters Y and Z from the Greek alphabet at the end of their own (therefore effectively reintroducing the Z) in order to be able to write greek words. This classical Latin alphabet was stable throughout the next one thousand years.

Most modern "Latin" alphabets usually include other letters. The letter W was introduced in the 11th century in french, and was then adopted in most other languages. The letters J and U were introduced in the 16th century in Italy, to distinguish them from I and V, which used to have both meanings. In other languages they are used for other sounds. Other Latin alphabets include more letters, such as the german *sharp s*, written ß, a contraction of 'sz', the nordic letters *thorn* and *eth*, taken from the futhark, and others signs.* Ref. 10

Similarly, lower case letters are not latin, in fact, and date only from the middle ages. Like most accents, who were also defined in the middle ages, they were introduced to save the then expensive paper surface by shortening words in writing or printing.

> Outside a dog, a book is a man's best friend.
> Inside a dog, it's too dark to read.
> Groucho Marx

* The runic script or *futhark*, a type of alphabet used in the middle ages in germanic countries, in the anglo-saxon sphere, and in the nordic countries, probably also derives from the etruscan alphabet. As the name says, the first letters were f, u, th, a, r, k (in other regions f, u, th, o, r, c). The third letter is the letter thorn mentioned above; it is often written 'Y' in old english, as in 'Ye Olde Shoppe.'

The greek alphabet

The greek alphabet in turn was derived from the phoenician or a similar north semitic alphabet in the 10th century BCE. In contrast to the etruscan and latin alphabets, each letter has a proper name, as was the case for the phoenician alphabet and many of its derivates. The greek letter names of course are the origin of the term *alphabet* itself. Ref. 9

In the tenth century BCE, the *ancient* greek alphabet consisted of the upper case letters only. In the 6th century BCE several letters were dropped, some new one and the lower-case versions were added, giving the *classic* alphabet. Still later, accents, subscripts and the breathings were introduced. The following table also gives the values the letters took when they were used as numbers. For this special use the obsolete ancient letters were kept also during the classical period.

ancient	classic	name	correspondence		ancient	classic	name	correspondence	
Α	Α α	Alpha	a	1	Ν	Ν ν	Nu	n	50
Β	Β β	Beta	b	2	Ξ	Ξ ξ	Xi	x	60
Γ	Γ γ	Gamma	g, n[a]	3	Ο	Ο ο	Omicron	o	70
Δ	Δ δ	Delta	d	4	Π	Π π	Pi	p	80
Ε	Ε ε	Epsilon	e	5	Ϡ ϡ		Sampi[c]	s	900
F ϛ		Digamma[b]	w	6			Qoppa	q	90
Ζ	Ζ ζ	Zeta	z	7	Ρ	Ρ ρ	Rho	r, rh	100
Η	Η η	Eta	e	8	Σ	Σ σ, ς	Sigma	s	200
Θ	Θ θ	Theta	th	9	Τ	Τ τ	Tau	t	300
Ι	Ι ι	Iota	i, j	10		Υ υ	Upsilon	y, u[d]	400
Κ	Κ κ	Kappa	k	20		Φ φ	Phi	ph, f	500
Λ	Λ λ	Lambda	l	30		Χ χ	Chi	ch	600
Μ	Μ μ	Mu	m	40		Ψ ψ	Psi	ps	700
						Ω ω	Omega	o	800

a. Only if before velars, i.e. before kappa, gamma, xi and chi.
b. ‘Digamma’ or ‘stigma’, as it is also called, are names deduced from the way the letter looks. The original letter name, also giving its pronounciation, was ‘waw’.
c. The letter sampi was positioned after omega in later times.
d. Only if second letter in diphtongs.

The latin correspondence in the list is the standard classical one, used in writing of greek words. The question of the *pronunciation* of greek has been a hot issue in specialist circles; the traditional *erasmian* pronunciation does not correspond to the results of linguistic research, nor to the modern greek one. (In modern Greek, pronunciation is different for β, which is now pronounced ‘v’, and for η, which is now pronounced ‘i:’.) The pronunciation of greek varied from region to region and with time. For attic, the main dialect spoken in the classical period, the question is now settled. Linguistic research showed that chi, phi, and theta were less aspirated than usually pronounced, and sounded like the initials of ‘cat’, ‘perfect’ and ‘tin’; moreover, the zeta seems to have been pronounced more like ‘zd’ as in ‘buzzed’. For the vowels, contrary to tradition, epsilon is closed and short whereas eta is open and long, omikron is closed and short, whereas omega is wide and long, and upsilon is really a ‘u’ sound like in ‘boot’, not a french ‘u’ or german ‘ü.’

The greek vowels can have rough or smooth *breathings*, as well as acute, grave, circumflex or dieresis *accents*, and *subscripts*. Breathings, used also on ρ, determine whether the letter is aspirated. Accents, taken only as stresses in the erasmian pronunciation, actually represented pitches. Classical greek could have up to three added signs per letter; modern greek never has more than one accent.

A descendant of the greek alphabet is the *cyrillic alphabet*, used with slight variations in many slavic languages, such as russian.* However, there is no standard transcription from cyrillic to latin, so that often the same author is spelled differently in different countries and at different occasions.

The hebrew alphabet

The phoenician alphabet is also at the origin of the hebrew alphabet, which starts in the following way. Ref. 3

letter	name	corr.
א	aleph	a
ב	beth	b
ג	gimel	g
ד	daleth	d
etc.	etc.	etc.

Only the first of these letters is used in mathematics.

There are a few additional alphabets in the world, some having a sign for each *sound*, such as Arabic and the Hieroglyphic script, and some having a sign for each *syllab*, such as Maya, Korean, or Japanese. In adition there are non-alphabetic writing systems, having signs for each *word*, such as Chinese. Even though there are about 7000 languages in total, there are only about *two dozen* writing systems. For physical and mathematical formulas, the sign system presented here, based on latin and greek letters, is a standard the world over, and is independent of the writing system used for text.

Digits and numbers

The digits and the method used in this text to write numbers stem from India. All was brought to the mediterranean by arabic mathematicians in the middle ages. The number system used in this text is thus much younger than the alphabet. The signs 0, 2, 3 and 7 still

* The greek alphabet also was at the origin of the *gothic alphabet*, which was defined in the 4th century by Wulfila for the gothic language, using also a few signs from the latin and futhark scripts. It is not to be confused with the so-called *gothic letters*, a style of the latin alphabet used all over Europe from the 11th century until the 16th century and later, when in latic countries the *antiqua* took over, the ancestor of the type in which this text is set. Gothic letters were used in type and handwriting in Germany until the government suddenly abolished them in 1941; sporadically, they still can be found in germanic countries. In many physics and mathematics books, gothic letters are used to denote vector quantities.

resemble closely those used in arabic writing, if one turns them clockwise by 90 degrees.*
The “arabic” numbers were made popular in Europe by Leonardo of Pisa, called Fibonacci, in his book *Liber Abaci*, which he published in 1202. From that day on mathematics was not the same any more. Everybody with paper and pen was now able to calculate and write down numbers as large as reason allows, and even larger. The new method had two innovations, both taken over from India: the *positional system* of writing numbers, and the digit zero. The positional system described by Fibonacci was so much more efficient to write numbers that it completely replaced the previous *roman number system*, which writes 1998 as IIMM or MCMIIC or MCMDCVIII, as well as the *greek number system*, in which the greek letters were used for numbers as shown above. Today we would say that compared to the previous systems, indian-arabic numbers are a much better technology. Indeed, the indian-arabic system is so practical that calculations done on paper completely eliminated calculations with help of the *abacus*, which therefore fell in disuse. The abacus is still in use only in those countries which do not use a positional system to write numbers. Similarly, only the positional number system allows mental calculations and made calculating prodigies possible.**

Calendars

The many ways to keep track of time differ greatly across the civilisations. The most common calendar, the one used in this text, is at the same time one of the most absurd, as it is a compromise between various political forces who tried to shape it.

In ancient times, independent localized entities, such as tribes or cities, preferred *lunar* calendars, because they are easily organized locally. This lead to the use of the month as timekeeping unit. Centralized states imposed *solar* calendars, to organize power control such as tax collection. Farmers, politicians, astronomers, and some, but not all, religious groups also wanted solar calendars following the solar year as precisely as possible. That is the origin of leap days. This structure was organized over 2000 years ago by the romans, and called the *julian calendar*.

The week is an invention of Babylonia, and was taken over and spread by various religious groups; even though about three thousand years old, it was added into the modern calendar only much later, towards the end of the roman empire. The last change took place towards the end of the middle ages, when more precise measurements of the solar year led to a new method to determine leap days, still in use today. Together with the fixation of the week rhythm, this standard is called the *gregorian calendar*.

Despite this complexity, the modern calendar allows to determine the day of the week of a given date in your head. Just do the following:

- take the last two digits of the year, divide by 4, discarding any fraction,
- add the last two digits of the year,
- subtract 1 for January of February of a leap year,

* The story of the development of the numbers is told most interestingly by G. Ifrah, *Histoire universelle des chiffres*, Seghers, 1981, which has been translated into several languages. He sums up the genealogy in ten beautiful tables, one for each digit, at the end of the book.

** About this last topic, see the fascinating book by Steven B. Smith, *The great mental calculators – The psychology, methods, and lives of the calculating prodigies*, Columbia University Press, 1983.

- add 6 for 2000's or 1600's, 4 for 1700's or 2100's,
  2 for 1800's and 2200's, and 0 for 1900's or 1500's,
- add the day of the month,
- add the month key value, namely 144 025 036 146 for JFM AMJ JAS OND.

The remainder after division by 7 gives the day of the week, with the correspondence 1 / 2 / 3 / 4 / 5 / 6 / 7 or 0 meaning sunday / monday / tuesday / wednesday / thursday / friday / saturday.*

Counting years is of course a matter of preference. The oldest method not attached to political power structures was the method used in ancient Greece, when years were counted in function of the olympic games. In those times, people used to say e.g. that they were born in the first year of the 23rd olympiad. Later, political powers always imposed counting years from some important event onwards.** Maybe reintroducing the olympic counting is worth considering?

Abbreviations, eponyms or concepts?

The scourge of modern physics are sentences like the following:

> The EPR paradox in the Bohm formulation can perhaps be resolved using the GRW approach, using the WKB approximation of the Schrödinger equation.

Using such vocabulary is the best method to make language unintelligible to outsiders. It uses abbreviations, of names in this case, which is already a shame. On top of this, the sentence uses people's names to characterize concepts, i.e. it uses *eponyms*. Originally, eponyms were intended as tributes to outstanding achievements. Today, when formulating new laws or variables has become nearly impossible, the spread of eponyms intelligible only to a steadily decreasing population simply reflects an increasingly ineffective drive to fame.

Eponyms are a lasting proof of the small imagination of scientists; they are avoided as much as possible in our walk. In this text, mathematical equations or entities are given

* Remembering the intermediate result for the current year can simplify things even more, especially since the dates 4.4., 6.6., 8.8., 10.10., 12.12., 9.5., 5.9., 7.11., 11.7., and the last day of february all fall on the same day of the week, namely on the year's intermediate result plus 4.

** The present counting of year was defined in the middle ages by setting the date for the foundation of Rome to the year 753 BCE, or 753 Before Common Era, and then counting backwards, implying that the BCE years behave like negative numbers. However, the year 1 follows directly after the year 1 BCE; there was no year 0.

Some other standards set by the roman empire also remain, and determine several abbreviations used in the text:

- ca. is a latin abbreviation for 'circa' and means 'roughly'.
- i.e. is a latin abbreviation for 'ita est' and means 'that is'.
- e.g. is a latin abbreviation for 'exempli gratia' and means 'for the sake of example'.

By the way, 'idem' means 'the same'. Also terms like frequency, acceleration, velocity, mass, force, momentum, inertia, gravitation, temperature, etc., are latin. In fact, there is a strong case to be made that the language of science has been latin for over two thousand years. In roman times it was latin with latin grammar, in modern times it switched to latin vocabulary and french grammar, and now it has switched to latin vocabulary and british/american grammar. Also many units of measurement date from roman times, as explained in the next appendix. Even the infatuation with greek technical terms, such as "gyroscope", "entropy", or "proton", dates from roman times.

Ref. 11

common names wherever possible. People's names are then used as appositions to these names. For example, "Newton's equation of motion" is never called "Newton's equation", "Einstein's field equations" is used instead of "Einstein's equations", "Heisenberg equation of motion" in place of "Heisenberg's equation".

However, exceptions are inevitable for certain terms within modern physics, for which no real alternatives exist. The Boltzmann constant, the Planck units, the Compton wavelength, the Casimir force, Lie algebras and groups, the Virasoro algebra are examples.

References

1 For a clear overview of the various sign conventions, see the front cover of Charles W. Misner, Kip S. Thorne & John A. Wheeler, *Gravitation*, Freeman, 1973. Cited on page 714.

2 See for example the voice 'Mathematical notation' in the *Encyclopedia of Mathematics*, 10 volumes, Kluwer Academic Publishers, Dordrecht, 1988,-1993. There is also the beautiful http://members.aol.com/jeff570/mathsym.html web site, and the extensive research by Florian Cajori, *A history of mathematical notations*, 2 volumes, The Open Court Publishing Co., 1928–1929. Cited on page 714, 714.

3 David R. Lide, editor, *CRC Handbook of Chemistry and Physics*, 78th edition, CRC Press, 1997. This classic reference work appears in a new edition every year. The full hebrew alphabet is given on page 2-90. Cited on page 714, 719.

4 Jan Tschichold, *Vormveranderingen van het &-teken*, Uitgeverij De Buitenkant, 1994. Cited on page 716.

5 Paul Lehmann, *Erforschung des Mittelalters – Ausgewählte Abhandlungen und Aufsätze*, Anton Hiersemann, Stuttgart, 1961, pp. 4–21. Cited on page 716.

6 M.B. Parkes, *Pause and effect: An Introduction to the history of punctuation in the West*, University of California Press, 1993. Cited on page 716.

7 This is explained by Berthold Louis Ullman, *Ancient writing and its influence*, 1932. Cited on page 716.

8 Bernard Bischoff, *Paläographie des römischen Altertums und des abendländischen Mittelalters*, Erich Schmidt Verlag, 1979, pp. 215–219. Cited on page 717.

9 Hans Jensen, *Die Schrift*, Berlin, 1969, translated into english as *Sign, Symbol and Script: an account of man's efforts to write*, Putnam's Sons, New York. Cited on page 718.

10 About the thorn and the eth, see the extensive report to be found on the web site http://www.indigo.ie/egt/standards/iso10646/wynnyogh/thorn.html. Cited on page 717.

11 The connections between greek roots and many french words, and thus many english ones can be used to get a good vocabulary of ancient greek without much study, as shown by the practical collection by J. Chaineux, *Quelques racines grecques*, Wetteren – De Meester, 1929. Cited on page 721.

Typeset in April 2001

Appendix B Measurements, Units, and Constants

Measurements are comparisons. The standard with which one compares is called a *unit*. Many different systems of units have been used throughout the world. Standards always confer a lot of power to the organization in charge of them, as can be seen most clearly in the computer industry; in the past the same happened for measurement units. To avoid misuse by authoritarian institutions, to eliminate at the same time all problems with differing, changing and irreproducible standards, and – this is not a joke – to simplify tax collection, already in the 18th century scientists, politicians, and economists have agreed on a set of units. It is called the *Système International d'Unités*, abbreviated *SI*, and is defined by an international treaty, the 'Convention du Mètre'. The units are maintained by an international organization, the 'Conférence Générale des Poids et Mesures', and its daughter organizations, the 'Commission Internationale des Poids et Mesures' and the 'Bureau International des Poids et Mesures', which all originated in the times just before the french revolution. Ref. 1

All SI units are built from seven *base units* whose official definitions, translated from french into english, are the following, together with the date of their formulation:

• "The *second* is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom." (1967) *

• "The *metre* is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second." (1983)

• "The *kilogram* is the unit of mass; it is equal to the mass of the international prototype of the kilogram." (1901) *

• "The *ampere* is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 metre apart in vacuum, would produce between these conductors a force equal to $2 \cdot 10^{-7}$ newton per metre of length." (1948)

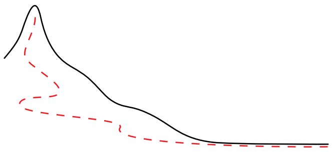
• "The *kelvin*, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water." (1967) *

• "The *mole* is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12." (1971) *

• "The *candela* is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency $540 \cdot 10^{12}$ hertz and has a radiant intensity in that

This is a section of the freely downloadable e-textbook

Motion Mountain



Hiking beyond space and time along the concepts of modern physics

available at
www.motionmountain.org

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Challenge - Did you find any mistakes?

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

direction of (1/683) watt per steradian." (1979) *

Note that both time and length units are defined as certain properties of a standard example of motion, namely light. This is an additional example making the point that the observation of motion as the fundamental type of change is a *prerequisite* for the definition and construction of time and space. By the way, the proposal of using light was made already in 1827, by Jacques Babinet. *

From these basic units, all other units are defined by multiplication and division. In this way, all SI units have the following properties:

• They form a system with *state-of-the-art precision*; all units are defined in such a way that the precision of their definition is higher than the precision of commonly used measurements. Moreover, the precision of the definitions are regularly improved. The present relative uncertainty of the definition of the the second is around 10^{-14}, for the metre about 10^{-10}, for the ampere 10^{-7}, for the kilogram about 10^{-9}, for the kelvin 10^{-6}, for the mole less than 10^{-6}, and for the candela 10^{-3}.

• They form an *absolute* system; all units are defined in such a way that they can be reproduced in every suitably equipped laboratory, independently, and with high precision. This avoids as much as possible any misuse by the standard setting organization. (At present, the kilogram, still defined with help of an artifact, is the last exception to this requirement; extensive research is underway to eliminate this artifact from the definition – an international race that will take a few more years.)

• They form a *practical* system: base units are adapted to daily life quantities. Frequently used units have standard names and abbreviations. The complete list includes the seven base units as well as:

- the *radian* (rad) and the *steradian* (sr) are *supplementary* SI units for angle, defined as the ratio of arc length and radius, and for solid angle, defined as the ratio of the subtended area and the square of the radius, respectively;

- the *derived* units with special names are, in their official english spelling, i.e. without capital letters and accents:

name	abbr. & definition	name	abbr. & definition
hertz	$\mathrm{Hz} = 1/\mathrm{s}$	newton	$\mathrm{N} = \mathrm{kg\,m/s^2}$
pascal	$\mathrm{Pa} = \mathrm{N/m^2} = \mathrm{kg/m\,s^2}$	joule	$\mathrm{J} = \mathrm{Nm} = \mathrm{kg\,m^2/s^2}$
watt	$\mathrm{W} = \mathrm{kg\,m^2/s^3}$	coulomb	$\mathrm{C} = \mathrm{As}$
volt	$\mathrm{V} = \mathrm{kg\,m^2/As^3}$	farad	$\mathrm{F} = \mathrm{As/V} = \mathrm{A^2s^4/kg\,m^2}$
ohm	$\Omega = \mathrm{V/A} = \mathrm{kg\,m^2/A^2s^3}$	siemens	$\mathrm{S} = 1/\Omega$
weber	$\mathrm{Wb} = \mathrm{Vs} = \mathrm{kg\,m^2/As^2}$	tesla	$\mathrm{T} = \mathrm{Wb/m^2} = \mathrm{kg/As^2}$
henry	$\mathrm{H} = \mathrm{Vs/A} = \mathrm{kg\,m^2/A^2s^2}$	degree Celsius	°C

∗ The international prototype of the kilogram is a platinum-iridium cylinder kept at the BIPM in Sèvres, in France. For more details on the levels of the cesium atom, consult a book on atomic physics. The Celsius scale of temperature θ is defined as: $\theta/°\mathrm{C} = T/\mathrm{K} - 273.15$; note the small difference with the number appearing in the definition of the kelvin. When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles or specified groups of such particle. In its definition, it is understood that the carbon 12 atoms are unbound, at rest and in their ground state. The frequency of the light in the definition of the candela corresponds to 555.5 nm, i.e. green colour, and is the wavelength for which the eye is most sensitive.

Ref. 2

∗ Jacques Babinet (1794–1874), french physicist who published important work in optics.

name	abbr. & definition	name	abbr. & definition
lumen	$\mathrm{lm} = \mathrm{cd\,sr}$	lux	$\mathrm{lx} = \mathrm{lm/m^2} = \mathrm{cd\,sr/m^2}$
becquerel	$\mathrm{Bq} = 1/\mathrm{s}$	gray	$\mathrm{Gy} = \mathrm{J/kg} = \mathrm{m^2/s^2}$
sievert	$\mathrm{Sv} = \mathrm{J/kg} = \mathrm{m^2/s^2}$	katal	$\mathrm{kat} = \mathrm{mol/s}$

- the *admitted* non-SI units are *minute*, *hour*, *day* (for time), *degree* $1° = \pi/180\,\mathrm{rad}$, *minute* $1' = \pi/10800\,\mathrm{rad}$, *second* $1'' = \pi/648\,000\,\mathrm{rad}$ (for angles), litre, and tonne.

All other units are to be avoided. One notes that in all these definitions of units, the kilogram only appears to the powers of 1, 0 and -1. The final explanation for this fact appeared only recently. See page 669

All SI units are made more practical by the introduction of standard names and abbreviations for the powers of ten, the so-called *prefixes*:*

	name	abbr.		name	abbr.		name	abbr.		name	abbr.
10^1	deca	da	10^{-1}	deci	d	10^{18}	Exa	E	10^{-18}	atto	a
10^2	hecto	h	10^{-2}	centi	c	10^{21}	Zetta	Z	10^{-21}	zepto	z
10^3	kilo	k	10^{-3}	milli	m	10^{24}	Yotta	Y	10^{-24}	yocto	y
10^6	Mega	M	10^{-6}	micro	μ						
10^9	Giga	G	10^{-9}	nano	n	unofficial:					
10^{12}	Tera	T	10^{-12}	pico	p	10^{27}	Xenna	X	10^{-27}	xenno	x
10^{15}	Peta	P	10^{-15}	femto	f	10^{33}	Vendeka	V	10^{-33}	vendeko	v

• They form a *complete* system; they cover in a systematic way the complete set of observables of physics. Moreover, they fix the units of measurements for physics and for all other sciences as well.

• They form a *universal* system; they can be used in trade, in industry, in commerce, at home, in education, and in research. They could even be used by other civilisations, if they existed.

• They form a *coherent* system; the product or quotient of two SI units is also a SI unit. This means that in principle, the same abbreviation 'SI' could be used for every SI unit.

The SI units are not the only possible set that fulfills all these requirements, but they form the only existing system doing so.**

* Some of these names are invented (yocto, to sound similar to latin octo 'eight', zepto to sound similar to latin septem, yotta and zetta to resemble them, exa and peta to sound like the greek words of six and five), some are from Danish/Norwegian (atto from atten 'eighteen', femto from femten 'fifteen'), some are from latin (from mille 'thousand', from centum 'hundred', from decem 'ten', from nanus 'dwarf'), some are from italian (from piccolo 'small'), some are greek (micro is from μικρός 'small', deca/deka from δέκα 'ten', hecto from ἑκατόν 'hundred', kilo from χίλιοι 'thousand', mega from μέγας 'large', giga from γίγας 'giant', tera from τέρας 'monster'). Translate: I was caught in such a traffic jam that I needed a microcentury for a picoparsec and that my car's fuel consumption was two tenths of a square millimetre. Challenge

** Most non-SI units still in use in the world are of roman origin: the mile comes from 'milia passum' (used to be one thousand strides of about 1480 mm each; today a nautical mile is a minute of arc), inch comes from 'uncia/onzia' (a twelfth – now of a foot); pound (from pondere 'to weigh') is used as a translation of 'libra' – balance – which is the origin of its abbreviation *lb*; even the habit of counting in dozens instead of tens is roman in origin. These and all other similarly funny units – like the system in which all units start with 'f', and which uses furlongs/fortnights as unit for velocity – are now officially defined as multiples of SI units.

Planck's natural units

Since the exact form of many equations depends on the used system of units, theoretical physicists often use other systems, optimized for producing simple equations. In microscopic physics, the system of *Planck's natural units* is frequently used. They are automatically introduced by setting $c = 1$, $\varepsilon_o = 1/4\pi$, $\mu_o = 4\pi$, $\hbar = 1$, $G = 1$, and $k = 1$ in equations written in SI units, such as those of this book. Planck units are thus defined from combinations of fundamental constants. The units corresponding to the fundamental SI units are given in the table.* The table is also useful for converting equations written in natural units back to SI units; one just substitutes every quantity X by $X/X_{\rm Pl}$.

Table 53 Planck's natural units

Name	definition	value
Basic units		
the Planck length	$l_{\rm Pl} = \sqrt{\hbar G/c^3}$	$= 1.61605(10) \cdot 10^{-35}\,{\rm m}$
the Planck time	$t_{\rm Pl} = \sqrt{\hbar G/c^5}$	$= 5.39056(34) \cdot 10^{-44}\,{\rm s}$
the Planck mass	$m_{\rm Pl} = \sqrt{\hbar c/G}$	$= 21.7671(14)\,\mu{\rm g}$
the Planck current	$I_{\rm Pl} = \sqrt{4\pi\varepsilon_o c^6/G}$	$= 3.47931(22) \cdot 10^{25}\,{\rm A}$
the Planck temperature	$T_{\rm Pl} = \sqrt{\hbar c^5/Gk^2}$	$= 1.41706(91) \cdot 10^{32}\,{\rm K}$
Trivial units		
the Planck velocity	$v_{\rm Pl} = c$	$= 0.3\,{\rm Gm/s}$
the Planck angular momentum	$L_{\rm Pl} = \hbar$	$= 1.1 \cdot 10^{-34}\,{\rm Js}$
the Planck action	$S_{\rm aPl} = \hbar$	$= 1.1 \cdot 10^{-34}\,{\rm Js}$
the Planck entropy	$S_{\rm ePl} = k$	$= 13.8\,{\rm yJ/K}$
Composed units		
the Planck mass density	$\rho_{\rm Pl} = c^5/G^2\hbar$	$= 5.2 \cdot 10^{96}\,{\rm kg/m^3}$
the Planck energy	$E_{\rm Pl} = \sqrt{\hbar c^5/G}$	$= 2.0\,{\rm GJ} = 1.2 \cdot 10^{28}\,{\rm eV}$
the Planck momentum	$p_{\rm Pl} = \sqrt{\hbar c^3/G}$	$= 6.5\,{\rm Nm}$
the Planck force	$F_{\rm Pl} = c^4/G$	$= 1.2 \cdot 10^{44}\,{\rm N}$
the Planck power	$P_{\rm Pl} = c^5/G$	$= 3.6 \cdot 10^{52}\,{\rm W}$
the Planck acceleration	$a_{\rm Pl} = \sqrt{c^7/\hbar G}$	$= 5.6 \cdot 10^{51}\,{\rm m/s^2}$
the Planck frequency	$f_{\rm Pl} = \sqrt{c^5/\hbar G}$	$= 1.9 \cdot 10^{43}\,{\rm Hz}$
the Planck electric charge	$q_{\rm Pl} = \sqrt{4\pi\varepsilon_o c\hbar}$	$= 1.9\,{\rm aC} = 11.7\,{\rm e}$
the Planck voltage	$U_{\rm Pl} = \sqrt{c^4/4\pi\varepsilon_o G}$	$= 1.0 \cdot 10^{27}\,{\rm V}$
the Planck resistance	$R_{\rm Pl} = 1/4\pi\varepsilon_o c$	$= 30.0\,\Omega$

* The natural units $x_{\rm Pl}$ given here are those commonly used today, i.e. those defined using the constant $\hbar$, and not, as Planck originally did, by using the constant $h = 2\pi\hbar$. A similar, additional freedom of choice arises for the electromagnetic units, which can be defined with other factors than 4π in the expressions; for example, using $4\pi\alpha$, with the *fine structure constant* α, one gets $q_{\rm Pl} = e$. For the explanation of the numbers between brackets, the standard deviations, see below, on page 731.

Name	definition	value
the Planck capacitance	$C_{\text{Pl}} = 4\pi\varepsilon_o\sqrt{\hbar G/c^3}$	$= 1.8\cdot 10^{-45}\,\text{F}$
the Planck inductance	$L_{\text{Pl}} = (1/4\pi\varepsilon_o)\sqrt{\hbar G/c^7}$	$= 1.6\cdot 10^{-42}\,\text{H}$
the Planck electric field	$E_{\text{Pl}} = \sqrt{c^7/4\pi\varepsilon_o\hbar G^2}$	$= 6.5\cdot 10^{61}\,\text{V/m}$
the Planck magnetic flux density	$B_{\text{Pl}} = \sqrt{c^5/4\pi\varepsilon_o\hbar G^2}$	$= 2.2\cdot 10^{53}\,\text{T}$

The natural units are important for another reason: whenever in daily life, one says of a quantity that it is "infinitely small (or large)", one has to substitute "small (or large) as the corresponding Planck unit". As explained on page 581, this substitution is correct because almost all Planck units provide, within a factor of order one, the extreme value for the corresponding observable. Exceptions are those quantities for which many particle systems can exceed single particle limits, such as mass and electrical resistance.

Other unit systems

In fundamental theoretical physics another system is also common. One aim of research being the calculation of the strength of all interactions, setting the gravitational constant G to unity, as is done when using Planck units, makes this aim more difficult to express in equations. Therefore one often sets only $c = \hbar = k = 1$ and $\mu_o = 1/\varepsilon_o = 4\pi$,* leaving only the gravitational constant G in the equations. In this system, only one fundamental unit exists, but its choice is still free.

Often a standard length is chosen as fundamental unit, length being the archetype of a measured quantity. For the most important physical observables one then gets the relations

$$\begin{aligned}
[l] &= 1/[E] = [t] = [C] = [L] \quad ,\\
1/[l] &= [E] = [m] = [p] = [a] = [f] = [I] = [U] = [T] \quad ,\\
[l]^2 &= 1/[E]^2 = [G] = [P] = 1/[B] = 1/[E_{\text{el.}}] \quad \text{and}\\
1 &= [v] = [q] = [e] = [R] = [S_{\text{action}}] = [S_{\text{entropy}}] = \hbar = c = k = [\alpha]
\end{aligned}$$

where we used the usual convention to write $[x]$ for the dimension or unit of quantity x. Using the same unit for speed and electric resistance is not to everybody's taste, however, and therefore electricians do not use this system.**

In many situations, in order to get an impression of the energies needed to observe the effect under study, a standard energy is chosen as fundamental unit. In particle physics the common energy unit is the *electron Volt* (eV), defined as the kinetic energy acquired by

* Other definitions for the proportionality constants in electrodynamics lead to the gaussian unit system often used in theoretical calculations, the Heaviside-Lorentz unit system, the electrostatic unit system, and the electromagnetic unit system, among others. For more details, see the standard text by J.D. JACKSON, *Classical electrodynamics*, 3rd edition, Wiley, 1998,

** In general relativity still another system is used sometimes, in which the *Schwarzschild radius* defined as $r_S = 2Gm/c^2$ is used to measure masses, by setting $c = G = 1$. In this case, in opposition to above, mass and length have the same dimension, and $\hbar$ has dimension of an area.

The web page http://www.chemie.fu-berlin.de/chemistry/general/units-en.html allows to convert various units into each other.

an electron when accelerated by an electrical potential difference of 1 Volt ("proton Volt" would be a better name). One has the relation $1\,\mathrm{eV}=1.6\cdot10^{-19}\,\mathrm{J}$, or roughly

$$1\,\mathrm{eV} \approx \tfrac{1}{6}\,\mathrm{aJ}$$

which is easily remembered. Together with the simplification $c = \hbar = 1$ one gets the value $G = 6.9\cdot10^{-57}\,\mathrm{eV}^{-2}$. The unit eV is then also used for mass, momentum, tempera- Challenge ture, frequency, time and length, with the respective correspondences $1\,\mathrm{eV} \mathrel{\widehat{=}} 1.8\cdot10^{-36}\,\mathrm{kg} \mathrel{\widehat{=}} 5.4\cdot10^{-28}\,\mathrm{Nm} \mathrel{\widehat{=}} 242\,\mathrm{THz} \mathrel{\widehat{=}} 11.6\,\mathrm{kK}$ and $1\,\mathrm{eV}^{-1} \mathrel{\widehat{=}} 4.1\,\mathrm{fs} \mathrel{\widehat{=}} 1.2\,\mu\mathrm{m}$.

To get some feeling for the unit eV, the following relations are useful. Room temperature, usually taken as 20 °C or 293 K, corresponds to a kinetic energy per particle of 0.25 eV. The Ref. 3 highest particle energy measured so far is a cosmic ray of energy of $3\cdot10^{20}$ eV. Down here on the earth, an accelerator with an energy of about 230 GeV for electrons and one with an energy of 10 TeV for protons will be built. Both are owned by CERN in Genève and have a circumference of 27 km.

The lowest temperature measured up to now is 280 pK for Rhodium nuclei in a special Ref. 4 cooling system. The interior of that cryostat is possibly the coolest point in the whole universe. At the same time, the kinetic energy per particle corresponding to that temperature is also the smallest ever measured; it corresponds to 24 feV or $3.8\cdot10^{-33}$ J. For isolated particles, the record seems to be for neutrons: kinetic energies as low as 10^{-7} eV have been achieved, corresponding to De Broglie wavelengths of 60 nm.

Curiosities

Now a few facts to put some more life into the idea of unit.

• Are you confused by the candela? The definition simply says that $683\,\mathrm{cd} = 683\,\mathrm{lm/sr}$ correspond to $1\,\mathrm{W/sr}$. The candela is thus a unit for light power per angle, except that it is corrected for the eye's sensitivity: the candela measures only *visible* power per angle. Similarly, $683\,\mathrm{lm} = 683\,\mathrm{cd}\cdot\mathrm{sr}$ correspond to 1 W, i.e. both the lumen and the watt measure power, or energy flux, except that the lumen measures only the *visible* part of the power. In english quantity names, the change is expressed by substituting 'radiant' by 'luminous'; e.g. the Watt measures *radiant* flux, whereas the lumen measure *luminous* flux.

The factor 683 is historical.A usual candle really emits a luminous intensity of about a candela. Therefore, at night, one can see a candle up to a distance of one or two dozen Challenge kilometers. For example, a 100 W incandescent light bulb produces 1700 lm, and the brightest light emitting diodes about 5 lm.

The *irradiance* of sunlight is about $1300\,\mathrm{W/m^2}$ on a sunny day; the *illuminance* is $120\,\mathrm{klm/m^2}$ or $170\,\mathrm{W/m^2}$, reflecting the fact that most energy radiated from the sun to the earth is outside the visible spectrum.

• The highest achieved light intensities are in excess of $10^{18}\,\mathrm{W/m^2}$, more than 15 orders of magnitude higher than the intensity of sunlight, and are achieved by tight focusing of pulsed lasers. The electric fields in such light pulses is of the same order of the field inside Ref. 5 atoms; such a beam ionizes all matter it encounters.

• The Planck length is roughly the de Broglie wavelength $\lambda_\mathrm{B} = h/mv$ of a man walk- Ref. 6 ing comfortably ($m = 80\,\mathrm{kg}$, $v = 0.5\,\mathrm{m/s}$); this motion is therefore aptly called the "Planck stroll."

• The Planck mass is equal to the mass of about 10^{19} protons. This is roughly the mass of a human embryo at about ten days of age.

• The second does not correspond to 1/86400th of the day any more (it did so in the year 1900); the earth now takes about 86400.002 s for a rotation, so that regularly the *International Earth Rotation Service* introduces a leap second to ensure that the sun is at the highest point in the sky at 12.00 o'clock sharp.* The time so defined is called *Universal Time Coordinate*. The velocity of rotation of the earth also changes irregularly from day to day due to the weather, its average changes from winter to summer due to the change in polar ice caps, and in addition that average decreases over time, due to the friction produced by the tides. The rate of insertion of leap seconds is therefore faster than every 500 days, and not completely constant in time.

• The most precisely measured quantities in nature are the frequency of certain millisecond pulsars,** the frequency of certain narrow atomic transitions and the Rydberg constant of *atomic* hydrogen, which can all be measured as exactly as the second is defined. At present, this gives about 14 digits of precision.

• The most precise clock ever built, using microwaves, had a stability of 10^{-16} during a running time of 500 s. For longer time periods, the record in 1997 was about 10^{-15}; but the area of 10^{-17} seems within technological reach. The precision of clocks is limited for short measuring times by noise, and for long measuring times by drifts, i.e. by systematic effects. The region of highest stability depends on the clock type and lies usually between 1 ms (optical clocks) and 5000 s for masers. Pulsars are the only clock for which this region is not known yet; it lies at more than 20 years, which is the time elapsed since their discovery. Ref. 7 Ref. 8

• The shortest times measured are the life times of certain "elementary" particles; in particular, the D meson was measured to live less than 10^{-22} s. Such times are measured in a bubble chamber, were the track is photographed. Can you estimate how long the track is? (Watch out – if your result cannot be observed with an optical microscope, you made a mistake in your calculation). Ref. 9 Challenge

• The longest measured times are the lifetimes of certain radioisotopes, over 10^{15} years, and the lower limit on of certain proton decays, over 10^{32} years. These times are thus much larger than the age of the universe, estimated to be twelve thousand million years. Ref. 10

• The least precisely measured fundamental quantities are the gravitational constant G and the strong coupling constant α_s. Other, even less precisely known quantities, are the age of the universe and its density (see the astrophysical table below). See page 733

• Variations of quantities are often much easier to measure than their values. For example, in gravitational wave detectors, the sensitivity achieved in 1992 was $\Delta l/l = 3 \cdot 10^{-19}$ for lengths of the order of 1 m. In other words, for a block of about a cubic metre of metal it is possible to measure length changes about 3000 times smaller than a proton radius. These set-ups are now being superseded by ring interferometers. Ring interferometers measuring frequency differences of 10^{-21} have already been built, and are still being improved. Ref. 11 Ref. 12

* Their web site at http://hpiers.obspm.fr gives more information on the details of these insertions, as does http://maia.usno.navy.mil, one of the few useful military web sites. See also http://www.bipm.fr, the site of the BIPM.

** An overview of this fascinating work is given by J.H. Taylor, *Pulsar timing and relativistic gravity*, Philosophical Transactions of the Royal Society, London A **341**, pp. 117–134, 1992.

• The swedish astronomer Anders Celsius (1701–1744) originally set the freezing point at 100 degrees and the boiling point of water at 0 degrees. But the switch to todays scale Ref. 13 was not successful. With the official definition of the Kelvin and the degree Celsius, at the standard pressure of 1013.25 Pa, water boils at 99.974 °C. Can you explain why it is not Challenge 100 °C any more?

• The size of SI units is adapted to humans: heartbeat, human size, human weight, human temperature, human substance, etc. In a somewhat unexpected way they realize the saying by Protagoras, 25 centuries ago: "Man is the measure of all things."

• It is well-known that the french philosopher Voltaire, after meeting Newton, publicized the now famous story that the connection between the fall of objects and the motion of the moon was discovered by Newton when he saw an apple falling from a tree. More than a century later, just before the french revolution, a committee of scientists decided to take as unit of force precisely the force exerted by gravity on a *standard apple*, and name it after the english scientist. After extensive study, it was found that the mass of the standard apple was 101,9716 g; its weight was called 1 newton. Since then, in the museum in Sèvres near Paris, one can observe the standard metre, the standard kilogram, and the standard apple.*

Precision and accuracy of measurements

As explained on page 143, *precision* measures how well a result is reproduced when the measurement is repeated; *accuracy* is the degree to which a measurement corresponds to the actual value. Lack of precision is due to accidental or *random errors*; they are best measured by the *standard deviation*, usually abbreviated σ; it is defined through

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (620)$$

where $\bar{x}$ is the average of the measurements x_i. (Can you imagine why $n-1$ is used in the Challenge formula instead of n?) By the way, for a gaussian distribution, $2.35\,\sigma$ is the full width at half maximum.

Lack of accuracy is due to *systematic errors*; usually they can only be estimated. This estimate is often added to the random errors to produce a *total* experimental error, sometimes Ref. 15 also called *total uncertainty*.

The following tables give the values of the most important physical constants and particle properties in SI units and in a few other common units, as published in the standard Ref. 16 references. The values are the world average of the best measurements up to December 1998. As usual, experimental errors, including both random and estimated systematic errors, are expressed by giving the one standard deviation uncertainty in the last digits; e.g. 0.31(6) means 0.31 ± 0.06. In fact, behind each of the numbers in the following tables there is a long story which would be worth telling, but for which there is not enough room here.**

* It is not a joke however, that owners of several apple trees in Britain and in the US claim descendance, by rerooting, from the original tree under which Newton had his insight. DNA tests have even been performed to decide if all these derive from the same tree, with the result that the tree at MIT, in contrast to the british ones, Ref. 14 is a fake – of course.

** Some of them can be found in the text by N.W. Wise, *The values of precision*, Princeton University Press, 1994. The field of high precision measurements, from which the results on these pages stem, is a very special

What are the limits on accuracy and precision? To put the quality of the following work in perspective, one can start by noting that there is no way, even in principle, to measure a quantity to a *precision* higher than about 61 digits, because $\Delta t/t \approx l_{\rm Pl}/d_{\rm horizon} = 10^{-61}$. In the third part of our escalation, studies of clocks and meter bars will reduce this limit, taking into account the average lifetime of a human being, to about 35 digits. See page 612

But it is not difficult to deduce more stringent limits. In practice, any machine cannot measure quantities with a better precision than given by the diameter of the earth divided by the smallest lengths measured, about 10^{-19} m; that makes about 26 digits. Using a more realistic limit of a 1000 m sized machine implies a limit of 22 digits. If, as predicted above, time measurements really achieve 17 digits of precision, then they are nearing the practical limit, because apart from size, there is an additional practical restriction: cost. Often an additional digit in measurement precision means an additional digit in equipment cost.

Basic physical constants

In principle, all experimental measurements of matter properties, be it the colour, the density, or the elastic properties, can be predicted using the values of the following constants, with help of quantum mechanics; more exactly, one needs the equations of the standard model of high energy physics. Ref. 16 See page 551

Table 54 Basic constants

Quantity	name	value in SI units	uncertainty
vacuum speed of light[a]	c	$299\,792\,458\,\mathrm{m/s}$	0
vacuum number of space-time dimensions		$3+1$	down to 10^{-19} m, up to 10^{26} m
vacuum permeability[a]	μ_o	$4\pi \cdot 10^{-7}\,\mathrm{H/m}$ $= 1.256\,637\,061\,435\,917\,295\,385\ldots\,\mu\mathrm{H/m}$	0
vacuum permittivity[a]	$\varepsilon_o = 1/\mu_o c^2$	$8.854\,187\,817\,620\ldots\,\mathrm{pF/m}$	0
Planck constant	h	$6.626\,068\,76(52) \cdot 10^{-34}\,\mathrm{Js}$	$7.8 \cdot 10^{-8}$
reduced Planck constant	$\hbar$	$1.054\,571\,596(82) \cdot 10^{-34}\,\mathrm{Js}$	$7.8 \cdot 10^{-8}$
positron charge	e	$0.160\,217\,6462(63)\,\mathrm{aC}$	$3.9 \cdot 10^{-8}$
Boltzmann constant	k	$1.380\,6503(24) \cdot 10^{-23}\,\mathrm{J/K}$	$1.7 \cdot 10^{-6}$
gravitational constant	G	$6.673(10) \cdot 10^{-11}\,\mathrm{Nm^2/kg^2}$	$1.5 \cdot 10^{-3}$
gravitational coupling constant	$\kappa = 8\pi G/c^4$	$2.076(3) \cdot 10^{-43}\,\mathrm{s^2/kg\,m}$	$1.5 \cdot 10^{-3}$
fine structure constant,[b]	$\alpha = \frac{e^2}{4\pi\varepsilon_o \hbar c}$	$1/137.035\,999\,76(50)$	$3.7 \cdot 10^{-9}$
e.m. coupling constant	$= g_{\rm em}(m_e^2 c^2)$	$= 0.007\,297\,352\,533(27)$	$3.7 \cdot 10^{-9}$
Fermi coupling constant,[b]	$G_{\rm F}/(\hbar c)^3$	$1.166\,39(1) \cdot 10^{-5}\,\mathrm{GeV^{-2}}$	$8.6 \cdot 10^{-6}$
weak coupling constant	$\alpha_{\rm w}(M_Z) = g_{\rm w}^2/4\pi$	$1/30.1(3)$	
weak mixing angle	$\sin^2\theta_{\rm W}(\overline{MS})$	$0.23124(24)$	$2.2 \cdot 10^{-3}$
strong coupling constant[b]	$\alpha_{\rm s}(M_Z) = g^2 s/4\pi$	$0.118(3)$	$25 \cdot 10^{-3}$

a. Defining constant.

b. All coupling constants depend on the four-momentum transfer, as explained in the section on renormalization. *Fine structure constant* is the traditional name of the electromagnetic coupling See page 510

world. A beautiful introduction to it is *Near Zero: Frontiers of Physics*, edited by J.D. FAIRBANKS, B.S. DEAVER, C.W. EVERITT & P.F. MICHAELSON, Freeman, 1988.

constant g_{em} in the case of a four momentum transfer of $Q^2 = m_e^2 c^2$, which is the smallest one possible. At higher momentum transfers it has larger values, e.g. $g_{\text{em}}(Q^2 = M_W^2 c^2) \approx 1/128$. The strong coupling constant has higher values at lower momentum transfers; e.g. one has $\alpha_s(34\,\text{GeV}) = 0.14(2)$.

Why do all these constants have the values they have? The answer depends on the constant. For constants without unit, such as the fine structure constant, the question about the value is central to its understanding. In these cases, the escalation provides the answer. For any constant having a unit, of such as the quantum of action $\hbar$, the numerical value has no intrinsic meaning. The numerical value is $1.054 \cdot 10^{-34}\,\text{Js}$ only because the joule and the second were defined in a particular way, leading to this value.

Note that the question why the value of a constant, even one with units, is not larger or smaller always means to understand the origin of some dimensionless number. For example, $\hbar$, G and c are not smaller or larger because the everyday world, in basic units, is of the dimensions we observe. The same happens when one asks about the size of atoms, people, trees, and stars, or about the duration of molecular and atomic processes, or about the mass of nuclei and mountains.

With the basic constants one deduces the following, frequently used constants.

Table 55 Derived constants

Quantity	name	value in SI units	uncertainty
Vacuum wave resistance	$Z_o = \sqrt{\mu_o/\varepsilon_o}$	$376.730\,313\,461\,77...\,\Omega$	0
Avogadro's number	N_A	$6.022\,141\,99(47) \cdot 10^{23}$	$7.9 \cdot 10^{-8}$
Rydberg constant[a]	$R_\infty = m_e c \alpha^2/2h$	$10\,973\,731.568\,549(83)\,\text{m}^{-1}$	$7.6 \cdot 10^{-12}$
mag. flux quantum	$\varphi_o = h/2e$	$2.067\,833\,636(81)\,\text{pWb}$	$3.9 \cdot 10^{-8}$
Josephson freq. ratio	$2e/h$	$483.597\,898(19)\,\text{THz/V}$	$3.9 \cdot 10^{-8}$
von Klitzing constant	$h/e^2 = \mu_o c/2\alpha$	$25\,812.807\,572(95)\,\Omega$	$3.7 \cdot 10^{-9}$
Bohr magneton	$\mu_B = e\hbar/2m_e$	$9.274\,008\,99(37) \cdot 10^{-24}\,\text{J/T}$	$4.0 \cdot 10^{-8}$
classical electron radius	$r_e = e^2/4\pi\varepsilon_o m_e c^2$	$2.817\,940\,285(31)\,\text{fm}$	$1.1 \cdot 10^{-8}$
Compton wavelength	$\lambda_c = h/m_e c$	$2.426\,310\,215(18)\,\text{pm}$	$7.3 \cdot 10^{-9}$
of the electron	$\bar{\lambda}_c = \hbar/m_e c = r_e/\alpha$	$0.386\,159\,2642(28)\,\text{pm}$	$7.3 \cdot 10^{-9}$
Bohr radius[a]	$a_\infty = r_e/\alpha^2$	$52.917\,720\,83(19)\,\text{pm}$	$3.7 \cdot 10^{-9}$
cyclotron frequency of the electron	$f_c/B = e/2\pi m_e$	$27.992\,4925(11)\,\text{GHz/T}$	$4.0 \cdot 10^{-8}$
nuclear magneton	$\mu_N = e\hbar/2m_p$	$5.050\,783\,17(20) \cdot 10^{-27}\,\text{J/T}$	$4.0 \cdot 10^{-8}$
proton electron mass ratio	m_p/m_e	$1\,836.152\,6675(39)$	$2.1 \cdot 10^{-9}$
Stephan-Boltzmann constant	$\sigma = \pi^2 k^4/60\hbar^3 c^2$	$5.670\,400(40) \cdot 10^{-8}\,\text{W/m}^2\text{K}^4$	$7.0 \cdot 10^{-6}$
Wien displacement law constant	$b = \lambda_{\max} T$	$2.897\,7686(51)\,\text{mmK}$	$1.7 \cdot 10^{-6}$
bits to entropy conv. const.		$10^{23}\,\text{bit} = 0.9569945(17)\,\text{J/K}$	
TNT energy content		$4.2\,\text{GJ/ton}$	

a. For infinite mass of the nucleus.

Some properties of the universe as a whole are listed in the following.

Table 56 Astrophysical constants

Quantity	name	value
gravitational constant	G	$6.67259(85) \cdot 10^{-11}\,\mathrm{m^3/kg\,s^2}$
cosmological constant	Λ	ca. $1 \cdot 10^{-52}\,\mathrm{m^{-2}}$
tropical year 1900[a]	a	$31\,556\,925.974\,7\,\mathrm{s}$
tropical year 1994	a	$31\,556\,925.2\,\mathrm{s}$
mean sidereal day	d	$23^h 56' 4.09053''$
astronomical unit[b]	AU	$149\,597\,870.691(30)\,\mathrm{km}$
light year	al	$9.460\,528\,173\ldots\,\mathrm{Pm}$
parsec	pc	$30.856\,775\,806\,\mathrm{Pm} = 3.261\,634\,\mathrm{al}$
age of the universe[c]	t_o	$> 3.5(4) \cdot 10^{17}\,\mathrm{s}$ or $> 11.5(1.5) \cdot 10^9\,\mathrm{a}$
(from matter, via galaxies and stars, using quantum theory: early 1997 results)		
age of the universe[c]	t_o	$4.7(1.5) \cdot 10^{17}\,\mathrm{s} = 13.5(1.5) \cdot 10^9\,\mathrm{a}$
(from spacetime, via expansion, using general relativity)		
universe's horizon's dist.[c]	$d_o = 3ct_o$	$5.2(1.4) \cdot 10^{26}\,\mathrm{m} = 13.8(4.5)\,\mathrm{Gpc}$
universe's topology		unknown
number of space dimensions		3
Hubble parameter[c]	H_o	$2.2(1.0) \cdot 10^{-18}\,\mathrm{s^{-1}} = 0.7(3) \cdot 10^{-10}\,\mathrm{a^{-1}}$
	$= h_o \cdot 100\,\mathrm{km/sMpc} = h_o \cdot 1.0227 \cdot 10^{-10}\,\mathrm{a^{-1}}$	
reduced Hubble par.[c]	h_o	$0.59 < h_o < 0.7$
critical density of the universe	$\rho_c = 3H_o^2/8\pi G$	$h_o^2 \cdot 1.878\,82(24) \cdot 10^{-26}\,\mathrm{kg/m^3}$
density parameter[c]	$\Omega_{Mo} = \rho_o/\rho_c$	ca. 0.3
luminous matter density		ca. $2 \cdot 10^{-28}\,\mathrm{kg/m^3}$
stars in the universe	n_s	$10^{22\pm1}$
baryons in the universe	n_b	$10^{81\pm1}$
baryon mass	m_b	$1.7 \cdot 10^{-27}\,\mathrm{kg}$
baryon number density		1 to $6\,/\mathrm{m^3}$
photons in the universe	n_γ	10^{89}
photon energy density	$\rho_\gamma = \pi^2 k^4/15T_o^4$	$4.6 \cdot 10^{-31}\,\mathrm{kg/m^3}$
photon number density	$\rho_\gamma = \pi^2 k^4/15T_o^4$	$400\,/\mathrm{cm^3}(T_o/2.7\,\mathrm{K})^3$, at present $410.89/\mathrm{cm^3}$
background temperature[d]	T_o	$2.726(5)\,\mathrm{K}$
Planck length	$l_{Pl} = \sqrt{\hbar G/c^3}$	$1.62 \cdot 10^{-35}\,\mathrm{m}$
Planck time	$t_{Pl} = \sqrt{\hbar G/c^5}$	$5.39 \cdot 10^{-44}\,\mathrm{s}$
Planck mass	$m_{Pl} = \sqrt{\hbar c/G}$	$21.8\,\mu\mathrm{g}$
instants in history[c]	t_o/t_{Pl}	$8.7(2.8) \cdot 10^{60}$
space-time points inside the horizon[c]	$N_o = (R_o/l_{Pl})^3 \cdot (t_o/t_{Pl})$	$10^{244\pm1}$
mass inside horizon	M	$10^{54\pm1}\,\mathrm{kg}$

a. Defining constant, from vernal equinox to vernal equinox; it was once used to define the second. (Remember: π seconds is a nanocentury.) The value for 1990 is about 0.7 s less, corresponding to a slowdown of roughly $-0.2\,\mathrm{ms/a}$. There is even an empirical formula available for the change of the length of the year over time. Challenge Ref. 17
b. Average distance earth-sun. The truly amazing precision of 30 m results from time averages of signals sent from Viking orbiters and Mars landers taken over a period of over twenty years.
c. The index *o* indicates present day values.

d. The radiation originated when the universe was between 10^5 to 10^6 years old and about 3000 K hot; the fluctuations ΔT_o which lead to galaxy formation are today of the size of $16 \pm 4\,\mu K = 6(2) \cdot 10^{-6}\, T_o$.

See page 251

Attention: in the third part of this text it is shown that many constants in this table are *not* physically sensible quantities. They have to be taken with lots of grains of salt. More specific constants, all sensible, are given in the following table.

Table 57 Astronomical constants

Quantity	name	value
earth'smass	M	$5.972\,23(8) \cdot 10^{24}$ kg
earth's gravitational length	$l = 2GM/c^2$	$8.870(1)$ mm
earth radius, equatorial[a]	R_{eq}	6378.140 km
earth radius, polar[a]	R_p	6356.7 km
equator pole distance[a]		10001.966 km (average)
earth flattening[a]	e	$1/298.257$
moon's radius	R_{mv}	1738 km in direction of earth
moon's radius	R_{mh}	17.. km in perpendicular direction
moon's mass	M_m	$7.35 \cdot 10^{22}$ kg
moon's mean distance[b]	d_m	384 401 km
moon's perigeon		typically 363 Mm, hist. minimum 359 861 km
moon's apogeon		typically 404 Mm, hist. maximum 406 720 km
moon's angular size[c] (average)		$0.5181° = 31.08'$, min. 0.49°, max. 0.55°
sun's mass	$M_\odot$	$1.988\,43(3) \cdot 10^{30}$ kg
sun's grav. length	$l_\odot = 2GM_\odot/c^2$	2.953 250 08 km
sun's luminosity	$L_\odot$	384.6 YW
solar radius, equatorial	$R_\odot$	$695.98(7)$ Mm
sun's angular size		0.53° average
sun's distance, average	AU	$149\,597\,870.691(30)$ km
solar velocity around centre of galaxy	$v_{\odot g}$	$220(20)$ km/s
solar velocity against cosmic background	$v_{\odot b}$	$370.6(5)$ km/s
distance to galaxy centre		$8.0(5)$ kpc $= 26.1(1.6)$ kal
most distant galaxy	0140+326RD1	$12.2 \cdot 10^9$ al $= 1.2 \cdot 10^{26}$ m, redshift 5.34

a. The shape of the earth is described most precisely with the World Geodetic System. The last edition dates from 1984. For an extensive presentation of its background and its details, see the http://www.eurocontrol.be/projects/eatchip/wgs84/start.html web site.
b. Measured centre to centre. To know the precise position of the moon at a given date, see the http://www.fourmilab.ch/earthview/moon-ap-per.html site, whereas for the planets see http://www.fourmilab.ch/solar/solar.html as well as the other pages on this site.
c. Angles are defined as follows: 1 degree $= 1° = \pi/180$ rad, 1 (first) minute $= 1' = 1°/60$, 1 second (minute) $= 1'' = 1'/60$. The ancient units 'third minute' and 'fourth minute', each 1/60th of the preceding, are not accepted any more. ('Minute' originally means 'very small', as it still does in modern english.)

Useful numbers

π	$3.14159\,26535\,89793\,23846\,26433\,83279\,50288\,41971\,69399\,37510_5$
e	$2.71828\,18284\,59045\,23536\,02874\,71352\,66249\,77572\,47093\,69995_9$
γ	$0.57721\,56649\,01532\,86060\,65120\,90082\,40243\,10421\,59335\,93992_3$
$\ln 2$	$0.69314\,71805\,59945\,30941\,72321\,21458\,17656\,80755\,00134\,36025_5$
$\ln 10$	$2.30258\,50929\,94045\,68401\,79914\,54684\,36420\,76011\,01488\,62877_2$
$\sqrt{10}$	$3.16227\,76601\,68379\,88935\,44432\,71853\,37195\,55139\,32521\,68268_5$

Ref. 18

References

1 *Le Système International d'Unités*, Bureau International des Poids et Mesures, Pavillon de Breteuil, Parc de Saint Cloud, 92310 Sèvres, France. All new developments concerning SI units are published in the journal *Metrologia*, edited by the same body. Showing the slow pace of an old institution, the BIPM was on the internet only in 1998; it is now reachable on its simple site at http://www.bipm.fr. The site of its british equivalent, http://www.npl.co.uk/npl/reference/si_units.html, is much better; it gives many other details as well as the english version of the SI unit definitions. Cited on page 723.

2 The bible in the field of time measurement are the two volumes by J. VANIER & C. AUDOIN, *The quantum physics of atomic frequency standards*, Adam Hilge, 1989. A popular account is ... …, *Splitting the second*, 2000.

The site http://opdaf1.obspm.fr/www/lexique.html gives a glossary of terms used in the field. On length measurements, see ... On mass and atomic mass measurements, see page 168. The precision of mass measurements of solids is limited by such simple effects as the adsorption of water on the weight. Can you estimate what a monolayer of water does on a weight of 1 kg? On electric current measurements, see ... On precision temperature measurements, see page 199. Cited on page 725. Challenge

3 DAVID J. BIRD & al., *Evidence for correlated changes in the spectrum and composition of cosmic rays at extremely high energies*, Physical Review Letters **71**, pp. 3401–3404, 1993. Cited on page 729.

4 PERTTI J. HAKONEN & al., *Nuclear antiferromagnetism in Rhodium metal at positive and negative nanokelvin temperature*, Physical Review Letters **70**, pp. 2818–2821, 1993. See also his article in the Scientific American, January 1994. Cited on page 729.

5 See e.g. K. CODLING & L.J. FRASINSKI, *Coulomb explosion of simple molecules in intense laser fields*, Contemporary Physics **35**, pp. 243–255, 1994. Cited on page 729.

6 A. ZEILINGER, *The Planck stroll*, American Journal of Physics **58**, p. 103, 1990. Cited on page 729.

7 The most precise clock ever built is ... Cited on page 730.

8 J. BERGQUIST, editor, *Proceedings of the fifth symposium on frequency standards and metrology*, World Scientific, 1997. Cited on page 730.

9 About short lifetime measurements, see e.g. the paper on D particle lifetime ... Cited on page 730.

10 About the long life of tantalum 180, see D. BELIC & al., *Photoactivation of* $^{180}\mathrm{Ta}^{\mathrm{m}}$ *and its implications for the nucleosynthesis of nature's rarest naturally occurring isotope*, Physical Review Letters **83**, pp. 5242–5245, 20 december 1999. Cited on page 730.

11 About the detection of gravitational waves, see ... Cited on page 730.

12 See the clear and extensive paper by G.E. STEDMAN, *Ring laser tests of fundamental physics and geophysics*, Reports on progress of physics **60**, pp. 615–688, 1997. Cited on page 730.

13 Following a private communication by Richard Rusby, this is the value of 1997, whereas it was estimated as 99.975 °C in 1989, as reported by Gareth JONES & Richard RUSBY, *Official: water boils at* 99.975 °C, Physics World, pp. 23–24, September 1989, and R.L. RUSBY, *Ironing out the standard scale*, Nature **338**, p. 1169, March 1989. For more on temperature measurements, see page 199. Cited on page 731.

14 See *Newton's apples fall from grace*, New Scientist, p. 5, 6 September 1996. More details can be found in R.G. KEESING, *The history of Newton's apple tree*, Contemporary Physics **39**, pp. 377–391, 1998. Cited on page 731.

15 The various concepts are even the topic of a separate international standard, ISO 5725, with the title *Accuracy and precision of measurement methods and results*. A good introduction is the book with the locomotive hanging out the window as title picture, namely John R. TAYLOR, *An introduction to error analysis: the study of uncertainties in physical measurements*, 2nd edition, University Science Books, Sausalito, 1997. Cited on page 731.

16 P.J. MOHR & B.N. TAYLOR, Reviews of Modern Physics **59**, p. 351, 2000. This is the set of constants resulting from an international adjustment and recommended for international use by the Committee on Data for Science and Technology (CODATA), a body in the International Council of Scientific Unions, which regroups the International Union of Pure and Applied Physics (IUPAP), the International Union of Pure and Applied Chemistry (IUPAC) and many more. The IUPAC has a horrible web site at http://chemistry.rsc.org/rsc/iupac.htm. Cited on page 731, 732.

17 The details are given in the well-known astronomical reference, P. Kenneth SEIDELMANN, *Explanatory Supplement to the Astronomical Almanac*, 1992. Cited on page 734.

18 For information about the number π, as well as about other constants, the web address http://www.cecm.sfu.ca/pi/pi.html provides lots of data and references. It also has a link to the pretty overview paper on http://www.astro.virginia.edu/~eww6n/math/Pi.html and to many other sites on the topic. Simple formulas for π are

$$\pi + 3 = \sum_{n=1}^{\infty} \frac{n\,2^n}{\binom{2n}{n}} \tag{621}$$

or the beautiful formula discovered in 1996 by Bailey, Borwein, and Plouffe

$$\pi = \sum_{n=0}^{\infty} \frac{1}{16n} \left(\frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6} \right) \quad . \tag{622}$$

The site also explains the newly discovered methods to calculate specific digits of π without having to calculate all the preceding ones. By the way, the number of (consecutive) digits known in 1999 was over 206 thousand million, as told in Science News **156**, p. 255, 16 October 1999. They pass all tests for a random string of numbers, as e.g. http://www.ast.univie.ac.at/~wasi/PI/pi_normal.html explains.

Another method to calculate π and other constants was discovered and published by David V. CHUDNOVSKY & Gregory V. CHUDNOVSKY, *The computation of classical constants*, Proc. Natl. Acad. Sci. USA, volume 86, pp. 8178–8182, 1989. The Chudnowsky brothers have built a supercomputer in Gregory's apartment for about 70 000 $, and for many years held the record for the largest number of digits for π. They battle already for decades with Kanada Yasumasa, who holds the record in 2000, calculated on a industrial supercomputer. New formulas to calculate π are still discovered regularly.

For the calculation of Euler's constant γ see also D.W. DeTemple, *A quicker convergence to Euler's constant*, The Mathematical Intelligencer pp. 468–470, May 1993.

Note that little is known about properties of numbers; e.g. it is still not known whether $\pi + e$ is a rational number or not! (It is believed that not.) Want to become a mathematician? Challenge

Cited on page 736.

Typeset in April 2001

Appendix C Particle Properties

The following table contains the *complete list of properties* of all elementary particles. The future should not change it much.* The header of this table therefore lists the *complete* set of properties, after the quantum number of colour is added, which characterize any particle. This list thus allows a complete characterization of the intrinsic properties of any moving entity, be it an object or an image.

Table 58 Elementary particle properties

Particle name and symbol	mass m	lifetime τ or energy width, main decay modes	isospin I, spin J, parity P, charge parity C	charge Q, isospin I, strangeness S, charm C, topness T, beauty B	lepton number L, baryon number B, R-parity
elementary radiation (bosons)					
photon γ	$0\ (< 6 \cdot 10^{-16}\,\mathrm{eV}/c^2)$	stable	$I(J^{PC}) = 0,1(1^{--})$	000000	$0,0,1$
$W^{\pm}$	$80.75(64)\,\mathrm{GeV}/c^2$	$2.06(6)\,\mathrm{GeV}$ 67.8(1.0)% hadrons, 32.1(2,0)% $l^{+}\nu$	$J=1$	±100000	$0,0,1$
Z	$91.187(7)\,\mathrm{GeV}/c^2$	$2.65(1) \cdot 10^{-25}\,\mathrm{s}$ 69.90(15)% hadrons	$J=1$	000000	$0,0,1$
gluon	0	stable	$I(J^{P}) = 0(1^{-})$	000000	$0,0,1$
elementary matter (fermions): leptons					
electron e	$9.109\,381\,88(72) \cdot 10^{-31}\,\mathrm{kg}$	$> 13 \cdot 10^{30}\,\mathrm{s}$	$J=\frac{1}{2}$	-100000	$1,0,1$
	$= 81.871\,0414(64)\,\mathrm{pJ}/c^2 = 0.510\,998\,902(21)\,\mathrm{MeV}/c^2 = 0.000\,548\,579\,9110(12)\,\mathrm{u}$				
	gyromagnetic ratio $g = \mu_e/\mu_B = -1.001\,159\,652\,1869(41)$				
	electric dipole moment $d = (-0.3 \pm 0.8) \cdot 10^{-29} e\,\mathrm{m}$				

* The official reference for all this data, worth a look by every physicist, is the massive collection by the particle data group, namely C. Caso & al., The European Physical Journal C **3**, p. 1, 1998. This review is published about every two years with updated data in one of the large journals on elementary particle physics. The web site http://pdg.web.cern.ch/pdg contains the most recent information. For stable particles, the official reference are the CODATA values, published by P.J. Mohr & B.N. Taylor, Reviews of Modern Physics **59**, p. 351, 2000.

Particle name and symbol	mass m	lifetime τ or energy width, main decay modes	isospin I, spin J, parity P, charge parity C	charge Q, isospin I, strangeness S, charm C, topness T, beauty B	lepton number L, baryon number B, R-parity
muon μ	$0.188\,353\,109(16)\,\mathrm{yg}$	$2.197\,03(4)\,\mu\mathrm{s}$ 99% $e^- \bar{\nu}_e \nu_\mu$	$J = \frac{1}{2}$	-100000	$1, 0, 1$
	$= 105.658\,3568(52)\,\mathrm{MeV}/c^2$ $= 0.113\,428\,9168(34)\,\mathrm{u}$				
	gyromagnetic ratio $g = \mu_\mu/(e\hbar/2m_\mu) = -1.001\,165\,916\,02(64)$				
	electric dipole moment $d = (3.7 \pm 3.4) \cdot 10^{-22} e\,\mathrm{m}$				
tau τ	$1.777\,05(29)\,\mathrm{GeV}/c^2$	$290.0(1.2)\,\mathrm{fs}$	$J = \frac{1}{2}$	-100000	$1, 0, 1$
el. neutrino ν_e	$< 7.2\,\mathrm{eV}/c^2$		$J = \frac{1}{2}$		$1, 0, 1$
muon neutrino ν_μ	$< 0.17\,\mathrm{MeV}/c^2$		$J = \frac{1}{2}$		$1, 0, 1$
tau neutrino ν_τ	$< 24\,\mathrm{MeV}/c^2$		$J = \frac{1}{2}$		$1, 0, 1$
elementary matter (fermions): quarks					
up quark u	$1.5 - 5\,\mathrm{MeV}/c^2$	see proton	$I(J^P) = \frac{1}{2}(\frac{1}{2}^+)$	$-\frac{1}{3} - \frac{1}{2}0000$	$0, 1/3, 1$
down quark d	$3 - 9\,\mathrm{MeV}/c^2$	see proton	$I(J^P) = \frac{1}{2}(\frac{1}{2}^+)$	$\frac{2}{3}\frac{1}{2}0000$	$0, 1/3, 1$
strange quark s	$60 - 170\,\mathrm{MeV}/c^2$		$I(J^P) = 0(\frac{1}{2}^+)$	$-\frac{1}{3}0 - 1000$	$0, 1/3, 1$
charm quark c	$1.25(15)\,\mathrm{GeV}/c^2$		$I(J^P) = 0(\frac{1}{2}^+)$	$\frac{2}{3}00 + 100$	$0, 1/3, 1$
bottom quark b	$4.25(15)\,\mathrm{GeV}/c^2$	$\tau = 1.33(11)\,\mathrm{ps}$	$I(J^P) = 0(\frac{1}{2}^+)$	$-\frac{1}{3}000 - 10$	$0, 1/3, 1$
top quark t	$173.8(5.2)\,\mathrm{GeV}/c^2$		$I(J^P) = 0(\frac{1}{2}^+)$	$\frac{2}{3}0000 + 1$	$0, 1/3, 1$
hypothetical, maybe elementary (boson)					
Higgs* H	$135(20)\,\mathrm{GeV}/c^2$		$J = 0$		
hypothetical elementary radiation (bosons)					
Selectron*			$J = 0$		$R = -1$
Smuon*			$J = 0$		$R = -1$
Stauon*			$J = 0$		$R = -1$
Sneutrinos*			$J = 0$		$R = -1$
Squark*			$J = 0$		$R = -1$
hypothetical elementary matter (fermions)					
Higgsino(s)*			$J = \frac{1}{2}$		$R = -1$
Wino* (a chargino)			$J = \frac{1}{2}$		$R = -1$
Zino* (a neutralino)			$J = \frac{1}{2}$		$R = -1$
Photino*			$J = \frac{1}{2}$		$R = -1$
Gluino*			$J = \frac{1}{2}$		$R = -1$

Notes:
* Presently a hypothetical particle.
- To keep the table short, the header does not explicitly mention *colour*, the charge of the strong interactions. It has to be added to the list of object properties.
- The electron has a radius of less than 10^{-22} m. It is possible to store single electrons in traps for many months. Ref. 1 Ref. 2
- See also the table of SI-prefixes on page 725. About the eV/c^2 mass unit, see page 728.

- Quantum numbers containing the word ‘parity’ are multiplicative; all others are additive.
- Time parity T, better called motion inversion parity, is equal to CP.
- The isospin I or I_Z is defined only for up and down quarks and their composites, such as the proton and the neutron. In the literature one also sees references to the so-called G-parity, defined as $G = (-1)^{IC}$.
- R-parity is a quantum number important in supersymmetric theories; it is related to the lepton number L, the baryon number B and the spin J through the definition $R = (-1)^{3B+L+2J}$. All particles from the standard model are R-even, whereas their superpartners are odd.
- The sign of the quantum numbers I_Z, S, C, B, T can be defined in different ways. In the standard assignment shown here, the sign of each of the nonvanishing quantum numbers is given by the sign of the charge of the corresponding quark.
- There is a difference between the half-life $t_{1/2}$ and the the lifetime τ of a particle; the *half-life* is is given by $t_{1/2} = \tau \ln 2$, where $\ln 2 \approx 0.693\,147\,18$, and is thus shorter than the lifetime. The *energy width* Γ of a particle is related to its lifetime τ by the uncertainty relation $\Gamma\tau = \hbar$.
- The unified *atomic mass unit* is defined as (1/12) of the mass of an Carbon atom of the isotope ^{12}C at rest and in its ground state. One has $1\,\mathrm{u} = \frac{1}{12} m(^{12}\mathrm{C}) = 1.6605402(10)\,\mathrm{yg}$.
- See page 541 for the precise definition and meaning of the quark masses.
- The electric polarizability is defined on page 339; it is predicted to vanish for all elementary particles.

Using the table of elementary particle properties, and using the standard model and the fundamental constants, one can in principle deduce *all* properties of composite matter and radiation, including all those encountered in everyday life. (Can you explain how the size of an object follows from them?) In a sense, this first table contains in it the complete results of the study of matter and radiation, i.e. all of quantum theory.

Challenge

A few examples of composites are grouped in the following table. It contains the most important cases, and whenever possible, the first, last, heaviest, lightest, longest and shortest lived example of each class of composites.

Table 59 Properties of selected composites

Composite	mass m, quantum numbers	lifetime τ, main decay modes	size
mesons (hadrons, bosons) out of the over 130 known types			
pion $\pi^0 (\mathrm{u\bar{u}} - \mathrm{d\bar{d}})/\sqrt{2}$	$134.9764(6)\,\mathrm{MeV}/c^2$	$84(6)\,\mathrm{as}$, 2γ 98.798(32)%	$\sim 1\,\mathrm{fm}$
	$I^G(J^{PC}) = 1^-(0^{-+})$, $S = C = B = 0$		
pion $\pi^+ (\mathrm{u\bar{d}})$	$139.56995(35)\,\mathrm{MeV}/c^2$	$26.030(5)\,\mathrm{ns}$, $\mu^+\nu_\mu$ 99.9877(4)%	$\sim 1\,\mathrm{fm}$
	$I^G(J^P) = 1^-(0^-)$, $S = C = B = 0$		
kaon K_S^o	$m_{K_S^o}$	$89.27(9)\,\mathrm{ps}$	
kaon K_L^o	$m_{K_S^o} + 3.491(9)\,\mu\mathrm{eV}/c^2$	$51.7(4)\,\mathrm{ns}$	
kaon $K^\pm$ $(\mathrm{u\bar{s}}, \mathrm{\bar{u}s})$	$493.677(16)\,\mathrm{MeV}/c^2$	$12.386(24)\,\mathrm{ns}$, $\mu^+\nu_\mu$ 63.51(18)% 21.16(14)% $\pi^+\pi^o$	
kaon K^o $(\mathrm{d\bar{s}})$ (50% K_S, 50% K_L)	$497.672(31)\,\mathrm{MeV}/c^2$	n.a.	$\sim 1\,\mathrm{fm}$
kaons $K^\pm, K^o, K_S^o, K_L^o$	$I(J^P) = \frac{1}{2}(0^-)$, $S = \pm 1$, $B = C = 0$		
baryons (hadrons, fermions) out of the over 100 types known			

Composite	mass m, quantum numbers	lifetime τ, main decay modes	size
proton p or N^+ (uud)	$1.672\,621\,58(13)\,\mathrm{yg}$ $= 1.007\,276\,466\,88(13)\,\mathrm{u}$ $= 938.271\,998(38)\,\mathrm{MeV}/c^2$ $I(J^P) = \frac{1}{2}(\frac{1}{2}^+)$, $S = 0$	$\tau_{\text{total}} > 1.6 \cdot 10^{25}$ a, $\tau(p \to e^+\pi^o) > 5.5 \cdot 10^{32}$ a	$0.89(1)$ fm Ref. 3
	gyromagnetic ratio $\mu_p/\mu_N = 2.792\,847\,337(29)$		
	electric dipole moment $d = (-4 \pm 6) \cdot 10^{-26} e\,\mathrm{m}$		
	electric polarizability $\alpha = 12.1(0.9) \cdot 10^{-4}\,\mathrm{fm}^3$		
	magnetic polarizability $\alpha = 2.1(0.9) \cdot 10^{-4}\,\mathrm{fm}^3$		
neutron n or N^o (udd)	$1.674\,927\,16(13)\,\mathrm{yg}$ $= 1.008\,664\,915\,78(55)\,\mathrm{u} = 939.565\,330(38)\,\mathrm{MeV}/c^2$ $I(J^P) = \frac{1}{2}(\frac{1}{2}^+)$, $S = 0$	$887.0(2.0)\,\mathrm{s}$, $pe^-\bar{\nu}_e$ 100 %	~ 1 fm
	gyromagnetic ratio $\mu_n/\mu_N = -1.913\,042\,72(45)$		
	electric dipole moment $d_n = (-3.3 \pm 4.3) \cdot 10^{-28} e\,\mathrm{m}$		
	electric polarizability $\alpha = 0.98(23) \cdot 10^{-3}\,\mathrm{fm}^3$		
omega Ω^- (sss)	$1672.43(32)\,\mathrm{MeV}/c^2$	$82.2(1.2)$ ps, ΛK^- 67.8(7)%, $\Xi^o\pi^-$ 23.6(7)%	~ 1 fm
	gyromagnetic ratio $\mu_\Omega/\mu_N = -1.94(22)$		
composite radiation: glueballs			
glueball $f_o(1500)$	$1503(11)$ MeV $I^G(J^{PC}) = 0^+(0^{++})$	full width $120(19)$ MeV	~ 1 fm
atoms out of the 115 known elements with over 2000 isotopes Ref. 4			
hydrogen (^{1}H) [smallest, and lightest of all]	$1.007825\,\mathrm{u} = 1.6735\,\mathrm{yg}$		$2 \cdot 0.032$ nm
helium (^{4}He)	$4.00260\,\mathrm{u} = 6.6465\,\mathrm{yg}$		... nm
carbon (^{12}C)	$12\,\mathrm{u} = 19.926\,482(12)\,\mathrm{yg}$		$2 \cdot 0.077$ nm
bismuth $^{209}_{83}$Bi [shortest living and rarest]	209 u	0.1 ps	
tantalum ^{180}Ta [longest living]	180 u	$> 10^{15}$ a Ref. 5	
francium [largest of all]	223 u	...	$2 \cdot 0.24$ nm
atom 112 [heaviest of all]	277 u		
molecules out of the over 10^7 known types			
hydrogen (H_2)	~ 2 u	$> 10^{25}$ years	
water (H_2O)	~ 18 u	$> 10^{25}$ years	
ATP (adenosinetriphosphate)	... zg	$> 10^{10}$ years	ca. 3 nm
human Y chromosome	... ag	$\sim 10^6$ years	ca. 50 mm
other composites			
whale nerve cell	~ 100 g	~ 50 years	20 m
cell (red blood)	... ng	4-100 days	$\sim 10\,\mu$m

Composite	mass m, quantum numbers	lifetime τ, main decay modes	size
cell (sperm)	10 pg	not fecundated: ~ 5 days	length 60 μm, head 3 μm times 5 μm
cell (ovule)	1 μg	fecundated: over 4000 million years	$\sim 120\,\mu$m
cell (E. Coli)	1 pg	4000 million years	body: 2 μm
typical adult human	$35\,\mathrm{kg} < m < 350\,\mathrm{kg}$	$\tau \approx 2.5 \cdot 10^9$ s Ref. 6 ≈ 600 million breaths $\approx 2\,500$ million heartbeats $\lesssim 122$ years, 60% H_2O and dust	~ 1.7 m
largest living thing	10^5 kg	ca. 200 years	~ 1 km

larger composites see table on page 107.

Notes (see also those of the previous table)
- Some nuclei are not yet discovered; in 1999 the known nuclei range from 1 to 118, but 113, 115 and 117 are still missing.
- Neutrons bound in nuclei have a lifetime of at least 10^{20} years.
- The number of existing molecules is several orders of magnitude larger than the number of analyzed and listed molecules.
- The $f_o(1500)$ resonance is now accepted as a glueball by the high energy community, as announced at the HEO conference in Warsaw in 1996.
- The charge parity C is defined only for certain neutral particles, namely those which are different from their antiparticles. For neutral mesons the charge parity is given by $C = (-1)^{L+S}$, where L is the orbital angular momentum.
- P is the parity under space inversion $\vec{r} \rightarrow -\vec{r}$. For mesons, one also has the relation $P = (-1)^{L+1}$, where L is again the orbital angular momentum.
- G parity is only defined for mesons, and is given by $G = (-1)^{L+S+I}$.
- The first *anti-atoms*, made of antielectrons and antiprotons, have been made in January 1996 at CERN in Genève. All properties for anti-matter checked so far are consistent with the predictions. Ref. 7

References

1 This is deduced from the $g-2$ measurements, as explained in his Nobel-prize talk by Hans DEHMELT, *Experiments with an isolated subatomic particle at rest*, Reviews of Modern Physics **62**, pp. 525–530, 1990, and in Hans DEHMELT, *Is the electron a composite particle?*, Hyperfine Interactions **81**, pp. 1–3, 1993. Cited on page 740.

2 G. GABRIELSE, H. DEHMELT & W. KELLS, *Observation of a relativistic, bistable hysteresis in the cyclotron motion of a single electron*, Physical Review Letters **54**, pp. 537–540, 1985. Cited on page 740.

3 The proton charge radius was determined by measuring the frequency of light emitted by hydrogen atoms to high precision by Th. UDEM, A. HUBER, B. GROSS, J. REICHERT, M. PREVEDELLI, M. WEITZ & T.W. HAUSCH, *Phase-coherent measurement of the hydrogen 1S-2S transition frequency with an optical frequency interval divider chain*, Physical Review Letters **79**, pp. 2646–2649, 1997. Cited on page 742.

This is a section of the freely downloadable e-textbook

Motion Mountain

Hiking beyond space and time along the concepts of modern physics

available at
www.motionmountain.org

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Challenge Did you find any mistakes?

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

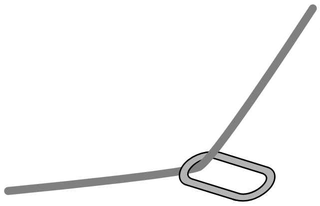
4 For a full list of isotopes, see R.B. Firestone, *Table of Isotopes, Eighth Edition, 1999 Update*, with CDROM, John Wiley & Sons, 1999. For a list of isotopes on the web, the standard is the korean web site by J. Chang, http://hpngp01.kaeri.re.kr/CoN/index.html, mirrored at http://www.nes.ruhr-uni-bochum.de/CoN. Cited on page 742.

5 About the long life of tantalum 180, see D. Belic & al., *Photoactivation of* $^{180}\mathrm{Ta}^{\mathrm{m}}$ *and its implications for the nucleosynthesis of nature's rarest naturally occurring isotope*, Physical Review Letters **83**, pp. 5242–5245, 20 december 1999. Cited on page 742.

6 Stephen J. Gould, 1980, *The Panda's thumb*, W.W. Norton & Co., New York. One of the several interesting and informative books on evolutionary biology by the best writer in the field. Cited on page 743.

7 For a good review, see the article by P.T. Greenland, *Antimatter*, Contemporary Physics **38**, pp. 181–203, 1997. Cited on page 743.

Typeset in
April 2001



APPENDIX D NUMBERS AND SPACES

A mathematician is a machine that transforms coffee into theorems.
Paul Erdös (1913, Budapest–1996)

Mathematical concepts can all be constructed from 'sets' and 'relations.' The most important ones were presented in the the first intermezzo. In the following a few more advanced concepts are presented as simply and vividly as possible,* for all those who want to smell the passion of mathematics. Ref. 1

In particular, we will expand the range of *algebraic* and the range of *topological* structures. Mathematicians are not only concerned with the exploration of concepts, but always also with their *classification*. Whenever a new mathematical concept is introduced, mathematicians try to classify all the possible cases and types. Most spectacularly this has been achieved for the different types of numbers, of simple groups, and for many types of spaces and manifolds.

More numbers

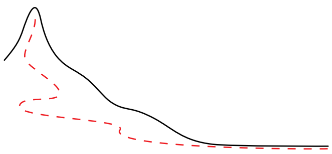
A person that can solve $x^2 - 92y^2 = 1$ in less than a year is a mathematician. Challenge
Brahmagupta, (598, Sindh–668) (implied: solve in *integers*)

The concept of 'number' is not limited to what was presented in the first intermezzo.** The simplest generalisation is achieved by extending them to manifolds of more than one dimension.

* The opposite approach is taken by the delightful text by Carl E. LINDERHOLM, *Mathematics made difficult*, Wolfe Publishing, 1971.
** An excellent introduction into number systems in mathematics is the book H.-D. EBBINGHAUS & al., *Zahlen*, 3. Auflage, Springer Verlag 1993. It is also available in english, under the title *Numbers*, Springer Verlag, 1990.

This is a section of the freely downloadable e-textbook

Motion Mountain



Hiking beyond space and time along the concepts of modern physics

available at
www.motionmountain.org

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Did you find any mistakes?

Challenge

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

Complex numbers

Complex numbers are defined by $z = a + ib$. The generators of the complex numbers, 1 and i, obey the well known algebra

$$\begin{array}{c|cc} \cdot & 1 & i \\ \hline 1 & 1 & i \\ i & i & -1 \end{array} \qquad (623)$$

often written as $i = +\sqrt{-1}$.

The *complex conjugate* z^*, also written $\bar{z}$, of a complex number $z = a + ib$ is defined as $z^* = a - ib$. The *absolute value* $|z|$ of a complex number is defined as $|z| = \sqrt{zz^*} = \sqrt{z^*z} = \sqrt{a^2 + b^2}$. It defines a *norm* on the vector space of the complex numbers. One has $|wz| = |w|\,|z|$, from which one deduces the *two-squares theorem*

$$(a_1^2 + a_2^2)(b_1^2 + b_2^2) = (a_1 b_1 - a_2 b_2)^2 + (a_1 b_2 + a_2 b_1)^2 \qquad (624)$$

valid for all real numbers a_i, b_i. It was already known, in its version for integers, to Diophantos of Alexandria.

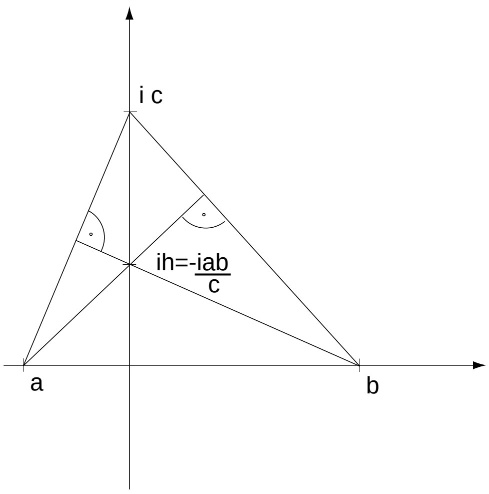


Figure 183 A property of triangles easily provable with complex numbers

This means that complex numbers can also be written as a a couple (a, A), with their addition defined as $(a, A) + (b, B) = (a + b, A + B)$ and their multiplication defined as $(a, A) \cdot (b, B) = (ab - AB, aB + bA)$. The two component writing allows to identify complex numbers with the points on a plane, and, once one has translated the definition of multiplication into geometrical language, to rapidly prove geometrical theorems, such as the one of figure 183.

Complex numbers can also be represented as 2×2 matrices of the form

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \quad \text{with} \quad a, b \in \mathbf{R} \quad . \qquad (625)$$

Usual matrix addition and multiplication then gives the same result as complex addition and multiplication. In this way, complex numbers can be represented by a special type of real matrices. What is $|z|$ in matrix language? Challenge

The set $\mathbf{C}$ of complex numbers with the mentioned multiplication forms a commutative two-dimensional field. In the field of complex numbers, quadratic equations for an unknown z, namely $az^2 + bz + c = 0$, always have two solutions. Challenge

Complex numbers can be used to describe the position of the points of a plane. Rotations around the origin can be described by multiplications by a complex number of unit length. Since complex numbers describe the two-dimensional plane, any two-dimensional quantity can be described with them. That is why electrical engineers describe quantities with phases, such as alternating current or electrical fields in space, with complex numbers.

By the way, there are as many complex numbers as there are real numbers. Are you able to show this? Challenge

Love is complex: it has real
and imaginary parts.

Quaternions

The position of the points on a line can be described by real numbers. Complex numbers can be used to describe the position of the points of a plane. If one tries to generalize the idea of a number to higher dimensional spaces, it turns out that no such number system can be defined for *three*-dimensional space. However a new number system, the *quaternions*, can be constructed from the points of *four*-dimensional space, but only if the requirement of commutativity of multiplication is dropped. In fact, no number system can be defined for dimensions other than 1, 2 and 4. The quaternions were discovered by several mathematicians in the 19th century, among them Hamilton,* who studied them for a long part of his life. In fact, Maxwell's electrodynamics was formulated with quaternions before it was with Ref. 2 three-dimensional vectors.

The quaternions form a 4-dimensional algebra over the reals with the basis $1, i, j, k$ satisfying

$$\begin{array}{c|cccc} \cdot & 1 & i & j & k \\ \hline 1 & 1 & i & j & k \\ i & i & -1 & k & -j \\ j & j & -k & -1 & i \\ k & k & j & -i & -1 \end{array} \qquad (626)$$

which is also often written $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$, $jk = -kj = i$, $ki = -ik = j$. The quaternions $1, i, j, k$ are also called *basic units* or *generators*. The missing symmetry along the diagonal of the table shows the lack of commutativity of quaternionic multiplication. It was the first time the idea of a non-commutative product appeared in mathematics. Despite this restriction, the multiplication of quaternions remains associative. As a consequence, polynomial equations in quaternions have many more solutions than in complex numbers, Challenge as the equation $X^2 + 1 = 0$ shows.

Every quaternion X can be written in the form

$$\begin{aligned} X &= x_0 + x_1 i + x_2 j + x_3 k = x_0 + \mathbf{v} \quad \text{or} \\ X &= (x_0, x_1, x_2, x_3) = (x_0, \mathbf{v}) \quad , \end{aligned} \qquad (627)$$

where x_0 is called the *scalar* part and $\mathbf{v}$ the *vector* part. The multiplication is thus defined as $(x, \mathbf{v})(y, \mathbf{w}) = (xy - \mathbf{v} \cdot \mathbf{w}, x\mathbf{w} + y\mathbf{v} + \mathbf{v} \times \mathbf{w})$. The *conjugate quaternion* $\overline{X}$ is defined as $\overline{X} = x_0 - \mathbf{v}$, and one has $\overline{XY} = \overline{Y}\,\overline{X}$. The *norm* $|X|$ of a quaternion X is defined as $|X|^2 =$

∗ William Rowan Hamilton (1805, Dublin–1865, Dunsink), irish enfant prodige, mathematician, named the quaternions after an expression from the vulgate (act. apost. 12, 4).

$X\overline{X} = \overline{X}X = x_0^2 + x_1^2 + x_2^2 + x_3^2 = x_0^2 + \mathbf{v}^2$. The norm is multiplicative, i.e. $|XY| = |X|\,|Y|$. The multiplication of two general quaternions can be written as

$$(a_1, b_1, c_1, d_1)(a_2, b_2, c_2, d_2) = (a_1a_2 - b_1b_2 - c_1c_2 - d_1d_2\ ,\ a_1b_2 + b_1a_2 + c_1d_2 - d_1c_2, \\ a_1c_2 - b_1d_2 + c_1a_2 + d_1b_2, a_1d_2 + b_1c_2 - c_1b_2 + d_1a_2) \quad (628)$$

From $|WZ| = |W|\,|Z|$ one deduces the *four-squares theorem*

$$(a_1^2 + a_2^2 + a_3^2 + a_4^2)(b_1^2 + b_2^2 + b_3^2 + b_4^2) \\ = (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4)^2 + (a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3)^2 \\ + (a_1b_3 + a_3b_1 + a_4b_2 - a_2b_4)^2 + (a_1b_4 + a_4b_1 + a_2b_3 - a_3b_2)^2 \quad (629)$$

valid for all real numbers a_i and b_i, and thus also for any set of eight integers. It was discovered in 1748 by Leonhard Euler (1707–1783) when trying to prove that each integer is the sum of four squares. (That proof was found only in 1770, by Joseph Lagrange.)

Hamilton thought that a quaternion with zero scalar part, which he simply called a *vector* – a term which he invented –, could be identified with an ordinary 3-dimensional translation vector; but this is wrong. Therefore, such a quaternion is now called a *pure*, or a *homogeneous*, or again, an *imaginary* quaternion. For two pure quaternions $V = (0, \mathbf{v})$ and $W = (0, \mathbf{w})$ one has the relation $VW = (-\mathbf{v}\cdot\mathbf{w}, \mathbf{v}\times\mathbf{w})$, where $\cdot$ denotes the scalar product and $\times$ denotes the vector product. Any general quaternion can be written as the ratio of two pure quaternions.

In reality, a pure quaternion $(0, \mathbf{v})$ does not behave under coordinate transformations like a (modern) vector; in fact, a pure quaternion represents a rotation by the angle π around the axis defined by the direction $\mathbf{v} = (v_x, v_y, v_z)$. Challenge

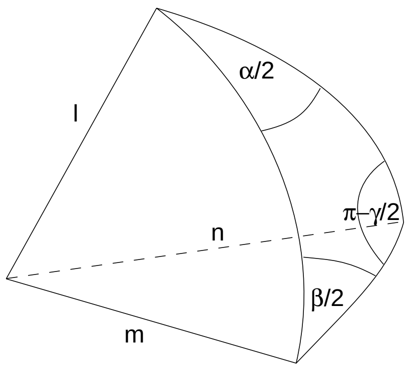


Figure 184 Combinations of rotations

It turns out that in three-dimensional space, a general rotation about the origin can be described by a *unit* quaternion, also called a *normed* quaternion, for which $|Q| = 1$. Such a quaternion can be written as $(\cos\theta/2, \mathbf{n}\sin\theta/2)$, where $\mathbf{n} = (n_x, n_y, n_z)$ is the normed vector describing the direction of the rotation axis, and θ is the rotation angle. Such a unit quaternion $Q = (\cos\theta/2, \mathbf{n}\sin\theta/2)$ rotates a pure quaternion $V = (0, \mathbf{v})$ into another pure quaternion $W = (0, \mathbf{w})$ given by

$$W = QVQ^* \quad . \quad (630)$$

In this case, if one uses pure quaternions such as V or W to describe positions, one can use unit quaternions to describe rotations and to calculate coordinate changes. The concatenation of two rotations is then given as the product of the corresponding unit quaternions. Indeed, a rotation by an angle α about the axis $\mathbf{l}$ followed by a rotation by an angle β about the axis $\mathbf{m}$ gives a rotation by an angle γ about axis $\mathbf{n}$, with the values determined by

$$(\cos\gamma/2, \sin\gamma/2\mathbf{n}) = (\cos\alpha/2, \sin\alpha/2\mathbf{l})(\cos\beta/2, \sin\beta/2\mathbf{m}) \quad , \quad (631)$$

shown graphically in figure 184.

Quaternions can teach something about the motion of hand and arm. Keeping the left arm straight, defining the three possible 90 degree motions as i, j, and k, and taking concatenation as multiplication, the motion of our arms follows the same "laws" as those of pure unit quaternions. Can you find out what -1 is? Challenge

The reason for this behaviour is the non-commutativity of rotations. But one can specify this noncommutativity more precisely, with mathematical language. The rotations in 3 dimensions around a point form the special orthogonal group in 3 dimensions, in short SO(3). But the motions of a hand attached to a shoulder via an arm form another group, isomorphic to the Lie group SU(2). The difference is due to the appearance of half angles in the parametrization of rotations; indeed, the above parametrizations imply that a rotation by 2π corresponds to a multiplication by -1! Only in the twentieth century it was realized that physical observables behaving in this way do exist: *spinors*. More on spinors can be found in the section on permutation symmetry, where belts are used as well as arms. In short, the group SU(2) of the quaternions is the *double cover* of the rotation group SO(3).

See page 465
Ref. 3

The easy description of rotations and positions with quaternions is used in robotics, in astronomy, and in flight simulators, because of the especially simple coding of coordinate transformations it provides. Also in the field of three-dimensional graphics quaternions are often used in software packages to calculate the path taken by repeatedly reflected light rays.

The algebra of the quaternions is the unique associative noncommutative finite-dimensional normed algebra with an identity over the field of real numbers. Quaternions form a noncommutative field, i.e. a skew field, in which the inverse of a quaternion X is $\overline{X}/N(X)$. Since one can thus define a division of quaternions, one also says that they form a *division algebra*. In fact the quaternions, the reals **R**, and the complex numbers **C**, form the only three examples of finite dimensional associative division algebras. In other words, the skew-field of quaternions is the unique finite-dimensional real associative non-commutative algebra without divisors of zero. The *center* of the quaternions, i.e. the set of those quaternions commuting with all quaternions, are the reals.

In a similar way as in the case of the complex numbers, quaternions can be represented as matrices of the form

$$\begin{pmatrix} A & B \\ -B^* & A^* \end{pmatrix} \quad \text{with} \quad A,B \in \mathbf{C}, \quad \text{or as} \quad \begin{pmatrix} a & b & c & d \\ -b & a & -d & c \\ -c & d & a & -b \\ -d & -c & b & a \end{pmatrix} \quad \text{with} \quad a,b,c,d \in \mathbf{R} \quad , \tag{632}$$

where $A = a + ib, B = c + id$ and the quaternion X is $X = A + Bj = a + ib + jc + kd$; usual matrix addition and multiplication then gives the same result as quaternionic addition and multiplication.

The generators of the quaternions can be realised for example as

$$1 : \sigma_0 \quad , \quad i : -i\sigma_1 \quad , \quad j : -i\sigma_2 \quad , \quad k : -i\sigma_3 \quad . \tag{633}$$

where the σ_n are the Pauli spin matrices.*

Other real 4×4 representations are also possible, such as

$$\begin{pmatrix} a & b & -d & -c \\ -b & a & -c & d \\ d & c & a & b \\ c & -d & -b & a \end{pmatrix} \tag{635}$$

but no representation by 3×3 matrices is possible. Challenge

These matrices contain real and complex elements, which poses no special problems. In contrast, when matrices with quaternionic elements are constructed, care has to be taken, because simple relations, such as $\mathrm{tr}AB = \mathrm{tr}BA$ are *not* fulfilled in general, since quaternionic multiplication is not commutative.

What do we learn from quaternions for the description of nature? The first idea is: binary rotations are similar to positions. One point is the similarity of binary rotations and positions, and thus translations. Are rotations the basic operations? Is it possible that translations are only shadows of rotations? The ways that translations are connected to rotations are investigated in the second and third part of the escalation.

As a remark, when Maxwell wrote down his equations of electrodynamics, he used quaternion notation. The now usual form was introduced later by other scientists, notably by Hertz and Heaviside. Maxwell's original equations of electrodynamics, in modern quaternion notation, read: See page 315 Challenge

$$dF = -\frac{Q}{\varepsilon_o} \tag{636}$$

where the quantities are defined as following:

$$\begin{aligned} F &= E + \sqrt{-1}cB \\ E &= iE_x + jE_y + kE_z \\ B &= iB_x + jB_y + kB_z \\ d &= \delta + \sqrt{-1}\partial_t / c \\ \delta &= i\partial_x + j\partial_y + k\partial_z \\ Q &= \rho + \sqrt{-1}J/c \end{aligned} \tag{637}$$

and where $\sqrt{-1}$ is the complex root of -1.

* The *Pauli spin matrices* are the complex, hermitian matrices

$$\sigma_0 = \mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{634}$$

whose eigenvalues are ± 1; they satisfy the relations $[\sigma_i, \sigma_k]_+ = 2\,\delta_{ik}$ and $[\sigma_i, \sigma_k] = 2i\varepsilon_{ikl}\,\sigma_l$. The linear combinations $\sigma_\pm = \frac{1}{2}(\sigma_1 \pm \sigma_2)$ are also frequently used. By the way, another possible representation of the quaternions is $i : i\sigma_3, j : i\sigma_2, k : i\sigma_1$.

Octonions

In the same way that the quaternions are constructed from complex numbers, octonions can be constructed from quaternions, as done by Arthur Cayley (1821–1895). *Octonions* or *octaves* are the elements of an 8-dimensional algebra over the reals with the generators $1, i_n$ with $n = 1 \dots 7$ satisfying

	1	i_1	i_2	i_3	i_4	i_5	i_6	i_7
1	1	i_1	i_2	i_3	i_4	i_5	i_6	i_7
i_1	i_1	-1	i_3	$-i_2$	i_5	$-i_4$	i_7	$-i_6$
i_2	i_2	$-i_3$	-1	i_1	$-i_6$	i_7	i_4	$-i_5$
i_3	i_3	i_2	$-i_1$	-1	i_7	i_6	$-i_5$	$-i_4$
i_4	i_4	$-i_5$	i_6	$-i_7$	-1	i_1	$-i_2$	i_3
i_5	i_5	i_4	$-i_7$	$-i_6$	$-i_1$	-1	i_3	i_2
i_6	i_6	$-i_7$	$-i_4$	i_5	i_2	$-i_3$	-1	i_1
i_7	i_7	i_6	i_5	i_4	$-i_3$	$-i_2$	$-i_1$	-1

(638)

Nineteen other, equivalent multiplication tables are also possible. This algebra is called the *Cayley algebra*; it has an identity and a unique division. The algebra is non-commutative and also non-associative. It is however, *alternative*, meaning that for all elements, one has $x(xy) = x^2 y$ and $(xy)y = xy^2$, a property somewhat weaker than associativity. It is the only 8-dimensional real alternative algebra without zero divisors. For this last reason, the set $\mathbf{\Omega}$ of all octonions does not form a field nor a ring, and the old designation of 'Cayley numbers' has been abandoned. The octonions are the most general hypercomplex 'numbers' whose norm is multiplicative. Associativity is not satisfied, since $(i_n i_m) i_l = \pm i_n (i_m i_l)$, where the minus sign is valid for combination of indices which belong to those triads, such as 1-2-4, which are not quaternionic.

Octonions can be represented as matrices of the form

$$\begin{pmatrix} A & B \\ -\bar{B} & \bar{A} \end{pmatrix} \quad \text{where} \quad A, B \in \mathbf{H} \quad , \quad \text{or as real } 8 \times 8 \text{ matrices.}$$

Matrix multiplication then gives the same result as octonionic multiplication.

One has $|wz| = |w|\,|z|$, from which one deduces the impressive *eight-squares theorem*

$$\begin{aligned}
&(a_1^2+a_2^2+a_3^2+a_4^2+a_5^2+a_6^2+a_7^2+a_8^2)(b_1^2+b_2^2+b_3^2+b_4^2+b_5^2+b_6^2+b_7^2+b_8^2) \\
&\quad = (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4 - a_5b_5 - a_6b_6 - a_7b_7 - a_8b_8)^2 \\
&\quad +(a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3 + a_5b_6 - a_6b_5 - a_7b_8 + a_8b_7)^2 \\
&\quad +(a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2 + a_5b_7 + a_6b_8 - a_7b_5 - a_8b_6)^2 \\
&\quad +(a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1 + a_5b_8 - a_6b_7 + a_7b_6 - a_8b_5)^2 \\
&\quad +(a_1b_5 - a_2b_6 - a_3b_7 - a_4b_8 + a_5b_1 + a_6b_2 + a_7b_3 + a_8b_4)^2 \\
&\quad +(a_1b_6 + a_2b_5 - a_3b_8 + a_4b_7 - a_5b_2 + a_6b_1 - a_7b_4 + a_8b_3)^2 \\
&\quad +(a_1b_7 + a_2b_8 + a_3b_5 - a_4b_6 - a_5b_3 + a_6b_4 + a_7b_1 - a_8b_2)^2 \\
&\quad +(a_1b_8 - a_2b_7 + a_3b_6 + a_4b_5 - a_5b_4 - a_6b_3 + a_7b_2 + a_8b_1)^2
\end{aligned} \tag{639}$$

valid for all real numbers a_i and b_i, and thus in particular also for all integers. It was discovered in 1818 by Carl Friedrich Degen (1766–1825), and then rediscovered in 1844 by John Graves and in 1845 by Cayley. There is no generalization to higher numbers of squares, a fact proven by Adolf Hurwitz (1859–1919) in 1898.

As a note, the octonions can be used to show that a vector product is not only possible in dimensions 3. A *vector product* or *cross product* is an operation satisfying

$$\begin{aligned} u \times v &= -v \times u \quad && \textit{anticommutativity} \\ (u \times v)w &= u(v \times w) \quad && \textit{exchange rule.} \end{aligned} \tag{640}$$

If one uses the definition

$$X \times Y = \frac{1}{2}(XY - YX) \quad , \tag{641}$$

then the $\times$-products of imaginary quaternions, i.e. of quaternions of the sort $(0, \mathbf{u})$, are again imaginary, and the $\mathbf{u}$'s obey the usual vector product, which indeed fulfills (640). It turns out that if one uses definition (641) for *octonions*, then for the imaginary octonions, i.e. for octonions of the sort $(0, \mathbf{U})$, that product also yields only imaginary octonions, and the $\mathbf{U}$'s also follow expression (640). In fact, this is the only other nontrivial example possible. Thus a vector product exists only in 3 and in 7 dimensions. Ref. 4 Challenge

Other types of numbers

The process of construction of a new system of hypercomplex 'numbers' or real algebras by 'doubling' a given one can be continued ad infinitum. However, octonions, *sedenions* and all the following doublings are neither rings nor fields, but only non-associative algebras with unity. Other finite-dimensional algebras with unit element over the field of the reals, once generally called hypercomplex 'numbers', can also be defined, such as 'dual numbers', 'double numbers','Clifford-Lifschitz numbers' etc. They play no special role in physics.

Mathematicians also have defined number fields which have "one and a half" dimensions, such as algebraic number fields. There is also a generalisation of the concept of integers to the complex domain, the *gaussian integers*, defined as $n + im$. Gauss even defined what now are known as *gaussian primes*. (Can you find out how?) They are not used in the description of nature, but are important in number theory. Ref. 5 Challenge

As a note, in the old days physicists used to call quantum mechanical operators 'q-numbers.' But this term has now fallen out of fashion.

Other extensions of the natural numbers are those which include numbers larger than the smallest type of infinity. The most important *transfinite numbers* are the *ordinals*, the *cardinals*, and the mentioned *surreals*. The ordinals are essentially the infinite integers (and the finite ones), whereas the surreals are the infinite (and finite) reals. The surreals were defined in the first intermezzo. They are to the ordinal numbers what the reals are to the integers: they fill up all the gaps in between. Interestingly, for the surreals, the summation of many divergent series in $\mathbf{R}$ converge. Can you find one example? Ref. 6 See page 390 Challenge

The surreals also include infinitely small numbers, as do the numbers of *nonstandard analysis*, also called *hyperreals*. In both number systems, in contrast to the case of the real numbers, the numbers $0.999\overline{9}$ and 1 do not coincide, but are separated by infinitely many other numbers. Ref. 4

Grassmann numbers

With the discovery of supersymmetry, another type of numbers became important, the *Grassmann numbers*.* They are in fact a special type of hypercomplex 'numbers'. In supersymmetric lagrangians, fields depend on two types of coordinates: on the usual real space-time coordinates and additionally on Grassmann coordinates.

Grassmann numbers, also called *fermionic coordinates*, θ have the defining properties

$$\theta^2 = 0 \quad \text{and} \quad \theta_i\theta_j + \theta_j\theta_i = 0 \quad . \tag{642}$$

Challenge You may want to look for a representation of these numbers. More about their use can be found in the section on supersymmetry.

Vector spaces

Vector spaces, also called linear spaces, are mathematical generalisations of certain aspects of the intuitive three-dimensional space. Any set of elements that can be added together and also be multiplied by numbers is called a vector space, if the result is again in the set and the usual rules of calculation hold.

More precisely, a *vector space* over a number field K is a set of elements, called *vectors* in this case, for which a vector addition and a *scalar multiplication* is defined for all vectors a, b, c and for all numbers s and r from K with the properties

$$\begin{aligned}
(a+b)+c = a+(b+c) = a+b+c \quad & \textit{associativity of vector addition} \\
n+a = n \quad & \textit{existence of null vector} \\
(-a)+a = n \quad & \textit{existence of negative vector} \\
1a = a \quad & \textit{regularity of scalar multiplication} \\
(s+r)(a+b) = sa+sb+ra+rb \quad & \textit{complete distributivity of scalar multiplication}
\end{aligned} \tag{643}$$

If the field K, whose elements are called *scalars* in this context, is taken to be the real (complex, quaternionic) numbers, one speaks of a real (complex, quaternionic) vector space. Vector spaces are also called *linear vector spaces* or simply *linear spaces*.

The complex numbers, the set of all functions defined on the real line, the set of all polynomials, the set of matrices of given number of rows and columns all form vector spaces. In mathematics, a vector is thus a more general concept than in physics. Physical vectors are more specialized objects, namely elements of normed inner product spaces.

To define them one needs the concept of metric space. A *metric space* is a vector space with a metric, i.e. a way to define distances between elements. A relation $d(a,b)$ between elements is called a metric if

$$\begin{aligned}
d(a,b) \geqslant 0 \quad & \textit{positivity of metric} \\
d(a,b)+d(b,c) \geqslant d(a,c) \quad & \textit{triangle inequality} \\
d(a,a) = 0 \quad & \textit{regularity of metric}
\end{aligned} \tag{644}$$

For example, measuring the distance between cities in France, i.e. points on a surface, by
Challenge the shortest distance of travel via Paris, except in the case if they both lie on a line already

* Hermann Günther Grassmann (1809–1877) mathematician.

going through Paris, defines a metric between the points in France.

A *normed* vector space is, obviously, a linear space with norm, or 'length' of a vector. A norm is a positive (or vanishing) number $||a||$ defined for each vector a with the properties

$$\begin{aligned} ||ra|| &= |r|\,||a|| && \textit{linearity of norm} \\ ||a+b|| &\leqslant ||a||+||b|| && \textit{triangle inequality} \\ ||a|| = 0 \quad \text{only if} \quad a &= 0 && \textit{regularity} \end{aligned} \tag{645}$$

Usually there are many ways to define a norm for a given space. Note that a norm can always be used to define a metric by setting Challenge

$$d(a,b) = ||a-b|| \tag{646}$$

so that all normed spaces are also metric spaces. The most special linear spaces are *inner product spaces*. They are vector spaces with an *inner product*, also called *scalar product* (not to be confused with the scalar multiplication!). For an inner product in the *real* case the properties of

$$\begin{aligned} ab &= ba && \textit{commutativity of scalar product} \\ (ra)(sb) &= rs(ab) && \textit{bilinearity of scalar product} \\ (a+b)c &= ac+bd && \textit{left distributivity of scalar product} \\ a(b+c) &= ab+ac && \textit{right distributivity of scalar product} \\ aa &\geqslant 0 && \textit{positivity of scalar product} \\ aa = 0 \quad \text{only if} \quad a &= 0 && \textit{regularity of scalar product} \end{aligned} \tag{647}$$

hold for all vectors a, b and all scalars r, s. A *real* inner product space (of finite dimension) is also called a *euclidean* vector space. The set of all velocities, the set of all positions, or the set of all possible momenta form such spaces.

In the *complex* case this definition is extended to*

$$\begin{aligned} ab &= (ba)^* && \textit{hermitean property} \\ (ra)(sb) &= r^*s(ab) && \textit{sesquilinearity of scalar product} \\ (a+b)c &= ac+bd && \textit{left distributivity of scalar product} \\ a(b+c) &= ab+ac && \textit{right distributivity of scalar product} \\ aa &\geqslant 0 && \textit{positivity of scalar product} \\ aa = 0 \quad \text{only if} \quad a &= 0 && \textit{regularity of scalar product} \end{aligned} \tag{648}$$

hold for all vectors a, b and all scalars r, s. A *complex* inner product space (of finite dimension) is also called a *unitary* or *hermitean* vector space. If the inner product space is *complete*, it is called, especially in the infinite-dimensional complex case, a *Hilbert space*. The space of all possible states of a quantum system form a Hilbert space.

* The term *sesquilinear* is latin for 'one-and-a-half-linear'. Sometimes however, the half-linearity is assumed in the other argument.

All inner product spaces are also metric spaces and thus normed spaces, if one defines the metric, as one usually does, by

$$d(a,b) = \sqrt{(a-b)(a-b)} \quad . \tag{649}$$

In inner product spaces, a *basis* can be defined, allowing to speak about the length and the direction of vectors, as we are used to in physics.

Algebras

The term *algebra* is used in mathematics with three different, but loosely related meanings. It denotes a part of mathematics, as in "I hated algebra at school"; it further denotes in general any formal rules that are obeyed by abstract objects, as e.g. in the expression 'tensor algebra'. Finally it denotes a specific mathematical structure, which is the only meaning used here.

An *algebra* $A = \{x, y, ...\}$ is a set of elements with an addition and a multiplication having the properties that for all elements

$$\begin{aligned} x+y &= y+x && \textit{commutativity of addition} \\ x(y+z) = xy+xz \quad , \quad (x+y)z &= xz+yz && \textit{distributivity of multiplication} \\ xx &\geqslant 0 && \textit{positivity} \\ xx = 0 \quad \text{only if} \quad x &= 0 && \textit{regularity of multiplication} \end{aligned} \tag{650}$$

As is clear from this definition, algebras are rather general mathematical structures. In physics, those special algebras related to symmetries play the most important role.

An *associative algebra* is an algebra whose multiplication has the additional property of

$$x(yz) = (xy)z \quad \textit{associativity} \tag{651}$$

Most physical algebras are associative.

A *linear algebra* is an algebra over a number field with the property that a multiplication by scalars c is defined such that

$$c(xy) = (cx)y = x(cy) \quad \textit{linearity} \tag{652}$$

For example, the set of all linear transformations in an n-dimensional linear space, such as the translations on a plane, in space or in time, are linear algebras. So is the set of observables of a quantum mechanical system.* Note that all linear algebras are themselves vector

* Linear transformations are mathematical objects which transform a vector into another with the property that sums and multiples of vectors are transformed into sums and the multiples of the transformed vectors. Are you able to give the set of all possible linear transformations of points on a plane? And in space? And in Minkowski space? Challenge

You will discover that all linear transformations transform some special vectors, called *eigenvectors* – from the german word 'eigen' meaning self' – into multiples of themselves. In other words, if for a transformation T one has

$$Te = \lambda e \tag{653}$$

spaces; the difference being that in addition a (linear) and associative multiplication among the vectors is defined.

A *star algebra*, also written $*$-*algebra*, is an algebra over the *complex* numbers for which there is a mapping $* : A \to A,\ x \mapsto x^*$, called an *involution*, with the properties

$$\begin{aligned} (x^*)^* &= x \\ (x+y)^* &= x^* + y^* \\ (cx)^* &= c^* x^* \quad \text{for all} \quad c \in \mathbf{C} \\ (xy)^* &= y^* x^* \end{aligned} \tag{654}$$

valid for all elements x, y of the algebra A. The element x^* is called the *adjoint* of x. Star algebras are the main structure used in quantum mechanics, since quantum mechanical observables form a $*$-algebra.

A C$*$-*algebra* is a Banach algebra over the complex numbers with an involution $*$ so that the norm $\|x\|$ of an element x can be defined as

$$\|x\|^2 = x^* x \tag{655}$$

and which it is a *complete* vector space, i.e. one in which Cauchy sequences converge. The name C comes from 'continuous functions'; they form such an algebra with a properly defined norm. Can you find it? Challenge

All C$*$-algebras contain a space of hermitean elements (which have a real spectrum), a set of normal elements, a multiplicative group of unitary elements and a set of positive elements (with nonnegative spectrum).

One important type of *mathematical* algebra deserves to be mentioned. A *division algebra* is an algebra for which $ax = b$ and $ya = b$ are uniquely solvable in x or y for all b and all $a \neq 0$. Division algebras are thus one way to generalize the concept of a number. One of the important results of modern mathematics states that division algebras can *only* have dimension 1, like the reals, or dimension 2, like the complex numbers, or dimension 4, like the quaternions, or dimension 8, like the octonions. There is thus no way to generalize the concept of number to other or to higher dimensions.

Lie algebras

A Lie algebra is special type of algebra and of vector space. A vector space L over the field $\mathbf{R}$ (or $\mathbf{C}$) with an additional binary operation $[\,,]$ called *Lie multiplication* or the *commutator*, is called a real (or complex) Lie algebra if this operation fulfills the properties

$$\begin{aligned} [X,Y] &= -[Y,X] && \textit{antisymmetry} \\ [aX+bY,Z] &= a[X,Z] + b[Y,Z] && \textit{linearity} \\ [X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] &= 0 && \textit{Jacobi identity} \end{aligned} \tag{656}$$

one calls e an *eigenvector*, and λ its associated *eigenvalue*. The set of all eigenvalues of a transformation T is called the *spectrum* of T. Physicists did not care for these mathematical concepts until they discovered quantum theory. Quantum theory showed that observables are transformations in Hilbert space, because any measurement interacts with a system and thus transforms it. Quantum mechanical experiments also showed that a measurement result for an observable can only be one of the eigenvalues of the corresponding transformation. Therefore every expert on motion must know what an eigenvalue is. Finally, the state of the system after the measurement is given by the eigenvector of the measured eigenvalue.

See page 448
See page 482

for all elements $X, Y, Z \in L$ and for all $a, b \in \mathbf{R}$ (or $\mathbf{C}$). A Lie algebra is called *commutative* if $[X, Y] = 0$ for all elements X and Y. The *dimension* of the Lie algebra is the dimension of the vector space. A subspace N of a Lie algebra L is called an *ideal* if $[L, N] \subset N$; any ideal is also a *subalgebra*. A *maximal ideal* M which satisfies $[L, M] = 0$ is called the *center* of L.

A Lie algebra is called a *linear* Lie algebra if its elements are linear transformations of another vector space V, simply said, if they are "matrices". It turns out that all finite dimensional Lie algebras are isomorphic to a linear Lie algebra. Therefore, by picturing the elements of Lie algebras in terms of matrices one covers all finite dimensional cases.

The name 'Lie algebra' derives from the fact that the *generators*, i.e. the infinitesimal elements of every Lie group form a Lie algebra. Since all important symmetries in nature form Lie groups, Lie algebras appear very frequently in physics. In mathematics, Lie algebras arise frequently because from any associative finite dimensional algebra in which $\cdot$ stands for its multiplication, one gets a Lie algebra by defining the commutator by

$$[X, Y] = X \cdot Y - Y \cdot X \quad ; \tag{657}$$

this fact gave the commutator its name. Therefore a Lie algebra can also be seen as a special type of associative algebra.

Since Lie algebras are vector spaces, the elements T_i of a *basis* of the Lie algebra always obey the relation:

$$[T_i, T_j] = \sum_k c_{ij}^k T_k \tag{658}$$

where the numbers c_{ij}^k are called the structure constants of the Lie algebra. They depend on the chosen basis. Structure constants determine the Lie algebra completely. For example, the algebra of the Lie group SU(2), with the three generators defined by $T_a = \sigma^a/2i$, where the σ^a are the Pauli spin matrices, has the structure constants $C_{abc} = \varepsilon_{abc}$. *

See page 751

* In the same ways as groups, Lie algebras can be represented by matrices, i.e. by linear operators. Representations of Lie algebras are important in physics because many continuous symmetry groups are Lie groups.

The *adjoint representation* of a Lie algebra with basis $a_1 ... a_n$ is the set of matrices $\mathrm{ad}(a)$ defined for each element a by

$$[a, a_j] = \sum_c \mathrm{ad}(a)_{cj} a_c \quad . \tag{659}$$

One easily finds that $\mathrm{ad}(a)_{jk} = c_{ij}^k$, where c_{ij}^k are the structure constants of the Lie algebra. For a real Lie algebra, all elements of $\mathrm{ad}(a)$ are real for all $a \in L$.

Note that for any Lie algebra, one can define a scalar product by

$$(X, Y) = \mathrm{Tr}(\mathrm{ad}X \, \mathrm{ad}Y) \quad . \tag{660}$$

This scalar product is symmetric and bilinear. The corresponding bilinear form is also called the *Killing form*, after the german mathematician Wilhelm Killing (1847–1923), the discoverer of the exceptional Lie groups. The Killing form is invariant under the action of any automorphism of the algebra L. In a given basis, one has

$$(X, Y) = \mathrm{Tr}(\, (\mathrm{ad}X)^i_k \, (\mathrm{ad}Y)^s_i) = c^i_{lk} c^k_{si} x^l y^s = g_{ls} x^l y^s \tag{661}$$

where $g_{ls} = c^i_{lk} c^k_{si}$ is called the *Cartan metric tensor* of the Lie algebra L.

Classification of Lie algebras

All Lie algebras can be divided in finite-dimensional and infinite dimensional ones. Every *finite-dimensional* Lie algebra turns out to be the (semidirect) sum of a semisimple and a solvable Lie algebra.

A Lie algebra is called *solvable* if , well, if it is not semisimple. Solvable Lie algebras have not been classified completely up to now. They are not important in physics.

A *semisimple* Lie algebra is a Lie algebra which has no non-zero solvable ideal. Other equivalent definitions are possible, depending on your taste:
- a semisimple Lie algebra does not contain non-zero abelian ideals,
- its Killing-form is non-singular, i.e. non-degenerate,
- it splits into the direct sum of non-abelian simple ideals (this decomposition is unique)
- every finite-dimensional linear representation is completely reducible
- the one-dimensional cohomology of g with values in an arbitrary finite-dimensional g-module is trivial.

All finite-dimensional semisimple Lie algebras have been completely classified. Every semisimple Lie algebra decomposes uniquely into a direct sum of *simple* Lie algebras. One has to distinguish complex and real simple algebras.

The simple finite-dimensional *complex* Lie algebras all belong to four infinite classes and to five exceptional cases. The infinite classes are also called *classical* and are A_n for $n \geqslant 1$, corresponding to the Lie groups $SL(n)$ and their compact “cousins” $SU(n)$, B_n for $n \geqslant 2$, corresponding to the Lie groups $SO(2n+1)$, C_n for $n \geqslant 3$, corresponding to the Lie groups $Sp(2n)$, and D_n for $n \geqslant 4$, corresponding to the Lie groups $SO(2n)$. These simple Lie algebras are defined as follows. A_n is the algebra of all skew-hermitian $n \times n$ matrices, B_n, C_n are the algebras of the symmetric $n \times n$ matrices, and D_n is the algebra of the traceless $n \times n$ matrices.

The exceptional Lie algebras are G_2, F_4, E_6, E_7, E_8. In all cases, the index gives the number of roots. The dimension of the algebras is $A_n : n(n+2)$, B_n and $C_n : n(2n+1)$, $D_n : n(2n-1)$, $G_2 : 14$, $F_4 : 32$, $E_6 : 78$, $E_7 : 133$, $E_8 : 248$.

The simple and finite-dimensional *real* Lie algebras are more numerous; they follow from the list of complex Lie algebras. Moreover, for each complex Lie group, there is always one compact real one. Real Lie algebras are not so important in fundamental physics. Ref. 7

Of the large number of *infinite dimensional* Lie algebras only few are important in physics, among them the *Poincaré algebra*, the *Cartan algebra*, the *Virasoro algebra*, and a few other Kac-Moody algebras.

For supersymmetry, i.e. for systems with anticommuting coordinates, the concept of Lie algebra has been extended, and so-called Lie-superalgebras have been defined.

The Virasoro algebra

The *Virasoro algebra* is the infinite algebra of operators L_n satisfying

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{c}{12}(m^3 - m)\delta_{m,-n} \tag{662}$$

where the number c is called the *central charge*, and the factor $1/12$ being introduced by historical convention. This rather specific algebra is important in physics because it is the algebra of conformal symmetry in *two* dimensions, as explained on page 689.* Are you able to find a representation in terms of infinite square matrices? Mathematically speaking, the Virasoro algebra is a special case of a Kac-Moody algebra.

Challenge

– CS – sections on topology, integration, and Lie groups to be added – CS –

Topology is group theory.
The Erlangen program

References

1 A general basis can be the *Encyclopedia of Mathematics*, in 10 volumes, Kluwer Academic Publishers, Dordrecht, 1988/–1993, It explains carefully all concepts used in mathematics. Spending an hour with it looking up related keywords is an efficient way to get an introduction into any part of mathematics, especially into the vocabulary and the main connections. Cited on page 746.

2 S.L. ALTMAN, *Rotations, quaternions and double groups*, Clarendon Press, 1986, and also S.L. ALTMAN, *Hamilton, Rodriguez, and the quaternion scandal*, Mathematical Magazine pp. 291–308, 1988. Cited on page 749.

3 See the fine book Louis H. KAUFFMAN, *Knots and physics*, World Scientific, second edition, 1994, which gives a clear and visual introduction to the mathematics of knots and its main applications to physics. Cited on page 751.

4 A good introduction to nonstandard numbers, quaternions, octonions, p-adic numbers, surreal numbers, and more is H.-D. EBBINGHAUS, H. HERMES, F. HIRZEBRUCH, M. KOECHER, K. MAINZER, J. NEUKIRCH, A. PRESTEL & R. REMMERT, *Zahlen*, Springer Verlag, 1993, also available in english as *Numbers*,, Springer Verlag, 1990. Cited on page 754, 754.

5 Gaussian numbers are presented and explained in ... Cited on page 754.

6 About transfinite numbers, see the delightful paperback by Rudy RUCKER, *Infinity and the mind – the science and philosophy of the infinite*, Bantam, Toronto, 1983. Cited on page 754.

7 M. FLATO, P. SALLY & G. ZUCKERMAN, (editors) *Applications of Group Theory in Physics and Mathematical Physics*, Lectures in applied mathematics, volume 21, American Mathematical Society 1985. This interesting book has been written before the superstring revolution, so that the latter topic is missing, in the otherwise excellent presentation. Cited on page 760.

See page 178

* Note that the conformal symmetry in *four* dimensions has 15 parameters, and thus its Lie algebra is finite (fifteen) dimensional.

Typeset in April 2001

Appendix E Information Sources on Motion

I only know that I know nothing.
Socrates (470–399 BCE)

In this text, interesting works introducing domains not covered here are given in footnotes, whereas the reference list at the end of each chapter collects general material satisfying any further curiosity about what is encountered in this escalation. All citations can also be found by looking up the author's name in the index. To find additional information, nowadays either libraries or the internet can help.

In a library, review articles of recent research appear in journals such as *Reviews of Modern Physics*, *Reports on Progress in Physics*, *Contemporary Physics*, and *Advances in Physics*. Pedagogical introductions are best found in the *American Journal of Physics*, the *European Journal of Physics*, and in *Physik in unserer Zeit*.

Actual overviews on research trends can be found irregularly in magazines such as *Physics World*, *Physics Today*, and *Physikalische Blätter*. For all sciences together, the best sources are the magazines *Nature*, *New Scientist*, *Naturwissenschaften*, *La Recherche*, and the cheap but excellent *Science News*.

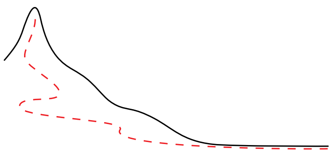
Research papers appear mainly in *Physics Letters B, Nuclear Physics B, Physical Review D, Physical Review Letters, Classical and Quantum Gravity*, *General Relativity and Gravitation*, *International Journal of Modern Physics*, and in *Modern Physics Letters*. The newest results and speculative ideas are found in conference proceedings, such as the *Nuclear Physics B Supplements*. Articles on the topic can also appear in *Fortschritte der Physik*, *Zeitschrift für Physik C, La Rivista del Nuovo Cimento, Europhysics Letters*, *Communications in Mathematical Physics*, *Journal of Mathematical Physics*, *Foundations of Physics*, *International Journal of Theoretical Physics*, and *Journal of Physics G*.

Papers on the description of motion without time and space which appear after this text can be found via the *Scientific Citation Index*. It is published in printed form or as compact disk and allows, given a paper, e.g. one from the references at the end of each chapter, to search for all subsequent publications which cite it. Then, using the bimonthly *Physics Abstracts*, which also exists both in paper and in electronic form, one can look up the abstract of the paper and check whether it is of interest.

But by far the simplest and most efficient way to keep in touch with ongoing research on motion is with help of the *internet*, the international computer network. To anybody with a personal computer connected to a telephone, most theoretical physics papers are available

This is a section of the freely downloadable e-textbook

Motion Mountain



Hiking beyond space and time along the concepts of modern physics

available at
www.motionmountain.org

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Did you find any mistakes?

Challenge

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

free of charge, as preprint, i.e. before official publication and check by referees. This famous service is available at:

Table 60 The Los Alamos e-print archive system for physics and related topics

Topic	server name	server address
general relativity and quantum cosmology	gr-qc	
theoretical high energy physics	hep-th	
computational high energy physics	hep-lat	
and lattice calculations		via e-mail, add
phenomenological high energy physics	hep-ph	@xxx.lanl.gov or
experimental high energy physics	hep-ex	@babbage.sissa.it
general physics	physics	to the server name, e.g.
theory, experiments and philosophy	quant-ph	hep-th@xxx.lanl.gov or
of quantum physics		gr-qc@xxx.lanl.gov
experimental nuclear physics	nucl-ex	
theoretical nuclear physics	nucl-th	
astrophysics	astro-ph	
condensed matter physics	cond-mat	
algebraic geometry	alg-geom	on the world-wide web, click on
differential geometry	dg-ga	http://xxx.lanl.gov or
functional analysis	funct-an	http://xxx.sissa.it or
quantum algebra and topology	q-alg	http://xxx.uni-augsburg.de or
adaptation, noise, and self-organizing systems	adap-org	http://xxx.yukawa.kyoto-u.ac.jp
chaotic dynamics	chao-dyn	
cellular automata and lattice gases	comp-gas	or simply on
nonlinear sciences	nlin-sys	
pattern formation and solitons	patt-sol	http://arXive.org
exactly solvable and integrable systems	solv-int	
computation and language	cmp-lg	

For details on how to use these servers via electronic mail, send a message to the server with the subject line consisting simply of the word “help”, without the quotes.

The internet expanded into a mix of library, media store, discussion platform, order desk, time waster, and much more. With a personal computer, a modem, and with free *browser* software, one can look for information in millions of pages of documents, using the mouse an the search engines. The various parts of the documents are located in various computers around the world, but the user does not need to be aware of this.*

* Decades ago, the provoking book by Ivan ILLICH, *Deschooling society*, Harper & Row, 1971, listed four basic ingredients for any educational system:

- access to *resources* for learning, e.g. books, equipment, games, etc. at an affordable price, for everybody, at any time in their life;

- for all who want to learn, access to *peers* in the same learning situation, for discussion, comparison, cooperation, competition;

- access to *elders*, e.g. teachers, for their care and criticism towards those who are learning;

- exchanges between student and *performers* in the field of interest, so that the latter can be models to the former. For example, there should be the possibility to listen to professional musicians, reading the works of specialists, as well as giving performers the possibility to share, to advertise, and to perform their skills.

To start using the web, send an electronic mail message consisting of the line 'HELP' to listserv@info.cern.ch, the server at the European Organisation for Particle Research, where the web was invented. Or ask a friend who knows.* Searching the web for authors, organizations, keywords, books, publications, companies and general information can be a rewarding experience. A few interesting servers are given below.

Table 61 Some interesting world wide web servers

Topic	web site address ("URL")
Physics and science	
"The Internet Pilot to Physics"	http://www.tp.umu.se/TIPTOP
A complete physics information site, organized with help from the European Physical Society including an encyclopedia, preprints, news, forum, student forum, conferences, job market, used machines market, web links, etc.	
Electronic preprints	http://xxx.lanl.gov and others – see above www.slac.stanford.edu/spires
High energy physics	http://mentor.lanl.gov/Welcome.html or http://info.cern.ch/hypertext/DataSources/bySubject/Physics/HEP.html
Particle data	http://pdg.web.cern.ch/pdg
Physics news, weekly	http://www.aip.org
Article summaries in 25 science magazines	http://www.mag.browse.com/science.html
Abstracts of physics journals	http://www.osti.gov
Science News	http://www.sciencenews.org
Pictures of physicists	http://www.if.ufrj.br/famous/physlist.html
Information on physicists	http://144.26.13.41/phyhist
Gravitation news	http://vishnu.nirvana.phys.psu.edu/mog.html
Living reviews in relativity	http://www.livingreviews.org
Information on relativity	http://math.ucr.edu/home/baez/relativity.html
Physics problems	http://star.tau.ac.il/QUIZ
Physics organizations	http://www.cern.ch/ http://info.cern.ch/ http://aps.org http://www.hep.net/documents/newsletters/pnu/pnu.html http://www.aip.org

Illich develops the idea that if such a system was informal, he then calls it a 'learning web' or 'opportunity web', it would be superior to any formal, state financed institutions, such as existing schools, for the development of mature human beings. The discussion is deepened in his following works, *Deschooling our lives*, Penguin, 1976, and *Tools for conviviality*, 1973. Today, any networked computer offers one or more of the following: the simple *e-mail* (electronic mail), the more sophisticated *ftp* (file transfer to and from another computer), the more rare access to *usenet* (the discussion groups on specific topics, such as particle physics), and the powerful *world-wide web*. (Simply speaking, each of the latter implies and includes the ones before.) In a rather unexpected way, all these facilities of the internet could transform it into the backbone of the opportunity web mentioned by Illich; it is a social development to follow closely. It depends on the user's discipline whether the world wide web actually does provide a learning web.

* To use ftp via electronic mail, send a message to archie@archie.mcgill.ca with 'help' as mail text. To get web pages via e-mail, send an e-mail message to w3mail@gmd.de consisting of the word 'help', or, for general instructions, to mail-server@rtfm.mit.edu with as body 'send usenet/news.answers/internet-services/access-via-email'.

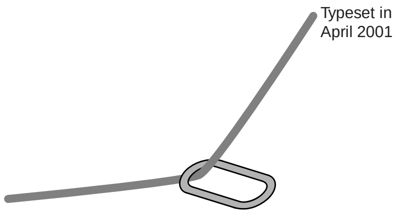
Topic	web site address (“URL”)
	http://www.nikhef.nl/www/pub/eps/eps.html
	http://www.het.brown.edu/physics/review/index.html
Physics textbooks on the web	http://www.plasma.uu.se/CED/Book
	http://biosci.umn.edu/biophys/OLTB/Textbook.html
Three beautiful french sets of notes on classical mechanics and particle theory	http://www.phy.ulaval.ca/enote.html
Math forum internet resource collection	http://forum.swarthmore.edu/ library/mathcoll.desc.html
Libraries	http://www.konbib.nl
	http://portico.bl.uk
	http://portico.bl.uk/gabriel/en/services.html
	http://www.niss.ac.uk/reference//opacsalpha.html
	http://www.bnf.fr
	http://www.laum.uni-hannover.de/iln/bibliotheken/kataloge.html
	http://www.loc/gov
	http://lcweb.loc.gov
Publishers	http://www.ioppublishing.com/
	http://www.aip.org
	http://www.amherts.edu/˜ajp
	http://www.elsevier.nl/
	http://www.nature.com/
Web related	
Good information search engines	http://www.altavista.com/cgi-bin/query?pg=aq
	http://www.metager.de
Search old usenet articles	http://www.dejanews.com
Information about the net	http://akebono.stanford.edu/yahoo/
	http://cuiwww.unige.ch/w3catalog
Frequently asked questions on various topics, also on physics	http://www.faqs.org
Computers	
File conversion	http://tom.cs.cmu.edu/intro.html
Download software and files	http://www.filez.com
Symbolic integration	http://www.integrals.com
	http://http.cs.berkeley.edu/ fateman/htest.html
Curiosities	
NASA	http://oel-www.jpl.nasa.gov/basics/bsf.html
The cosmic mirror	http://www.astro.uni-bonn.de/ dfischer/mirror
Observable satellites	http://liftoff.msfc.nasa.gov/RealTime/JPass/20/
Optical illusions	http://www.sandlotscience.com
Petit’s science comics	http://www.jp-petit.com/science/index.html
Physical toys	http://www.e20.physik.tu-muenchen.de/˜cucke/toylink.htm
Physics humor	http://www.escape.ca/˜dcc/phys/humor.htm
Literature on magic	http://www.faqs.org/faqs/magic-faq/part2/
Algebraic surfaces	http://www.mathematik.uni-kl.de/˜hunt/drawings.html
Making paper airplanes	http://pchelp.inc.net/paper_ac.htm
	http://www.ivic.qc.ca/˜aleexpert/aluniversite/ klinevogelmann.html

Topic	web site address (“URL”)
Postmodern culture	http://jefferson.village.virgina.edu/pmc/contents.all.html
Pseudoscience	suhep.phy.syr.edu/courses/modules/PSEUDO/pseudo_main.html
Crackpots, english language	www.crank.net
Mathematical quotations	http://math.furman.edu/ mwoodard/mquot.html
The World Question Center	http://www.edge.org

Do you want to study physics without actually going to university? Nowadays it is possible to study via e-mail, in german, at the University of Kaiserslautern.* In the near future, a nationwide project in Britain should allow the same for english speaking students. Perhaps you want to read or to recommend as introduction the last update of this physics text.

Si tacuisses, philosophus mansisses.**
After Boethius.

∗ Write to the Universität Kaiserslautern, Postfach 3049, 67653 Kaiserslautern, Germany.
∗∗ ‘If you had kept quiet, you would have remained philosopher.’ After the story Boethius tells in *De consolatione philosophiae*, 2,7, 67 ff.

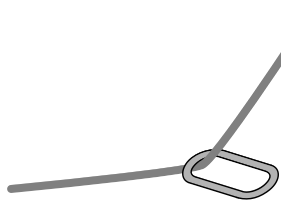


APPENDIX F   LIST OF FIGURES

Picture credits

Typeset in April 2001

Appendix G List of Tables



Typeset in April 2001

Appendix H Challenge Solutions and Hints

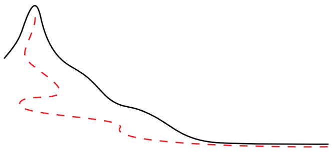
Let me know the challenge for which you want a hint or a solution given here, and I will send it to you and include it in the text in the future.

Page **125**: The height to which an animal can jump is given by the ratio between its mass, proportional to the length l cubed, and its leg muscle strength l^2 times their length l.

Page **754**: For a gaussian integer $n + im$ to be prime, the integer $n^2 + m^2$ must be prime, and in addition, a condition on n mod 3 must be satisfied; which one and why?

This is a section of the freely downloadable e-textbook

Motion Mountain



Hiking beyond space and time along the concepts of modern physics

available at
www.motionmountain.org

To the kind reader

In exchange for getting this section for free, I ask you for a short email on some of the following topics:

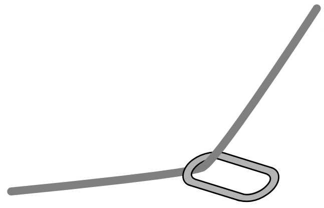
- What was hard to understand?
- What was boring?
- What were you or your friends expecting?
- Challenge - Did you find any mistakes?

Of course, any other suggestion is welcome. This section is part of a physics text written over many years. The text lives and grows through the feedback from its readers, who help to improve and to complete it. For a particularly useful contribution (send it in english, italian, dutch, german or french) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
cs@motionmountain.org

Typeset in
April 2001



Appendix I Subject and Name Index

Typeset in
April 2001

Page numbers in ordinary typeface refer to pages where the keyword is used or the name mentioned. Page numbers in *italic* typeface refer to pages where the keyword is defined or presented in detail.

C

E

G

L

M

Q

X

Y

...
ma la religione di voi è qui
e passa
di generazione in generazione
ammonendo
che Scienza è Libertà.

G. Carducci

Dalla lapide nell'atrio dell'Università di Bologna.*

* '... but the religion of you all is here and passes from generation to generation, admonishing that SCIENCE IS FREEDOM.' From Carducci's text in the entry hall of the University of Bologna, the oldest university of the world.