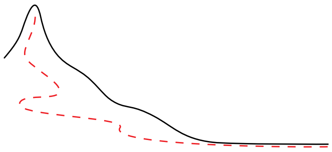
Christoph Schiller



MOTION MOUNTAIN

Hiking beyond space and time

along the concepts of modern physics

available at

www.motionmountain.net

Nunc editio secunda emendata.

Second revised edition

Picture credits are listen on page 933.

The author can be reached at mm@motionmountain.net by email.

To T.

τῷ ἐμοὶ δαὶμονι

Die Menschen stärken, die Sachen klären.

CONTENT OVERVIEW

FIRST PART : CLASSICAL PHYSICS
HOW DO THINGS AND IMAGES MOVE?

SECOND PART : QUANTUM THEORY
WHAT IS MATTER? WHAT ARE INTERACTIONS?

THIRD PART : MOTION WITHOUT MOTION
WHAT ARE SPACE, TIME AND PARTICLES?

APPENDICES

CONTENTS

FIRST PART : CLASSICAL PHYSICS
HOW DO THINGS AND IMAGES MOVE?

SECOND PART : QUANTUM THEORY
WHAT IS MATTER? WHAT ARE INTERACTIONS?

THIRD PART : MOTION WITHOUT MOTION
WHAT ARE SPACE, TIME AND PARTICLES?

APPENDICES

Foreword

Primum movere, deinde docere.

When watching the intensity with which small children explore their environment, we 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, were allowed to thrive in an environment without bounds, reaching from the atoms to the stars? Probably each adolescent would know more about nature than most senior physics teachers. 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 world 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 a focus on concepts and understanding, and a reduction of mathematics to the necessary minimum. It is my aim to focus on the essence of problems and to avoid unnecessary detail. Learning the language of physics is 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 led to numerous results not yet found in books accessible to undergraduates. But also the standard topics of mechanics, electricity, light, quantum theory, particle physics and general relativity are greatly enriched by a systematic collection of modern results scattered around the scientific literature and not yet commonly found in textbooks. However, the topic of this text being the fundamentals of motion, some interesting domains like material physics, biophysics, statistical physics, hydrodynamics, robotics, nanoscience and self-organization are only introduced, but not treated. All topics are in the reach of an undergraduate.

Being *vivid* means to challenge, to question, and to dare. The text is everywhere as provocative as possible without being wrong. I try to make the intellectual interest of the physical description of nature as apparent as possible. All topics should fascinate the normal undergraduate. Many physics texts tend to be commented formula collections which drown the beauty of the topic in pages and pages of formalism. My experience with colleagues, physics students and laymen has shown that the best way to produce interest in a topic is to make a simple and correct, but surprising statement. Surprises confuse, anger, open the mind and thus reveal something about oneself and the world. Surprises widen the inner horizon. Therefore this text is built around a collection of the most astonishing facts known on the topic of motion. The aim is to startle as much as possible. In my view, reading a book on general physics should be similar to a visit to a magician's show. We watch, are astonished, do not believe our eyes, think and finally understand the trick. Nature is similar: many things are different from what they appear. Like a magic show, 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 they may be in 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 to think without using these concepts. The constant exposition to surprises prepares for this aim.

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 made 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.

Books are an old-fashioned means of communication; their possibilities for interactivity are limited. Nevertheless, I want to make it clear that the study of motion, like the rest of science, must be seen as part of the entertainment industry. To achieve this, I tried to make the reader discover and enjoy conceptual physics in the same way that he enjoys a good story. All good stories are adventure stories, and the most intense adventures are those encountered during youth. This text is modelled on the way that children and young people discover and enjoy the world around them. Numerous challenges are proposed: some of research level difficulty (marked r), some difficult (d), some of normal student level (n), and some easy (e). Hints or answers are given in the appendix. In addition, the walk is as structured and as complete as possible. Every topic can be enjoyed by itself and builds only on what was told previously. The idea is to learn, step by step, from what we encounter in everyday life.

Young people are often refreshingly direct. In this adventure, a clear and consistent line of thought is presented, even though it may often be somewhat controversial. On the other hand, young people are flexible; they like to turn situations around and explore them from new and fresh viewpoints. In physics likewise, viewpoints must be changed regularly to get to the bottom of things. For example, the definitions of terms such as ‘object’, ‘particle’, ‘state’, ‘mass’, ‘space’, ‘vacuum’, or ‘time’ are changed several times during this adventure, as they were in the history of the subject.

Young people also like to dare. Courage is needed to drop space and time as tools for the description for the world and to approach nature with the openness of complete freedom from preconceptions. Just ask a physicist whether the world is deterministic or space continuous, and whether he would put his hand into fire about the answer. Emotions will quickly run high: changing thinking habits produces fear. But nothing is more challenging, intense and satisfying than overcoming one’s own fears.

Challenge 1 n

The literature, lectures, and mass media regularly produce statements on physics 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 sometimes amazing, sometimes saddening and sometimes hilarious; but in most cases it blocks the advance of understanding. This text does its best to cut down the number of lies, simply because lies reduce the intensity of life and the strength of people. With the freedom of thought thus achieved, we are ready to intensely enjoy our curiosity and to be fully entertained by its discoveries.

Many people who have kept the flame of their childhood curiosity alive have helped to make this project come true. Saverio Pascazio, Fernand Mayné, Anna Koolen, Ata Masafumi, Roberto Crespi, Serge Pahaut, Valentin Altarez Menendez, Frank van Heyningen and Maria Scali, Luca Bombelli, Herman Elswijk, Marcel Krijn, Marc de Jong, Martin van der

Mark, Kim Jalink, my parents Peter and Isabella Schiller, Mike van Wijk, Renate Georgi, Paul Tegelaar, M. Jamil, Ron Murdock, and most of all my wife Britta have given me encouragement, maybe the most precious of presents. The project and the collection of material owes much to 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, Jon Thaler, Luca Bombelli, Douglas Singleton, George McQuarry, Tilman Hausherr, Brian Oberquell, Peer Zalm, Martin van der Mark, Vladimir Surdin, Julia Simon, Antonio Fermani, Don Page, Stephen Haley Peter Mayr, Allan Hayes, Norbert Dragon, Igor Ivanov, Doug Renselle, Wim de Muynck, Steve Carlip, Tom Bruce, Ryan Budney, Gary Ruben, Chris Hillman, Olivier Glassey, Jochen Greiner, squark, Martin Hardcastle, Mark Biggar, Pavel Kuzin, Douglas Brebner, Luciano Lombardi, Franco Bagnoli, Lukas Fabian Moser, Dejan Corovic, Paul Vannoni, John Haber, Saverio Pascazio, Klaus Finkenzeller, Leo Volin, Jeff Aronson, Roggie Boone, Lawrence Tuppen, Quentin David Jones, Arnaldo Uguzzoni, Frans van Nieuwpoort, Alan Mahoney, Britta Schiller, Petr Danecek, Ingo Thies, Vitaliy Solomatin, Carl Offner, the continuous help by Jonatan Kelu and in particular by the extensive, passionate and conscientious help by Adrian Kubala. The software tools were refined with help from Danie Els, Sebastian Rahtz, Don Story, Vincent Darley, Johan Linde, Joseph Hertzlinger, Rick Zaccone and John Warkentin; the most valuable help in this domain came from Donald Arsenau. The improvements in the graphic design owe much to the suggestions and the experience of my wife Britta.

In a simple view of the world there are three types of adventures: those 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. The gamble of writing paid off if the text can purvey this.

Eindhoven and other places, 25 January 2003

CS

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, amazing, and exciting journey we can make in a lifetime? What is the most interesting place to visit? We can travel to places as remote as possible, like explorers or cosmonauts, we can look into places as far as imaginable, like astronomers, we can visit the past, like historians or archaeologists, or we can delve as deeply as possible into the human soul, like artists or psychologists. All these voyages lead either to other places or to other times (or nowadays, to other servers on the internet). However, we can do better; the most daring trip is not the one leading to the most inaccessible place, but the trip 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 we are free of all restrictions, but also of any security of thought. There, discoveries are still to be made and adventures to be fought. Almost nobody has ever been there; humanity took 2500 years to complete the trip and has achieved it only recently.

To venture into this part of nature, we need to be curious about the essence of travel itself, in particular about its most tiny details. The essence of any travel is *motion*. In principle, the quest to understand motion in all its details can be pursued behind a desk, with a book, some paper and a pen. But to make the adventure more apparent, this text tells the story of the quest as the ascent of a high mountain. Every step towards the top corresponds to a step towards higher precision in the description of motion. At the same time, each step will increase the pleasure and the delights encountered. At the top of the mountain we shall arrive in the domain we were looking for, where 'space' and 'time' are words which have lost all content, and where the sight of the world's beauty is overwhelming and unforgettable.

Try to answer the following questions without ever referring to either space or time: Challenge 2 n

- Can you *prove* that two points extremely close to each other always leave room for a third one in between?
- Can you describe the shape of a knot through the telephone?
- Have you ever tried to make a telephone appointment with a friend without using any time or position term, such as clock, hour, place, where, when, 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 of *any observation at all* which you can describe without concepts from the domains 'space', 'time' or 'object'?
- 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'?
- Can you explain what time is? And what clocks are?

---

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

- Have you ever tried to understand why things move? Or why motion exists at all?

This book tells how to achieve these and other feats, bringing to completion an ancient dream 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 we have 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? We often hear: ‘nobody knows!’ That would imply that nobody knows what space, time and motion really are. However, this is not the case any longer. 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 – at least in big lines.

Why does everything fall downwards? Why do colours of objects differ? 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 but fire not? Why is the universe so big? Why can birds fly but men can’t? Why is lightning not straight? Why are atoms neither square, nor the size of cherries? Why does the floor not fall? Why are computers not faster? Why is all this not different? All these questions seem to have little in common; but the impression is wrong. They are all about motion. Indeed, they all appear and are answered in what follows.

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

What types of bodies are there? What types 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, or even be distinguished exactly, and that they are only approximations of a single, deeper layer of nature, of which in fact they 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, frightening and mysterious sides, is divided into three parts. Don’t panic. All topics will be introduced step by step, in such a way that they all can be understood and enjoyed.

How do things move? The usual answer states that motion is an object changing position over time. This seemingly boring statement encompasses general relativity, one of the most incredible descriptions of nature ever imagined. We find that space is warped, that light does not usually travel straight, and that time is not the same for everybody. We also discover that gravity is not an interaction, but rather the change of time with position in space, that the surface of the earth can be seen as 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? Things are composites of a few types of particles. In addition, 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 made of particles as well. 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 all variations of motion of particles. This story is told in more detail in the second part, that on quantum mechanics. Here we'll learn 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 you incur when buying a can of beans. At that point the path is prepared for the central theme of this mountain ascent:

What are particles, position, and time? The recent results of an age-long search are making it possible to start answering this question. 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 is often told. See page 726
- It seems that position, time, and every particle are aspects of a complex, extended entity incessantly varying in shape.
- 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 we introduce these two concepts in the description of nature. On the other hand, it is impossible *not* to introduce them, because they automatically appear when we divide nature into parts, an act we cannot avoid due to the mechanisms of our senses and of our thinking.
- Research is also presently uncovering that the final description of nature, with complete precision, does not use any form of infinity. We find, 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, we find 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.

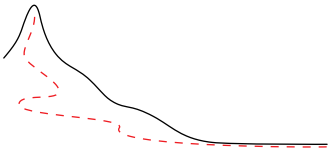
This final part develops the present state of the search for a unified description of general relativity and of quantum mechanics which overcomes their mutual contradictions. This will be one of the most astonishing successes of physics; it will complete the description of motion in all its aspects, from the motion of electrons in the brain to the motion of stars on the other end of the universe. The secrets of space, time, matter and forces have to be unravelled to achieve it. It is a fascinating story, assembled piece-by-piece by thousands of researchers.

In any mountain ascent, every now and then the hiker finds something particularly interesting. Often, he or she then takes a small detour in order to have a closer look. Sometimes an interesting route to be tried in a following trip appears. In this text, the ‘intermezzi’, the sections entitled ‘curiosities’, and the footnotes correspond to such detours. Intermezzi present topic an the border of physics to other sciences, curiosities list physical puzzles and other noteworthy details about motion. 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, both the printed ones and the worldwide web sites, list sources for material used or mentioned in the text. In the electronic version of this text, clicking web site names 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 of printed and electronic information. Lists of tables, figures and challenge solutions are given. The index points to all concept definitions, major topics, mentioned physicists and cited authors. Each of these pages is also interesting to read.

At the end of the mountain ascent, 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

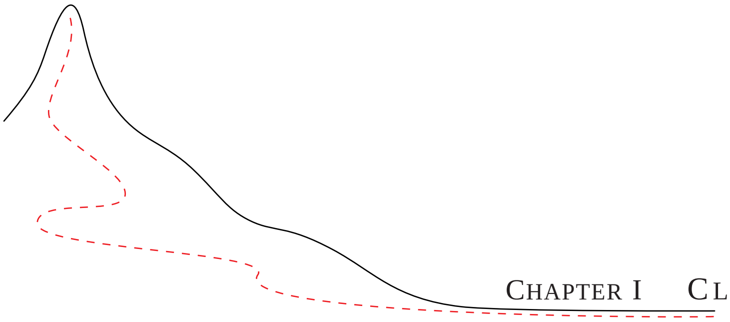


CLASSICAL PHYSICS

HOW DO THINGS AND IMAGES MOVE?

Where the experience of hiking and other motion
leads us to introduce for its description
the concepts of velocity, time, length, mass, and charge,
as well as action, field, and manifold,
allowing us to understand – among others –
why we have legs instead of wheels,
how empty space can be bent, wobble, and move,
what sex has to do with magnets and amber,
and why we can see the stars.

Typeset in
January 2003



CHAPTER I CLASSICAL MOTION

2. Why care about motion?

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

Wham! The lightning crashing in the tree nearby violently disrupts our quiet forest walk and causes our hearts to 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.

Figure 1 An example of motion observed in nature

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 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. The fascination of motion has always made it a favourite subject of curiosity. By the sixth century BCE in ancient Greece, its study had been given Ref. 1 a name: *physics*.

Motion is also important to the human condition. Who are we? Where do we come from? What will we do? What should we do? What will the future bring? Where do people come from? Where

---

* Plants for example cannot move (much); for their self-defence, 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.

** 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.

do they go to? What is death? Where does the world come from?
Where does life lead to? All these questions are about motion. Studying motion will provide answers which are both deep and surprising.

Motion is mysterious. Though found everywhere – in the stars, in the tides, in our eyelids – neither the ancient thinkers nor myriads of others in the following twenty-five centuries have been able to shed some light on the central mystery: *what is motion?* We will discover that the standard reply, ‘motion is the change of place in time’, is inadequate. Just recently an answer has finally been found; this is the story of the way to reach it.

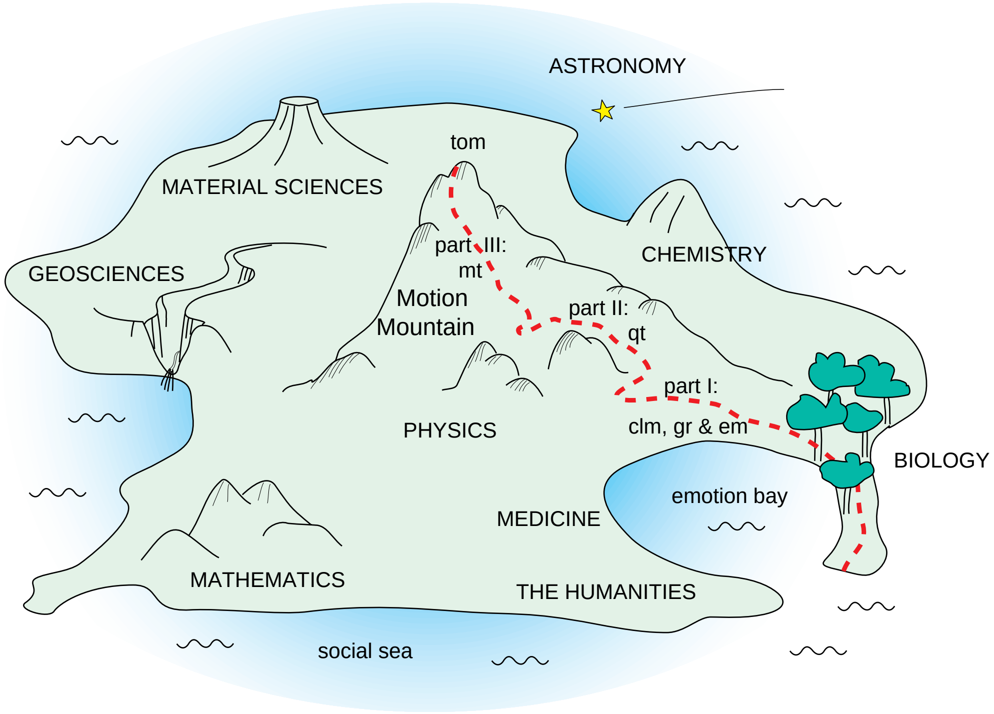


Figure 2 Experience Island, with Motion Mountain and the trail to be followed (clm: classical mechanics, gr: general relativity, em: electromagnetism, qt: quantum theory, mt: M-theory, tom: the theory of motion)

Motion is a part of human experience. If we imagine human experience as an island, then destiny, or the waves of the sea, carried us to its shore. Near the centre of the island an especially high mountain stands out. From its top we can oversee the whole landscape and get an impression of the relationships between all human experiences, in particular between all examples of motion. This is a guide to the top of Motion Mountain, as it is called. Clearly, the first question to ask is:

Does motion exist?

> *Das Rätsel* gibt es nicht. Wenn sich eine Frage überhaupt stellen läßt, so *kann* sie beantwortet werden.[*]

---

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

Ludwig Wittgenstein, *Tractatus*, 6.5

To sharpen the mind for the issue of motion's existence, have a look at Figure 3, and slightly move the page. The figure seems to rotate. How can one make sure that real motion is different from this or some similar illusion?[*] Challenge 3 n

Many scholars simply argued that motion does not exist at all. Their arguments deeply influenced the investigation of motion. For example, the Greek philosopher Parmenides of Elea (born ca. 515 BCE) argued that since nothing comes from nothing, change cannot exist. He underscored the *permanence* of nature and thus consistently maintained that all change and thus all motion is an illusion. Ref. 2 Ref. 3

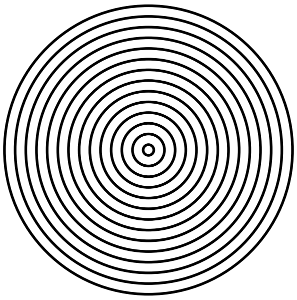


Figure 3 Illusion of motion: fake rotation

Heraclitos (ca. 540–ca. 480 BCE) held the opposite view. He expressed in his sentence πάντα ῥεῖ 'panta rhei' or 'everything flows'.[**] He saw change as the essence of nature, in full contrast to Parmenides. These two equally famous opinions induced many scholars to investigate in more detail whether in nature there are *conserved* quantities or whether *creation* is possible. We will uncover the answer later on; what is your favourite option? Challenge 4 n

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 first has 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. Zeno's argument focuses on the relation between *infinity* and its opposite, finitude, in the description of motion. In modern quantum theory, a similar issue troubles many scientists up to this day. Ref. 4

Zeno also maintained that by looking at a moving object at a *single* instant of time, one cannot maintain that it moves. Zeno argued that at a single instant of time, there is no difference between a moving and a resting body. He thus questioned whether motion can clearly be distinguished from its opposite, *rest*. Like in the history of physics, also during our enquiry we will switch back and forward between a positive to a negative answer. It was this very question that led Albert Einstein to the development of general relativity, one of the high points of our journey. Later on, we will even ask whether single instants do exist at all. This far-reaching question is central to the last part of our adventure.

When we consider quantum theory, we will discover that motion is indeed to a large extend an illusion, as Parmenides claimed. More precisely, we will show that motion is observed only due to the limitations of the human condition. We will find that we experience motion only because we evolved on earth, with a finite size, made of a large but finite number of atoms, with a finite but moderate temperature, electrically neutral, large compared

---

[*] Solution to *challenges* are either given on page 937 or later on in the text. Challenges are classified as research level (r), difficult (d), normal student level (n) and easy (e).

[**] Appendix A explains how to read greek text.

to a black hole of our same mass, large compared to our quantum mechanical wavelength, small compared to the universe, with a limited memory, forced by our brain to approximate space and time as continuous entities, and forced by our brain to describe nature as made of different parts. If any one of these conditions were not fulfilled, we would not observe motion; motion then would not exist. All these results can be uncovered most efficiently if we start with the following question:

How should we 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*.[*]

Like any science, also the approach of physics is double: we advance *precision* and with *curiosity*. Precision makes meaningful communication possible, and curiosity makes it worthwhile.[**] Whenever one talks about motion and aims for increasing precision or for more detailed knowledge, one is engaged, whether knowingly or not, in the ascent of Motion Mountain. With every increase in the precision of description, one gains some height towards the top.

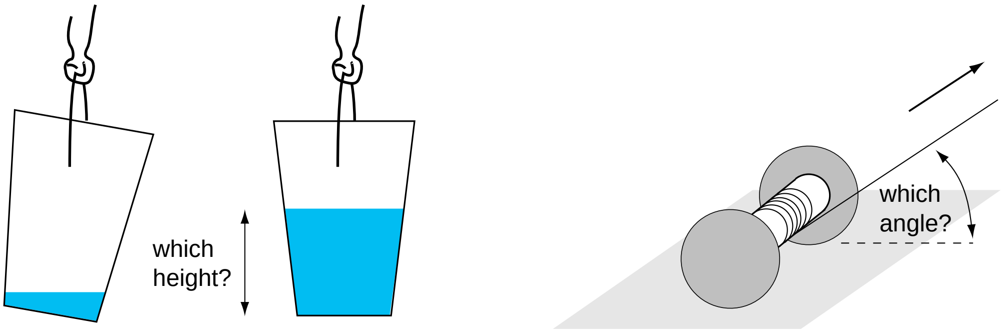


Figure 4 Which water height makes a bucket hang vertically? At which angle does the pulled reel change direction of motion?

High precision means going into small details. This method actually *increases* the pleasure of the adventure.[***] The higher we get on Motion Mountain, the further we can see and the more our 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 side of biology, and directly enters the forest lying at the foot of the mountain. Ref. 6

---

[*] Charles Baudelaire (1821, Paris–1867, Paris) *Beauty*: 'I hate movement, which changes shapes, and never do I cry and never do I laugh.' 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 by JEARL WALKER, *The flying circus of physics*, Wiley, 1975. Ref. 5

[***] Distrust anybody who wants to talk you *out* of investigating details. He is trying to get you. Be vigilant also during *this* walk. Challenge 5 n

Motion type
motion pictures
motion perception
motion for fitness and wellness
motion control in sport
perpetual motion
motion as proof of existence of various gods Ref. 7
economic efficiency of motion
locomotion
motions in parliament
movements in watches
movement teaching and learning
musical movements
religious movements
moves in chess
motion therapy
motion sickness
motion for meditation
motion ability as health check
motion in dance, music and other arts
motion of insects, horses, robots, stars and angels
emotion
commotion
movements in art, sciences and politics
movements in the stock market
movement development in children
troop movements
bowel movements

Table 1 What one can find about motion in a library

Intense curiosity implies to go straight to the limits: understanding motion means to study the largest distances, the highest velocities, the smallest particles, the strongest forces, and the strangest concepts. Let us start.

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 are rather varied. Already in ancient greece people had the suspicion that all types of motion, as well as other types of change, are related. It is usual to distinguish at least three categories.

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

A second category of change 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
Ref. 8 of character is (mostly) a type of transformation.

Ref. 9 The third and especially important category of change is *growth*; it is observed for animals, plants, bacteria, crystals, mountains, stars, and even galaxies. 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 is one of complexity and of time scale.

Figure 5 An example of transport

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 willpower, 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 the 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 volitional movements[**] indeed have 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. Since the two types of motion have the same properties, through the study of motion of non-living objects we can learn something about the human condition. This is most evident when one touches the topics of determinism, causality, probability, infinity, time, and sex, to name but a few of the themes we will encounter on the way.

Ref. 10

---

[*] 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 all the deception and self-deception involved in creating and maintaining 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 his http://www.randi.org web site for more details.

[**] 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.

With the accumulation of observations in the nineteenth and twentieth centuries, more and more restrictions on the study of motion were put into question. Extensive observations showed that all transformations and all growth phenomena, including behaviour change and evolution, are also examples of transport. In the middle of the twentieth century this culminated in the confirmation of an idea already formulated 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*. (Do you agree?) Challenge 6 n 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 path to this result.

Then, in the third part, the particle idea 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 only tell us that classifying the various appearances of motion is not productive. Classifying motion does not allow talking about it with precision. To achieve precision, we need to select a few specific examples of motion, and study them in full detail. Only by trying to achieve maximum precision can we hope to arrive at the fundamental properties of motion.

It is intuitively obvious that the most precise description is achievable for the *simplest* possible examples. In everyday life, this is the case for the motion of any non-living, solid, rigid body in our environment, such as a stone thrown through the air. Indeed, like all humans, each of us learned to throw objects long before learning to walk. Throwing was the Ref. 11 first act we performed ourselves in a chain of events that led 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 explored the properties of motion; first of all we learned that in order to describe and to understand motion we needed 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 Challenge 7 n If you dislike formulas, use the following three-minute method to change the situation. It is worth trying it, as it will make you enjoy this book much more. Life is short, and reading this text should be a pleasure, shouldn't it?

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

---

[*] The importance of throwing is also seen from the terms derived from it: in Latin, words like subject or 'thrown below', object or 'thrown in front', and interjection or 'thrown in between'; in Greek, it led to terms like symbol or 'thrown together', problem or 'thrown forward', emblem or 'thrown into', and – last but not least – devil or 'thrown through'.

[**] The world is independent of my will.

- Open your eyes for a second or two and look at page 229 or any other page of your choice which contains many formulas. Then close your eyes and return to your marvellous experience.
- Open your eyes again and look at page 229, then close them again and return to that marvellous experience.
- Repeat 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 229. 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. Perception starts already before birth, and we continue enjoying it as long as we can. That is why television, even when devoid of content, is so successful. During our walk through this forest at the feet of Motion Mountain, we cannot avoid perceiving. Perception is first of all the ability to *distinguish*. We use the basic mental act of distinguishing in almost every instant of life; for example, during childhood we first learned to distinguish familiar from unfamiliar observations. This is possible only together with another basic ability, namely the capacity to *memorize* experiences. Memory gives us the ability to experience, to talk and thus to explore nature. Perceiving, classifying and memorizing together form *learning*. Without any one of these three abilities, we could not study motion.

Children rapidly learn to distinguish *permanence* from *variability*. aspects. Infants learn to *recognize* human faces, even though faces never look exactly the same each time they are seen. From recognition of faces, children extend recognition to all other observations. Recognition works pretty well in everyday life; it is nice to recognize friends even at night, and even after many beers (not a challenge).

Sitting on the grass in a clearing of the forest at the feet of Motion Mountain, surrounded by the trees and the silence typical of such places, a feeling of calmness and tranquillity envelops us. Suddenly, something moves in the bushes; immediately our eyes turn, the attention focuses. The nerve cells which detect motion are part of the most ancient piece of our brain, shared with birds and reptiles: the brain stem. Then the cortex, or modern brain, takes over to analyse the type of motion and 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 recognize it as a deer, we relax again. But how did we distinguish between landscape and deer? Ref. 12

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 corners of these pages. Each image shows only 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 Ref. 13

object at any given 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.

Like many similar experiments, the flip movie in the lower left corner shows two central connections. First, we perceive motion only if we are able to distinguish an *object* from a *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 an abstraction of – among others – the idea of background. The background is extended; the moving entity is localized.[**] Does his seem boring? Wait a second.

Ref. 13

We call 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. Also the concept of object is therefore a relative concept. The basic conceptual distinction between localized, isolable objects and the extended environment is not trivial or unimportant. First, it smells of a circular definition. (Do you agree?) This question will keep us busy for a while. Second, we are so used to our ability of isolating local systems from the environment 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.

Challenge 8 n

See page 745

> 'Wisdom is one thing: to understand the thought
> which steers all things through all things.'
> Heraclitos of Ephesos

Ref. 15

Does the world need states?

> 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 topics for the other senses are the domain of perception research.

Ref. 14

[**] The topic of motion perception is full of additional aspects. An excellent 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. His motion illusions can be experienced and explored on the associated http://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	
objects	*states*	*images*	*interactions*	*phase space*	*space-time*
impenetrable	global	penetrable	local	composed	simple
aspects:					
mass	instant	intensity	source	dimension	curvature
size	position	colour	domain	distance	topology
charge	momentum	appearance	strength	volume	distance
spin	energy	disappearance	direction	subspaces	area
etc.	etc.	etc.	etc.	etc.	etc.

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

Table 2 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 contradiction. But the state not only distinguishes situations in time. The state contains *all* those aspects of a system (a group of objects) which set it apart from all *similar* systems. Two objects can have the same mass, the same shape, the same colour, the same composition and 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 us 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.* These properties are not boring. Just ponder this: does the universe have a state? Challenge 9 n

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, we can say that *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. The first approximation is based on the precise description of what all children know about motion.

Curiosities and challenges on motion

Motion is not always a simple topic.[**]

- Is the motion of a ghost an example of motion? Challenge 10 n
- Can the universe move? Challenge 11 n
- Can something stop moving? If yes: how would you show it? If not: does this mean that nature is infinite? Challenge 12 n
- To talk about precision with precision, we need to measure it. How would you do that? Challenge 13 n
- Would we observe motion if we had no memory? Challenge 14 n
- What is the lowest speed you have observed? Challenge 15 n
- According to legend, Sessa, the Indian inventor of the game of chess, asked for the following reward for his invention: he wanted one grain of rice for the first square, two for the second, four for the third, eight for the fourth, and so on. How much time do all the rice fields of the world take to produce the necessary rice? Challenge 16 n
- When moving a burning candle, the flame lags behind. How does the flame behave if the candle is inside a glass, still burning, and the glass is accelerated? Challenge 17 n
- A frictionless ball lies near the edge of a perfectly flat and horizontal table. What happens? Challenge 18 n
- You step into a closed box without windows. The box is moved by outside forces unknown to you. Can you determine how you move from inside the box? Challenge 19 n
- What is the length of rope one has to pull in order to lift a mass by a height h with a block and tackle with four wheels? Challenge 20 n

3. Galilean physics – motion in everyday life

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

The simplest description of motion is the one we all, like cats or monkeys, use unconsciously in everyday life: *only one thing at a time can be at a given spot*. (Do you agree?) Challenge 21 n This general description can be separated into three assumptions: matter is *impenetrable* and *moves*, time is made of *instants*, and space is made of *points*. This description of nature

---

[*] 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 not permanent; it shrinks and bends with time, due to processes at the molecular level. To be precise, the length of a piece of wood is 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 – at least for quite while.

[**] Sections entitled ‘curiosities’ are collections of different topics that allow to check and to expand the usage of concepts introduced before.

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

is called *Galilean physics*, or also *Newtonian physics*. Galileo Galilei (1564–1642), Tuscan professor of mechanics, was a founder of modern physics and famous for advocating the importance of observations as checks of statements about nature. By requiring and performing these checks throughout his life, he was led to continuously increase the accuracy in the description of motion. For example, Galileo studied motion by measuring change of position with a self-constructed stopwatch. His approach changed the speculative description of ancient Greece into the experimental physics of Renaissance Italy. The English alchemist, occultist, 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 demonstrate this idea.[*]

What is velocity?

> There is nothing else like it.
> Jochen Rindt[**]

We observe that objects can move in various ways; in particular, they can overtake each other. We also observe that objects can move in different directions. We then observe that velocities can be added. The list of all properties found in Table 3 is summarized by mathematicians with a special term; they say that velocities form a *Euclidean vector space*.[***] More details about this strange term will be given shortly. For now we just note that in describing nature, mathematical concepts offer the most accurate vehicle. See page 55.

When velocity is assumed to be an Euclidean vector, it is called *Galilean* velocity. Do not assume that velocity needs space and time measurements to be defined first. Ref. 18 Are you able to find a means to measure velocities without measuring space and time? Challenge 22 n If so, you probably want to continue reading on page 199, jumping 2000 years of inquiries. If you cannot do so, consider this: whenever we measure a quantity we assume that everybody is able to do so, and that everybody will get the same result. That is, we take *measurement* to be a comparison with a standard. We thus implicitly assume that such a standard exists, i.e. that an example of a 'perfect' velocity can be found. Historically, the study of motion did not investigate this question first, because for many centuries nobody could find such a standard velocity.

---

[*] The best and most informative book on the life of Galileo and his times is by Pietro Redondi (see the footnote on page 165). Galileo was born in the year the pencil was invented. Before his time, it was impossible to do paper and pencil calculations. For the curious, the http://www.mpiwg-berlin.mpg.de web site allows you to read an original manuscript by Galileo. Newton's other hobby, as master of the mint, was to supervise personally the Ref. 17 hanging of counterfeiters. About Newton's infatuation with alchemy, see the books by Dobbs. Among others, Newton believed himself to be chosen by god; he took his latin name, Isaacus Neuutonus, and formed the anagram Jeova santus unus. About Newton and his importance for classical mechanics, see the text by Clifford Ref. 16 Truesdell.

[**] Jochen Rindt, (1942–1970), famous Austrian Formula One racing car driver.

[***] It is named after Euclid, or Eukleides, the great Greek mathematician who lived in Alexandria around 300 BCE. Euclid wrote a monumental treatise of geometry, the *Elements*, which is a milestone of human thought. It sums up all knowledge on geometry of his time. For the first time, Euclid introduces two approaches which are now common use: all statements are deduced from a small number of basic 'axioms', and for every statement a 'proof' is given. The book, still in print, has been the reference geometry text for over 2000 years.

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 3 Properties of Galilean velocity

Velocity is a profound subject for a second reason: we will discover that all properties of Table 3 are only approximate; *none* is actually correct. Improved descriptions will lead us to relativity, quantum theory, and more. But for now, we'll stick with Galilean velocity, and continue with the next Galilean concept: time.

> Without the concepts *place*, *empty* and *time*, change cannot be. [...]
> It is therefore clear [...] that their investigation has to be carried out,
> by studying each of them separately.
> Aristotle

What is time?

> Time does not exist in itself, but only through the perceived objects,
> 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. The term 'object' in fact is just a Latin word meaning 'that which has been thrown in front.' Developmental psychology has shown experimentally that from this experience children extract the concepts of time and space. Adult physicists do the same when studying motion. Ref. 11

If we throw a stone through the air, we can define a *sequence* of observations. Our memory and our senses give us this ability. The sense of hearing registers the various sounds during the rise, the fall and the landing of the stone. Our eyes track the location of the stone from one point to the next. All observations have their place in a sequence, with some observations preceding them, some observations simultaneous to them, and still others succeeding them. We say that observations are perceived to happen at various *instants*, and we call the sequence of all instants *time*. An observation that is considered the smallest part of a sequence, i.e. not itself a sequence, is called an *event*. Events are central to the definition of time; in particular, starting or stopping a stopwatch are events. But do events exist? Keep this question in the back of your head as we move on. Challenge 24 n See page 729

Sequential phenomena have an additional property known as stretch, extension, or duration. *Duration* expresses the idea that sequences *take* time. A sequence takes time means that other sequences can take place together with it.

Observation	Velocity
Stalagmite growth	ca. 0.3 pm/s
Can you find something slower?	challenge Challenge 23 n
Typical motion of continents	10 mm/a=0.3 nm/s
Human growth during childhood	ca. 4 nm/s
Hair growth	ca. 5 nm/s
Tree growth	up to 30 nm/s
Sperm motion	60 to 160 μm/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. 19
Speed of snowflakes	0.5 m/s to 1.5 m/s
Signal speed in human nerve cells	0.5 m/s to 120 m/s Ref. 20
Speed of rain drops, depending on radius	2 m/s to 8 m/s
Fastest swimming fish	ca. 22 m/s
Fastest running animal	ca. 30 m/s
Speed of air in throat when sneezing	ca. 42 m/s
Fastest bird	ca. 85 m/s
Average speed of oxygen molecule in air at room temperature	280 m/s
Sound speed in dry air at sea level and standard temperature	ca. 330 m/s
Record car speed	ca. 340 m/s
Cracking whip end	ca. 500 m/s
Speed of a rifle bullet	ca. 3 km/s
Speed of crack propagation 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. 21
Speed of electrons inside a colour tv	ca. $1 \cdot 10^8$ m/s
Speed of radio messages in space	299 972 458 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 by man	$7 \cdot 10^{13}$ m/s
Highest possible velocity for a light spot	infinite

Table 4 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
Period of caesium ground state hyperfine transition	108.782 775 707 78 ps
Beat of wings of fruit fly	ca. 1 ms
Period of pulsar (rotating star) PSR 1913+16	0.059 029 995 271(2) s
Human 'instant'	ca. 20 ms
Shortest lifetime of living being	ca. 0.3 d
Average length of day 400 million years ago	79 200 s
Average length of day today	86 400.002(1) s
Your 1000 million seconds anniversary	31.7 a
Age of oldest living tree	4600 a
Use of human language	ca. 200 000 a
Age of Himalaya	ca. 35 to $55 \cdot 10^6$ 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 tantalum nucleus ^{180}Ta	ca. 10^{15} a

Table 5 Some time measurements

How exactly is the concept of time, including sequence and duration, deduced from observation? Many people have looked into this question: astronomers, physicists, watchmakers, psychologists, and philosophers. All find that *time is deduced by comparing motions*. Children, beginning at a very young age, develop the concept of 'time' from the comparison of motions in their surroundings. When grown-ups take as a standard the motion of the sun they call the resulting type of time *local time*; from the moon they deduce a *lunar calendar*; if they take a particular village clock on a European island they calls it *universal time coordinate* (UTC), once known as 'Greenwich mean time.'* Astronomers use 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. 11

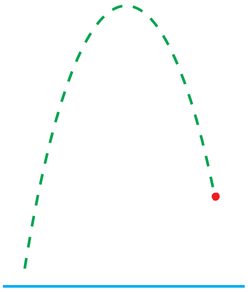


Figure 6 A typical path followed by a stone thrown through the air

See page 879

Note that in the year 2000 an earth rotation does not take 86 400 seconds any more, as it did in the year 1900, but 86 400.002 seconds. Can you deduce in which year your birthday will have shifted by a whole day? Challenge 26 n

---

[*] 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, aeroplanes and trucks. For more information, see the http://www.gpsworld.com web site. The time-keeping infrastructure is also important for other parts of the modern economy as well. Can you spot the most important ones? Challenge 25 n

All methods for the definition of time are based on a common approach; in order to make the concept as precise as possible, a *standard* motion is picked, and with it a standard sequence and a standard duration is defined. The device which performs this task is called a *clock*.[*] We can thus answer the question of the section 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. Note as a curiosity that the word 'moment' is indeed derived from the word 'movement'. Language follows physics in this case. 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. Instead, they should investigate the nature of motion! But this is the aim of our walk anyhow. We are thus set to learn all secrets of time as a side result.

In short, 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? We will continue to study these questions throughout this work and come to a conclusion eventually. 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 do exist; this means that in nature there somehow is an intrinsic, natural and *ideal* way to measure time. Can you see which one? Challenge 27 n

Time is not only an aspect of observations, it is also a facet of personal experience. Even in ones's innermost private life, in one's thoughts, feelings and dreams, one experiences sequences and durations. Children learn to relate this internal experience of time with external observations, and to make use of the sequential property of events in their actions. Studies of the origin of psychological time show that it coincides – apart from its lack of accuracy – with clock time.[**] Every living human necessarily uses in his daily life the concept of time as a combination of sequence and duration; this fact has been checked in numerous investigations. For example, the term 'when' exists in all human languages. Ref. 23

Time is a concept *necessary* to distinguish among observations. In any sequence, we observe 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. Yet between two instants, as close as we can observe them, there is always room for other events. Durations, or *time intervals*, measured by different people with different clocks agree in everyday life; moreover, all observers agree on the order in a sequence of events. Time is thus unique.

---

[*] 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 of a few per cent. Only recently did it became clear what type of clock forms 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, conscious feeling of time. Many other clocks are also part of the human body; the time keepers for shorter times are electrical oscillators at cellular level, such as for the heart beat, and for longer times chemical reactions, such as in the monthly period. Ref. 22

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 6 Properties of Galilean time

These properties, listed in Table 6, form what is called *Galilean time*; it corresponds to the precise version of our everyday experience of time. All these properties can be expressed simultaneously by describing time with 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 can be described by a real number, often abbreviated t, and the duration of a sequence of events is given by the difference between the values for the starting and the final event.

See page 470

Ref. 24

When Galileo studied motion in the 17th century, there were no stopwatches 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?

Challenge 28 n

We will have quite some fun with Galilean time. However, hundreds of years of close scrutiny have shown that *every single* property of time just listed is approximate, and none is strictly correct.

Challenge 29 n

What time is it at the north pole now?

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 expression 'the flow of time' is is often used to convey that nature change follows after change, in a continuous manner. Though the hands of a clock 'flow',** time itself does not.

---

[*] 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-handed, and the right hand has more freedom at the outside of a circle. Since chariot races in stadia went counter-clockwise already thousands

Time is a concept introduced specially to describe the flow of events around us; it does not itself flow, it *describes* 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 389

The misleading use of the expression 'flow of time', propagated first by some Greek thinkers and then again by Newton, continues. Aristotle (384–322 BCE), careful to think logically, pointed out its misconception. Ref. 25 Nevertheless, expressions such as 'time reversal', the 'irreversibility of time', and the much-abused 'time's arrow' are still common. 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 arrows' and compare them with each other. Ref. 26 Predictably, no tangible or new result is extracted.

What is space?

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

Why can we distinguish one tree from another? We see that they are in different positions. Distinguish positions is the main ability of our sense of sight. Whenever 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 as both a container and a background.

Closely related with space and position is *size*, the set of positions an objects occupies. Small objects occupy only subsets of the positions occupied by large ones. We will discuss size shortly.

How do we deduce space from observations? During childhood, humans (and most higher animals) learn to bring 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 description. The result of this learning process is a certain 'image" of space in the brain. Indeed, the question 'where?' can be asked and answered in all languages of the world. Being more precise, adults derive space from distance measurements. The concepts of length, area, volume, angle, and solid angle are all deduced with their help. Geometers,

---

of years ago, all races continue to do so to this day. Also every supermarket leads its guests anticlockwise through the hall. (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 every clock implicitly continues to tell on which hemisphere it was invented.

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

surveyors, architects, astronomers, carpet salesmen, and producers of 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.

By the way, meter bars work well only if they are straight. But 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. Can you describe how humans achieved this? Challenge 30 n

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. Objects can take positions in an apparently *continuous* manner: there are more positions than can be counted.* 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 touched, which is called its *surface*. Length and surface can be measured with help of a meter bar. The length of objects is independent of the person measuring it, of the position of the objects, and of their orientation. In daily life the sum of angles in any triangle is equal to two right angles. There are no limits in space.

Space has three dimensions; we can define sequences of positions in precisely three independent ways. Indeed, the inner ear of (practically) all vertebrates has three semicircular canals that sense the body's position in the three dimensions of space.** Similarly, each human eye is moved by three pairs of muscles. Another proof that space has three dimensions is given by the problems solved by shoe-laces: if space had more than three dimensions, shoe-laces would not be useful, because knots exist only in three-dimensional space. Why three? This is perhaps the most difficult question of physics; it will be answered only in the very last part of our walk.

Challenge 31 n It is often said that thinking in four dimensions is impossible. That is wrong. Just try. For example, can you confirm that in four dimensions knots are impossible?

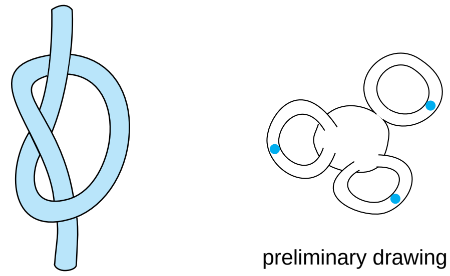


Figure 7 Two proofs of the three-dimensionality of space: a knot and the inner ear of a mammal

Like time intervals, length intervals can be described most precisely with the help of *real numbers*. In order to simplify communication, standard *units* are used, so that everybody uses the same numbers for the same length. Units allows to explore the general properties of *Galilean space* experimentally: space, the container of objects, is continuous, three-dimensional, isotropic, homogeneous, infinite, Euclidean, and unique or 'absolute'. In mathematics, a structure or mathematical concept with all the properties just mentioned is called a three-dimensional *Euclidean space*. Its elements, *(mathematical) points*, are described by three real parame-

---

* For a definition of uncountability, see page 463.

** Note that saying that space has three dimensions *implies* that space is continuous; the mathematician L.E.J. Brower showed that dimensionality is only a useful concept for continuous sets.

Points	Physical property	Mathematical name (see index 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 7 Properties of Galilean space

ters. They are usually written

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

and are called *coordinates*, which specify or order the location of a point in space. (For the precise definition of Euclidean spaces, see page 55.)

What is described here in just half a page actually took two thousand years to be worked out, mainly because the concepts of ‘real number’ and ‘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 of expression (1) are named *Cartesian*.

Like time, space is a *necessary* concept to describe the world. Indeed, space is automatically introduced when we describe situations with many objects. For example, when many spheres lie on a billiard table, we cannot avoid using space to describe the relations among them. In everyday life we cannot avoid using spatial concepts.

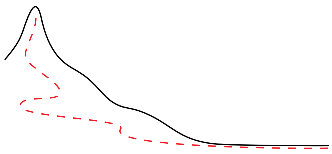
By the way, even though we need space to talk about nature, it is still interesting to ask *why* this is possible. For example, since length measurement methods are possible, there must be a *natural* or *ideal* way to measure distances, sizes and straightness. Can you find it? Challenge 32 n

Like in the case of time, each of the properties of space just listed has to be checked. And again, careful observations will show that each of them is an approximation. In more simple and drastic words, *all* of them are wrong. This confirms Weyl’s statement at the beginning of this section; we will discover that this story is told by every forest in the world, and of course also by the one at the foot of Motion Mountain we are crossing now. We need only listen carefully to what the trees have to tell.

Are space and time absolute or relative?

In everyday life, the concepts of Galilean space and time include two opposing aspects that have coloured every discussion about them for several centuries. On one hand, space and time express something invariant and permanent; they both act like big containers for all

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.net

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

Observation	Distance
Galaxy Compton wavelength	ca. 10^{-85} m (prediction only)
Planck length	ca. 10^{-35} m
Shortest measurable length	ca. 10^{-32} m
Proton diameter	ca. 1 fm
Smallest eardrum oscillation detectable by human ear	ca. 5 pm
Electron Compton wavelength	2.426 310 215(18) pm
Hydrogen atom size	ca. 30 pm
Wavelength of visible light	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
Length of earth's equator	40 075 014.8(6) m
Total length of human nerve cells	$8 \cdot 10^5$ km
Average sun's distance	149 597 870 691(30) m
Light year	9.5 Pm
Distance to typical star at night	ca. 10 Em
Distance Andromeda galaxy	ca. 28 Zm
Size of galaxy	ca. 10 Zm
Most distant visible object	ca. 100 Ym

Table 8 Some distance measurements

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*. Between these two viewpoints, which one do you prefer?The results of physics have alternatively favoured one over the other. We will repeat this alternation throughout our adventure, until we find the solution. Challenge 34 e Ref. 27

Size: why area exists, but volume does not

We saw that a central aspect of objects was their size. All children learn the details of the shape and size of their own body. During this development, which takes place mainly before school age, every human learns how to use the properties of size and space in his actions. As precision-aiming adults, it seems obvious that with the definition of *distance* as the difference between coordinates it is possible to define *length* in a reliable way. It took hundreds of years to discover that this is *not* the case. Several investigations both in physics and in mathematics led to complications.

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

Following the coastline on a map with an odometer, a device shown in the Figure 8, one finds that the length l of the coastline depends on the scale s (say 1/10,000 or 1/500,000) of the map used:

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

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? Can a coastline really have *infinite* length? Yes, it can. In fact, mathematicians have described many such curves, called *fractals*. An infinite number of them exists, and Figure 9 shows one example.* Can you construct another example? Challenge 35 e

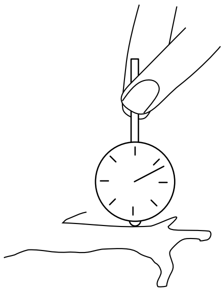


Figure 8 A curvemeter or odometer

Length has other strange properties. The Italian mathematician Giuseppe Vitali discovered that it is possible to cut a line segment of length 1 into pieces which can be reassembled – merely 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 only possible using infinitely many pieces? Challenge 36 d

In summary, length is well-defined for straight and nicely-curved lines, but not for intricate lines, or for lines made of infinitely many pieces. We therefore avoid fractals and other strangely-shaped curves in the following, while being 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 come back to these assumptions in the third part of our adventure.

But all these problems pale when compared to the following one. Commonly, area and volume are defined using length. You think it's easy? You're wrong, as well as a victim of prejudices spread by schools around the world. To define area and volume with precision, their definitions must have two properties: the values must 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 must 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?

---

[*] Most of these curves are *selfsimilar*, i.e. they follow scaling laws similar to the above-mentioned, and are nowadays 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. It is available also in several other languages.

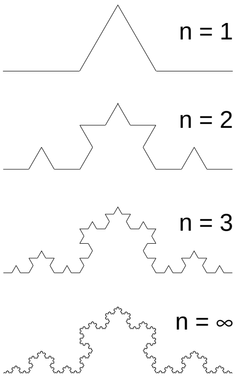


Figure 9 A fractal: a self-similar curve of infinite length, and its construction

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 finite number of straight cuts, one can define an area for all polygons. One can then define area for nicely curved shapes as limit of the sum of infinitely many polygons. This method is called *integration*, and is introduced in detail in the section on physical action.

Challenge 37 n

See page 132

However, integration does not allow us to define area for arbitrarily bounded regions. (Can you imagine such a region?) For a complete definition, more sophisticated tools are needed. They were discovered in 1923 by the famous Polish mathematician Stefan Banach (Krakow, 1892–Lvov, 1945) who 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 the fractal curve just mentioned. Today this generalized concept of area, technically a 'finitely additive isometrically invariant measure,' is called a *Banach measure* in his honour. 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 measure – for any set of points whatsoever.[*]

Challenge 38 n

figure to be finished

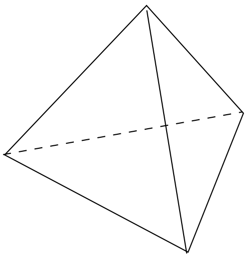


Figure 10 Polyhedra and dihedral angles

What is the situation in *three* dimensions, i.e. for volume? One can start in the same way as for areas, by defining the volume V of a rectangular polyhedron with sides a, b, c as $V = abc$. But then one encounters a first problem: a general polyhedron cannot be cut into a cube by straight cuts! The limitation was discovered in 1900 and 1902 by the German mathematician Max Dehn (1878–1952). He found that the possibility depends on the values of the edge angles, or dihedral angles, as the mathematicians call them. 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 Dehn found that the sum of all the numbers for all the edges of a solid does not change under dissection, provided that the function fulfils $g(\alpha+\beta) = g(\alpha) + g(\beta)$ and $g(\pi) = 0$. An example of such a strange function g is the one assigning the value 0 to any rational multiple of π and the value 1 to a basis set of irrational multiples of π. The values for all other dihedral angles of the polyhedron can then be constructed by combination of rational multiples of these basis angles. Using this function, you may then deduce for yourself that a cube cannot be

Challenge 39 n

---

[*] 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. In addition, 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 level of details looked at, one finds area values from a few square metres up to over 100 m^2.

dissected into a regular tetrahedron because their Dehn invariants are different.[*]

Despite the problems with Dehn invariants, one *can* define a rigid and additive concept of volume for polyhedra, since for all of them and in general for all 'nicely curved' shapes, one can again use integration for the definition of their volume.

Now let us consider general shapes and general cuts in three dimensions, not just the 'nice' ones mentioned so far. One then gets the famous *Banach-Tarski theorem* (or paradox). Ref. 28 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 their theorem says: take any two sets not extending to infinity and containing a solid sphere each; then 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![**] Volume is not a useful concept at all.

The Banach-Tarski theorem raises two questions: can we do this with gold or with bread? That would solve many problems. Can we do this blowing up with empty space? In other Challenge 40 n words, are matter and empty space continuous? Both topics will be studied in the rest of our walk. For the moment, we eliminate the troubling issue altogether by restricting our interest to smoothly-curved shapes. With this restriction, volumes of matter and of empty space behave nicely: they are additive and rigid, and show no paradoxes. Nevertheless, we 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 you see a solid object with a straight edge, it is a 99%-safe bet to come to the conclusion that 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 flat, in the city most objects are. How is it possible for us to make straight objects while there are none to be found in nature?

Traditionally we call a line *straight* if it touches either a plumb-line or a light ray along its whole length. In fact, the two definitions are equivalent. Can you confirm this? Can you Challenge 41 n find another definition? Obviously, we call a surface *flat* if for any chosen orientation and position it touches a plumb-line or a light ray along its whole extension.

We note that straightness and flatness are concepts defined with help of bodies or radiation. In nature, spatial concepts, like temporal concepts, require bodies or motion for their definition.

---

[*] This is also told in the beautiful book by M. AIGLER & G.M. ZIEGLER, *Proofs from the Book*, Springer Verlag, 1999. The title is due to the famous habit of the great mathematician Paul Erdös to imagine that all beautiful mathematical proofs are assembled in the 'book of proofs'.

[**] The proof of the result does not need much mathematics; it is explained beautifully by Ian Stewart in *Paradox of the spheres*, New Scientist, 14 January 1995, pp. 28–31. 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 beautiful book by Ref. 29 Wagon.

[***] Exceptions are some crystalline minerals. Other candidates which might come to mind, such as certain Ref. 30 bacteria which have (almost) square or (almost) triangular shapes are not real exceptions.

A hollow earth?

Space and straightness pose subtle challenges. Some strange people maintain that all humans live on the *inside* of a sphere; they call this the *hollow earth theory*. They pretend that the moon, the sun and the stars are all near the centre of the hollow sphere. They also explain that light follows curved paths in the sky and that when usual physicists talk about a distance r from the centre of the earth, the real hollow earth distance is $r_{\mathrm{he}} = R^2_{\mathrm{earth}}/r$. Can you show that this model is wrong? Challenge 42 n The great Austrian physicist Roman Sexl (1939–1986) used to ask this question to his students and fellow physicists. The answer is simple: if you think you have an argument to show that this view is not correct, you made a mistake! There is *no way* to show that such a view is wrong. It is possible to explain the horizon, the appearance of day and night and the satellite images of the round earth.To explain what happened during Challenge 43 e the flight to the moon is also fun. A coherent hollow earth view is *equivalent* to the usual picture of an infinitely extended space. We will come back to this problem in the section on general relativity. See page 337

Curiosities and fun challenges on everyday space and time

Here are a few questions to make you think.

- How often in 24 hours do the hour and minute hands of a clock lie on top of each other? Challenge 44 n How often does this happen for clocks having also a hand for seconds?
- How many times in twelve hours can the two hands of a clock be *exchanged* with the Ref. 5 result that the new situation shows a *valid* time? What happens for clocks having also a third Challenge 45 n hand for seconds?
- How many minutes does the earth turn in one minute? Challenge 46 n
- In 1996 the smallest experimentally probed distance was 10^{-19} m, achieved between Ref. 72 quarks at Fermilab. What does this mean for the continuity of space? Challenge 47 n
- ‘Where am I?’ is a common question; ‘when am I?’ is never used, not even in other languages. Why? Challenge 48 n
- Is there a smallest time interval in nature? A smallest distance Challenge 49
- Given that you know what straightness is, how would you characterize the curvature of a line using numbers? And that of a surface? Challenge 50 n
- What is the speed of your eyelid? Challenge 51 n
- Zeno studied what happens to a moving object at a given instant of time. To discuss with him, you decide to build the fastest possible shutter for a photographic camera that you can imagine. You have all the money you want. What is the shortest shutter time you would achieve? Challenge 52 n
- Can you prove Pythagoras’s theorem by geometrical means alone, without using coordinates?(There are more than 30 possibilities.) Challenge 53 e
- Why are most planets and moons (almost) spherical? Challenge 54 n
- A rubber band connects the tips of the two hands of a clock. What is the path followed by the middle point of the band? Challenge 55
- Both the sun and the moon seem larger when they are on the horizon. Ptolemy already explains this illusion by an unconscious apparent distance change of the human brain. In fact, the moon is further away from the observer when it is just above the horizon, and thus

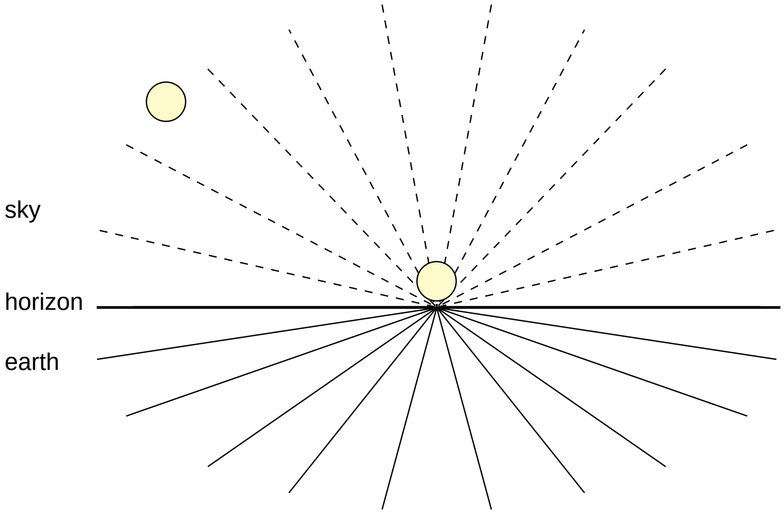


Figure 11 How the apparent size of the moon and the sun changes

its image is smaller than a few hours earlier, when it was high in the sky. Can you confirm this? Challenge 56 n

- Cylinders can be used to move a plank over the floor; they keep the plank always at the same distance from the floor. What cross sections *other* than a circle allow to realize the same feat? How many examples can you find? Challenge 57 n
- Also Galileo made mistakes. In his famous book, the *Dialogues*, he says that the curve formed by a thin chain hanging between two nails is a parabola. That is not correct. What is the correct curve? Challenge 58 n
- How does a *vernier* work? It is called *nonius* in other languages, derived from the Latin word for 'nine'. Can you design a vernier which instead of increasing the precision tenfold, does so by an arbitrary factor? Is there a limit? Challenge 59 n
- Draw three circles, of different sizes, that touch each other. Now draw a fourth circle in the space between, touching the outer three. What simple relation do the inverse radii of the four circles obey? Challenge 60 n
- With two rulers, one can add and subtract numbers by lying them side by side. Are you able to design rulers which allow to multiply and divide in the same manner? Challenge 61 n More elaborate devices using this principle were called *slide rulers* and were the precursors of electronic calculators; they were in use all over the world until the 1970s.
- How many days would a year have if the earth turned the other way with the same rotation frequency? Challenge 62 n

How to describe motion: kinematics

> Il libro della natura è scritto nella lingua della matematica.*

---

* The book of nature is written in the language of mathematics.

Galileo Galilei

Experiments show that the properties of Galilean time and space are extracted from the environment by most higher animals and by young children. Later, when children learn to speak, they put these experiences into concepts, as was just done above. With help of the concepts just introduced, grown up children then say that *motion is change of position with time*. This description is illustrated by flipping rapidly the lower left corners of the pages of this book. 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 other variations from one picture to the next, due to the imperfections of printing techniques, even simulate 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, this rephrasing is useful because it allows for high *precision*, as we will find out soon. After all, precision is our guiding principle during this promenade. The detailed description of changes in position is traditionally called *kinematics*.

The set of all positions taken by an object over time forms a *path* or *trajectory*. The origin of this concept is evident when one watches fireworks* or again the flip movie in the lower left corner of this part of the mountain ascent. With the description of space and time by real numbers, a 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 464.) 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, 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 about one part in 300, for all falling bodies on all points of the surface of the earth. Where do 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 will occupy us during the larger part of this hike. Ref. 31

Equation (3) allows to determine the depth of a well given the time a stone takes to reach is bottom. The equation 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 yields a velocity of 27 km/h. The velocity is thus proportional only to the square root of the height; does it mean that fear of falling results from an overestimation of its actual effects? Challenge 63 n Challenge 64 n

---

* 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.

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

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

a *complete* description for the path followed by thrown stones results. A path of this shape is called a *parabola* and is shown in Figure 12.[*] A (rotated) parabola is also the shape used for light reflectors inside pocket lamps or car headlights. Can you show why? Challenge 65

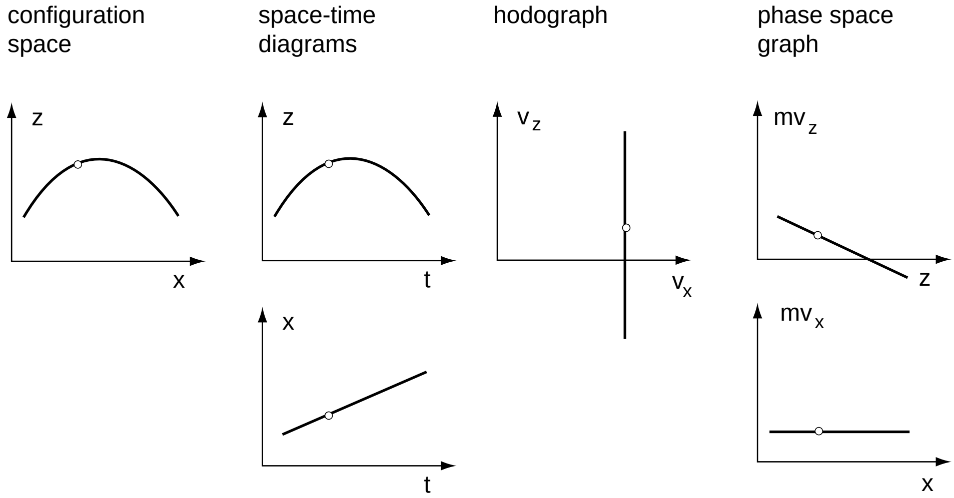


Figure 12 Various types of graphs describing the same flying stone

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

- What is the distance one can reach with a stone, given the speed and the angle from the horizontal at which it is thrown? Challenge 66 n
- How can the speed of falling rain be measured with an umbrella? Challenge 67 n
- What is the maximum numbers of balls that could be juggled at the same time? Challenge 68 n
- What is an upper limit for the long jump record? Use as input that the running speed world record in 1997 is 12 m/s ≈ 43 km/h by Ben Johnson, and the women's record is 11 m/s ≈ 40 km/h. In fact, long jumpers never run much faster than about 9.5 m/s. How much could they win in jump distance if they could run full speed? How could they achieve that? In addition, long jumpers take off at angles of about 20 °, as they are not able to achieve a higher angle at the speed they are running. How much would they gain if they could achieve 45 °? Ref. 32 Ref. 73 Challenge 69 n
- Are gun bullets falling back after being fired into the air dangerous? Challenge 70 n
- Is it true that rain drops would kill if it weren't for the air resistance of the atmosphere? Challenge 71 n

---

[*] Apart from the graphs shown in Figure 12, 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. The phase space diagram is also called state space diagram.

The last two questions arise because equation (3) does not hold in all cases. For example, leaves or potato chips do not follow it. As already Galileo knew, this is a consequence of air resistance; we will discuss it shortly.

In fact, even without air resistance, the path of a stone would not always be a parabola; can you specify such situations? Challenge 72 n

What is rest?

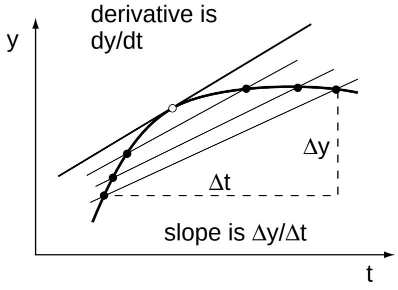


Figure 13 Derivatives

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 impressions, 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 can define the *velocity* **v** of an object as the change of its position **x** with time t, 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 in order to remind the reader 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 13. From this definition follow the working rules Challenge 73 e

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

c being any number. This is all one ever needs to know about derivatives. The quantities dt and dy, sometimes useful by themselves, are called *differentials*. These concepts are due to the Saxon lawyer, physicist, mathematician, philosopher, diplomat, and historian Gottfried Wilhelm Leibniz (1646, Leipzig–1716, Hannover). Derivatives lie at the basis of all calculations based on the continuity of space and time.

Indeed, the definition of velocity assumes that it makes sense to take the limit $\Delta t \to 0$, in other words, that infinitely small time intervals do exist in nature. The definition of velocity with derivatives is possible only because both space and time are described by sets which are *continuous*, or in mathematical language, *connected*. In the rest of our walk we should never forget that right from the beginning of classical physics, *infinities* are present in its description of nature. In fact, differential calculus can be defined as the study of infinity and its uses. We thus discover straight away that the appearance of infinity does not automatically render a description impossible or imprecise. (In fact, we will only use the smallest two of the various types of infinities. They and several other types are introduced in the intermezzo following this chapter.) See page 462

The appearance of infinity in the usual description of motion was first criticized in his famous ironical arguments by Zeno of Elea (around 445 BCE), a disciple of Parmenides. In his well-known 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 criticized by Zeno actually works quite well in everyday life. The reason is simple but deep: changes in daily life are continuous. Large changes are made up of many small changes. Ref. 33 Challenge 74 e

This property of nature is not obvious. For example, we note that we have tacitly assumed that the path of an object is not a fractal nor some other badly behaved entity. Is this correct? For everyday life, it is. But the result does not apply in all domains of nature. In fact, Zeno will be partly rehabilitated later in our walk, and the more so the more we will proceed. For the moment though, we have no choice: we continue with the basic assumption that in nature changes happen smoothly. See page 729

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. The concept of velocity is obviously a member of this list. Continuing in the same way, we call *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. They add little to the description of nature, as 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. 34

---

* 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 the definition of a unique null vector and of a single negative vector for each vector in it. Challenge 75 n

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

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

Observation	acceleration
Acceleration at equator due to earth's rotation	$0.34\,\mathrm{mm/s^2}$
Centrifugal acceleration due to the 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 the earth's surface, depending on location	$9.8 \pm 0.1\,\mathrm{m/s^2}$
Standard gravitational acceleration	$9.806\,65\,\mathrm{m/s^2}$
Fastest car accelerated by wheels	ca. $15\,\mathrm{m/s^2}$
Gravitational acceleration on Jupiter's surface	$240\,\mathrm{m/s^2}$
Acceleration of cheetah	ca. $32\,\mathrm{m/s^2}$
Fastest leg acceleration (insects)	ca. $2\,\mathrm{mm/s^2}$
Tennis ball against wall	ca. $0.1\,\mathrm{Mm/s^2}$
Bullet acceleration in rifle	ca. $5\,\mathrm{Mm/s^2}$
Fastest centrifuges	ca. $0.1\,\mathrm{Gm/s^2}$
Acceleration of protons in large accelerator	ca. $90\,\mathrm{Tm/s^2}$
Acceleration of protons inside nucleus	ca. $10^{31}\,\mathrm{m/s^2}$
Highest possible acceleration in nature, $\sqrt{c^7/\hbar G}$	$5.6 \cdot 10^{52}\,\mathrm{m/s^2}$

Table 9 Some acceleration values found in nature

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 from being 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. Challenge 76 e

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 its parts no longer need to be taken into account is called a *particle*. (The older term *corpuscule* has fallen out of fashion.) Particles are thus idealized little

---

In coordinate notation, the standard scalar product is given by the number $a_x b_x + a_y b_y + a_z b_z$. Whenever 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: $a = \sqrt{\mathbf{aa}}$.

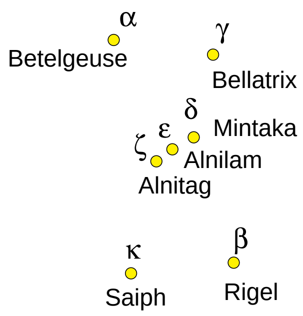
[*] If I know an object I also know all its possible occurrences in states of affairs.

Ref. 35 [**] Matter is a word derived from the Latin 'materia', which originally meant 'wood' and was derived via intermediate steps from 'mater', meaning 'mother'.

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

Do point-like 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 measure angular sizes as small as $2\,\mu$rad can be measured, a limit given by the fluctuations of the air in the atmosphere. In space, such as 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 'point-like", even when seen with the most powerful telescopes.

One can even see the difference between 'point-like' sources and finite size ones with the naked eye alone: at night, stars twinkle, planets do not. (Check it!)This effect is due to the turbulence of air. Turbulence makes the almost point-like stars twinkle because it deflects light rays by very small amounts; but air turbulence is too weak to lead to twinkling of sources of larger angular size, such as planets or satellites. Challenge 77 e



picture to be included

Figure 14 Betelgeuse

In fact, 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. Ref. 36 Of course, all other stars have a finite size as well, like the sun has, but one cannot prove this by measuring dimensions on pictures. (True?) Challenge 78 n

An object is *point-like for the naked eye* if its angular size is smaller than about $2' = 0.6$ mrad. Can you estimate the size of a 'point-like' dust particle? Challenge 79 n By the way, an object is *invisible* to the naked eye if it is point-like 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 from the moon, or from the space shuttle? Challenge 80 n

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

---

[*] The web site http://www.astro.uiuc.edu/~kaler/sow/sowlist.html gives an introduction to the different types of stars. The http://www.astro.uiuc.edu/~dolan/constellations/constellations.html web site provides detailed and interesting information about constellations.

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 le système solaire*, Springer Verlag, 1993. The most beautiful pictures of the stars can be found in D. MALIN, *A view of the universe*, Sky Publishing and Cambridge University Press, 1993. Fascinating stories about what people do to take such pictures are told by P. MANLY, *Unusual telescopes*, Cambridge University Press, 1991.

question will be central in the last two parts of our walk. In fact we have also forgotten to ask and to check whether points in space do exist. Our walk will lead us to the astonishing result that all the answers to these questions are negative. Can you imagine how this could be proven? Do not be disappointed if you find this difficult; many brilliant minds have had the same problem. Challenge 81 n

However, many particles, such as electrons, quarks, or photons are point-like for all practical purposes. 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. The simplest description, the *continuum approximation*, describes extended bodies as an infinite collection of point particles. It allows to understand and to predict the motion of milk and honey, the motion of the air in hurricanes, and of perfume in rooms. Also the motion of fire and all other gaseous bodies, the bending of bamboo in the wind, the shape changes of chewing gum, and the growth of plants and animals can be described in this way. Ref. 37

A better approximation than the continuum one is described shortly. Nevertheless, all observations have confirmed that the motion of large bodies can be described to high precision as the result of the motion of their parts. This approach will guide us through the first two parts of our mountain ascent. Only in the third part we will discover that at a fundamental scale, this decomposition cannot be possible.

Legs and wheels

Shape is an important aspect of bodies: among others, it allows us to count them. For example, one finds 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 or 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 argument is not correct, as Figure 15 shows. Can you find an example for this kind of motion in your own body? Are you able to see how many cables may be attached to the rotating body of the figure without hindering the rotation? See Appendix D Ref. 38 Challenge 82 n Challenge 83 n

Despite this possibility, such a rotating part still cannot make a wheel. Can you see why? Challenge 84 n

In summary, whenever one observes a construction in which some part is turning continuously – and without the ‘wiring’ of the figure – one knows immediately that it is an artefact: 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, seeds from various trees, some animals, children and dancers sometimes move by rotating as a whole. Ref. 39

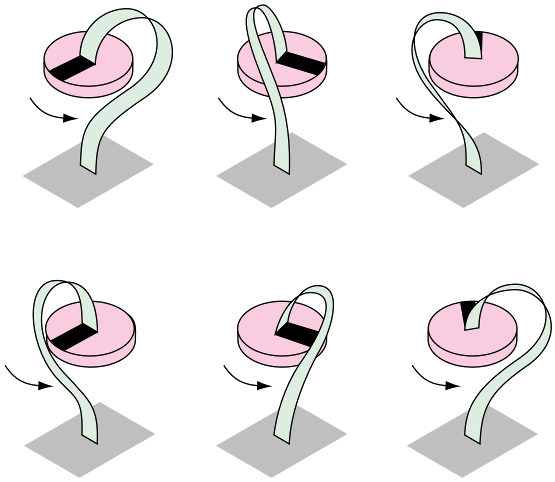


Figure 15 How an object can rotate continuously without tangling up the connection to a second one

Single bodies, and thus all living beings, can only move through *deformation* of their shape: therefore they are limited to walking or running, to crawling, and to flapping wings or fins. In contrast, systems of several bodies, such as bicycles, pedal boats or other machines, 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 16 Legs and wheels in living beings

However, like so many statements about living creatures, this one also has exceptions. The distinction between one and two bodies is poorly defined if the whole systems has only a few molecules. This happens most clearly inside bacteria. Organisms such as Escherichia coli, the well-known bacterium found in the human gut, or bacteria from the Salmolella family, all swim using flagella. *Flagella* are thin filaments, similar to tiny hairs sticking out of the cell membrane. In the nineteen seventies it was shown that each flagellum, made of one or a few long molecules with a diameter of a few tens of nanometres, does in fact turn See page 710 about its axis. A bacterium is able to turn its flagella in both clockwise and anticlockwise directions, can achieve more than a thousand turns per second, and can turn all its flagella Ref. 40 in perfect synchronization. Therefore wheels actually do exist in living beings, albeit only tiny ones. But let us now continue with our study of simple objects.

---

[*] 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, often outperform human built machines. As an example, compare the size of the smallest flying systems built by evolution with those built by humans. The discrepancy has two reasons. First of all, nature's systems have integrated repair and maintenance systems. Second, nature can build large structures inside containers with small openings. In fact, nature is very good at building sailing ships Challenge 85 n inside glass bottles. The human body is full of such examples; can you name a few?

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. Both objects and images are able to move. Running tigers, falling snowflakes, and material ejected by volcanoes are examples of motion, as they change position over time. For the same reason, the shadow following our body, the beam of light circling the tower of a lighthouse on a misty night, and the rainbow that constantly keeps the same apparent distance from the hiker are examples of motion.

Everybody who has ever seen an animated cartoon in the cinema knows that images can move in more surprising ways than objects. Images can change their size, shape, and even colour, a feat only few objects are able to perform.* Images can appear and disappear without trace, multiply, interpenetrate, go backwards in time, and defy gravity or any other force. Images, even usual shadows, can even move faster than light. Images can float in space and keep the same distance to approaching objects. Objects cannot do almost anything of this. In general, the 'laws of cartoon physics' are rather different from those in nature. In fact, the motion of images does not seem to follow any rules at all, in contrast to the motion of objects. Together, both objects and images differ from their environment in that they have *boundaries* defining their size and shape. We feel the need for precise criteria allowing the two cases to be distinguished. Ref. 42

The clearest distinction between images and objects is made with the same method that children or animals use when they stand in front of a mirror for the first time: they try to *touch* what they see. Indeed, if we are able to touch what we see – or more precisely, if we are able to move it – we call it an *object*, otherwise an *image*.** One cannot touch images, but one can touch objects. And as everybody knows, touching something means to feel it resisting movement. Certain bodies, such as butterflies, pose little resistance and are moved with ease, others, such as ships, resist more, and are moved 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 575

Summing up, for the description of motion one must distinguish bodies, which can be touched and are impenetrable, from images, which cannot and are not. Everything visible is either an object or an image; there is no third possibility. (Do you agree?) If the object is so far away that it cannot be touched, such as a star or a comet, it can be difficult to Challenge 87 n

---

* 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, television, computer displays, heated objects, and certain lasers can do it. Do you know more examples? 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 colour*, 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 colour in the outdoors*, Springer, 1993, an updated version based on his wonderful book series, *De natuurkunde van 't vrije veld*, Thieme & Cie, Zutphen. Reading it is a must for all natural scientists. Challenge 86 n Ref. 41

** 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 552, and with elementary particles.

decide whether one is dealing with an image or an object; we will encounter this difficulty repeatedly. For example, how would you show that comets are objects and not images? Challenge 88 n

Moving images are made of *radiation* in the same way that objects are made of *matter*. Images are the domain of shadow theatre, cinema, television, computer graphics, belief systems and drug experts: photographs, motion pictures, ghosts, angels, dreams, and many hallucinations are images. To understand images, we need to study radiation; but due to the importance of objects – after all we are objects ourselves – we will study material bodies first. Ref. 43

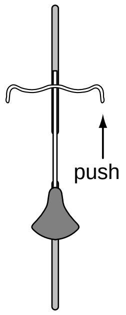


Figure 17 In which direction does the bicycle turn?

Motion and contact

> Democritus affirms that there is only one type of movement:
> That resulting from collision.
> Aetius, *Opinions*.

Ref. 44

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

But contact is not the only way to put something into motion; a counter-example is an apple falling from a tree or a magnet pulling another. Non-contact influences are more fascinating: nothing is hidden, but nevertheless something mysterious happens. Contact motion seems easier to grasp, and that is why one usually starts with it. However, we will soon find out that taking this choice one makes a similar experience that bicycle riders make. When riding a bicycle at sustained speed and trying to turn left by pushing the right side of the steering bar, one takes a *right* turn.[*] In other words, despite our choice 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.[**]
> Archimedes (ca. 283–212), as cited by Pappus.

Ref. 45

When we push something we do not know, such as when we kick some object on the street, we automatically pay attention to the same two aspects that children explore when they stand before a mirror for the first time, or when they see a red laser spot for the first time. We all check whether the unknown entity can be pushed, and pay attention to how much the

---

Challenge 89 n

[*] This surprising effect obviously works only above a certain minimal speed. Can you determine which one? Be careful! Too strong a push will make you fall.

[**] 'Give me a place to stand, and I'll move the earth.' Thus already Archimedes knew that the distinction used by lawyers between movable and immovable property makes no sense.

unknown object moves. Higher precision is possible with experiments like the one shown in Figure 18. Repeating the experiment with various pairs of objects, one notes that a fixed quantity m_i can be ascribed 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 change produced by the collision. The number m_i is called the *mass* of the object.

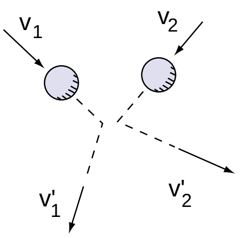


Figure 18 Collisions define mass

In order to get mass values common to everybody, the mass of one particular, selected object has to be fixed in advance. This special object 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 object in the world is determined.

The mass thus measures the difficulty of getting something moving. High masses are harder to move than low masses. Obviously, only objects 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. 35

Experiments also show 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 system is the sum of the mass of the components. In short, *Galilean mass is a measure for the quantity of matter.*

The definition of mass can also be given in another way. We can ascribe a number m_i to every object i such that for collisions free of outside interference the following sum is unchanged *throughout* the collision:

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

The product of the velocity $\mathbf{v}_\mathrm{i}$ and the mass m_i is called the *momentum* of the body. The sum, or *total momentum* of the system, is the same before and after the collision; it is a *conserved* quantity. The two conservation principles (13) and (14) were first stated in this way by the important Dutch physicist Christiaan Huygens.*

As a consequence, if a moving sphere hits a resting one of the same mass, a simple rule determines the angle between the directions the two spheres take after the collision. Can you find this rule? It is useful when playing billiards. Challenge 90 n

---

* Christiaan Huygens (1629, 's Gravenhage –1695, Hofwyck) was one of the main physicists of his time; he clarified the concepts of mechanics; he also was one of the first to show that light is a wave.

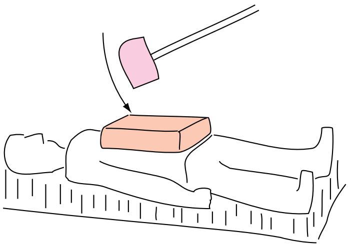


Figure 19 Is this dangerous?

Another consequence was shown on the cover photograph of the CERN Courier in 1994. It 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. As the impact is mostly absorbed by the concrete, there is no pain and no danger – except if the concrete is missed. Why? Challenge 91 n

The above definition of mass has been generalized by the physicist and philosopher Ernst Mach* in such a way that it is valid even if the two objects interact without contact, as long as they do so along the line connecting their positions. The mass ratio between two bodies is defined as negative acceleration ratio, thus as

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

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

- The definition of mass *implies* the conservation of momentum $\sum mv$. Momentum conservation is *not* a separate principle. 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 the products $m_1 a_1$ and $-m_2 a_2$. Such products are called *forces*. The equality of acting and reacting forces is not a separate principle; mass is defined in such a way that it holds.
- The definition of mass is *independent* of whether contact is involved or not, and 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 measuring masses of the bodies around us, we discover its main properties. Mass is *additive* in everyday life,, as the mass of two bodies combined is equal to the sum of the two separate masses. Furthermore, mass is *continuous*; it can seemingly take any positive value. Finally, mass is *conserved*; the mass of a system, defined as the sum of the mass of all constituents, does not change over time if the system is kept isolated from the rest of the world. Mass is conserved not only in collisions: also during melting, evaporation, digestion, and all other processes mass does not appear or disappear.

---

[*] Ernst Mach (1838, Chrlice–1916), Austrian physicist and philosopher. The *mach* unit for aeroplane speed as a 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 led to Mach's principle, which we will discuss later on. He was also proud to be the last scientist denying – humorously, and against all evidence – the existence of atoms.

[**] As mentioned above, only *central* forces obey the relation (15) used to define mass. Central forces act between the centre of mass of bodies. We give a precise definition later on. But since all fundamental forces are central, this is not a restriction. There seems to be one notable exception: magnetism. Is the definition of mass valid in this case? See page 71 Challenge 92 n

Mass values	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
do not change	conservation	invariance
do not disappear	impenetrability	positivity

Table 10 Properties of Galilean mass

Later we will find that also in the case of mass all these properties are only approximate. Precise experiments show that none of them are correct.[*] For the moment we continue with the present, Galilean concept of mass, as we have no better one at our disposal.

In a famous experiment in the 19th century, for several weeks a man lived with all his food and drink supply, and also his toilet, on a large balance. How did the measured weight change with time? Challenge 93 n

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 measurement is impossible. Can you find a better way to determine the weight of the earth? Challenge 94 n

In summary, 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. Indeed, a *negative* (inertial) mass would mean that such a body would move in the opposite direction of any applied force or acceleration. 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 would still fall downwards in the field of a large positive mass (though slower than an equivalent positive mass). Are you able to confirm this?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 value would stay at constant distance and spontaneously accelerate away along the line connecting the two masses.Note that both energy and momentum are conserved in all these situations.[**] Negative-mass bodies have never been observed. Antimatter, which will be discussed later, also has positive mass.

Challenge 95 e

Challenge 96 e

See page 226 and 550

---

[*] In particular, in order to define mass we must be able to *distinguish* bodies. This seems a trivial requirement, but we discover that this is not always possible in nature.

[**] 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 walls, and accelerating, i.e. losing energy, at the same time. They would allow building of a perpetuum mobile of the second kind, i.e. a device circumventing the second principle of thermodynamics.Moreover, such a system would have no thermodynamical equilibrium, because its energy could decrease forever. The more one thinks about negative mass, the more one finds strange properties contradicting observations. By the way, what is the range of possible mass values for tachyons?

Challenge 97

Challenge 98 n

Observation	Mass
Mass increase through absorption of one green photon	$3.7 \cdot 10^{-36}$ kg
Lightest known object: electron	$9.1 \cdot 10^{-31}$ kg
Atom of argon	$39.962\,383\,123(3)$ u $=$ $66.359\,1$ yg
Human at early age	10^{-11} kg
Water adsorbed onto a kilogram metal weight	ca. 10^{-8} kg
Planck mass	$2.2 \cdot 10^{-8}$ kg
Fingerprint	ca. 10^{-7} kg
Typical ant	ca. 10^{-7} kg
Water droplet	ca. 10^{-6} kg
Honey bee	$1 \cdot 10^{-4}$ kg
Largest living being	ca. 10^{6} kg
Largest ocean goingship	ca. $400 \cdot 10^{6}$ kg
Largest object moved by man (Troll gas rig)	$687.5 \cdot 10^{6}$ kg
Large antarctic iceberg	10^{15} kg
Water on earth	10^{21} kg
Solar mass	$2.0 \cdot 10^{30}$ kg
Our galaxy	ca. 10^{41} kg
Total mass visible in the universe	ca. 10^{54} kg

Table 11 Some mass values

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 during 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 it can be said of both 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 in collisions is always useful when thinking about the description of moving objects.

Momentum is conserved. That explains the limitations you might experience when being on a perfectly frictionless surface, such as ice or a Polished, oil covered marble: you cannot propel yourself forward by patting 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. Neither are humans. Can you imagine why?)

Challenge 99 n

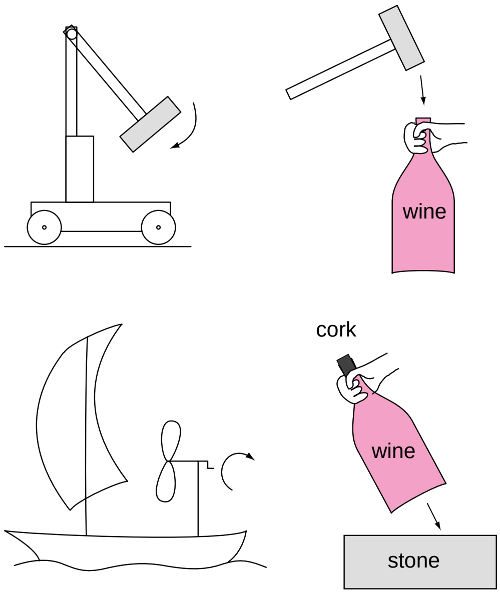


Figure 20 What happens?

The conservation of momentum and mass also means that teleportation ('beam me up') is impossible in nature. Can you explain this to a non-physicist? Challenge 100 n

Momentum conservation implies that motion never stops; it is only *exchanged*. On the other hand, motion often disappears in our environment, as in the case of a stone dropped to the ground, or of a ball left rolling on grass. Moreover, in daily life we often observe creation of motion, such as every time we open 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. A muscle only *transforms* one type of motion, namely that of the electrons in certain chemical compounds* 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 wheels. Both systems need fuel and get warm in the process.

We must also study the microscopic behaviour when a ball rolls on grass until it stops. The disappearance of motion is called *friction*. Studying the situation carefully, one finds that the grass and the ball heat up a little during this process. During friction, visible motion is transformed into heat. Later, when we discover the structure of matter, it will become 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 object is at rest, but is warm. Heat is a form of motion. Friction thus only seems to be disappearance of motion; in fact it is a transformation of ordered into unordered motion.

Despite momentum conservation, *macroscopic* perpetual motion does not exist, since friction cannot be eliminated completely.** Motion is eternal only at the microscopic scale.

Ref. 46 * Usually adenosinetriphosphate (ATP), the fuel of most processes in animals

** Some funny examples of past attempts to built a *perpetual motion machine* are described in STANISLAV MICHEL, *Perpetuum mobile*, VDI Verlag, 1976. Interestingly, the idea of eternal motion came to Europe from India, via the islamic world, around the year 1200, and became popular as it opposed the then standard view that all motion on earth disappears over time. See also the www.geocities.com/mercutio78_99/pmm.html web site. The conceptual mistake made by all eccentrics and used by all crooks is always the same: the hope of overcoming friction.

If the machine is well constructed, i.e. with little friction, it can take the little energy it needs for the sustenance of its motion from very subtle environmental effects. For example, in the Victoria and Albert Museum in London one can admire a beautiful clock powered by the variations of air pressure over time. Ref. 47

Small friction means that motion takes a long time to stop. One immediately thinks of the motion of the planets. In fact, there *is* friction between the earth and the sun. (Can you guess one of the mechanisms?) But the Challenge 101 n

The disappearance and also the spontaneous appearance of motion in everyday life is an illusion, due to the limitations of our senses. For example, the motion proper to every living being exists before its birth, and stays 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', or living force, by Leibniz and many others.

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

More on conservation

When collisions are studied even more, a second conserved quantity turns up. Experiments show that in the case of perfect, or elastic collisions – collisions without friction – the following quantity, called the *kinetic energy* T of the system, is also conserved:

$$T = \sum_i \frac{1}{2} m_i \mathbf{v}_i^2 = \text{const} \quad . \qquad (16)$$

The factor 1/2 and the name 'kinetic energy' were introduced by the French engineer and mathematician Gustave-Gaspard Coriolis (Paris, 1792– Paris, 1843) so that the relation $dT/dv = p$ would be obeyed. (Why?) Energy is a word taken from ancient Greek; originally it was used to describe character, and meant 'intellectual or moral vigour'. It was taken into physics by William Thomson and William Rankine around 1860 because its literal meaning is 'force within.' Challenge 102 n

(Physical) *energy* measures the ability to generate motion. A body has a lot of energy if it has the ability to move many other bodies. Energy is a number; it has no direction. The total momentum of two equal masses moving with opposite velocities is zero; the total energy depends on the velocity values. Energy thus also measures motion, but in a different way than momentum does. Energy measures motion in a more global way.

Do not be surprised if you do not grasp the difference between momentum and energy straight away: physicists took about two centuries to figure it out. For some time they even insisted on using the same word for both of them, and often they didn't know which situation required which concept. So you are allowed take a few minutes to think about the topic.

Both energy and momentum measure how systems change. Momentum tells how systems change over distance, energy measures how systems change over time. Momentum is needed to compare motion here and there. Energy is needed to compare motion now and later.

One way to express the difference between energy and momentum is to think about the following challenges. Is it 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 103 e

value is so small that the earth has already circled around the sun for thousands of millions of years, and will do so for quite some time.

Observation	Energy value
Kinetic energy of oxygen in air	ca. 10^{-21} J
Green photon energy	ca. 10^{-19} J
X-ray photon energy	ca. 10^{-15} J
γ photon energy	ca. 10^{-12} J
Highest particle energy in accelerators	ca. 10^{-7} J
Comfortably walking human	20 J
Flying arrow	50 J
Right hook in boxing	ca. 50 J
Energy in flashlight battery	ca. 1 kJ
Flying rifle bullet	10 kJ
Apple digestion	0.2 MJ
Car on highway	1 MJ
Highest laser pulse energy	1.8 MJ
Lightning flash	up to 1 GJ
Planck energy	2.0 GJ
Small nuclear bomb (20 kton)	84 TJ
Earthquake of magnitude 7	2 PJ
Largest nuclear bomb (50 Mton)	210 PJ
Impact of meteorite with 2 km diameter	ca. 1 EJ
Rotation energy of earth	$2 \cdot 10^{29}$ J
Supernova explosion	ca. 10^{44} J
Gamma ray burst	up to 10^{47} J
Energy content of sun's mass	$1.8 \cdot 10^{47}$ J

Table 12 Some energy measurements

Another distinction is taught by athletics: the *real* long jump world record, almost 10 m, is still kept by an athlete who in the early 20th century ran with two weights in his hands, and then threw the weights behind him in the moment he took off. Can you explain the feat? Challenge 104 n

When a car travelling at 100 m/s runs frontally into a parked car of the same make, which car has the larger damage? What changes if the parked car has its brakes on? Challenge 105 n

To get a better 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 420 EJ, for a world population of about 6000 million people. Ref. 48 To see what this energy consumption means, translate it into a personal power consumption; one gets about 2.2 kW. The Watt W is the unit of power, and is simply defined as 1 J/s, reflecting the definition of *(physical) power* as energy per time. As a working person can produce mechanical work for about 100 W, the average human energy consumption corresponds to about 22 humans working 24 hours a day. In particular, if one looks at the energy consumption in countries of the first world, the average inhabitant there has machines working for him equivalent to several hundred 'servants'. Can you point out some of these machines? Challenge 106 n

Kinetic energy is thus not conserved in everyday life. For example, in non-elastic collisions, like that of a chewing gum and a wall, kinetic energy is lost. *Friction* destroys kinetic energy, as it destroys momentum. At the same time, friction produces heat. It was one of the important conceptual discoveries of physics that *total* energy is conserved if one includes the discovery that heat is a form of energy. Friction is thus in fact a process transforming kinetic energy, i.e. the energy connected with the motion of a body, into heat. On a microscopic scale, energy is conserved.* Indeed, without energy conservation, the concept of time would not be definable. We will show this connection shortly.

Rotation

Rotation keeps us alive. Without the change of day and night, we would be either fried or frozen to death, depending on our location on our planet. A short summary of rotation is thus appropriate. We saw before that a body is described by its reluctance to move; 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 *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 outside world:

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

---

* In fact, the conservation of energy was stated in its modern form only in 1842, by Julius Robert Mayer. He was a medical doctor by training, and the journal 'Annalen der Physik' refused to publish his paper, as it supposedly contained 'fundamental errors.' What the editors called errors were in fact the contradictions with their prejudices. Later on, Helmholtz, Kelvin, Joule, and many others acknowledged Mayer's genius. Today, energy conservation is one of the pillars of physics, as it is valid in all its domains.

Observation	Power
Power of flagellar motor in bacterium	...
Incandescent light bulb light output	1 to 5 W
Incandescent light bulb electricity consumption	25 to 100 W
A human, during one work shift	ca. 100 W
One horse, for one shift	ca. 300 W
Eddy Merckx, the great bicycle athlete, during one hour	ca. 500 W
Official horse power	ca. 735 W
Large motor bike	100 kW
Electrical power station	100 to 6000 MW
World's electrical power production in 2000	450 GW
Input on earth surface: sun's irradiation of earth Ref. 49	0.17 EW
Input on earth surface: thermal energy from inside of earth	32 TW
Input on earth surface: power from tides (i.e. from earth's rotation)	3 TW
Input on earth surface: power generated by man from fossil fuels	8 to 11 TW
Lost from earth surface: power stored by plants' photosynthesis	40 TW
World's record laser power	ca. 1 PW
Output of earth surface: sunlight reflected into space	0.06 EW
Output of earth surface: power radiated into space at 287 K	0.11 EW
Sun's output	384.6 YW

Table 13 Some power measurements

Quantity	Linear motion		Rotation	
State	time	t	time	t
	position	x	angle	φ
	momentum	$p = mv$	angular momentum	$L = \Theta\omega$
	energy	$mv^2/2$	energy	$\Theta\omega^2/2$
Motion	velocity	v	angular velocity	ω
	acceleration	a	angular acceleration	α
Reluctance to move	mass	m	moment of inertia	Θ
Motion change	force	ma	torque	$\Theta\alpha$

Table 14 Correspondence between linear and rotational motion

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

$$\Theta = \sum_{\mathrm{n}} m_{\mathrm{n}} \rho_{\mathrm{n}}^2 \quad , \tag{18}$$

where r_{n} is the distance from the mass element m_{n} to the axis of rotation. Can you confirm the expression?Therefore, the moment of inertia of a body depends on the chosen axis of rotation. Can you confirm this for a brick? Challenge 107 e Challenge 108

Obviously, the value of the moment of inertia also depends on the location of the axis used for its definition. For each axis direction, one distinguishes *intrinsic* moment of inertia, when the axis passes through the centre of mass of the body, from *extrinsic* moment of

Observation	Angular velocity
Galactic rotation	$= 2\pi\,/220\,000\,000\,\mathrm{a}$
Average sun rotation around its axis	ca. $2\pi\cdot 3.8\cdot 10^{-7}/\,\mathrm{s} = 2\pi\,/\,30\mathrm{d}$
Typical lighthouse	ca. $2\pi\cdot 0.08/\,\mathrm{s}$
Jumping ballet dancer	ca. $2\pi\cdot 3/\,\mathrm{s}$
Ship's diesel engine	$2\pi\cdot 5/\,\mathrm{s}$
Helicopter motor	$2\pi\cdot 5.3/\,\mathrm{s}$
Washing machine	up to $2\pi\cdot 20/\,\mathrm{s}$
Bacterial flagella	ca. $2\pi\cdot 100/\,\mathrm{s}$
Racing car engine	up to $2\pi\cdot 600/\,\mathrm{s}$
Fastest turbine built	ca. $2\pi\cdot 10^3/\,\mathrm{s}$
Fastest pulsars (rotating stars)	up to $2\pi\cdot 10^3/\,\mathrm{s}$
Ultracentrifuge	¿ $2\pi\cdot 2\cdot 10^3/\,\mathrm{s}$
Proton rotation	ca. $2\pi\cdot 10^{20}/\,\mathrm{s}$
Highest possible, Planck angular velocity	ca. $2\pi\cdot 10^{35}/\,\mathrm{s}$

Table 15 Some rotation speeds

Challenge 110 n

inertia, when it does not.* In the same way, one distinguishes intrinsic and extrinsic angular momenta. (By the way, the *centre of mass* of a body is that imaginary point which moves straight during vertical fall, even if the body is rotating. Can you find a way to determine its location for a specific body?)

Challenge 111 n

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

$$\mathbf{L} = \mathbf{r}\times\mathbf{p} = \frac{2\mathbf{A}(T)m}{T} \quad \text{so that} \quad L = r\,p = \frac{2A(T)m}{T} \tag{20}$$

where $\mathbf{A}(T)$ is the surface swept by the position vector of the particle during time T.** The angular momentum thus points along the rotation axis, following the right hand rule.

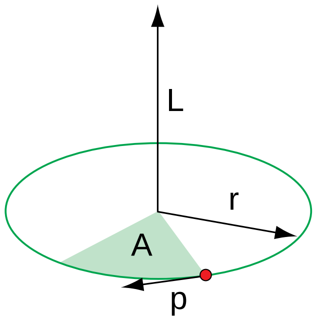


Figure 21 Angular momentum

We then define a corresponding *rotational energy* as

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

---

[*] Extrinsic and intrinsic moment of inertia are related by

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

Challenge 109 n

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

[**] For the curious, the result of 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 orientation is given by the *right hand rule*, and whose length is given by $ab\sin\sphericalangle(\mathbf{a},\mathbf{b})$, i.e. by the surface area of the parallelogram spanned by the two vectors. From the

Can you guess how much larger the rotational energy of the earth is compared with the yearly electricity usage of humanity? If you can find a way to harness this energy, you will become famous. Challenge 114 n

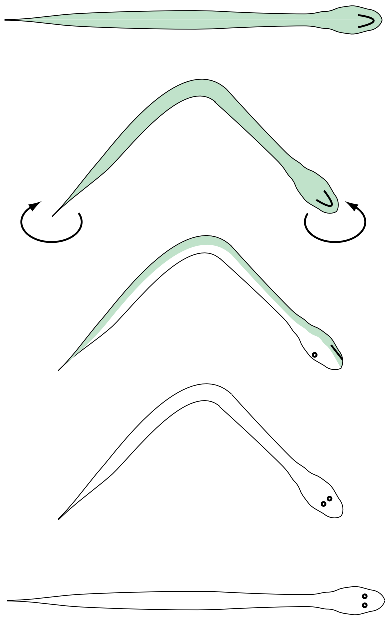


Figure 22 How a snake turns itself around its axis

As in the case of linear motion, rotational energy and angular momentum are not always conserved in the macroscopic world, due to friction; but they are always conserved on the microscopic scale.

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, we need to push *against* something. Is this also the case for rotation?

Surprisingly, it is possible to turn even *without* pushing against something. You can check this on a well-oiled rotating office chair: simply rotate 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 land feet first. Also snakes know how to rotate themselves, as Figure 22 shows. Ref. 50 During the Olympic games one can watch board divers and gymnasts perform similar tricks. Rotation is thus different from translation in this aspect. Why? Challenge 115

Rolling wheels

Rotation is an interesting phenomenon in many ways. A rolling wheel does *not* turn around its axis, but around its point of contact. Let us show this.

A wheel of radius R is *rolling* if the speed of the axis v_{axis} is related to the angular velocity by

$$\omega = \frac{v_{\text{axis}}}{R} \quad . \tag{23}$$

Challenge 112 e definition you can show that the vector product has the properties

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

See page 911 The vector product exists (almost) only in three-dimensional vector spaces. (See Appendix D) The cross product vanishes if and only if the vectors are parallel. The parallelepiped spanned by three vectors **a**, **b** and **c** has Challenge 113 e thevolume $V = \mathbf{c}(\mathbf{a}\times\mathbf{b})$. The pyramid or tetrahedron formed by the three vectors has one sixth of that volume.

For any point P on the wheel, with distance r from the axis, the velocity v_P is the sum of the motion of the axis and the motion around the axis. Figure 23 shows that v_P is orthogonal to d, the distance between the point P and the contact point of the wheel. The figure also Challenge 116 e showsthat the length ratio between v_P and d is the same as between v_axis and R. As a result, we can write

$$\mathbf{v}_\mathrm{P} = \boldsymbol{\omega} \times \mathbf{d} \tag{24}$$

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

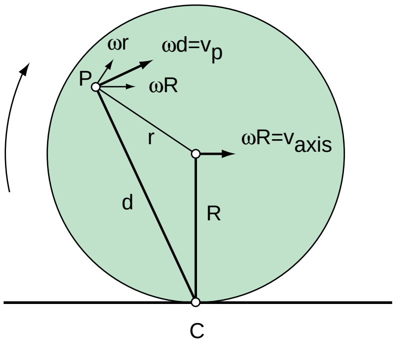


Figure 23 The velocities and unit vectors for a rolling wheel

Surprisingly, when a wheel rolls, some points on it move towards the wheel's axis, some stay at fixed distance, and others move away from it. Can you determine where these various Challenge 117 n points are located? Together, they lead to interesting pictures when a rolling wheel with spokes, such as a bicycle wheel, is Ref. 51 photographed.

With these results you can tackle the following beautiful Ref. 52 challenge. When a turning bicycle wheel is deposed on a slippery surface, it will slip for a while and then end up rolling. How does the final speed depend on the initial speed and on Challenge 118 d the friction?

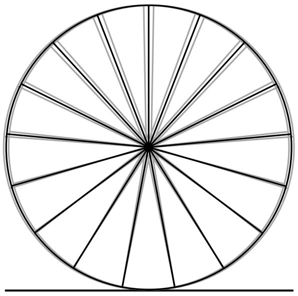


Figure 24 A simulated photograph of a rolling wheel with spokes

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, when a body movement is performed with as little energy as possible, it is natural and graceful. (This can indeed be taken as the actual definition of grace. The connection is common Ref. 10 knowledge in the world of dance; it is also a central aspect of the methods used by actors to learn how to move their bodies as beautifully as possible.)

To convince yourself about the energy savings, try walking or running with your arms fixed or moving in the opposite direction than usual: the effort is considerably higher. In fact, when a leg is moved, it produces a torque around the body axis which has to be counterbalanced. The method using the 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 those of the legs, the hips. Animals on two legs but without arms, such as penguins or pigeons, have more difficulty walking; they have to move their whole torso with every step.

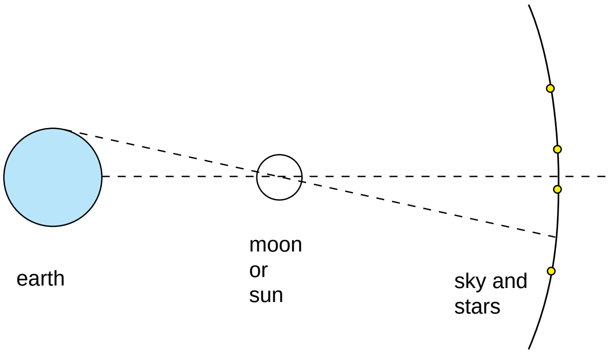


Figure 25 The parallaxis - obviously not drawn to scale

Which muscles do most of the work when walking, the motion which experts call *gait*? In 1980, Serge Gracovetsky found that in human gait most power comes from the *spine* muscles, not from the legs. (Note that people without legs are also able to walk.) When you take a step, the lumbar muscles straighten the spine; this automatically makes it turn a bit to one side, so that 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. In fact, one can experience the increase in tension in the *back* muscles when walking without moving the arms,thus confirming where the human engine is located. Ref. 53 Challenge 119 e

Is the earth rotating?

Eppur si muove![*]

The search for answers to this question gives a beautiful cross-section of the history of classical physics. Already around the year 265 BCE, the Greek thinker Aristarchos of Samos maintained that the earth rotates. He had 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 by an observer on the surface of the earth and when seen by a hypothetical observer at its centre. Aristarchos noticed that the moon and the sun wobble across the sky, and this wobble has a period of 24 hours. He concluded that the earth rotates. Ref. 54

Measurements of the the aberration of light also show the rotation of the earth; it can be detected with telescopes while looking at the stars. The *aberration* is a change of the expected light direction which we will discuss shortly. At the equator, earth rotation adds an angular deviation of $0.32'$, changing sign every 12 hours, to the aberration due to the See page 200

---

[*] ‘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 165).

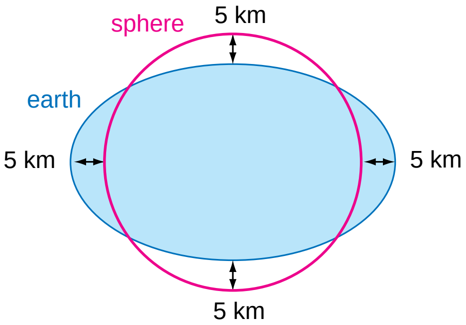


Figure 26 Earth's deviation from spherical shape due to its rotation

motion of the earth around the sun, about $20.5'$. In modern times, astronomers had found a number of additional proofs, but none was accessible to the man on the street.

Also the measurements showing that the earth is not a sphere, but *flattened* at the poles, showed the rotation of the earth. Again however, this measurement result by Maupertuis* in the 18th century is not accessible to direct observation.

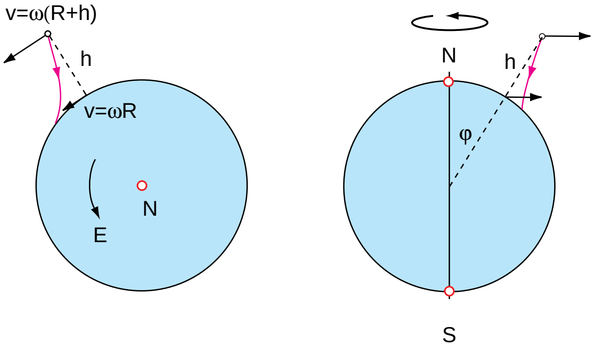


Figure 27 The deviations of free fall towards the east and towards the equator, both due to the rotation of the earth

Then, in the years 1790 to 1792 in Bologna, Giovanni Battista Guglielmini (1763–1817) finally succeeded to measure what Galileo and Newton had predicted to be the simplest proof for the earth's rotation. On the earth, objects do not fall vertically, but are slightly deviated to the east. This deviation appears because an object keeps the larger horizontal velocity it had at the height from which it started falling, as shown in Figure 27. Guglielmini's result was the first non-astronomical proof of the earth's rotation. The experiments were repeated in 1802 by Johann Friedrich Benzenberg (1777–1846). Using metal balls which he dropped from the Michaelis tower in Hamburg – a height of 76 m – Benzenberg found

---

* Pierre Louis Moreau de Maupertuis (1698–1759), French physicist and mathematician. He was one of the key figures in the quest for the principle of least action, which he named in this way. He was also founding president of the Berlin Academy of Sciences.

that the deviation to the east was 9.6 mm. Can you confirm that the value measured by Benzenberg almost agrees with the assumption that the earth turns once every 24 hours?(There is also a much smaller deviation towards the equator, not measured by Guglielmini, Benzenberg or anybody after them up to this day; however, it completes the list of effects on free fall by the rotation of the earth.) Both deviations are easily understood if we remember that falling objects describe an ellipse around the centre of the rotating earth. The elliptical shape shows that the path of a thrown stone does not lie on a plane for an observer standing on earth; for such an observer, the exact path thus cannot be drawn on a piece of paper. Challenge 120

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', found a closely related effect that nobody had noticed in everyday life up to then. An object travelling in a rotating background does not move on a straight line. If the rotation is counterclockwise, as is the case for the earth on the northern hemisphere, the velocity of objects is slightly turned to the right, while its magnitude stays constant. This so-called *Coriolis acceleration* (or Coriolis force) is due to the change of distance to the rotation axis. Can you deduce the analytical expression for it, namely $\mathbf{a}_{\mathrm{C}} = 2\omega \times \mathbf{v}$? Challenge 121

The Coriolis acceleration determines the handiness of many large scale phenomena with a spiral shape, such as the directions of cyclones and anticyclones in meteorology, the general wind patterns on earth and the deflection of ocean currents and tides. Most beautifully, the Coriolis acceleration explains why icebergs do not follow the direction of the wind as they drift away from the polar caps. The Coriolis acceleration also plays a role in the flight of canon balls (that was the original interest of Coriolis), in satellite launches, in the motion of sunspots and even in the motion of electrons in molecules. All these phenomena are of opposite sign on the northern and southern hemisphere and thus prove the rotation of the earth. (In the first world war, many ship cannons missed their targets in the southern hemisphere because the engineers had compensated them for the Coriolis effect on the northern hemisphere.) Ref. 56 Ref. 55

Only in 1962, after several earlier attempts by other researchers, Asher Shapiro was the first to verify that the Coriolis effect has a tiny influence on the direction of the vortex formed by the water flowing out of a bathtub. More than a normal bathtub he had to use a carefully-designed experimental set-up, because contrary to an often-heard assertion, no such effect can be seen in real bathtubs. He succeeded only by carefully eliminating all disturbances from the system; for example, he waited 24 hours after the filling of the reservoir (and never actually stepped in or out of it!) in order to avoid any left-over motion of water which would disturb the effect, and built a carefully designed, completely rotationally-symmetric opening mechanism. Others have repeated the experiment on the southern hemisphere, confirming the result. In other words, the handiness of usual bathtub vortices is *not* caused by the rotation of the earth, but results from the way the water starts to flow out. But let us go on with the story about earth's rotation. Ref. 57 Ref. 57

Finally, in 1851, the French physician turned physicist Jean Bernard Léon Foucault (1819, Paris–1868, Paris) performed an experiment which cleared all doubts and which

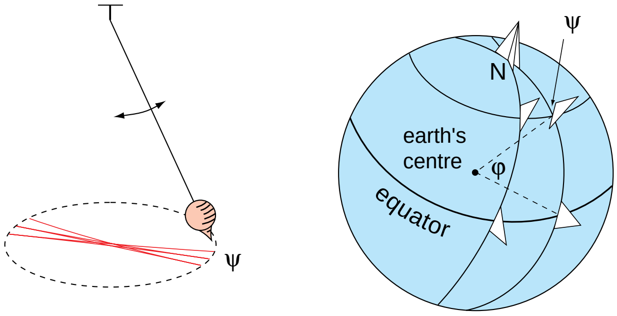


Figure 28 The motion of a pendulum on the rotating earth

rendered him world-famous practically overnight. He suspended a 67 m long pendulum* in the Panthéon in Paris and showed the astonished public that the direction of its swing changed over time, rotating slowly. To everybody with a few minutes of patience to watch the change of direction, the experiment proved that the earth rotates. More precisely, the rotation period T_F of the oscillation plane seen on earth is given by Challenge 123 n

$$T_\mathrm{F} = \frac{24\,\mathrm{h}}{\sin\varphi} \qquad (25)$$

where φ is the latitude of the location of the pendulum, e.g. 0° at the equator and 90° at the north pole. This formula is perhaps the most beautiful result of Galilean kinematics. Ref. 58

Foucault is also the inventor and namer of the *gyroscope*. He built the device, shown in Figure 29, in 1852, one year after his pendulum. With it, he again demonstrated the rotation of the earth. Once it rotates, the axis stays fixed in space, but only when seen from far of the earth. For an observer on earth, the axis direction changes regularly with a period of 24 hours. Gyroscopes are now routinely used in ships and in aeroplanes to give the direction of north, because they are more precise and more reliable than magnetic compasses. In the most modern versions, one uses laser light running in circles instead of rotating masses.**

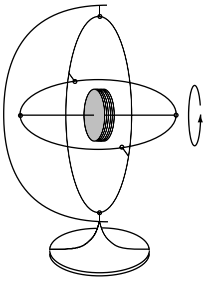


Figure 29 A gyroscope

In 1909, Roland von Eötvös measured a simple effect: due to the rotation of the earth, the weight of a object depends on the direction in which it moves. As a result, a balance in rotation around the vertical axis does not stay perfectly horizontal: the balance starts to oscillate slightly. Can you explain the origin of the effect? Challenge 125 n

In 1910, Eduard Hagen published the results of an even simpler experiment, proposed by Louis Poinsot in 1851. If two masses on a horizontal bar are slowly moved towards the support, as shown in Figure 30, and if the friction is kept low enough, the bar rotates. Obviously, this would not happen if the earth were not rotating. Can you explain the observation? Challenge 126 n

---

Challenge 122 * Why was such a long pendulum necessary?Understanding the reasons allows one to repeat the experiment at home, using a pendulum as short as 70 cm, with help of a few tricks. Ref. 59

Challenge 124 n ** Can you guess how rotation is detected in this case?

This not-so-known effect is also useful for winning bets among physicists.

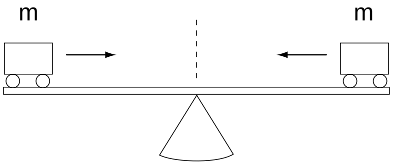


Figure 30 Showing the rotation of the earth through the rotation of an axis

In 1925, Albert Michelson* and his collaborators in Illinois constructed a vacuum interferometer with the incredible perimeter of 1.9 km. Interferometers produce bright and dark fringes of light; the position of the fringes depends on the way the interferometers rotates. The fringe shift is due to an effect first measured in 1913 by the French physicist Georges Sagnac: the rotation of a complete ring interferometer with angular frequency $\mathbf{\Omega}$ produces a fringe shift with a phase $\Delta\varphi$ given by

See page 537

Challenge 127

$$\Delta\varphi = \frac{8\pi\,\mathbf{\Omega}\mathbf{A}}{c\lambda} \tag{26}$$

where $\mathbf{A}$ 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*, even though it had been predicted already 20 years earlier by Oliver Lodge.** Michelson and his team found a fringe shift with a period of 24 hours and of exactly the magnitude predicted by the rotation of the earth. Modern high precision versions use ring lasers with areas of only a few square metres, but are able to measure variations of the rotation rates of the earth of less than one part per million. Indeed, over the course of a year the length of a day varies irregularly by a few milliseconds, mostly due to influences from the sun or the moon, due to weather changes, and due to hot magma flows deep inside the earth. All these effects can be studied with such precision interferometers; they can also be used for research into the motion of the soil due to lunar tides, to earth quakes, and for checks of the theory of relativity. Ref. 60 Ref. 61

In summary, observations show that the earth surface rotates at 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? That is a hard question. If you find a method leading to an answer, publish it! (The same is valid for the question whether the length of the year is constant.) Only few methods are known, as we will find out shortly. Ref. 62

But 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. Ref. 65

Rotation is not the only motion of the earth; it performs other motions as well. This was known already long ago. In 128 BCE, the Greek astronomer Hipparchos discovered what

---

[*] Albert Abraham Michelson (1852, Strelno–1931, Pasadena) Prussian-Polish-US-American physicist, Nobel prize in physics in 1907, obsessed by the precise measurement of the speed of light.

[**] Oliver Lodge (1851–1940) was a British physicist who studied electromagnetic waves and tried to communicate with the dead. A strange but influential figure, his ideas are often cited when fun needs to made of 'physicists'; for example, he was one of those few physicists who believed that at the end of the 19th century physics was complete.

is today called the *(equinoctial) precession*. He compared a measurement he made himself with another made 169 years before. Hipparchos found that the earth's changes direction over time. He concluded that the sky was moving; today we prefer to say that the axis of the earth is moving. During a period of 23 000 years the axis draws a cone with an opening angle of 23.5 °. This motion is generated by the tidal forces of the moon and the sun on the equatorial bulge of the earth, which itself is due to its rotation.

In addition, the axis of the earth is not even fixed compared to the earth's surface. In 1884, by measuring the exact angle above the horizon of the celestial north pole, Friedrich Küstner (1856–1936) found that the axis of the earth *moves* with respect to the earth's crust, as Bessel had suggested forty years earlier. As a consequence of Küstner's discovery, the International Latitude Service was created. The *polar motion* Küstner discovered turned out to consist of three components: a small linear drift – not yet understood – a yearly elliptical motion due to seasonal changes of the air and water masses, and a circular motion[*] with a period of about 1.2 years due to fluctuations in the pressure at the bottom of the oceans. In practice, the north pole moves with an amplitude of 15 m around an average central position. Ref. 63

In 1912, the German meteorologist and geophysicist Alfred Wegener (1880–1930) discovered an even larger effect. After studying the shapes of the continental shelves and the geological layers on both sides of the Atlantic, he conjectured that the continents *move*. Even though derided at first, his discoveries were genuine. Satellite measurements confirm this model; for example, the American continent moves away from the European continent by about 10 mm every year. 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. Following the modern version of the model, called *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 mechanism for the motion. Ref. 64

Does the earth move?

The centre of the earth is not at rest in the universe. In the third century BCE Aristarchos of Samos had already maintained that the earth turns around the sun. However, 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? The distance between the earth and the sun was known since the 17th century, but only in 1837, Friedrich Wilhelm Bessel[**] was the first to observe the *parallax* of a star. This was a result of extremely careful measurements and complex calculations: 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

---

[*] The circular motion, a wobble, was predicted by the great Swiss mathematician Leonhard Euler (1707–1783); using this prediction and Küstner's data, in 1891 Seth Carlo Chandler claimed to be the discoverer of the circular component.

[**] Friedrich Wilhelm Bessel (1784–1846), westphalian astronomer who left a successful business career to dedicate his life to the stars, and became the foremost astronomer of his time.

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 to the star to be 105 Pm, or 11.1 light years. Challenge 128 n

Bessel had thus managed for the first time to measure the distance of a star. By doing so he also proved that the earth is not fixed with respect to the stars in the sky and that the earth indeed revolves around the sun. The motion itself was not a surprise; it was predicted by universal gravity. In addition, the mentioned aberration of light, discovered in 1728 by James Bradley and to be discussed shortly, had already shown, albeit indirectly, that the earth moves around the sun. See page 200

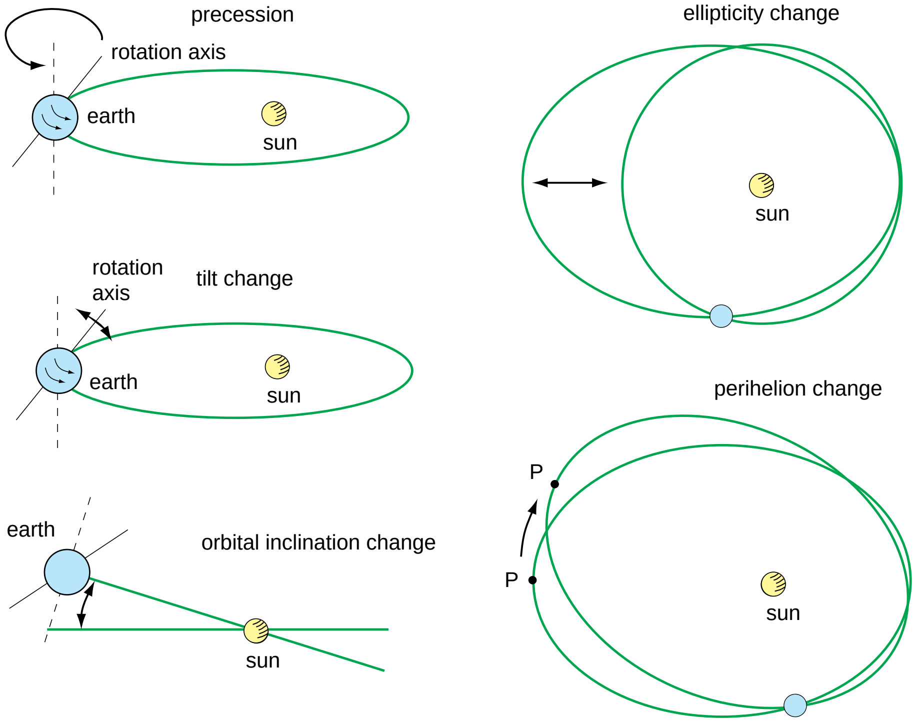


Figure 31 Changes in the earth's motion around the sun

With the improvement of telescopes, other motions of the earth were discovered. In 1748, James Bradley announced that there is a small regular *change* of the precession, which he called *nutation*, with a period of 18.6 years and an amplitude of 19.2 arc seconds. It appears because the plane of the moon's orbit around the earth 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 129

Astronomers also discovered that the 23.5 degree tilt – or *obliquity* – of the earth's axis, the angle between its intrinsic and its orbital angular momentum, actually changes from 22.1 to 24.5 degrees with a period of 41 000 years. This motion is due to the attraction of the sun and the deviations of the earth from a spherical shape. During the second world war, in 1941, the Serbian astronomer Milutin Milankovitch (1879–1958) retracted into solitude and studied the consequences. In his studies he realized that this 41 000 year period of the tilt, together with the precession period of 23 000 years,[*] gives rise to the over twenty *ice ages* in the last two million years. This happens through stronger or weaker irradiation of the poles by the sun. The changing amounts of melted ice then lead to changes in average temperature. The last ice age had is peak about 20 000 years ago and finished around 10 000 years ago; the next is still far away. A spectacular confirmation of the ice age cycles, in addition to the many geological proofs, came through measurements of oxygen isotope ratios in sea sediments, which allow to track the average temperature in the past million years. Ref. 66

The earth's orbit also changes its *eccentricity* with time, from completely circular to slightly oval and back. However, this happens in very complex ways, not with periodic regularity. The typical time scale is 100 000 to 125 000 years.

In addition, the earth's orbit changes in *inclination* with respect to the orbits of the other planets; this seems to happen regularly every 100 000 years. In this period the inclination changes from 2.5 degrees to minus 2.5 degrees and back.

Even the direction in which the ellipse points changes with time. This so-called *perihelion shift* is due in large part to the influence of the other planets; a small remaining part will be important in the chapter on general relativity. It was the first piece of data confirming the theory.

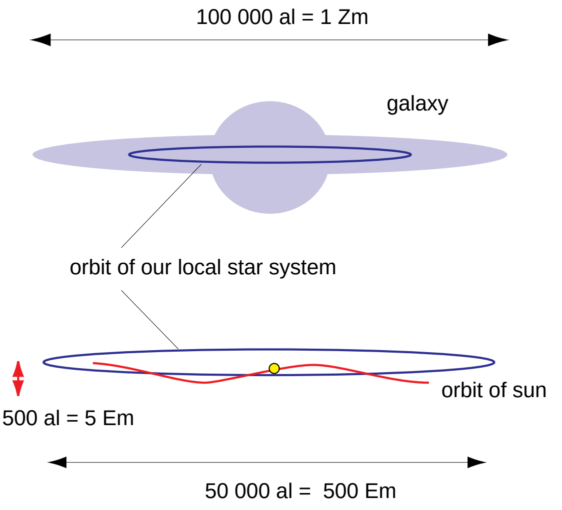


Figure 32 The motion of the sun around the galaxy

The next step is to ask whether the sun moves. It indeed does. Locally, it moves with a speed of 19.4 km/s towards the constellation of Hercules. This was already shown by William Herschel in 1783. But globally, the motion is even more interesting. The diameter of the galaxy is 100 000 light years, and we are located 25 000 light years from the centre. At our distance, the galaxy is 1 300 light years thick; we are 68 light years 'above' 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. It seems that the sun will continue moving away from the galaxy plane until it is about 250 light years above the plane, and then move back. The oscillation period is estimated to be around 60 million years, and has been brought into relation with the mass extinctions of animal life on earth, possibly because some gas cloud is encountered on the way. The issue is still a hot topic of research. Ref. 67

---

[*] In fact, precession has two periods, one of 23 000 and one of 19 000 years due to the interaction between precession and perihelion shift.

We turn around the galaxy centre because the formation of galaxies, like that of solar systems, always happens in a whirl. By the way, are you able to confirm by your own observation that our galaxy itself rotates? Challenge 130 n

Finally, we can ask whether the galaxy itself moves. Its motion can indeed be observed 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.) Ref. 68 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 gravitational attraction of the other, nearby galaxies in our local group of galaxies.*

Why don't we feel all these motions of the earth? Again, this question was answered by the master. Galileo explained in his lectures and books that only *relative* velocities between bodies produce effects, not the absolute values of the 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 the accelerations it produces at everyday scale are tiny. Nevertheless, many of these motions do induce measurable effects in atomic clocks. See page 266

By the way, every physicist knows that the statement by Wittgenstein

> Daß die Sonne morgen aufgehen wird, ist eine Hypothese; und das heißt: wir wissen nicht, ob sie aufgehen wird.**

is *wrong*. Can you explain why? Challenge 131 n

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 kilometres away. This is one fact that would makes time travel extremely difficult in practice, even if it were possible in theory! But we stop this discussion at this point, and we have a look at motion in everyday life.

Curiosities and fun challenges 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.

- A surprising effect is used in home tools such as hammer drills. We remember that when a small ball hits a large one at rest, both balls move after the hit, and the small one obviously moves faster than the large one. Ref. 74

---

* This is roughly the end of the ladder. Note that the expansion of the universe, to be studied later on, produces no motion.

** 'It is an hypothesis that the sun will rise tomorrow; and this means that we do not *know* whether it will rise.' This well-known statement is found in Ludwig Wittgenstein, *Tractatus*, 6.36311.

Challenge 132 n *** Do you agree with the quotation?

Despite this result, when a short cylinder hits a long one of the same diameter and material, but with a length which is some *integer* multiple of that of the short one, something strange happens. After the hit, the small cylinder remains almost at rest, whereas the large one moves. Momentum conservation seems not to hold at all in this case. (In fact this is the reason that momentum conservation demonstrations in schools are always shown with spheres.) What is the reason of this effect? Challenge 133

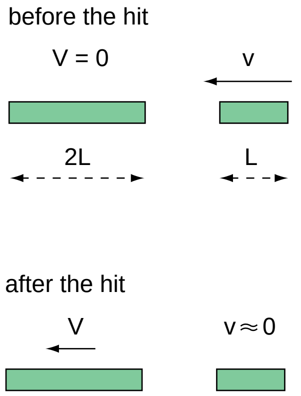


Figure 33 Momentum seems not to be conserved in this situation

- Does a wall get a stronger jolt when it is hit by a ball rebounding from it or when it is hit by a ball which remains stuck to it? Challenge 134 n
- Housewives know how to extract a cork from the inside of a wine bottle using a cloth. Can you imagine how? Challenge 135 n
- The sliding ladder problem, shown schematically in Figure 34, asks for 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? Challenge 136
- Ref. 69 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, one of which should not cost more than 2000 Euro? Challenge 137 n
- The level of acceleration a human can survive depends on the duration one is subjected to it. For a tenth of a second, $30\ g = 300\ \mathrm{m/s^2}$, as generated by ejector seats in aeroplanes, is acceptable. (It seems that the record acceleration a human survived is about $80\ g = 800\ \mathrm{m/s^2}$.) But as a rule of thumb it is said that accelerations of $15\ g = 150\ \mathrm{m/s^2}$ or more are fatal.
- The highest *microscopic* accelerations are observed in particle collisions, where one gets values up to $10^{35}\ \mathrm{m/s^2}$. The highest *macroscopic* accelerations are probably found in the collapsing interiors of *supernovae*, the exploding stars which can be so bright as to be visible in the sky even during the daytime. A candidate on earth is 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 at 1500 m/s in as little as a few μm, giving an acceleration of several $10^{11}\ \mathrm{m/s^2}$. Ref. 70

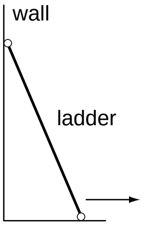


Figure 34 How does the ladder fall?

- If a canon located at the equator shoots a bullet in the vertical direction, where does the bullet fall back? Challenge 138
- Is travelling 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 exposure to the vacuum. So far only one man experienced vacuum without protection. He lost consciousness after 14 seconds, but survived unharmed. Ref. 71

no figure yet

Figure 35 Observation of sonoluminescence with a diagram of the experimental set-up

- How does the kinetic energy of a rifle bullet compare with that of a running man? Challenge 139 n
- In which direction does a flame lean if it burns inside a jar on a rotating turntable? Challenge 140 n
- 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 accelerated. In which direction does the ball move? Challenge 141
- What happens to the size of an egg when one places it into a jar of vinegar for a few days? Challenge 142 n

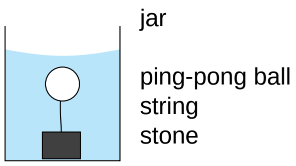


Figure 36 How does the ball move when the jar is accelerated?

- Does centrifugal acceleration exist? Most students at the 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 See page 489 the definition applies to the term and make up your own mind. Challenge 143
- Rotation holds a surprise for everybody studying it carefully. Angular momentum is a quantity with a magnitude and a direction. However, it is not a vector, as any mirror shows. The angular momentum of body circling in a plane parallel to a mirror behaves differently from a usual arrow: its mirror image is not reflected if it points towards the mirror! You can easily check this by yourself.For this reason, angular momentum is called a *pseudovector*. Challenge 144 e The fact has no important consequences in classical physics; but we have to keep it in mind for latter occasions.
- What is the best way to transport full coffee or tea cups while at the same time avoiding spilling any precious liquid? Challenge 145 n
- The moon recedes from the earth by 3.8 cm a year, due to friction. Can you find the responsible mechanism? Challenge 146
- What is the amplitude of a pendulum oscillating in such a way that the absolute value of its acceleration at the lowest point and at the return point are equal? Challenge 147
- Can you confirm that the value of the acceleration of a drop of water falling through vapour is $g/7$? Challenge 148
- Figure 37 shows the so-called *Celtic wiggle stone*, a stone that starts rotating on a plane surface when it is put into oscillation. The size can vary between a few centimetres and a few Ref. 74 metres. Simply by bending a spoon one can realize a primitive form of this strange device, if the bend is not completely symmetrical. The rotation is always in the same direction. If the stone is put into rotation in the wrong direction, after a while it stops and starts rotating in the other sense! Can you explain the effect? Challenge 149

- What is the motion of the point below the sun on a map of the earth? Challenge 150
- The moment of inertia of a body does depend on the shape of a body; usually, angular momentum and the angular velocity do not point in the same direction. Can you confirm this with an example? Challenge 151
- Can it happen that a satellite dish for geostationary tv satellites focuses the sunshine onto the receiver? Challenge 152 n
- Why is it difficult to fire a rocket from an aeroplane in direction opposite to the motion of the plane? Challenge 153 n
- You have two hollow spheres: they have the same weight, the same size, and painted the same colour. One is made of copper, the other of aluminium. Obviously, they fall with the same speed and acceleration. What happens if they both roll down a tilted plane? Challenge 154

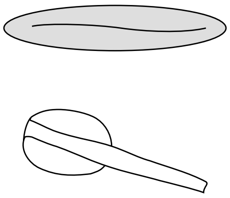


Figure 37 The famous Celtic stone and a version made with a spoon

- An ape hangs on a rope. The rope hangs over a wheel and is attached to a mass of equal weight hanging down on the other side. The rope is massless, the wheel massless and frictionless. What happens when the ape climbs the rope? Challenge 155 n
- What is the shape of a rope when rope jumping? Challenge 156
- How can you determine the speed of a rifle bullet only with a scale and a meter? Challenge 157
- Why does a gun make a hole in a door and cannot push it open, in contrast to what a finger can? Challenge 158
- Can a waterskier move with a higher speed than the boat pulling him? Challenge 159 n
- Take two cans of the same size and weight, one full of ravioli and one full of peas. Which one rolls faster on an inclined plane? Challenge 160
- What is the moment of inertia of a homogeneous sphere? Challenge 161 n
- Is it true that the moon in the first quarter in the northern hemisphere looks like the moon in the last quarter in the southern hemisphere? Challenge 162 n

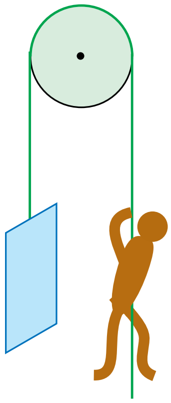


Figure 38 How does the ape move with respect to the mass?

- An impressive confirmation that the earth is round can be seen at sunset, if one turns, against usual habits, the back to the sun. On the eastern sky one can see the impressive rise of the earth's shadow. (In fact, more precise investigations show that it is not the shadow of the earth alone, but the shadow of its ionosphere.) One can admire a vast shadow rising over the whole horizon, clearly having the shape of a segment of a huge circle.
- Since the earth is round and space has three dimensions, there are many ways to drive from one point on the earth to another along a circle segment. This has interesting consequences for volleyballs and for looking after women. Take a volleyball and focus at its air inlet. If you want to move the inlet to a different direction with a simple rotation, you can choose the rotation axis in may different ways. Can you confirm this?In other words, when we look into a given direction and then want to change to another, the eye can realize this change in different ways. The option chosen by the human eye has been studied by medical Challenge 163 e

scientist already in the 18th century. It is called Listing's 'law'.[*] It states that all axes that nature choses lie in one plane. Can you imagine its position in space? Men have a deep interest that this mechanisms is being followed; if it were not, on the beach, when men look towards one woman after the other, the muscles moving the eyes could get knotted up... Challenge 164 n

Legs or wheels? – again

The acceleration and deceleration of standard wheel-driven cars is never much higher than about $1g = 9.8\,\mathrm{m/s^2}$, the acceleration due to gravity on our planet. Higher accelerations are achieved by motor bikes and racing cars through the use of suspensions which divert 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 pushing a car 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, backward and sideways accelerations (necessary for speed increase, 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 engineers 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, but high performance radio controlled model cars also achieve such values.

How do all these efforts compare to legs? High-jump athletes can achieve peak accelerations of about 2-4 times g, cheetahs over $3g$, 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 – a cheetah easily beats any car or motorbike – and evolution developed legs, instead of wheels, to improve the chances of an animal in danger to get to safety. In short, legs *outperform* wheels. Ref. 75

There are other reasons to use legs instead of wheels. (Can you name some?)For example, legs, in contrast to wheels, allow walking on water. Most famous for this ability is the *basilisk*,[**] a lizard living in Central America. This reptile is about 50 cm long and has a mass of about 90 g. It looks like a miniature Tyrannosaurus Rex and can actually run over water surfaces on its hind legs. The motion has been studied in detail with high-speed cameras and by measurements using aluminium models of the animal's feet. The experiments show that the feet slapping on the water provides only 25% of the force necessary to run above water; the other 75% is provided by a pocket of compressed air that the basilisks create between Challenge 165 Ref. 76

---

[*] If you are interested in learning how nature copes with the complexities of three dimensions, see the http://schorlab.berkeley.edu/vilis/whatisLL.htm and http://www.med.uwo.ca/physiology/courses/LLConsequencesWeb/ListingsLaw/perceptual2.htm web sites.

[**] In the Middle Ages, the term 'basilisk' referred to a mythical monster supposed to appear shortly before the end of the world. Today, it is a small reptile in the Americas.

their feet and the water. In fact, basilisks mainly walk on air.[*] It was calculated that a human is also able to 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 so.

By the way, all animals have an Challenge 166 n even number of legs. Why? Do you know an exception? In fact, one can argue that no animal has less than four legs. Why?

After this short overview of 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.

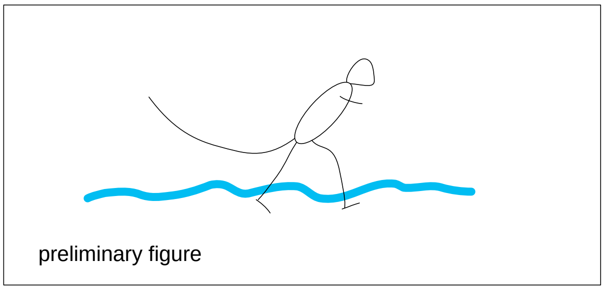


Figure 39 A basilisk lizard (Basiliscus basiliscus) running over water

The dynamics of gravitation

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

The first and main contact-free method to generate motion we discover in our environment 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. *Gravitation* produces an acceleration along the line connecting the centres of gravity of the two bodies. Note that in order to make this statement, it is necessary 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 centre and a mass. Today, these statements are common knowledge, but they are by no means evident from everyday personal experience.[***]

How does gravitation change when two bodies are far apart? The experts for distant objects are the astronomers. Over the years they performed numerous measurements of the movements of the moon and the planets. The most industrious of all, the Dane Tycho Brahe,[****] who organized an industrial search for astronomical facts sponsored by his king. His measurements were the basis for the research of his young assistant, the Swabian astronomer See page 489 Johannes Kepler[*****] who found the first precise description of planetary motion. In 1684,

---

Ref. 77 [*] Both effects used by basilisks are also found in fast canoeing.

[**] '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 Challenge 167 boundaries or suspension method. Are you ableto convince a friend that the earth is round and not flat? Can you find another argument apart from the roundness of the earth's shadow when it is visible on the moon? If the Challenge 168 n earth is round, the top of two buildings is further apart than their base. Can this effect be measured?

[****] Tycho Brahe (1546–1601), famous Danish astronomer, builder of Uraniaborg, the astronomical castle. He consumed almost 10% of the Danish gross national product for his research, which produced the first star catalogue and the first precise position measurements of planets.

[*****] Johannes Kepler (1571, Weil der Stadt–1630); after helping his mother defend herself in a trial where she was¡accused of witchcraft, he studies protestant theology, and became a teacher of mathematics, astronomy and rhetoric. His first book on astronomy made him famous, and he became assistant of Tycho Brahe and then,

all observations of planets and stones were condensed into an astonishingly simple result by the English physicist Robert Hooke:[*] every body of mass M attracts any other body towards its centre with an acceleration whose magnitude a is given by

$$a = G\frac{M}{r^2} \qquad (27)$$

where r is the centre-to-centre distance of the two bodies. This is called the *universal 'law' of gravitation*, or universal gravity for short, 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. 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' of gravitation, because the English physicist Isaac Newton proved more elegantly than Hooke that it agreed with all astronomical and terrestrial observations. Above all, however, he organized a better public relations campaign, in which he claimed to be the originator of the idea. Ref. 78

Newton published a simple proof showing that this description of astronomical motion also gives the correct description for stones thrown through the air, down here on 'father earth'. To achieve 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 $a_\mathrm{m}/g = R^2/d_\mathrm{m}^2$, where R is the radius of the earth. The moon's distance d_m can be measured by triangulation, comparing the position of the moon against the starry background from two different points on earth.[**] The result is $d_\mathrm{m}/R = 60 \pm 3$, depending on the orbital position of the moon, so that an average ratio $a_\mathrm{m}/g = 3.6 \cdot 10^3$ is predicted from universal gravity. 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 $a_\mathrm{m} = d_\mathrm{m}(2\pi/T)^2$, where $T = 2.4\,\mathrm{Ms}$ is the time the moon takes for one 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 $d_\mathrm{m} = 0.38\,\mathrm{Gm}$. One thus has $a_\mathrm{m}/g = 3.6 \cdot 10^3$, in agreement with the above prediction.

---

at his teacher's death, the Imperial Mathematician. He was the first to use mathematics in the description of astronomical observations, and introduced the concept and field of 'celestial physics'.

[*] Robert Hooke, (1635–1703), important English physicist, secretary of the Royal Society.

[**] The first precise – but not the first – measurement was realized in 1752 by the French astronomers Lalande and La Caille, who simultaneously measured the position of the moon seen from Berlin and from 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 $v = 2\pi r/T$. The drawing of the vector $\mathbf{v}$ over time, the so-called *hodograph*,shows 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 corresponding expression, namely $a = 2\pi v/T$. Challenge 169

[****] This is the hardest quantity to measure oneself. The most surprising way to determine the earth's size is the following: watch a sunset in the garden of 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's radius R is then given by $R = kh/t^2$, with $k = 378 \cdot 10^6\,\mathrm{s}^2$. Ref. 79 Challenge 170

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. Ref. 80

With this famous ‘moon calculation’ we have thus shown that the inverse square property of gravitation indeed describes both the motion of the moon and that of stones. You might Challenge 172 n want to deduce the value of GM.

From the observation that on the earth all motion eventually comes to rest, whereas in the sky all motion is eternal, Aristotle and many others had concluded that motion in the sublunar world has different properties than motion in the translunar world. Several thinkers had criticized this distinction, notably the French philosopher and rector of the University Ref. 81 of Paris, Jean Buridan.* The moon calculation was the most important result showing this distinction to be wrong. This is the reason for calling the expression (27) the *universal* ‘law’ of gravitation.

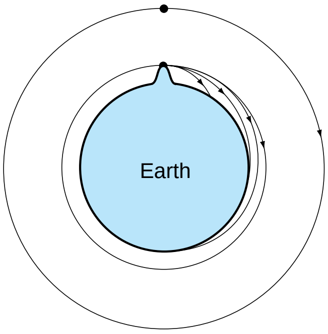


figure to be inserted

Figure 40 A physicist’s and an artist’s view of the fall of the moon: a graph by Christiaan Huygens and a marble by Auguste Rodin

This result allows us 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 peculiarity that instead of falling *towards* the earth, it is continuously falling *around* it. The moon is continuously missing the earth.**

Properties of gravitation

Gravitation implies that the path of a stone is not a parabola, as stated earlier, but actually an ellipse around the centre of the earth. This happens for exactly the same reason that the Challenge 174 planets move in ellipses around the sun. Are you able to confirm this statement?

Make another photograph of the moon a few hours later, when it is just above the horizon. On this picture, See page 50 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 Challenge 171 d_{m} is given by the relation $d_{\mathrm{m}}^2 = R^2 + [2Rq\cos\theta/(1-q^2)]^2$. Enjoy its derivation from the drawing.

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 °.

* Jean Buridan (ca. 1295– ca. 1366) was also one of the first modern thinkers to speculate on a rotation of the earth about 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 84, this explanation is nowadays out of favour at most universities.

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

$$g = GM_{\mathrm{earth}}/R^2_{\mathrm{earth}} \quad . \tag{28}$$

It can be deduced from equation (27) 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 on a plateau.

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 175 n

Note that 9.8 is roughly π^2. This is *not* a coincidence: the metre has been chosen in such a way to make this correct. The period of a swinging pendulum, i.e. a back and forward swing, is givenby* Challenge 176

$$T = 2\pi\sqrt{\frac{l}{g}} \quad . \tag{29}$$

If the meter had been defined 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 metre 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.

Then the proposal was made to define the metre as $1/40,000,000$ of the circumference of the earth through the poles, a so-called *meridian*. This proposal was almost identical to – but much more precise than – the pendulum proposal. The meridian definition of the metre was then adopted by the French national assembly on the 26th of March 1791, with the statement that 'a meridian passes under the feet of every human being, and all meridians are equal.'

But one can still ask: 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; it is still unanswered and a topic of research. The second question is addressed in Appendix B.

If all objects attract each other, that should also be the case for objects in everyday life. 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. Measuring this effect allows to determine the gravitational constant G.

Ref. 82 There is a beautiful problem connected to the left part of the figure: Which points of the surface of the earth Challenge 173 can be reached by shooting from a mountain? And which points can be reached by shooting only horizontally?

[*] Formula (29) is noteworthy because 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.

Challenge 177 n A leg also moves like a pendulum, when one walks normally. Why then do taller people tend to walk faster?

Note that measuring G is also the only way to determine the mass of the *earth*. The first to do so, in 1798, was the English physicist Henry Cavendish (1731–1810); therefore he called the result of his experiments 'weighing the earth'. Are you able to imagine how he did it? The value found in experiments is Challenge 178 n

$$G = 6.7 \cdot 10^{-11}\,\mathrm{Nm}^2/\mathrm{kg}^2 \quad . \tag{30}$$

Cavendish's experiments were thus the first to confirm that gravity works also sideways.

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 we usually do not notice the attraction to other people. When we notice 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*. Challenge 179

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

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

The gradient is just a learned term for 'slope along the steepest direction'. It is defined for any point on a slope, is long for a steep one and short for a flat one, and it points in the direction of steepest ascent, as shown in Figure 41. The gradient is abbreviated ∇, pronounced 'nabla', and is mathematically 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 higher potential values at larger heights.* For a point-like or a spherical body of mass M, the potential is

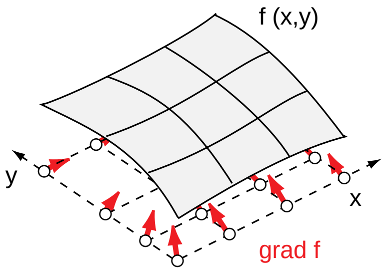


Figure 41 The potential and the gradient

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

A potential considerably simplifies 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.** The potential φ is an interesting quantity because with a single number at every position in space we can describe the vector aspects

---

[*] 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. The details lie outside the scope of this walk.

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

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

of gravitational acceleration. It automatically describes that gravity in New Zealand acts in the opposite direction to gravity in Paris. In addition, the potential suggests the introduction of the so-called *potential energy*

$$U = m\varphi \qquad (34)$$

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

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

In other words, the *total energy*, defined as the sum of kinetic and potential energy, is *conserved* in motion due to gravity. This is a characteristic property of gravitation. Not all accelerations can be derived from a potential; systems with this property are called *conservative*. The accelerations due to friction are not conservative, but those due to electromagnetism are.

Interestingly, the number of dimensions of space d is coded into the potential of a spherical mass: its dependence on the radius r is in fact $1/r^{d-2}$.The exponent $d - 2$ has been checked experimentally to high precision; no deviation of d from 3 has ever been found. Challenge 181 Ref. 84

The concept of potential helps in understanding *shape* of the earth. Since most of the earth is still liquid when seen on a large scale, its surface is always horizontal with respect to the direction determined by the combination of the accelerations of gravity and rotation. In short, the earth is *not* a sphere. The mathematical shape following from this requirement is called a *geoid*. That shape is given approximately, with an error of less than about 50 m, by an ellipsoid. Can you describe the geoid mathematically?The geoid is an excellent approximation to the actual shape of the earth; sea level differs from it by less than twenty metres. The 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 equatorial plane than the north pole by about 30 m. This is probably due to the large land masses in the northern hemisphere. Ref. 85 Ref. 86 Challenge 182

The inertia of matter, through 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. The precise flattening $(a - b)/a$ has the value $1/298.3 = 0.0034$. As a result, the top of Mount Chimborazo in Ecuador, even though its height is only 6267 m above See Appendix B

---

where $\rho = \rho(\mathbf{x},t)$ is the mass volume density of the body and the operator Δ, pronounced ‘delta’, is defined as $\Delta f = \nabla\nabla f = \partial^2 f/\partial x^2 + \partial^2 f/\partial y^2 + \partial^2 f/\partial z^2$. Equation (33) is called the *Poisson equation* for the potential φ, after Siméon-Denis Poisson (1781–1840), eminent French mathematician and physicist. The positions at which ρ is not zero are called the *sources* of the potential. The source term $\Delta\varphi$ 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 $\Delta\varphi \approx \bar{\varphi} - \varphi(x)$?)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 180

Often the concept of *gravitational field* is introduced, defined as $\mathbf{g} = -\nabla\varphi$. We avoid this in our walk, because we will discover that following the theory of relativity gravity is not due to a field at all; in fact even the concept of gravitational potential turns out to be only an approximation.

sea level, is about 20 km farther away from the centre of the earth than the top of Mount Qomolangma[*] in Nepal, whose height above sea level is 8850 m. The top of Mount Chimborazo is in fact the surface point most distant from the centre of the earth.

As a consequence, if the earth stopped rotating, 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 4 km. The northern parts of Europe would be covered by between 6 km and 10 km of water. Mount Qomolangma would be over 11 km above sea level. The resulting shape change 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 tomorrow, despite what some philosophers might pretend. See page 82

Gravitation in the sky

The expression $a = GM/r^2$ also describes the motion of all the planets around the sun. Anyone can check that the planets always stay within the *zodiac*, a narrow stripe across the sky. The centre line of the zodiac gives the path of the sun and is called the *ecliptic*, since the moon must 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 movements of the planets in the zodiac. The three main ones are:

- Planets move on ellipses with the sun located at one focus (1609);
- Planets sweep out equal areas in equal times (1609);
- For all planets, calling the duration of the orbit T and semimajor axis d, the ratio T^2/d^3 is the same (1619).

The sheer work required to deduce them was enormous. Kepler had no calculation machine available, not even a slide rule. The calculation technology he used was the recently discovered logarithms. Anyone who has used tables of logarithms to actually perform calculations can get a feeling for the large amount of work behind these three discoveries.

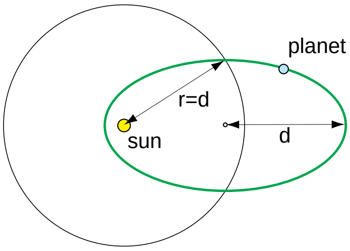


Figure 42 The motion of a planet around the sun, showing its semimajor axis d, which is also the spatial average of its distance from the sun

Can you show that all three laws follow from the expression of universal gravity? Challenge 183

Even Newton was not able to write down, let even to handle, differential equations at the time he published his results on gravitation. Ref. 16 In fact his notation and calculation methods were poor. The English mathematician G.H. Hardy[***] used to say that insistence to use Newton's integral and differential notation, which he developed much

---

[*] Mount Qomolangma is sometimes also called Mount Everest.

[**] The apparent height of the ecliptic changes with the time of the year and is the reason for the changing seasons. Therefore seasons are a gravitational effect as well.

[***] G.H. Hardy (1877–1947) important English number theorist. He also 'discovered' the famous Indian mathematician Srinivasa Ramanujan, bringing him to Britain, and wrote the well-known *A Mathematician's Apology*.

later – instead of using the one of his rival Leibniz, common today – threw back English mathematics by 100 years.

Kepler, Hooke and Newton became famous because they brought order to the description of planetary motion. This achievement, though of small practical significance, was widely publicized because of the age-old prejudices linked to astrology.

However, there is more to gravitation. Universal gravity 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?)In fact, universal Challenge 184 gravity explains much more about the moon.

The moon

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 appear mainly because the moon does not describe a circular, but an elliptical orbit around the earth and because the axis of the moon is slightly inclined compared to that of its rotation around the earth. As a result, only around 45% of the moon's surface are permanently hidden from earth.

The first photographs of the hidden areas were taken in the 1960s by a Soviet satellite. The surface is much more irregular than the visible one, as the hidden side is the one which intercepts most asteroids attracted by the earth. Thus the gravitation of the moon helps to deflect asteroids from the earth. The number of animal life extinctions is thus brought to a small, but not negligible number. In other words, the gravitational attraction of the moon saved the life of humans already many times over.[*]

The trips to the moon in the 1970s also showed that the moon originated from the earth itself: long ago, an object hit the earth almost tangentially and threw a sizeable fraction of material up into the sky. This is the only mechanism able to explain the large size of the moon, its low iron content, as well as its general material composition. Ref. 87

The moon is receding from the earth at 3.8 cm a year. This result confirms the old deduction that the tides slow down the earth's rotation. Can you imagine how this measurement wasperformed?[**] Since the moon slows down the earth, the earth also changes shape due Challenge 185 to this effect. (Remember that the shape of the earth depends on its rotation speed.) These changes in shape influence the tectonic activity of the earth, and maybe also the drift of the continents.

The moon has many effects on animal life. A famous example is the midge Clunio, which lives on sea coasts with pronounced tides. Clunio lives between six and twelve weeks as a Ref. 88 larva then hatches and lives only one or two hours as adult flying insect, during which it

---

[*] The web pages http://cfa-www.harvard.edu/iau/lists/Closest.html and InnerPlot.html give an impression of the number of objects which almost hit the earth every year. Without the moon, we would have many additional catastrophes.

[**] 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.

reproduces. The reproduction is only successful if the midge hatches during the low tide phase of a spring tide. Spring tides are the especially strong tides during the full and new moons, when the solar and lunar effects add, and occur only every 14.8 days. In 1995, Dietrich Neumann showed that the larvae have two built-in clocks, a circadian and a circalunar one, which together control the hatching to precisely those few hours when the insect can reproduce. He also showed that the circalunar clock 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: they are the smallest known astronomers.

If insects can have circalunar cycles, it should come as no surprise that women also have such a cycle. However, in this case the origin of the cycle length is still unknown. Ref. 89

The moon also helps to stabilize the tilt of the earth's axis, keeping it more or less fixed relative to the plane of motion around the sun. Without the moon, the axis would change its direction irregularly, we would not have a regular day and night rhythm, we would have extremely large climate changes, and the evolution of life would have been impossible. Ref. 90 Without the moon, the earth would also rotate much faster and we would have much more unfriendly weather. Ref. 92 The moon's main remaining effect on the earth, the precession of its axis, is responsible for the ice ages. See page 78

We will also see that the moon shields the earth from cosmic radiation by greatly increasing the earth's magnetic field. In other words, the moon is of central importance for the evolution of life. Understanding how often planets have moon-sized moons is thus important for the estimation of the probability that life exists on other planets. Ref. 91 So far, it seems that large moons are rare; the issue is still an area of research. But let us return to the effects of gravitation in the sky.

Orbits

The path of a body orbiting another under the influence of gravity is an ellipse with the central body at one focus. A circular orbit is also possible, a circle being a special case of an ellipse.

Gravitation implies that comets return. 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 named after him. 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.

Depending on the initial energy and the initial angular momentum of the body with respect to the central planet, there are two additional possibilities: *parabolic* paths and *hyperbolic* paths. Can you determine the conditions on the energy and the angular momentum for these paths to appear?

Some comets follow parabolas when moving around the sun, but most follow elliptical paths. Hyperbolic paths are less common; they are often used to change the direction of artificial satellites on their way through the solar system.

For more than two gravitating objects, many more possibilities for motions of bodies appear. In fact, the many-body problem is still a topic of research, and the results are fascinating mathematically, even though a bit less physically. Thus we cover only a few examples here.

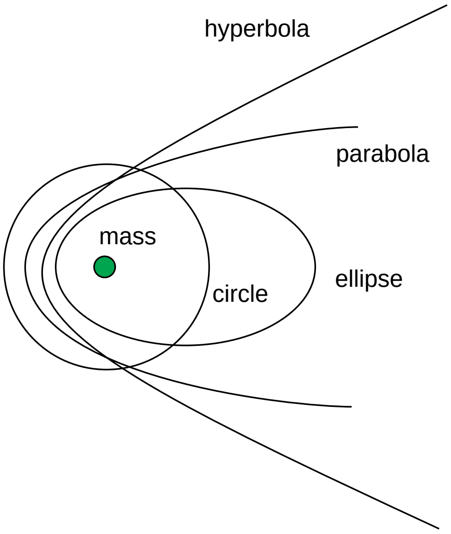


Figure 43 The possible orbits due to universal gravity

When several planets circle a sun, they also attract each other. Planets thus do not move in perfect ellipses. The largest deviation is a perihelion change. It is observed for Mercury and a few other planets, including the earth. Other deviations from elliptical paths appear during a single orbit. In 1846, the observed deviations of the motion of the planet Uranus from the path predicted by universal gravity were used to predict the existence of another planet, Neptune, which was discovered shortly afterwards. See Figure 31

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, maybe using more than two bodies? Yes; there are two examples.* The first, the geostationary satellites, is used for easy transmission of television and other signals from and towards earth. See page 64

The *Lagrangian libration points* are the second example. 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 locate them, not forgetting to take rotation into account?There are three additional Lagrangian points on the earth-moon axis. Challenge 186 How many are of them stable? Challenge 187

There are thousands of asteroids, called *Trojan asteroids*, at and around the Lagrangian points of the Sun-Jupiter system. In 1990, a Trojan asteroid for the Mars-Sun system was discovered. Finally, in 1997, a Trojan asteroid was found which follows the earth in its way around the sun. This second companion of the earths has a diameter of 5 km. Similarly, on the main Lagrangian points of the earth-moon system a high concentration of dust has been observed. Ref. 93

To sum up, the single equation $\mathbf{a} = -GM\mathbf{r}/r^3$ correctly describes a large number of phenomena in the sky. The first person to make clear that the expression describes *everything*

---

* Levitation is discussed in detail on page 420.

happening in the sky was Pierre Simon Laplace[*] famous treatise *Mécanique céleste* He summarized the book in the famous answer he gave to Napoleon, when the latter told him 'I do not read anything about the creator in your book'; Laplace answered: 'I did not need this hypothesis any more'.

These results are quite a feat for such a simple expression. How precise is it? Since astronomy allows the most precise measurements of gravitational motion, it also provides the most stringent tests. Simon Newcomb (1835–1909) repeated Laplace's analysis and concluded after intensive study that there was only one known example of discrepancy from universal gravity, namely one observation for the planet Mercury. (Nowadays a few more are known.) The point most distant from the sun of the orbit of planet Mercury, its perihelion, changes with a rate slightly smaller than the predicted one: the tiny difference is around 43 arc seconds per century. The study of motion had to wait for Albert Einstein to explain it. Ref. 94

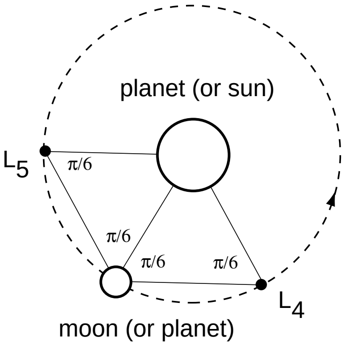


Figure 44 The two stable Lagrangian points

Tides

Why do physics texts always talk about tides? Because, as general relativity shows, tides prove that space is curved! It is thus useful to study them a bit in more detail. Gravitation describes the sea tides as results of the attraction of the ocean water by the moon and the sun. Tides are interesting; 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? Challenge 188 n The *soil* is also lifted and lowered by the sun and the moon, by about 0.3 m, as satellite measurements show. Ref. 31 Even the *atmosphere* is subject to tides, and the corresponding pressure variations can be filtered out from the weather pressure measurements. Ref. 95

Tides appear for any *extended* body moving in the gravitational field of another. To understand the origin of tides, it suffices to picture a body in orbit, like the earth, and to imagine its components, such as the segments of Figure 45, as being kept together by springs. Universal gravity implies that orbits are slower the more distant they are from a central body. As a result, the segment on the outside of the orbit would like to be slower than the central one; through the springs it is *pulled* by the rest of the body. In contrast, the inside segment would like to orbit more rapidly and is thus *retained* by the others. Being slowed down, the inside segments wants to fall towards the sun. In sum, both segments feel a pull away from the centre of the body, until the springs stop the deformation. Therefore, *extended bodies are deformed in direction of the field inhomogeneity.*

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

---

[*] Pierre Simon Laplace (1749, Beaumont–en-Auge-1827, Paris), important French mathematician. His treatise appeared in 5 volumes between 1798 and 1825.

radius perpendicular to it. If the inner springs are too weak, the body is torn into pieces; in this way a *ring* of fragments can form, such as the asteroid ring between Mars and Jupiter or the rings around Saturn.

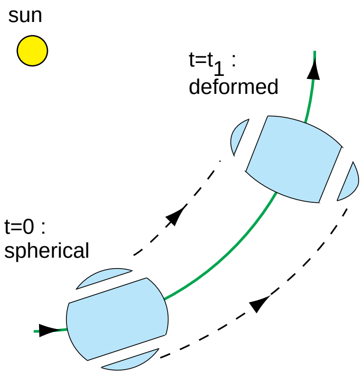


Figure 45 Tidal deformations due to gravity

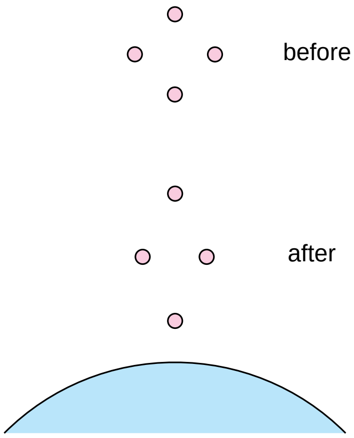


Figure 46 The origin of tides

Let us return to the earth. If a body is surrounded by water, it will form bulges in direction of the applied gravitational field. In order to measure and compare the strength of the tides from the sun and the moon, we reduce tidal effects to their bare minimum, as shown in Figure 46. Tides appear because nearby points falling together approach or diverge, depending on their relative position. Tides thus depend on the change of acceleration with distance; in other words, this *relative* acceleration is the derivative of gravitational acceleration.

Using the numbers from Appendix B, the gravitational accelerations from the sun and the moon are

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

and thus the attraction from the moon is about 178 times weaker than that from the sun.

The *relative* acceleration $b = \nabla a$ of any two nearby point masses falling together near a large spherical mass M is given by Challenge 189

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

which yields the values

$$b_{\text{sun}} = -\frac{2GM_{\text{sun}}}{d_{\text{sun}}^3} = -0.8 \cdot 10^{-13}\,/\text{s}^2$$
$$b_{\text{moon}} = -\frac{2GM_{\text{moon}}}{d_{\text{moon}}^3} = -1.7 \cdot 10^{-13}\,/\text{s}^2 \quad . \qquad (38)$$

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. When sun, moon and earth are aligned, the two tides add up; these so-called *spring tides* are especially strong and happen every 14.8 days, at full and new moon.

Ref. 61

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 due to other effects, such as the weather, are often larger). The results fit well with fossil results showing that 400 million years ago, in the Devon, a year had 400 days, and a day about 22 hours. It is also estimated that 900 million years ago, each of the 481 days of a year were 18.2 hours long. The friction at the basis of this slowdown also results in an increase of the distance of the moon by about 3.8 cm per year. Are you able to explain why?

Challenge 190 n

See page 252

See page 286

In summary, tides are due to relative accelerations of nearby mass points. This has an important consequence. In the chapter on general relativity we will find that time multiplied by the speed of light plays the same role as length. Time then becomes an additional dimension, as shown in Figure 47. Using this similarity, two free particles moving in the same direction correspond to parallel lines in space-time. Two particles falling side-by-side also correspond to parallel lines. Tides show that such two particles approach each other. In other words, tides imply that parallel lines approach each other. But parallel lines can approach each other *only* if space-time is curved. In short, tides imply *curved* space-time and space. This simple reasoning could have already been performed in the 18th century; however, it took another 200 years and Albert Einstein's genius to uncover it.

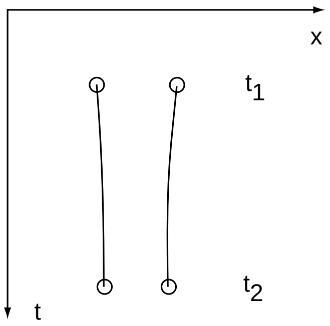


Figure 47 Particles falling side by side approach over time

Can light fall?

See page 199

Towards the end of the 17th century people found out that light has a finite velocity – a story which we will tell in detail later on. An entity which moves with infinite velocity cannot be affected by gravity, as there is no time to produce an effect. An entity with a finite speed, however, should feel gravity and thus fall.

Ref. 96

Does the speed increase when light reaches the surface of the earth? For almost three centuries people had no measurement means to detect any effect; so the question was not investigated. Then, in 1801, the Prussian astronomer Johann Soldner (1776–1833) was the first to put the question in a different way. Being an astronomer, he was used to measuring stars and their observation angles. He realized that due to gravity, light passing near a massive body would be *deflected*.

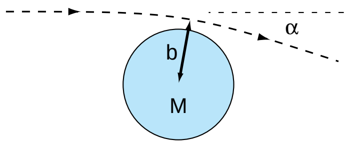


Figure 48 Masses bend light

For a body on a hyperbolic path, moving with velocity c past a body of mass M at distance b, Soldner deduced the deflection angle Challenge 191

$$\alpha_{\text{univ.grav.}} = \frac{2}{b}\frac{GM}{c^2} \quad . \tag{39}$$

In his time, this angle was far too small to be measured even for light deflected by the mass of the sun, where it turns out to be at most a tiny $0.88\,''= 4.3\,\mu$rad. Thus the issue was forgotten. Had it been 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 value. The first measurement took place in 1919 and found a deflection up to $1.75\,''$, exactly the double of expression (39). The reason is not easy to find; it is due to the curvature of space, as we will see.* In summary, light can fall, but the issue bears some surprises. See page 279

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 us 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 lift things.**

Is the gravitational mass of a body equal to the inertial mass? A rough answer is given by the experience that an object which is difficult to move is also difficult to lift. The simplest 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 in the 17th century Galileo had made widely known an even older argument showing without a single experiment that the acceleration is indeed the same for all bodies. If larger masses fell more rapidly than smaller ones, then the following paradox would appear. Any body can be seen as composed from a large fragment attached to a small fragment. If small bodies really fell 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 break into pieces). 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.

Challenge 192 * By the way, how would you measure the deflection of light near the bright sun?

Challenge 193 ** What are the values shown by a balance for a person of 85 kg juggling three balls of 0.3 kg each?

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 has been confirmed with the precision expected. Ref. 97 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 disinformation, now common across the whole world of physics education. See page 61 Let us go back to the definition of mass as negative inverse acceleration ratio. We mentioned that the physical origins of the accelerations do not play a role in the definition because it does not appear in the expression. In other words, the value of the mass is by definition independent of the interaction. That means in particular that inertial mass, based on electromagnetic interaction, and gravitational mass are identical *by definition*.

We also note that we have never defined a separate concept of 'passive gravitational mass'. The mass being accelerated by gravitation is the inertial mass. Worse, there is no way to define a 'passive gravitational mass' at all. Try it! Challenge 194 All methods, such as weighing an object, cannot be distinguished from those which determine inertial mass from its reaction to acceleration. Indeed, all methods to measure mass use non-gravitational mechanisms. Scales are good examples.

If the 'passive gravitational mass' were different from inertial mass, we would have strange consequences. For those bodies for which it were different we would get into trouble with energy conservation. Also assuming that 'active gravitational mass' differs from inertial mass get us into trouble.

Another way to look at the issue is the following. How could 'gravitational mass' differ from inertial mass? Would the difference depend on relative velocity, time, position, composition, or on mass itself? Each of these possibilities contradicts either energy or momentum conservation.

No wonder that all measurements confirm the equality of all mass types. The issue is usually resurrected in general relativity, with no new results. See page 268 'Both' masses remain equal. Mass is a unique property of bodies. Another issue remains, though. What is the *origin* of mass? Why does it exist? This simple but deep question cannot be answered by classical physics.

Curiosities and fun challenges about gravitation

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

- Do all objects on earth fall with the same acceleration of $9.8\,\mathrm{m/s^2}$, 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 195 n

---

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

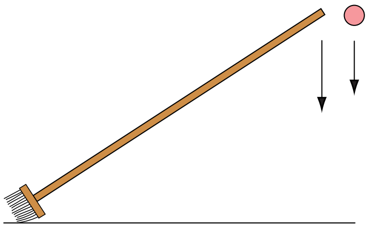


Figure 49 Brooms fall more rapidly than stones

- How can you use your observations made during travels to show that the earth is not flat? Challenge 196 n
- Is the acceleration due to gravity constant? Not really. Every day, it is estimated that 10^8 kg fall onto the earth as meteorites.
- Both the earth and the moon attract bodies. The centre of mass of the moon-earth system is 4800 km away from the centre of the earth, quite near its surface. Why do bodies on earth still fall towards the centre of the earth? Challenge 197
- 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 this statement?What then is wrong about Galileo's argument about the constancy of acceleration of free fall? Challenge 198
- It is easy to lift a mass of a kilogram on a table. Twenty kilograms is tougher. A thousand is impossible. However, $6 \cdot 10^{24}$ kg is easy. Why? Challenge 199 n

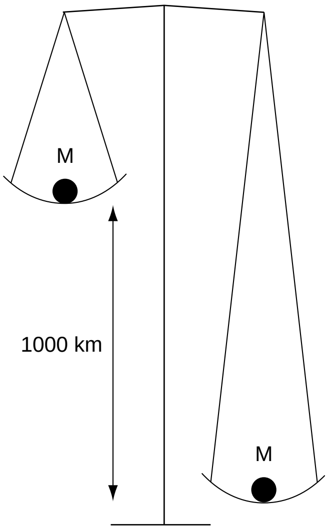


Figure 50 A honest balance?

- The strength ratio between the tides of moon and sun is roughly $7/3$. Is it true that this is also the ratio between the mass densities of the two bodies? Challenge 200
- The friction between the earth and the moon slows down the rotation of both. The moon stopped rotating already millions of years ago, and the earth is on its way to do so as well. When the earth will have stopped rotating, the moon will stop moving away from earth. How far will the moon be at that time?Afterwards however, even more in the future, Challenge 201 the moon will move back to the earth, due to the friction between the earth-moon system and the sun. Even though this effect would only take place if the sun burned forever, which is known to be false, can you explain it? Challenge 202 n
- When you run towards the east, you *loses weight*. There are even two different reasons for this: the 'centrifugal' acceleration increases and thus the force with which we are pulled down diminishes, and the Coriolis force appears, with a similar result. Can you estimate the size of the two effects? Challenge 203
- What is the time ratio between a stone falling through a distance l and a pendulum swinging though half a circle of radius l? (This problem is due Challenge 204 n to Galileo.) How many digits of the number π can one expect to determine in this way?
- Why can a spacecraft accelerate through the *slingshot effect* when going round a planet, despite momentum conservation? Challenge 205 n
- The orbit of a planet around the sun has many interesting properties. What is the hodo- Ref. 83 graph of the orbit? What is the hodograph for parabolic and hyperbolic orbits? Challenge 206
- A simple, but difficult question: if all bodies attract each other, why don't or didn't all stars fall towards each other? Challenge 207 n
- The acceleration g due to gravity at a depth of 3000 km is $10.05\,\mathrm{m/s^2}$, over 2% higher than at the surface of the earth. How is this possible? Also on the Tibetan plateau, g is Ref. 98

higher than the sea level value of $9.81\,\mathrm{m/s^2}$, even though the plateau is more distant from the centre of the earth than sea level is. How come? Challenge 208 n

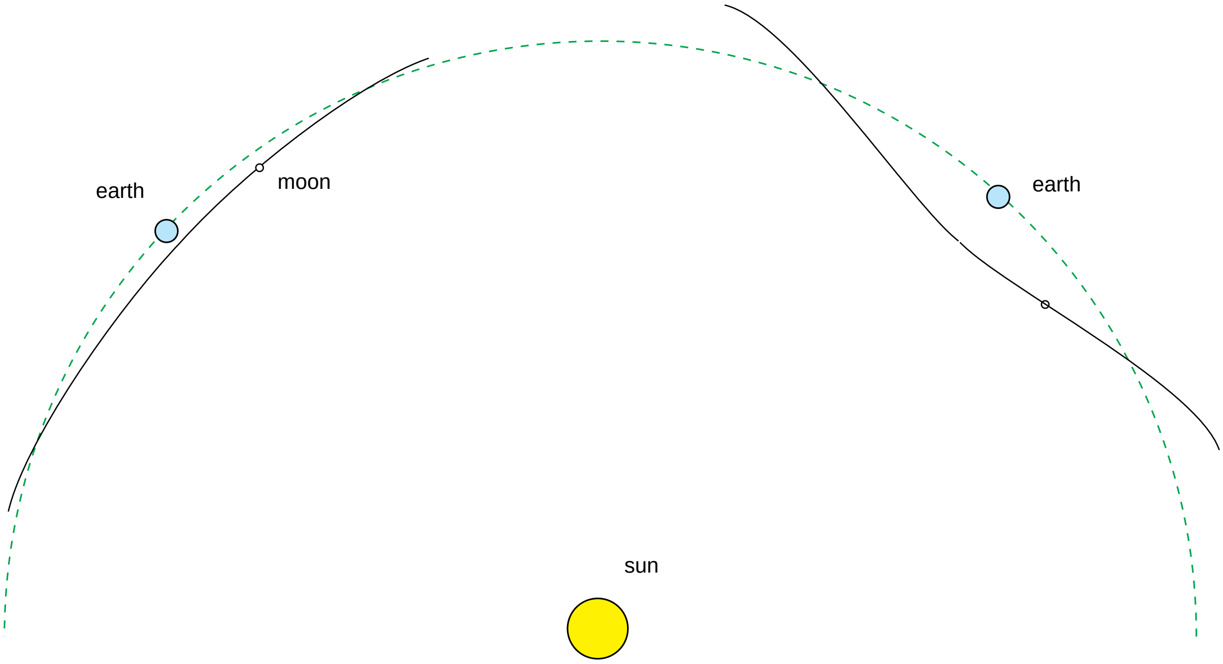


Figure 51 Which of the two moon paths is correct?

- When the moon circles the sun, does its path have sections concave towards the sun, as shown in the right part of the figure, or not, as shown on the left part? Challenge 209 n
- 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 $11\,\mathrm{km/s}$. What is the escape velocity for the solar system? By the way, the escape velocity of our galaxy is $129\,\mathrm{km/s}$. What would happen if a planet or a system were so heavy that its escape velocity would be larger than the speed of light? Challenge 210 n
- Can gravity produce repulsion? What happens to small test body on the inside of a large C-shaped mass? Is it pushed towards the centre of mass? Challenge 211
- For bodies of irregular shape, the centre of gravity of a body is *not* the same as the centre of mass. Are you able to confirm this? (Hint: find and use the simplest example possible.) Challenge 212 n
- Ref. 99 The *shape* of the earth is not a sphere. As a consequence, a plumb line usually does not point to the centre of the earth. What is the largest deviation in degrees? Challenge 213
- What is the largest asteroid one can escape from by jumping? Challenge 214 n
- The constellation in which the sun stands at noon (at the centre 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's axis. A check with a map of the star sky shows that the twelve See page 80 signs do not have the same length, and that there are fourteen of them, not twelve (there is *Ophiuchus*, the snake, between Scorpius and Sagittarius, and *Cetus*, the whale, between Acquarius and Pisces). 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. Ref. 100

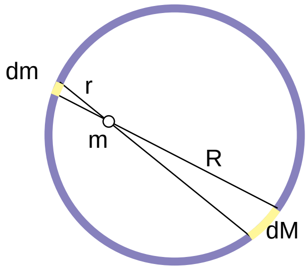


Figure 52 The vanishing of gravitational force inside a spherical shell of matter

▪ 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 argument using the picture?This works only because of the Challenge 215 $1/r^2$ dependence of gravity. Can you show that the result does not hold for non-spherical shells?Note that the vanish- Challenge 216 ing of gravity inside a spherical shell usually does not hold if other matter is found outside the shell. How could one eliminate the effects of outside matter? Challenge 217

▪ There is no planet X, i.e. no tenth planet in our solar system outside Neptune and Pluto. But there are many small Ref. 101 objects beyond them, in the so-called Kuiper belt and Oort cloud. Sometimes they change trajectory due to the attraction of a planet: that is the birth of a new comet.

▪ In astronomy many new examples of motion are regularly discovered even in the present century. Sometimes there are also false alarms. One example was the alleged fall of *mini comets* on the earth. They were supposedly made of a few dozens of kilograms of ice and hitting the earth every few seconds. It is now known not to happen. On the other hand, it is Ref. 102 known that many tons of asteroids fall on the earth every day, in the form of tiny particles. By the way, discovering objects hitting the earth is not at all easy. Astronomers like to point out that an asteroid as large as the one which led to the extinction of the dinosaurs could hit the earth without any astronomer noticing beforehand, if the direction is slightly unusual, such as from the south, where few telescopes are located.

▪ Universal gravity allows only elliptical, parabolic, or hyperbolic orbits. It is impossible that a small object approaching a large one is captured. At least, that is what we have learned so far. Nevertheless, all astronomy books tell stories of capture in our solar system, e.g. about several outer satellites of Saturn. How is this possible? Challenge 218

▪ Is it true that the centre of mass of the solar system is always inside the sun? Challenge 219

▪ All points on the earth do not receive the same number of daylight hours during a year. Can you see why? Challenge 220

▪ Can the phase of the moon have a measurable effect on the human body? What about the tidal effects of the moon? Challenge 221

---

[*] 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.

- There is an important difference between the heliocentric system and the old idea that all planets turn around the earth. The heliocentric system states that certain planets, such as Mars or Venus, can be *between* the earth and the sun at certain times, and *behind* the sun at other times. In contrast, the geocentric system states that they are always in between. Why did such an important difference not invalidate the geocentric system right away? Challenge 222

- Ref. 103 The strangest reformulation of the description $m\mathbf{a} = \nabla U$ is the almost absurd looking

$$\nabla v = d\mathbf{v}/ds \tag{40}$$

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 223

- Seen from Neptune, the size of the sun is the same as that of Jupiter seen from the earth at the time of its closest approach. True? Challenge 224 n

- Ref. 105 What is gravity? This is not a simple question. Already in 1747, Georges-Louis Lesage proposed an explanation for the $1/r^2$ dependence. He argued that the 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. Can you show that such an attraction has a $1/r^2$ dependence?The argument only works if the collisions are inelastic. Why? That would mean that all bodies would heat up with time, as Jean-Marc Levy-Leblond explains. Challenge 225 Ref. 1

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 mountain ascent will we settle the issue.

- For which bodies does gravity decrease when approaching them? Challenge 226
- Could one put a satellite into orbit using a cannon? Does the answer depend on the direction in which one shoots? Challenge 227
- How often does the earth rise and fall when seen from the moon? Does the earth show phases? Challenge 228 n
- What is the weight of the moon?How does it compare to the weight of the Alps? Challenge 229
- Due to the slightly flattened shape of the earth, the source of the Mississippi is about 20 km nearer to the centre of the earth than its mouth; the water effectively runs uphill. How can this be? Challenge 230 n
- If a star is made of high density material, the orbital speed of a planet circling it close by could be larger than the speed of light. How does nature avoid this strange possibility? Challenge 231 n
- What will happen to the solar system in the future? This question is surprisingly hard to answer. The main expert of this topic, U.S. 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 there is clear evidence of chaos in the evolution of the solar system, at a small level. The various planets influence each other in subtle and still poorly understood ways. Also effects in the past are being studied, such as energy of Jupiter due to its ejection of smaller asteroids from the solar system, or energy gains of Neptune. There is still a lot of research to be done in this field. Ref. 104 See page 190
- One of the great mysteries of the solar system is the description of planet distances discovered in 1766 by Johann Daniel Titius (1729–1796) and publicized by Johann Elert Bode

Planet	n	predicted distance in AU	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 16 One of the big mysteries of nature: planet distances and the values resulting from the Titius-Bode rule

(1747–1826). Titius discovered that planet distances from the sun can be approximated by

$$d = a + b\,2^n \quad \text{with} \quad a = 0.4\,\text{AU}\;,\; b = 0.3\,\text{AU} \tag{41}$$

when distances are measured in astronomical units. The resulting approximation is compared with observations in Table 231.

Interestingly, the last three planets, as well as the planetoids, were discovered *after* Titius' death; the rule had successfully predicted Uranus' distance, 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 Titius' rule. Even more astonishing is that the rule works also well for the distances of several moons around Jupiter. Explaining the rule is one of the great challenges remaining in classical mechanics. It is known that the rule must be a consequence of the formation of satellite systems. The bodies not following the rule, such as the outer planets of the sun or the outer moons of Jupiter, are believed not to be part of the original system but to have been captured later on.

▪ In 1722, the great mathematician Leonhard Euler made a calculation mistake which led him to conclude that if a tunnel were built from one pole of the earth to the other, a stone falling into it would arrive at the earth's centre and then turn back up directly. Voltaire made fun of this conclusion for many years. Can you find Euler's correction and show that the real motion is an oscillation from one pole to the other, and can you calculate the time a pole-to-pole fall would take (assuming homogeneous density)? Challenge 232

What would be the oscillation time for an arbitrary straight surface-to-surface tunnel of length l, thus not going from pole to pole? Challenge 233 n

What is classical mechanics?

All types of motion that can be described when the mass of a body is its only permanent property form what is called *mechanics*. The same name is also given the experts studying the field. We can think of mechanics as the athletic part of physics;* like in athletics, also in mechanics only lengths, times and masses are measured.

More specifically, our topic of investigation so far is called *classical* mechanics, to distinguish it from *quantum* mechanics. The main difference is that in classical physics arbitrary small values are assumed to exist, whereas this is not the case in quantum physics. The use of real numbers for observable quantities is thus central to classical physics.

Classical mechanics is often also called *Galilean physics* or *Newtonian physics*. The basis of classical mechanics, the description of motion using only space and time, is called *kinematics*. An example is the description of free fall by $z(t) = z_o + v_o(t - t_o) - \frac{1}{2}g(t - t_o)^2$. The other, main part of classical mechanics is the description of motion as a consequence of interactions between bodies; it is called *dynamics*. An example of dynamics is the formula of universal gravity.

The distinction of kinematics and dynamics can also be made in relativity, thermodynamics and electrodynamics. Even though we have not explored these fields of enquiry yet, we know that there is more to the world than gravity. A simple observation makes the point: friction. Friction cannot be due to gravity, because friction is not observed in the skies, where motion follows gravity rules only.** Moreover, on earth, friction is independent of gravity, as you might want to check.There must be another interaction responsible for friction. We shall study it shortly. But one issue merits a discussion right away.

Challenge 234 e

Should one use force?

Everybody has to take a stand on this question, even students of physics. Indeed, many types of forces are used and observed in daily life. One speaks of muscular, gravitational, characterial, sexual, satanic, supernatural, social, political, economic and many other types of forces. Physicists see things in a simpler way. They call the different types of forces observed between objects *interactions*. The study of the details of all these interactions will show that in everyday life, they are of electrical origin.

For physicists, all change is due to motion. The term force then also gets a more restrictive definition. *(Physical) force* is defined as the *flow of momentum*, i.e. as

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

Force is the flow of motion. If a force acts on a body, momentum flows into it. Using the Galilean definition of linear momentum $\mathbf{p} = m\mathbf{v}$, we can rewrite the definition of force as

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

---

[*] 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 lies at 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 favour in physics since about a century.

[**] 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.

See page 274

where $\mathbf{F} = \mathbf{F}(t, \mathbf{x})$ is the force acting on an object of mass m and where $\mathbf{a} = \mathbf{a}(t, \mathbf{x}) = d\mathbf{v}/dt = d^2\mathbf{x}/dt^2$ is the acceleration of the same object, that is to say 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* all aspects to muscular force. For example, the more force is used, the further a stone can be thrown.

However, whenever the concept of force is used, it should be remembered that *physical force is different from everyday force or everyday effort*. Effort is probably best approximated by the concept of *(physical) power*, usually abbreviated P, and defined as Ref. 106

$$P = \frac{dW}{dt} = \mathbf{F} \cdot \mathbf{v} \qquad (44)$$

in which *(physical) work* W is defined as $W = \mathbf{F} \cdot \mathbf{s}$. Physical work is a form of energy, as you might want to check. Work has to be taken into account when the conservation of energy is checked. Note that a man walking carrying a heavy rucksack is not doing (almost) any work at all. With the definition of work you can solve the following puzzle: what happens to the electricity consumption of an escalator if you walk on it instead of standing still? Challenge 235 n

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 an incorrect view, namely the so-called *Aristotelian view*, in which force is proportional to velocity. Sometimes it is even stated that they use a different concept of *state* of motion. It is then added with a tone of superiority how wrong all this is. This is a typical example of intellectual disinformation. Every student knows from riding a bicycle, from throwing a stone or from pulling objects that increased effort results in increased speed. The student is right; wrong are those theoreticians who deduce that the student has a mistaken concept of *force*. In fact, the student is just using, instead of the *physical* concept of force, the *everyday* version, namely effort. Indeed, the effort exerted by gravity on a flying stone is smallest at the highest point of the trajectory. Understanding the difference between physical force and everyday effort is the main hurdle in learning mechanics.** Ref. 107

Often the flow of momentum, equation (42), is not recognized as the definition of force. This is mainly due to an everyday observation: there seem to be forces without any associated acceleration or momentum change, such as in a string under tension or in water of high pressure. Pushing against a tree, there is no motion, yet a force is applied. If force is momentum flow, where does the momentum go? It flows into the slight deformations of the arm and the tree. In fact, when one starts pushing and thus deforming, the associated momentum change of the molecules, the atoms, or the electrons of the two bodies can be observed. After the deformation is established, and looking at even higher magnification, one

---

[*] 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 or fluid bodies. Surprisingly and in contrast to frequently made statements, equation (43) is even correct in relativity, as shown on page 232. Ref. 16

[**] 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 at the same time that mass is defined using force (the latter part of the sentence being a fundamental mistake).

indeed finds that a continuous and equal flow of momentum is going on in both directions. By the way, the nature of this flow will be clarified in the part on quantum theory.

Since force is net momentum flow, force is needed as a separate concept only in everyday life, where it is useful in situations where net momentum flows are smaller than the total flows. At the microscopic level, momentum alone suffices for the description of motion. For example, the concept of *weight* describes the flow of momentum due to gravity. Thus we will almost not use the term 'weight' in our adventure.

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 by pumping momentum into objects? Table 17 gives an overview.

Every example of motion, from the one which lets us choose the direction of our gaze to the one which carries a butterfly through the landscape, can be put into one of the two leftmost columns of Table 17. Physically, the two columns are separated by the following criterion: in the first class, the 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. Just check. Challenge 237 e

Friction can be so strong that all motion of a body against its environment is made impossible. This type of friction, called *static friction* or *sticking friction*, is common and important: without it, the turning wheels in bicycles, trains and cars would have no effect. Not a single screw would stay tightened. We could neither run nor walk in a forest, as the soil would be more slippery than polished ice. 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. Without static friction, the propellers in ships, aeroplanes and helicopters would not be effective and the wings of aeroplanes would produce no lift to keep them in the air. In short, static friction is required in every case that we want to move against the environment.

Once an object moves through its environment, it is hindered by another type of friction; it is called *dynamic friction* and acts between bodies in relative motion. Without it, falling bodies would always rebound to the same height without ever stopping on the floor; neither parachutes nor brakes would work; worse, we would have no memory, as we will see later on.*

As the motion examples in the second column of Table 17 include friction, in those examples macroscopic energy is not conserved; the systems are *dissipative*. In the first column, macroscopic energy is constant; the systems are *conservative*.

The first two columns can also be distinguished using a more abstract, mathematical criterion: on the left are accelerations which can be derived from a potential, on the right,

---

* 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.

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 in certain directions of motion. This so-called 'superlubrification' is presently a topic of research. Ref. 108

Situations which can lead to acceleration	**Situations which only lead to deceleration**	**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
growth of mountains	meteorite crash	swimming of larvae
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		
bow and arrow	trouser suspenders	ultrasound motor
bent trees standing up again	pillow, air bag	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 passing trains	shock absorbers	water turbine
nuclei		
radioactivity	plunging into sun	supernova explosion
biology		
bamboo growth	find example! Challenge 236	molecular motors

Table 17 A selection of processes and devices changing the motion of bodies

decelerations 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 to calculate the trajectory of an object, instead of the three values of the acceleration or the force. Moreover, the magnitude of the velocity of an object at any point can be calculated directly from energy conservation.

In the processes from the second column this is impossible. These are the cases where we necessarily have to use force if we wants to describe the motion of the system. For example, the wind resistance of a body is *roughly* given by

$$F = 1/2 c_w \rho A v^2 \qquad (45)$$

where A is the area of its cross section, v its velocity relative to the air, ρ is the density of air, and where the *drag coefficient* c_w is a pure number which depends on the shape of the moving object. You may check that aerodynamic resistance cannot be derived from apotential.[*] Challenge 238

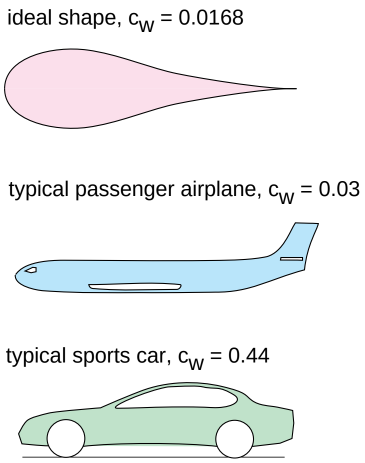


Figure 53 Shapes and air resistance

The drag coefficient c_w is found experimentally to always be 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 of 0.44 and higher.[**]

Ref. 110 Wind resistance is also of importance to humans, in particular in athletics. It is estimated that 100 m sprinters spend between 3% and 6% of their power overcoming drag. This leads Challenge 240 to varying sprint times t_w when wind of speed w is involved, related bythe expression

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

where 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 the formula?) Challenge 241

In contrast, static friction has different properties. It is proportional to the force pressing Ref. 111 the two bodies together. Why? Studying the situation in more detail, sticking friction is found to be proportional to the actual contact area. It turns out that putting two solids into contact is rather like turning Switzerland upside down and putting it onto Austria; the area

---

[*] Such a statement about friction is correct only in three dimensions, as is the case in nature; in the case of a Challenge 239 n single dimension, a potential can *always* be found.

[**] The topic of aerodynamic shapes is even more interesting for fluid bodies. They are kept together by *surface tension*. For example, surface tension 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 millimetres, and that larger rain drops are *lens* shaped, with the flat part towards the bottom. The usual tear shape is *not* encountered Ref. 109 in nature; something vaguely similar to it appears during drop detachment, but *never* during drop fall.

of contact is much smaller than the one estimated macroscopically. The important point is that 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; researchers are investigating the issues using instruments such as atomic force microscopes, lateral force microscopes and triboscopes. One result of these efforts are computer hard disks with longer lifetime, as the friction between disk and reading head is a central quantity determining the lifetime.

All examples of friction are accompanied by an increase in 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 and friction itself are due to motion of large numbers 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 – then also appear, and motion is found to be conserved. As a result, all motion is conservative on microscopic scale. Therefore, on 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 microscopic details.[**]

> Et qu'avons-nous besoin de ce moteur, quand l'étude réfléchie de la nature
> nous prouve que le mouvement perpétuel est la première de ses lois ?[***]
> Donatien de Sade *Justine, ou les malheurs de la vertu*.

Complete states: initial conditions

> Quid sit futurum cras, fuge quaerere ...[****]
> Horace, *Odi*, lib. I, ode 9, v. 13.

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

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

---

[*] 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 strongly criticized at that time; only a generation later, when quantum mechanics quietly got rid of the concept for good, did the idea become 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 one of the pillars of general relativity, as we will see later on.

[**] 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.

[***] 'And whatfor do we need this motor, when the reasoned study of nature proves us that perpetual motion is the first of its laws?'

Ref. 45 [****] 'What future be tomorrow, never ask ...' Horace is Quintus Horatius Flaccus (65–8 BCE), the great Roman poet the quantities with an index o, such as $\mathbf{x}_\mathrm{o}, \mathbf{v}_\mathrm{o}$, etc., are called *initial conditions*. They are necessary for any description of motion. Different physical systems have different initial conditions. Initial conditions thus specify the *individuality* of a given system. Initial conditions also allow us to distinguish the present situation of a system from that at any previous time: initial conditions specify the *changing aspects* of a system. In other words, they summarize the past of a system.

See page 33 Initial conditions are thus precisely the properties we sought for a description of the *state* of a system. To find a complete description of states we thus only need a complete description of initial conditions. It turns out that for gravitation, like for all other microscopic interactions, there is *no* need for initial acceleration $\mathbf{a}_\mathrm{o}$, initial jerk $\mathbf{j}_\mathrm{o}$, or higher-order initial quantities. Indeed, in nature acceleration and jerk depend on the properties of objects alone, and do not depend on their past. For example, the expression $a = GM/r^2$, giving the acceleration of a small body near a large one, does not depend on the past at all, but only on the environment. The same happens for the other fundamental interactions, as we will find shortly.

See page 56 The *complete state* of a moving mass point is thus described by specifying its position and its momentum for all instants of time. Thus we have achieved a complete description of the intrinsic properties 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 we also need orientation, angular velocity, and angular momentum. Can you specify the necessary quantities in the case of extended elastic bodies or fluids? Challenge 242

The set of all possible states of a system is given a special name: it is called the *phase space*. We will use the concept repeatedly. Like any space, it has a number of dimensions. Can you specify it for a system made of N point particles? Challenge 243

However, there are situations in nature where the motion of an object depends on other characteristics than its mass; motion can depend on its colour (can you find an example?), Challenge 244 n on its temperature, and on a few other properties which we will soon discover. Can you give an example of an intrinsic property we have missed so far? Challenge 245 n And for each intrinsic property there are state variables to discover. These new properties are the basis of field of physical enquiry beyond mechanics. We must therefore conclude that we do not have a complete description of motion yet.

It is interesting to recall an older challenge and ask again: does the universe have initial conditions? Challenge 246 n 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 the history of a system.

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, we jump down, we cannot stop the jump in the middle of the trajectory; once the jump is began, it is unavoidable and determined, like all passive motion. However, when we start moving an arm, we 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? Challenge 247 e

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

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

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

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

we then can determine the motion, i.e. the complete trajectory $\mathbf{x}(t)$. Due to this possibility, an equation such as (48) is called an *evolution equation* for the motion of the object. (Note that the term ‘evolution’ has different meanings in physics and in biology.) An evolution equation always expresses the observation that not all types of change are observed in nature, but only certain specific cases. Not all imaginable sequences of events are observed, but only a limited number of them. In particular, equation (48) expresses that from one instant to the next, objects change their motion based on 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 different meanings, let us distinguish it carefully from several similar concepts, to avoid misunderstandings.

Motion can be deterministic and at the same time still be *unpredictable*. The latter property can have four origins: an impracticably large number of particles involved, the complexity of the evolution equations, insufficient information on initial conditions, or strange shapes of space-time. The weather is an example where the first three conditions are fulfilled at the same time.** Nevertheless, its motion is still deterministic. Near black holes all four cases apply together. We will discuss black holes in the section on general relativity. Nevertheless, near black holes, motion is still deterministic.

---

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

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

Motion can be both deterministic and 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, such as an drop of ink spreading in clear water. In all such cases the randomness and the irreproducibility are only apparent; they disappear when the description of states and initial conditions in the microscopic domain are included. In short, determinism does not contradict *(macroscopic) irreversibility*. However, on the microscopic scale, deterministic motion is always reversible.

A final concept to be distinguished from determinism is *acausality*. Causality is the requirement that cause must precede the effect. This is trivial in Galilean physics, but becomes of importance in special relativity, where causality implies that the speed of light is a limit for the spreading of effects. Indeed, it seems impossible to have deterministic motion (of matter and energy) which is *acausal*, i.e. faster than light. Can you confirm this? This topic will be deepened in the section on special relativity. Challenge 248 n

Saying that motion is 'deterministic' means that it is fixed in the future *and also 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 *past* observations correctly. It is our lack of freedom to change the past which results in our lack of choice in the description of nature that is so central to physics. In this sense, the term 'initial condition' is an unfortunate choice, because it automatically leads us 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 specific state. This state can be either initial, intermediate, or final. Deterministic motion is uniquely specified into the past and into the future.

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 first that we are able to define sequences among observations, and second, that unrestricted change is impossible. This is in contrast to movies, where one person can walk through a door and exit into another continent or another century. In nature we do not observe metamorphoses, such as 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 observation that *sequential changes are restricted to a single possibility*. Challenge 249 n

Since determinism is connected to the use of the concept of time, new questions arise whenever 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.

---

* Mathematicians have developed a large number of tests to determine whether 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 Monte Carlo test. Ref. 112

In summary, every description of nature which uses the concept of time, such as that of everyday life, that of classical physics and 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 **??**

The idea that motion is determined often produces fear, because we are taught to associate determinism with lack of freedom. On the other hand, 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 opposed to happiness.

But what is free will precisely? Much ink has been consumed trying to find a precise definition. One can try to define free will as the arbitrariness of the choice of initial conditions. However, 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 these definitions face the same simple problem: whatever the definition, there is *no way* to experimentally prove that an action was performed freely. The possible definitions are useless. In short, free will *cannot* be observed. (Psychologists also have a lot of their own data to underline this, but that is another topic.)

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. Of course, we still have to investigate this question in other domains, in the very small and in the very large. Indeed, we will change our opinion several times. On the other hand, we know the result of everyday life: the concept of curiosity is based on the idea that everything discovered is useful afterwards. If nature continually surprised us, curiosity would make no sense.

Another observation speaks against surprises: in the beginning of our walk we defined time using the continuity of motion; later on we expressed this by saying that time is a consequence of the conservation of energy. Conservation is the opposite of surprise. By the way, a challenge remains: can you show that time would not be definable even if surprises existed only *rarely*? Challenge 250

In summary, so far we have no evidence that surprises exist in nature. Time exists because nature is deterministic. Free will cannot be defined with the precision required by physics. Given that there are no sudden changes, there is only one consistent definition of free will: it is a *feeling*, in particular of independence of others, of independence from fear, and of accepting the consequences of one's actions. Free will is a feeling of satisfaction. This solves the apparent paradox; free will, being a feeling, exists as a human experience, even Ref. 113

---

[*] That can be a lot of fun though.

though all objects move without any possibility of choice. There is no contradiction.[*]

Ref. 114 Even if human 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 Challenge 252 e imply for your life, for the actions, the choices,the responsibilities and the pleasures you encounter?[**] If you conclude that being determined is different from being free, you should change your 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.

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

A strange summary about motion

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

Classical mechanics describes nature in a rather simple way. *Objects* are permanent and massive entities localized in space-time. *States* are changing properties of objects, described by position in space and instant in time, by energy and momentum, and by their rotational equivalents. *Time* is the relation between events measured by a clock. *Clocks* are devices in undisturbed motion whose position can be observed. *Space* and position is the relation between objects measured by a meter bar. *Meter bars* are devices whose shape is subdivided by some marks, fixed in an invariant and observable manner. *Motion* is change of position with time (times mass); it is determined, does not show 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 Challenge 254 n spot it? Each of the two central concepts of motion is defined with help of the other. Physicists had worked for about two hundred years on classical mechanics without noticing or wanting to notice the situation. Even thinkers with an interest to discredit science did not Challenge 255 n point it out. Can an exact science be based on a circular definition? Obviously, physics has done quite well so far. Some even say the situation is unavoidable in principle. Despite these opinions, undoing this logical loop is one of the aims of the rest of our walk. To achieve it, we need to substantially increase the level of precision in our description of motion.

---

[*] That free will is a feeling can also be confirmed by careful introspection. The idea of free will always appears *after* an action has been started. It is a beautiful experiment to sit down in a quiet environment, with the plan to make, within an unspecified number of minutes, a small gesture, such as closing a hand. If you carefully observe, Challenge 251 e in all detail, what happens inside yourself around that very moment of decision,you find either a mechanism which led to the decision, or a diffuse, unclear mist. You never finds free will. Such an experiment is a beautiful way to experience deeply the wonders of the self. Experiences of this kind might also be one of the origins of human spirituality, as they show the connection everybody has with the rest of nature.

Challenge 253 n [**] If nature's 'laws' are deterministic, are they in contrast with moral or ethical 'laws'? Can people still be held responsible for their actions?

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

References

Aiunt enim multum legendum esse, non multa.
Plinius*

Typeset in January 2003

1 A beautiful book explaining physics and its many applications in nature and technology vividly and thoroughly is PAUL G. HEWITT, JOHN SUCHOCKI & LESLIE A. HEWITT, *Conceptual Physical Science*, Second Edition, 1999.

A book famous for its passion for curiosity is RICHARD P. FEYNMAN, R.B. LEIGHTON & M. SANDS, *The Feynman Lectures on Physics*, Addison Wesley, 1977.

A lot on motion can be learned from quiz books. A good collection is JEAN-MARC LÉVY-LEBLOND, *La physique en questions - mécanique*, Vuibert, Paris, 1998.

Another excellent quiz collection is YAKOV PERELMAN, *Oh, la physique*, Dunod, Paris, 2000, a translation from the Russian original.

A good problem book is W.G. REES, *Physics by Example: 200 Problems and Solutions*, Cambridge University Press, New York 1994.

A good history of physical ideas is given in the excellent text by DAVID PARK, *The How and the Why*, Princeton University Press, 1988. Cited on pages 25, 105 and 163.

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.

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.

See page 490

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 33. 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 pages 28 and 50.

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 28.

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.

---

[*] 'Read much, but not anything.' Ep. 7, 9, 15. Gaius Plinius Secundus (23 or 24, Novum Comum–79, Vesuve eruption), roman writer, especially famous for his large, mainly scientific work *Naturalis historia*, which has been translated and read for almost 2000 years.

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 pages 29 and 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 29.

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, 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 pages 30 and 73.

11 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 449. At http://vanbc.wimsey.com/~chris/JPS/JPS.html you can find the web site maintained by the Jean Piaget Society. Cited on pages 31, 37 and 39.

12 A description of the reptile brain in comparison to the mammalian and the human one can be found in ... Cited on page 32.

13 The lower left corner movie can be reproduced on a computer after typing the following lines in the Mathematica software package: Cited on pages 32 and 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.

14 An introduction into perception research is E. BRUCE GOLDSTEIN, *Perception*, Books/Cole, 5th edition, 1998. Cited on page 33.

15 All fragments from Heraclitos are from JOHN MANSLEY ROBINSON, *An introduction to early Greek philosophy*, Houghton Muffin 1968, chapter 5. Cited on pages 33 and 144.

16 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 pages 36, 93 and 108.

17 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 36.

18 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.

19 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 256 n

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 pages 38 and 218.

20 RAINER FLINDT, *Biologie in Zahlen – Eine Datensammlung in Tabellen mit über 10.000 Einzelwerten*, Spektrum Akademischer Verlag, Heidelberg, 2000. Cited on page 38.

21 Two jets with that speed have been observed by I.F. Mirabel & L.F. Rodríguez, *A superluminal source in the Galaxy*, Nature **371**, pp. 46–48, 1994, as well as the comments on p. 18. Cited on page 38.

22 The clocks in our brain are described in ... The clocks in our body are described in ... Cited on page 40.

23 This has been shown among others by the work of Anna Wierzbicka mentioned in more detail in the intermezzo following this chapter, on page 457. 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.

24 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 41.

25 Aristotle rejects the idea of the flow of time in ... Cited on page 42.

26 Perhaps the most informative of the books about the 'arrow of time' is Hans Dieter Zeh, *The Physical Basis of the Direction of Time*, Springer Verlag, 4th edition, 2001. It is still the best book on the topic. Most other texts – have a look on the internet – lack clarity of ideas.

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 42.

27 On the issue of absolute and relative motion there are many books about few issues. Examples are John Barbour, *Absolute or Relative Motion? A Study from the Machian Point of View of the Discovery and the Structure of Spacetime Theories*, Cambridge U. Press, Cambridge, 1989, or John Earman, *World Enough and Spacetime: Absolute vs Relational Theories of Spacetime*, MIT Press, Cambridge, 1989. Cited on page 46.

28 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. Gelbaum & J.M.H. Olmsted, *Counterexamples in Analysis*, Holden-Day, 1964, and *Theorems and Counterexamples in Mathematics*, Springer Verlag, 1990. Cited on page 49.

29 The beautiful but not easy text is Steve Wagon, *The Banach Tarski Paradox*, Cambridge University Press, 1993. Cited on pages 49 and 938.

30 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 49.

31 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 Burša & Karel Pěč, *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 interesting effects found by these investigations. Cited on pages 52 and 97.

32 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 $12.048\,\mathrm{m/s} = 43.372\,\mathrm{km/h}$ by Ben Johnson for men, and $10.99\,\mathrm{m/s} = 39.56\,\mathrm{km/h}$ for women.

The 1998 world record for ball juggling is nine balls. Cited on page 53.

33 The arguments of Zeno can be found in Aristotle, *Physics*, VI,9. It can be found translated in almost any language. Cited on pages 55 and 118.

34 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 55.

35 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 pages 56 and 62.

36 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 57.

37 A good biology textbook on growth is ... Cited on page 58.

38 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, December 1975. 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 58.

39 James A. Young, *Tumbleweed*, Scientific American **264**, pp. 82–87, March 1991. The tumbleweed is in fact quite rare, except in in Hollywood westerns, where all directors feel obliged to give it a special appearance. Cited on page 58.

40 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 59.

41 There is also the beautiful book by Penelope Farrant, *Colour in Nature*, Blandford, 1997. Cited on page 60.

42 The laws of cartoon physics can be found in various places on the internet. Cited on page 60.

43 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 61.

44 AETIUS, Opinions, I, XXIII, 3. See JEAN-PAUL DUMONT, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 426, 1991. Cited on page 61.

45 GIUSEPPE FUMAGALLI, *Chi l'ha detto?*, Hoepli, Milano, 1983. The sentence is also the motto of the cities of the Hansa. Cited on pages 61, 112 and 129.

46 For the role and chemistry of adenosinetriphosphate (ATP) in cells and in living beings, see ... Cited on page 66.

47 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 66.

48 A Shell study estimates the world's total energy consumption in 2000 to be 500 EJ. The US department of energy estimates it to be around 416 EJ. We took the lower value here. A discussion and a breakdown into electricity usage (14 EJ) and other energy forms, with variations per country, can be found in S. BENKA, *The energy challenge*, Physics Today **55**, pp. 38–39, April 2002, and in E.J. MONITZ & M.A. KENDERDINE, *Meeting energy challenges: technology and policy*, Physics Today **55**, pp. 40–46, April 2002. Cited on page 69.

49 For an overview, see the paper by J.F. MULLIGAN & H.G. HERTZ, *An unpublished lecture by Heinrich Hertz: 'On the energy balance of the earth'*, American Journal of Physics **65**, pp. 36–45, 1997. Cited on page 70.

50 For a beautiful photograph of this feline feat, see the cover of the journal and the article of J. DARIUS, *A tale of a falling cat*, Nature **308**, p. 109, 1984. Cited on page 72.

51 NATTHI L. SHARMA, *A new observation about rolling motion*, European Journal of Physics **17**, pp. 353–356, 1996. Cited on page 73.

52 C. SINGH, *When physical intuition fails*, American Journal of Physics **70**, pp. 1103–1109, 2002. Cited on page 73.

53 SERGE GRACOVETSKY, *The spinal engine*, Springer Verlag, 1990. Cited on page 74.

54 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 pages 74 and 201.

55 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*, the current of warm water flowing from the caribbean to the north sea, is influenced by it. Cited on page 76.

56 The influence of the Coriolis effect on icebergs was studied most thoroughly by the swedish physicist turned oceanographer Walfrid Ekman (1874–1954); The topic was suggested by the great explorer Fridtjof Nansen, who also made the first observations. In his honour, one speaks of Ekman layer, Ekman transport, and Ekman spirals. Any text on oceanography or physical geography will tell more about them. Cited on page 76.

57 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; the first southern hemisphere test was L.M. TREFETHEN & al., *The bath-tub vortex in the southern hemisphere*, Nature **201**, pp. 1084–1085, 1965. 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 76.

58 The calculation of the period of Foucault's pendulum ... Cited on page 77.

59 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 77.

60 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 78.

61 About the length of the day, see the http://maia.usno.navy.mil web site, or the books by K. Lambeck, *The earth's variable rotation: geophysical causes and consequences*, Cambridge University Press, 1980, and by W.H. Munk & G.J.F. MacDonald, *The rotation of the earth*, Cambridge University Press, 1960. Cited on pages 78 and 99.

62 One example of data is 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 78.

63 The motion of the earth's axis in the last ten years is shown on the mentioned http://maia.usno.navy.mil web site. The International Latitude Service founded by Küstner, is now part of the International Earth Rotation Service; more information can be found on the http://www.iers.org web site. The latest idea is that the circular component of the polar motion, which in the US is called 'Chandler wobble' after the scientist who has expanded on the original discovery by Küstner, is due by two thirds to fluctuations of the ocean pressure at the bottom of the oceans and by one third due to 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 79.

64 For more information about Alfred Wegener, read ...; about plate tectonics, see the http://www.scotese.com web site. On earthquakes, see the http://www.geo.ed.ac.uk/quakexe/quakes web site. On volcanoes, see the http://vulcan.wr.usgs.gov and http://www.dartmouth.edu/~volcano/ web sites. Cited on page 79.

65 The rotation and history of the solar system ... Cited on page 78.

66 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. Note that the web is full of information on the ice ages. Just look up Milankovitch in a search engine. Cited on page 81.

67 Roberta Humphreys & Jeffrey Larsen, *The sun's distance above the galactic plane*, Astronomical Journal **110**, pp. 2183–2188, November 1995. Cited on page 81.

68 C.L. Bennet, M.S. Turner & M. White, *The cosmic rosetta stone*, Physics Today **50**, pp. 32–38, November 1997. Cited on page 82.

69 The equilibrium of ships, so important in car ferries, is an interesting part of ship building; an introduction is given by ... Cited on page 83.

70 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 83.

71 On http://www.sff.net/people/geoffrey.landis/vacuum.html you 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 83.

72 The smallest distances are probed in particle accelerators; the distance can be determined from the energy of the particle beam. Cited on page 50. Challenge 257 n

73 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. In particular, world class long jumpers run at $9.35 \pm 0.15\,\mathrm{m/s}$, with vertical takeoff speeds of $3.35 \pm 0.15\,\mathrm{m/s}$, giving takeoff angles of about (only) $20\,°$. A new technique for achieving higher takeoff angles would allow to increase the world long jump record dramatically. Cited on page 53.

74 This and many other physics surprises are described in the beautiful lecture script by Josef Zweck, *Physik im Alltag*, Skript zur Vorlesung im WS 1999/2000 der Universität Regensburg. Cited on pages 82 and 84.

75 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 86.

76 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 the http://rjf2.biol.berkeley.edu/Full_Lab/FL_Personnel/J_Glasheen/J_Glasheen.html web site.

Several shore birds also have the ability to run over water, using the same mechanism. Cited on page 86.

77 A. Fernandez-Nieves & F.J. de las Nieves, *About the propulsion system of a kayak and of Basiliscus basiliscus*, European Journal of Physics **19**, pp. 425–429, 1998. Cited on page 87.

78 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 88.

79 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 88.

80 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 88.

81 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 thoroughly researched texts provide numerous details and explain various philosophical viewpoints, but lack clear statements and conclusions on the accurate description of nature, and thus are not of help on 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 89.

82 D. Topper & D.E. Vincent, *An analysis of Newton's projectile diagram*, European Journal of Physics **20**, pp. 59–66, 1999. Cited on page 90.

83 This is explained for example by D.K. Firpić & 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 102.

84 About the measurement of spatial dimensions via gravity, see ... Cited on page 92.

85 There are many books explaining the origin of the precise shape of the earth, such as the pocket book S. Anders, *Weil die Erde rotiert*, Verlag Harri Deutsch, Thun, Frankfurt am Main, 1985. Cited on page 92.

86 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.

See also the web site of the *International Earth Rotation Service* at http://hpiers.obspm.fr. Cited on page 92.

87 W.K. Hartman, R.J. Phillips & G.J. Taylor, editors, *Origin of the Moon*, Lunar and Planetary Institute, Houston, 1986. Cited on page 94.

88 Dietrich Neumann, *Physiologische Uhren von Insekten – Zur Ökophysiologie lunarperiodisch kontrollierter Fortpflanzungszeiten*, Naturwissenschaften **82**, pp. 310–320, 1995. Cited on page 94.

89 The origin of the duration of the menstrual cycle is not yet settled; however, there are explanations on how it gets synchronized with other cycles. For a general explanation see Arkady Pikovsky, Michael Rosenblum & Jürgen Kurths, *Synchronization: a Universal Concept in Nonlinear Science*, Cambridge University Press, New York, 2002. Cited on page 95.

90 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 95.

91 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 95.

92 Neil F. Comins, *What if the moon did not exist? – Voyages to earths that might have been*, Harper Collins, 1993. Cited on page 95.

93 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. Cited on page 96.

94 Simon Newcomb, Astronomical Papers of the American Ephemeris **1**, p. 472, 1882. Cited on page 97.

95 A beautiful introduction is the classic G. Falk & W. Ruppel, *Mechanik, Relativität, Gravitation – ein Lehrbuch*, Springer Verlag, Dritte Auflage, 1983. Cited on page 97.

96 J. Soldner, *Berliner Astronomisches Jahrbuch auf das Jahr 1804*, 1801, p. 161. Cited on page 99.

97 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, *The equivalence of inertial and passive gravitational mass*, Annals of Physics (NY) **26**, pp. 442–517, 1964, one of the most interesting and entertaining research articles in experimental physics, and by V.B. Braginsky & V.I.

PANOV, Soviet Physics - JETP **34**, pp. 463–466, 1971. Modern results, with errors less than one part in 10^{12}, are by Y. SU & al., *New tests of the universality of free fall*, Physical Review **D50**, pp. 3614–3636, 1994. Several experiments have been proposed to test the equality in space to less than one part in 10^{16}. Cited on page 101.

98 See L. HODGES, *Gravitational field strength inside the earth*, American Journal of Physics **59**, pp. 954–956, 1991. Cited on page 102.

99 P. MOHAZZABI & M.C. JAMES, *Plumb line and the shape of the earth*, American Journal of Physics **68**, pp. 1038–1041, 2000. Cited on page 103.

100 From NEIL DE GASSE TYSON, *The universe down to earth*, Columbia University Press, 1994. Cited on page 104.

101 GERALD QUINLAN, Nature **363**, pp. 18–19, 1993. See 1

102 See ROBERT MATTHEWS, *Not a snowball's chance ...*, 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 104.

Cited on page 104.

103 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 105.

104 G.J. SUSSMAN & J. WISDOM, *Chaotic Evolution of the Solar System*, Science **257**, pp. 56–62, 1992. Cited on page 105.

105 GEORGE-LOUIS LESAGE, *Lucrèce Newtonien*, Nouveaux mémoires de l'academie royale des sciences et belles letters pp. 404–431, 1747. Cited on page 105.

106 Knowledge is power. Time is money. Now, power is defined as work per time. Inserting the previous equations and transforming them yields

$$\text{money} = \frac{\text{work}}{\text{knowledge}} \quad ; \tag{50}$$

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 108.

107 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 field of physics teaching. Cited on page 108.

108 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 109.

109 C.D. AHRENS, *Meteorology today: an introduction to the weather, climate, and the environment*, West Publishing Company, St. Paul, 1991. Cited on page 111.

110 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://www.arxiv.org/physics/9705004 together with other papers on similar topics by the same author. Cited on page 111.

111 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 111.

112 The test of randomness can be found in the text ... The topic has many philosophical implications, discussed e.g. by ... Cited on page 115.

113 For one aspect of the issue, see for example the book by Bert Hellinger, *Zweierlei Glück*, Carl Auer Systeme Verlag, 1997. The author explains how to live serenely and with the highest possible responsibility of one's actions, by reducing entanglements with the destiny of others. He describes a simple and powerful technique to realize this goal.

For another aspect, see the text by peace Nobel price winner Aung San Suu Kyi, *Freedom from Fear*, Penguin, 1991. Cited on page 116.

114 Henrik Walter, *Neurophilosophie der Willensfreiheit*, ... Cited on page 117.

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

Πλεῖν ἀνάγϰε, ζῆν οὐϰ ἀνάγϰη.
Navigare necesse, vivere non necesse.*
Pompeius

Typeset in January 2003

The discovery of the universal law of gravity teaches an important lesson. All over the earth, even in Australia, people observe that stones fall 'down.' It is thus necessary to look for a description of gravity valid *globally*. It then is only a small additional step to deduce the result $a = GM/r^2$. In short, thinking globally provides an efficient way to increase the precision of motion description. How can we be as global as possible? It turns out there are six approaches to follow. Each of them is of help on our way to the top of Motion Mountain. We first give an overview of these approaches and present them in detail afterwards.

- The first global approach is a reaction to a limitation of what we learned so far. Whenever we calculate the motion of a particle by calculating the acceleration it is subjected to, we are using the most *local* description of motion possible. Indeed, the acceleration at a certain place and instant of time only determines the position of the particle *just after* that moment and the motion *just following* that place.

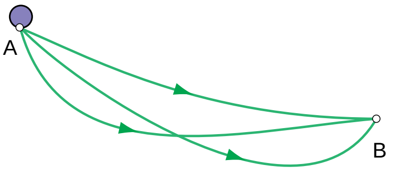


Figure 54 What shape of rail allows the black stone to glide most rapidly from point A to the lower point B?

Evolution equations thus have a mental horizon of radius zero. The opposite approach is shown in the famous question of Figure 54. The challenge is to find that path which allows the fastest motion between two points.The motion as a whole, for all times and positions, is sought. Challenge 258

The global approach required by approaches such as this one will lead to a description of motion which is simple, powerful and fascinating.

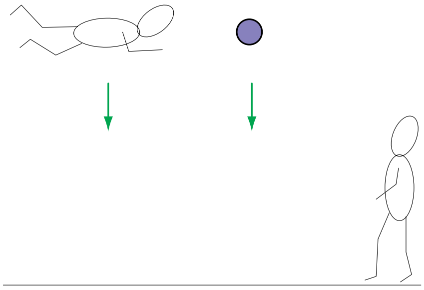
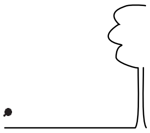


Figure 55 Can motion be described in a manner common to all observers?

- The second global approach to motion emerges when comparing the various descriptions produced by different observers of the same system. For example, observations by somebody falling from a cliff, a passenger in a roller coaster and an observer on the ground will usually differ. Studying their connections and finding a global description, valid for everybody, will lead us to the theory of relativity.
- The third way to look at motion globally is to turn towards the investigations of *extended and rigid* bodies. Their motion can be surprising, as the experiment in Figure 56 shows. In addition,

---

[*] 'To navigate is necessary, to live is not.' Gnaeus Pompeius Magnus (106–48 BCE), as cited by Plutarchus (ca. 45–ca. 125). Ref. 45



studying the interactions among several rigid bodies is essential for the design of machines. As an example, the mechanism in Figure 57 connects the motion of points C and P. It implicitly defines a circle such that one always has the relation $r_C = 1/r_P$ for the distances r from its centre. Are you able to find that circle?

Challenge 259

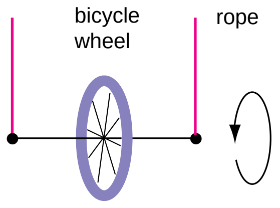


Figure 56 What happens when one rope is cut?

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 in understanding general relativity, as we will see.

Ref. 115

Challenge 260

A final example is the research into the way that human movements, such as the general motion of an arm, are built from a small number of basic motor primitives. These are a few fascinating topics of engineering; unfortunately, we won't have time to explore them in our hike.

Ref. 116

- The description of *non-rigid extended bodies* is the fourth generalization of the study of motion. One part, *fluid mechanics*, studies the flow of honey, water, or air around solid bodies such as spoons, ships, sails, and wings. For example, it investigates how insects, birds and aeroplanes fly,[*] why sail 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.challengenoreggb

Ref. 117

Challenge 261 n

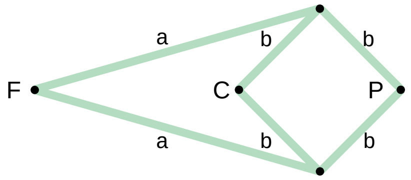


Figure 57 How to draw a straight line with a compass: fix point F, put a pencil into joint P, and move C with a compass along a circle

The other part of the study of extended bodies, the behaviour of deformable *solid* bodies, is called *continuum mechanics*. It studies deformations and oscillations of extended bodies. Among others it explains why bells are made in the shape they are or where materials break when under load. We will mention a few issues from this field in special relativity and quantum theory.

- A fifth general viewpoint, related to the preceding, arises when we ask for the motion of large numbers of particles. In these cases the study of motion is called *statistical mechanics*. We will visit it only briefly and introduce only those concepts we need for our further path.

---

[*] The mechanisms of insect flight are still a research subject. Traditionally, fluid dynamics concentrated on large systems like boats, ships and aeroplanes. 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 required and much more knowledge involved in letting small things fly than large things. More about this topic on page 708.

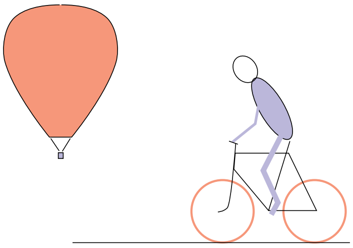


Figure 58 Why do balloons stay inflated? How do you measure the weight of a bicycle rider with a ruler only?

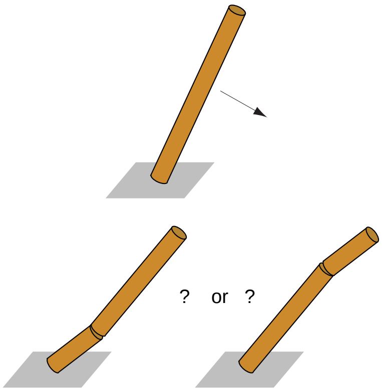


Figure 59 How and where does a falling brick chimney break?

- The sixth and final set of moving systems are those requiring for their description most of the mentioned viewpoints *at the same time*. Such systems form an important part of everyday experience, since *life* itself requires this approach. The formation of a specific number of petals in flowers, the differentiation of the embryo in the womb and the origin of the heartbeat in the human body are examples of such situations. Other examples are the emergence of mountains ridges and cloud patterns, or the formation of sea waves by the wind.

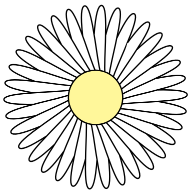
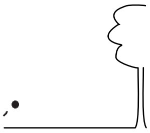


Figure 60 What determines the number of petals in a daisy?

All these are examples of *growth processes*. Physicists speak of *self-organization*. Its topics are the spontaneous appearance of patterns, shapes and cycles. Self-organisation and growth are a common research theme across many sciences, from biology, chemistry and medicine to the geosciences and engineering.

In the following, we give a short introduction into these global descriptions of motion. We will start with the first of the six global descriptions just mentioned, namely the global description of moving point-like objects. This beautiful method, the result of several centuries of collective effort, is the highlight of mechanics. It also provides the basis for all further descriptions of motion we will encounter.



The principle of least action

Motion can be described by numbers. For a single particle, the time dependence of coordinates does precisely that. Writing an expression $(x(t), y(t), z(t))$ for the path of a moving particle is one of the pillars of modern physics. In addition, motion is a type of change. Can change be described by numbers? Yes, it can. A single number is sufficient.

The way to *measure* change was discovered by chance, as a by-product of other studies. Physicists took almost two centuries to find it. Therefore the quantity measuring it has a strange name: it is called *(physical) action*.[*] To remember the connection of action with change, just think about Hollywood movies: a lot of action means that a lot is going on; a large action means a large amount of change.

We are now ready to define action. Imagine taking two snapshots of a system, and attempting to define the amount of change that occurred in between. When do things change a lot, and when do they change only a bit? First of all, a system with a lot of motion shows a lot of change. The action of a system is (usually) the sum of the actions of its subsystems. Secondly, change often builds up over time; in other cases, change can compensate some change which happened just before. Change can increase or decrease. Third, for a system in which motion is stored, transformed or shifted from one subsystem to the other, the change is smaller than for a system where this is not the case.

These properties leave only one possibility: change is measured by the average of kinetic minus potential energy, multiplied by the elapsed time. This product has all properties just mentioned: it (usually) is larger if the system is larger; the product generally builds up with time, except if the evolution compensates something that happened earlier; finally, the product decreases if the system transforms motion into potential energy. Challenge 262 e

The so-called *action* S, measuring the change observed in a system, is thus defined as

$$S = \int_{t_\mathrm{i}}^{t_\mathrm{f}} (T - U)\, dt = \int_{t_\mathrm{i}}^{t_\mathrm{f}} L\, dt \quad . \qquad (51)$$

Let us explore the notation. T is the kinetic energy and U the potential energy we already know. Their difference L, a quantity also called the *Lagrangian (function)* of the system,[**] See page 91 describes what is being added over time, whenever things change. The sign $\int$ is a stretched

---

[*] Note that the action measuring change is not the same as the 'action' appearing in sentences such as 'action equals reaction.' This last expression, coined by Newton, has not stuck; therefore the term has been recycled. After having been used for some intermediate meaning, it was given the modern meaning used above. That is the only meaning used in the following.

Similarly, even the 'principle of least action' used to be different from the one of this chapter. The name has been recycled as well, to designate what was often called *Hamilton's principle* in the Anglo-Saxon world, even though it is (mostly) due to others, primarily to Leibniz. The old names and meanings are falling into disuse and are not continued here.

Behind all these shifts in vocabulary is the story of a two-centuries-long, intense search to describe motion with so-called *extremal* or *variational principles*; the game was to complete and improve the work by Leibniz. These searches are only of historical value today, because all these historical principles are special cases of the Ref. 118 one described here, which is the most general one. It provides the key to all the others.

[**] It is named after Giuseppe Luigi Lagrangia (Torino 1736–Paris 1813), better known as Joseph Louis Lagrange. He was the most important mathematician of his time. He developed most of the mathematical tools used nowadays for calculations in classical mechanics and classical gravitation.

Observation	Action value
Smallest measurable change	$0.5 \cdot 10^{-34}$ Js
Exposure of photographic film	$1.1 \cdot 10^{-34}$ Js to 10^{-9} Js
Wing beat of a fruit fly	ca. 1 pJs
Flower opening in the morning	ca. 1 nJs
Getting a red face	ca. 10 mJs
Held versus dropped glass	0.8 Js
Average tree bent by the wind from one side to the other	500 Js
Making a white rabbit vanish by 'real' magic	100 PJs
Hiding a white rabbit	ca. 0.1 Js
Maximum brain change in a minute	ca. 5 Js
Levitating yourself within a minute by 1 m	ca. 40 kJs
Car crash	ca. 2 kJs
Birth	ca. 2 kJs
Change due to a human life	ca. 1 EJs
Driving car stops within an eyelash	20 kJs
Large earthquake	ca. 1 PJs
Driving car disappears within an eyelash	1 ZJs
Sun rise	ca. 0.1 ZJs
Gamma ray burster before and after explosion	ca. 10^{56} Js
Universe after one second has elapsed	undefined and undefinable

Table 18 Some action values for changes either observed or imagined

'S' for 'sum', is pronounced 'integral of' and designates the operation of adding up in tiny time steps dt. The initial and the final times are written below and above the integration sign. The adding up operation, called *integration*, thus simply means the measurement of the grey area shown in Figure 61.

Challenge 263 e We can see integration also as an abbreviation, namely

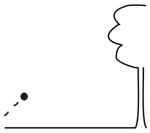
$$\int_{t_\mathrm{i}}^{t_\mathrm{f}} L(t)\, dt = \lim_{\Delta t \to 0} \sum_{\mathrm{m}=\mathrm{i}}^{\mathrm{f}} L(t_\mathrm{m}) \Delta t \qquad (52)$$

defining it as the sum one gets when the time slices get as small as possible.[*] Since the $\sum$ sign also means a sum, and since a tiny Δt is written dt, we can understand why integration is written the way it is. Remembering these meanings will allow you to understand every formula with integration sign you will ever see. Also this ingenious notation, by the way, is due to Leibniz. Physically, the integral measures the *effect* that L builds up over time. Indeed, the action is called 'effect' in other languages, such as German. It is then said that the action is the integral of the Lagrangian over time.

The unit of action, and thus of physical change, is therefore the unit of energy, the Joule, times the unit of time, the second. Change is measured in Js. A large value means a big change. Some examples are given in Table 119.

---

[*] Of course, there are more details to integration. They can be found in Appendix D.



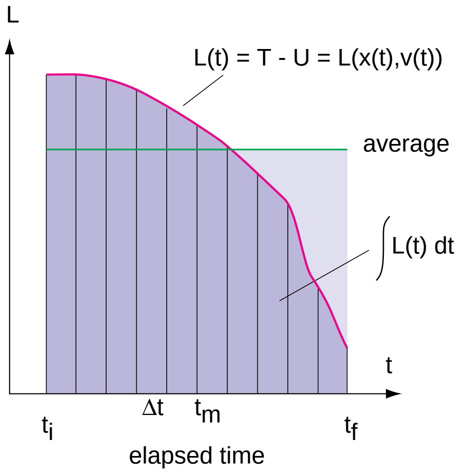


Figure 61 Defining a total effect by adding up small effects over time

To understand the definition of action, let's first take the simplest case; we take a system for which the potential energy is zero, such as a particle moving freely. Obviously, a large kinetic energy means lots of motion. If we observe the particle at time t_i and again at time t_f, the more distant the initial and the final instants, the larger the change. In addition, the observed change is larger if the particle moves more rapidly, as its kinetic energy is larger.

Now take the next case: assume that there is a potential involved. For example, a falling particle exchanges potential energy for kinetic energy during its motion. The more energy is exchanged, the less change there is. Hence the minus sign in the definition of L.

When drawing the curve for $L(t)$ in the case of free fall, we find that the definition of integration makes us count the grey surface *below* the time axis *negatively*. This allows the compensation of previously build up change, as we expected.

Change can also be measured for a system made of several components. We simply add all kinetic and subtract all potential energies. This allows us to define actions even for gases, liquids, and solid matter. In short, action is an *additive* quantity.

The action thus measures all changes observed in a system between two instants using a single number. Whatever happens, be it an explosion, a caress or a colour change, one number is sufficient. We will discover that this approach is also possible in relativity and quantum theory. Any change going on in a system can be measured by a single number.

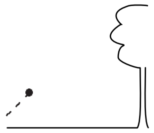
Now that we have defined a precise method to measure change, we can specify the way in which it allows the description of motion. In nature, the change happening between two instants is always the *smallest* possible. *In nature, action is minimal.*[*] Nature always chooses that trajectory, that path or that way to move among all possible and imaginable ones for which the change is *minimal*. Let us study a few examples.

In the simple case of a free particle, when no potentials are involved, the principle of minimal action implies that it moves in a *straight* line with *constant* velocity. All other paths would lead to larger actions. Can you confirm this? Challenge 264 e

Similarly, a thrown stone flies along a parabola, – or more precisely, along an ellipse, as we found out – because any other path, say one in which the stone makes a loop in the air, would imply a *larger* action. You might want to check for yourself that such a weird stone would not keep change to aminimum. Challenge 265 e

---

[*] In fact, in some rare, academic situations the action is maximal, so that the snobbish form of the principle states that the action is 'stationary,' or an 'extremum,' meaning minimal *or* maximal. The condition of vanishing variation encompasses both cases.



All observations confirm the simple statement: things always move producing the smallest possible value for the action. The statement applies to the full path and to each of its segments.

It is customary to express the idea of minimal change in the following way. The action varies when the path is varied. You remember from high school that at a minimum, the variation of a quantity vanishes; a minimum has a horizontal slope. In the present case, we do not vary a variable, we vary complete paths; hence we do not say slope, but variation and write it δS.

The *principle of least action* thus states:

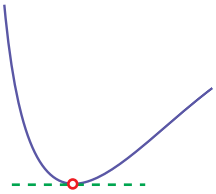


Figure 62 The minimum of a curve has vanishing slope

$$\triangleright\ \textit{The actual trajectory between specified end points is given by } \delta S = 0 \tag{53}$$

Mathematicians call this a *variational principle*. The end points are mentioned to make clear that for the comparison of actions we have to compare motions with the same initial and final situations.

Before we discuss the result, we check that the principle is equivalent to the evolution equation. We will see that this is always the case if motion is described with sufficient precision.[*] The equivalence is always shown in the same, standard procedure. (This procedure

---

[*] 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 note. First of all, Lagrangians do not exist for non-conservative, or *dissipative* systems. In other words, for any motion involving *friction*, as there is no potential, there is no action. One way out is to use a generalized formulation of the principle of least action. Whenever there is no potential, we can express the *work* variation δW between different trajectories as

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

Motion is then described in the following way:

$$\triangleright\ \textit{The actual trajectory is given by } \int_{t_i}^{t_f} (\delta T + \delta W)\, dt = 0 \quad \textit{provided} \quad \delta x(t_i) = \delta x(t_f) = 0 \quad . \tag{55}$$

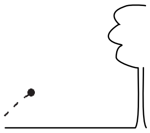
The quantity being varied has no name; it could be called a generalized version of change. You might want to
Challenge 266 check that it leads to the correct evolution equations.In other words, *proper* Lagrangian descriptions exist only for *conservative* systems; however, for dissipative systems the principle can be generalized and remains useful.

Physicists will disagree with this classification and prefer another way out. What a mathematician calls a generalization is a special case for a physicist; principle (55) hides that *all* friction derives from the usual principle of minimal action, if we include the complete microscopic details. There is no friction in the microscopic domain. Friction is an approximate concept.

Nevertheless, additional, more mathematical viewpoints are useful; for example, they lead to interesting discoveries such as further limitations on the use of Lagrangians. These limitations, which apply only if the world is viewed as purely classical – which it isn't – were discovered in times when computers where not available, and when such studies were fashionable. Here are a few results.

The generalized coordinates used in Lagrangians are not necessarily the Cartesian ones. Generalized coordinates are especially useful when there are *constraints*, such as in the case of a pendulum, where the weight always has to be at the same distance from the suspension, or in the case of an ice skater, where the skate has to
Ref. 119 move in the direction it is pointing. Generalized coordinates may even be mixtures of positions and momenta. They can be divided into a few general cases.

Generalized coordinates are called *holonomic-scleronomic* when they are related to Cartesian coordinates in a fixed way, independently of time; the pendulum is an example, as is a particle in a potential. Coordinates are



is part of the so-called *calculus of variations*.) To start with, the condition $\delta S = 0$ implies that the action, i.e. the grey area under the curve, is a minimum. A little bit of thinking shows that a Lagrangian $L(x_n, v_n) = T(v_n) - U(x_n)$ implies that all motionfollows Challenge 267

$$\frac{d}{dt}\left(\frac{\partial T}{\partial v_n}\right) = \frac{\partial U}{\partial x_n} \tag{56}$$

where n counts all coordinates of all particles.[*] For a single particle, these *Lagrange's equations of motion* reduce to Challenge 268

$$m\mathbf{a} = \nabla U \quad . \tag{58}$$

This is the original evolution equation; the principle of least action thus implies the equation of motion. (Can you show the converse?) Challenge 269

In other words, *all systems evolve in such a way that the necessary change is as small as possible.* Nature is economic. Nature is thus the opposite of a Hollywood thriller, where the action is maximized; nature is more like a wise old man who keeps his actions to a minimum. Or, if you prefer, nature is a Dr. Dolittle.

The principle of minimal action also states that the actual trajectory is the one for which the *average* of the Lagrangian over the whole trajectory is minimal. The graph also shows the connection. Can you confirm it?This way to look at the action allows to deduce Lagrange's equations (56) directly. Challenge 270

The description of motion with the principle of least action thus distinguishes the actual trajectory from all other imaginable ones. This fact lead to Leibniz's famous interpretation that the world is the 'best of all possible worlds.'[**] We may dismiss this somewhat metaphysical speculation, but not the inherent fascination of the result. Leibniz was so excited about it because expression (53) 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? What do you think? At the present point, we have a hint *against* this possibility. A final answer to Challenge 271 n

---

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 two terms rheonomic and scleronomic are due to Ludwig Boltzmann. See page 176

See page 112 The more general situation is called *anholonomic*; the term is due, like the term holonomic, to Heinrich Hertz. Lagrangians work well only for holonomic systems.

To sum up, even though the use of Lagrangians and of action 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.

[*] The most general form for a Lagrangian $L(q_n, \dot{q}_n, t)$, using generalized coordinates q_n, leads to Lagrange equations of the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i} \quad . \tag{57}$$

In order to deduce them, we also need the relation $\delta\dot{q} = d/dt(\delta q)$; it is valid only for the *holonomic* coordinates introduced above and explains their importance.

It should also be noted that the Lagrangian for a moving system is not unique; however, the study of how the various Lagrangians for a given moving system are related is not part of this walk. Ref. 120

[**] This idea was ridiculed by the French philosopher Voltaire (1694–1778) in his lucid writings, notably in the brilliant book *Candide ou l'optimisme*, 1759, available in paperback from Folio-Gallimard, 1992.

this question will emerge only in the last part of our adventure.

Compared to description of motion with evolution equations, description with a Lagrangian has several advantages. First of all, it is usually more *compact* than writing the corresponding evolution equations. For example, only *one* Lagrangian is needed for one system, *independently* of the number of particles. One also makes fewer mistakes, especially sign mistakes, as one rapidly learns when performing calculations: just try to write down the evolution equations for a chain of masses connected by springs, and then compare the effort with a derivation using a Lagrangian.We will discover another example shortly: Challenge 272 David Hilbert took only a few weeks to deduce the equations of motion of general relativity using a Lagrangian, after Albert Einstein had worked for ten years searching for them directly.

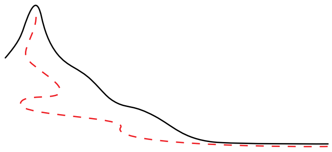
In addition, the description with a Lagrangian is valid with *any* type of coordinates describing the objects of investigation. Coordinates do not have to be Cartesian, they can be chosen as one prefers, cylindrical, spherical, hyperbolical, etc. The advantage of using these so-called *generalized coordinates* will not be studied in our walk; they allow one to rapidly calculate the behaviour of many mechanical systems which are too complicated to be described with Cartesian coordinates. For example, for the programming of the motion of robot arms, joint angles provide a clearer description than Cartesian coordinates of arms ends. Angles are non-Cartesian coordinates. They simplify calculations considerably, such as the task of finding the most economical way to move the hand of a robot from one point to the other.

More importantly, the Lagrangian allows one to quickly deduce the key properties of a system, namely its *symmetries* and its *conserved quantities*. We will develop this important ability shortly, and then use it regularly throughout our walk. See page 150

Finally, the Lagrangian formulation can be generalized to encompass *all types of interactions*. The concepts of kinetic and potential energy are interaction independent. Indeed, the principle of least action can also be used in electricity, magnetism, and optics. It is central to general relativity and to quantum theory, and allows one to easily relate both fields to classical mechanics.

When the principle of least action became well-known, people applied it to an ever-increasing number of problems. Ref. 118 Today, Lagrangians are used in everything from the study of elementary particle collisions to the programming of motion in artificial intelligence. However, we 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 type of motion, but only a different view of it. Challenge 273 In fact, the correspondence is so close that we can say that the search of a new physical 'law' of motion is 'simply' or 'only' the search for a new Lagrangian. This is not a surprise, as the description of nature requires the description of change, and change is described by actions and Lagrangians.

Even though the principle of least action is not an explanation, it calls for one. We need some patience, though. *Why* nature follows the principle of least action and *how* it realizes it will become clear in the part on quantum theory.



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

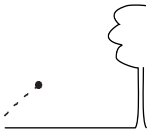
In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net



Why is motion so often bound?

> The optimist thinks this is the best of all possible worlds,
> and the pessimist knows it.
> Robert Oppenheimer

Looking around oneself on earth and in the sky, we find 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 listed in Figure 63 and Table 19. Obviously, the stronger the interaction, the smaller the aggregate. But why is matter mainly found in lumps at all?

Ref. 123

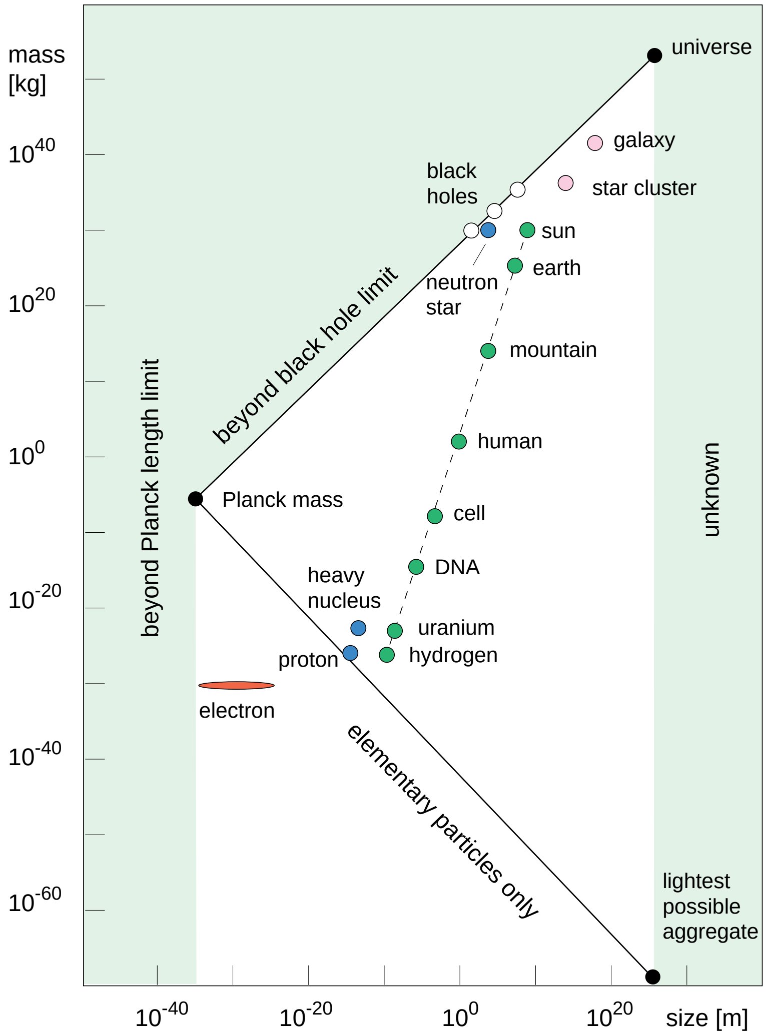


Figure 63 Aggregates in nature

First of all, aggregates form because of the existence of *attractive* interactions between objects. Secondly, they stay together because of *friction*, that is to say because the energy released when the objects approach can be changed into heat, which prevents the objects from leaving again. Thirdly, aggregates exist because of *repulsive* effects which prevent the components from collapsing completely. Together, these three characteristics ensure that bound motion is much more frequent than unbound, ‘free’ motion.

Only three types of attractions lead to aggregates: gravity, the electric interaction, and the strong nuclear interaction. Similarly, only three types of repulsive effects are observed: rotation, pressure, and the Pauli exclusion principle (which we will encounter later on). Of the nine combinations, only some appear in nature. Can you find out which ones are missing from Figure 63 and Table 19, and why?

See page 573

Challenge 275

Together, attraction and friction imply that change and action are minimized when objects come and stay together. The principle of least action thus implies the stability of aggregates. By the way, the same arguments also explain why so many aggregates *rotate*. Can you provide the connection?

Challenge 276

But why does friction exist at all? And why do attractive and repulsive interactions exist? In addition, the above answers assume that in some distant past matter was *not* found in lumps. Is this correct? In order to find out, we must first study another global property of motion.

Table 19 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 and helium atoms, galaxy super clusters
quasar		$20 \cdot 10^6$	baryons and leptons
galaxy supercluster	ca. 3 Ym	10^6	galaxy groups
galaxy group/cluster	ca. 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	1 to 70	sun, planets, moons, planetoids, comets, asteroids, dust, gas
Oort cloud	6 to 30 Pm	1	comets, dust
Kuiper belt	60 Tm	1	planetoids, 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; in particular, heavy atoms
planetoids (Varuna, etc)	50- 1000km	ca. 10	(estimates go up to 10^9) solids
moons	10 - 1000km	ca. 50	solids
electromagnetically bound aggregates[c]			
asteroids, mountains[d]	1 m to 930 km	$>$26 000	(10^9 estimated) solids, usually monolithic
comets	10 cm to 50 km	$10^{\cdots}$	ice and dust

Interaction aggregate	size (diameter)	observed number	constituents
planetoids, solids, liquids, gases, cheese	1 nm to > 100 km	n.a.	molecules, atoms
animals, plants, kefir	5 μm to 1 km	$10^{\cdots}$	organs, cells
brain	0.15 m	10^{10}	neurons and other cell types
cells		$10^{28\pm2}$	organelles, membranes, molecules
smallest: (...)	ca. 5 μm		molecules
amoeba	600 μm		molecules
largest: (whale nerve, single cell plants)	ca. 30 m		molecules
molecules		ca. $10^{78\pm2}$	atoms
H_2	ca. 50 pm		atoms
DNA (human)	2 m (total)		atoms
atoms, ions	30 pm to 300 pm	$10^{80\pm2}$	electrons and nuclei
aggregates bound by the weak interaction [c]			
none			
aggregates bound by the strong interaction [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 was the first evidence found for objects circling other stars than our sun; most of the over 70 *extrasolar planets* found so far were found around F, G, and K stars, including neutron stars. For example, three objects circle the pulsar PSR1257+12 and a matter ring circles the Ref. 124 star β Pictoris. The objects seem to be dark stars, brown dwarfs or large gas planets like Jupiter. None of the systems found so far form solar systems of the type we live in.

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%. These are from the rare blue O class or 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% are of the yellow G class, like Alpha Centauri, Capella or the sun. Exceptions are the few K giants, such as Arcturus and Aldebaran, and See page 323 the M supergiants, such as Betelgeuse and Antares. More on stars later on..

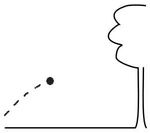
c. For more details on *microscopic* aggregates, see the table of composites in Appendix C.

d. It is estimated that there are about 10^9 asteroids (or planetoids) larger than 1 km and about 10^{20} Ref. 125 heavier than 100 kg.

Curiosities and challenges about Lagrangians

Lagrangians are a fascinating topic.

- When Lagrange published his book *Mécanique analytique*, in 1788, it formed one of the high points in the history of mechanics. He was very proud of having 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 come into general use.



- Given that action is the basic quantity describing motion, we can define energy as action per time, and momentum as action per distance. The *energy* of a system thus describes how much it changes over time, and the *momentum* how much it changes over distance. What are angular momentum and rotational energy? Challenge 277
- Effects by telekinesis or praying are impossible, as in most cases the change inside the brain is usually much smaller than the changes claimed in the outside world. Is the argument correct? Challenge 278
- In Galilean physics, the Lagrangian is given by the difference between kinetic and potential energy. Later on, this definition will be generalized in a way that sharpens the understanding of this distinction: the Lagrangian becomes the difference between a term for free particles and a term due to their interactions. In other words, particle motion is a continuous compromise between what the particle would do if it were free and what other particles want it to do. In this, particles behave a lot like humans beings.
- In nature, the sum $T+U$ of kinetic and potential energy is *constant* during motion (for closed systems), whereas the average of the difference $T-U$ is *minimal*. Is it possible to deduce, by combining the two expressions, that systems tend to a minimum potential energy? Challenge 279
- There is a principle of *least effort* describing the growth of trees. When a tree grows, almost 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. This is the reason why not all leaves are at the very top of a tree.
- 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 Ref. 121 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 vessels: a few large ones, and increasingly more the smaller they are. Second, the vessels all have a universal minimum size. And third, the networks are optimized in order to minimize the energy needed for transport. Together, these relations explain many additional scaling rules; they might also explain why animal life span scales as $m^{-1/4}$, or that most mammals have roughly the same number of heart beats.

A competing explanation, using a different minimum principle, states that quarter powers arise in any network built so that the flow arrives to the destination by the most direct path. Ref. 122

- The minimum principle for the motion of light is even more beautiful; light always takes the path which requires the shortest travel time. It was found already long ago that this idea describes exactly how light changes direction when it moves from air to water. In water, light moves more slowly; the speed ratio between air and water is called the *refractive index* of water. The refractive index, usually abbreviated n, is material-dependent. The value for water is about 1.3. This speed ratio, together with the minimum time principle, leads to the 'law' of refraction, a simple relation between the sine of the two angles. Can you deduce it? Challenge 280 n In fact, the exact definition of the refractive index is with respect to vacuum, not to air. But the difference is negligible; can you imagine why? Challenge 281 n

For diamond, the refractive index is 2.4. The high value is one reason for the sparkle of diamonds with the so-called *brilliant* cut. Can you specify some additional reasons? Challenge 282 n

- Are you able to confirm that each of these minimum prin-
Challenge 283 ciples is a special case of the principle of least action?In fact this is true for *all* known minimum principles in nature. Each of them, like the principle of least action, is a principle of least change.

- In Galilean physics, the value of the action depends on the observer. It is the same for observers with different orientations and positions, but not the same for observers with dif-
Challenge 284 ferent speeds. What does special relativity require?How will the action look in that case?

- Measuring all change going on in the universe presup-
Challenge 285 n poses that the universe is a system. Is that correct?

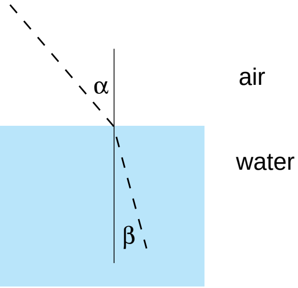
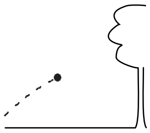


Figure 64 Refraction of light is due to travel time optimization

Motion and symmetry

Figure 65 Forget-me-not, also called *myosotis* (Barraginaceae)

The second way to describe motion globally is to describe it in such a way that all observers *agree*. An object under observation is called *symmetric* if it looks the same when seen from different points of view. For example, the forget-me-not of Figure 65 is symmetrical because it looks the same after turning around it by 72 degrees; many fruit tree flowers 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 a *low* symmetry. For example, a four-leaf clover has a higher symmetry than a usual, three-leaf one. Different points of view imply different observers; in physics, the viewpoints are often called *frames of reference* and are described mathematically by coordinate systems.

High symmetry means many agreeing observers. 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, we will deduce that nature as a whole is symmetric from the simple fact that we have the ability to talk about it!Moreover, the symmetry Challenge 286 of nature is considerably higher than that of a forget-me-not. This large symmetry is at the basis of the famous expression $E_o = mc^2$, as we will see.

Why can we think and talk?

> The hidden harmony is stronger than the apparent.
> Heraclitos of Ephesos, about 500 B.C. Ref. 15

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 already had similar experiences *beforehand*.

‘Similar’ means that what *we* observe and what *others* observe somehow correspond. Many aspects of observations do not depend 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 the highest possible symmetry are called *scalars* in physics. Other aspects change from observer to observer, such as apparent size variations with distance. However, the actual size is observer-independent. In general terms, any type of *viewpoint independence* is a form of symmetry, and the observation 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 will be explored in this section and during most of the rest of our hike.

In the world around us, we note another general property: not only does the same phenomenon look similar to different observers, but *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. Nature shows *reproducibility*. Nature shows no surprises. In fact, our memory and our thinking are only possible because of this basic property of nature. (Can you confirm this?)As we will see, Challenge 287 reproducibility leads to additional strong restrictions on the description of nature

Without viewpoint independence and reproducibility, talking to others or to oneself would be impossible. Even more importantly, we will discover that viewpoint independence and reproducibility do more than determining 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 a young human starts to meet other people in childhood, it 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. In our adventure, we concentrate on a section of the first type of experiences, *physical* observations. However, even among these, distinctions are to be made. In daily life we are used to assuming that weights, volumes, lengths, and time intervals are independent of the viewpoint of the observer. We can talk about these observed quantities to anybody, and there are no disagreements over 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 its value.

In the case of viewpoint dependent observations, understanding is still possible with help of little effort: each observer can *imagine* observing 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 due only to the different viewpoints; otherwise, the difference is fundamental, and they cannot talk to each other. Using this approach, you might even argue whether human feelings, judgments or taste lead to fundamental differences. Challenge 288

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, we must answer the following questions in order to find a *complete* description of the state of physical systems:

- Which viewpoints are possible?
- How are descriptions transformed from one viewpoint to another?
- Which observables do these symmetries admit?
- What do these results tell us about motion?

In the discussion so far, we have studied viewpoints differing in location, in orientation, in time and, most importantly, in motion. With respect to each other, observers can be at rest, move with constant speed, or accelerate. These ‘concrete’ changes of viewpoint are those 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 with this type of invariance are also called *external* symmetries. They are listed in Table 21. See page 154

---

* Humans develop the ability to imagine that others can be in situations *different* from their own at the age of about four years. Therefore, before the age of four, humans are unable to understand special relativity; afterwards, they can. Ref. 127

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, it is required that all statements be consistent across different mathematical descriptions. This requirement of consistency is called the *principle of gauge invariance*. The associated symmetries are 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.

The third principle, whose importance is also not evident from everyday life, is the behaviour of a system under exchange of its parts. The associated invariance is called *permutation symmetry*. It is a *discrete* symmetry, and we will encounter it in the second part of our adventure.

Later on we will discover that looking for a complete description of the state of objects will also yield a complete description of their *intrinsic* properties. But enough of introduction; let us come to the heart of the topic.

Symmetries and groups

Since we are looking for a complete description of motion, we need to understand the symmetries of nature. 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 that: the concatenation of two elements, namely 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* or *multiplication* and pronounced ‘after’ or ‘times’, in which the following properties hold for all elements a, b, c :

$$\begin{aligned} \textit{associativity}, \text{ i.e.} \quad & (a \circ b) \circ c = a \circ (b \circ c) \\ \text{a } \textit{neutral element } e \text{ exists such that} \quad & e \circ a = a \circ e = a \\ \text{an } \textit{inverse element } a^{-1} \text{ exists such that} \quad & a^{-1} \circ a = a \circ a^{-1} = e \quad . \end{aligned} \tag{59}$$

Any set which fulfils 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 defined in a completely abstract manner. Can you give an example of a group taken from daily life? Groups appear frequently in physics and mathematics, because symmetries are almost everywhere, as we will see.[*] Are you able to list the symmetry operations

Challenge 289 n

---

[*] 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 concatenation/multiplication is commutative, i.e. if $a \circ b = b \circ a$ for all couples of elements. In this case the multiplication is sometimes called *addition*. A subset $G_1 \subset G$ 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.

Challenge 290 ofFigure 66?*

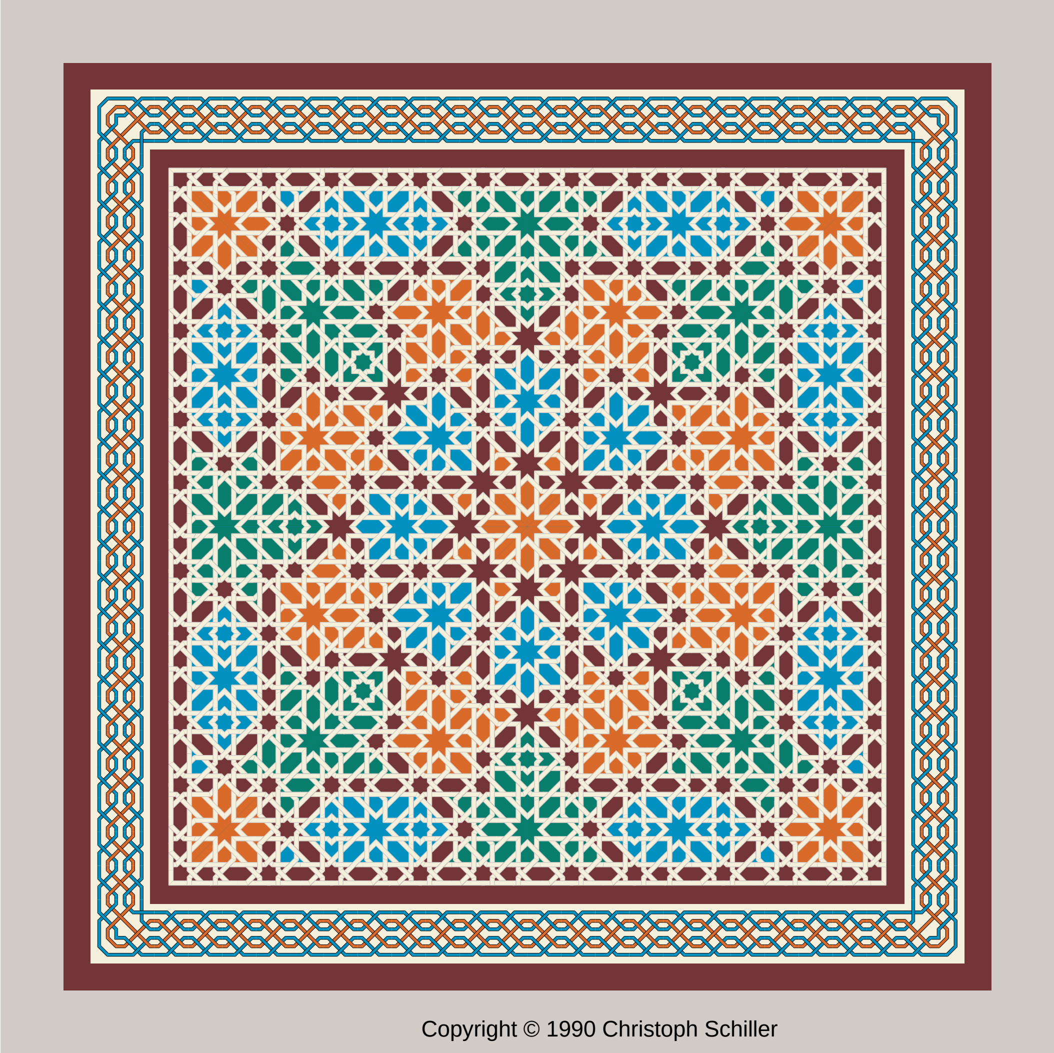


Figure 66 A Hispano-Arabic ornament from the Governor's Palace in Sevilla

Representations

Observing a symmetric and composed system such as the one shown in Figure 66, we noticethat each of its parts, for example each red patch, belongs to a set of similar objects, usually called a *multiplet*. Taken as a whole, the multiplet has (at least) the symmetry properties of the whole system. For some coloured patches we need four objects to make up

Challenge 291 e

---

* 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.

a full multiplet, whereas for others we need two or only one, such as in the case of the central star. In fact, in any symmetric system each part can be classified by saying to what type of multiplet it belongs. Throughout our mountain ascent we will perform the same classification with every part of nature, with ever-increasing precision.

A *multiplet* is a set of parts which transform into each other under all symmetry transformations. Mathematicians often call abstract multiplets *representations*. By specifying to which multiplet a component belongs, we describe in which way the component is part of the whole system. Let us see how this classification is achieved.

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 292 e

$$D(\mathrm{refl}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \tag{60}$$

since every point (x,y) becomes transformed to (y,x) when multiplied by the matrix $D(\mathrm{refl})$.Therefore, for a mathematician a *representation* of a symmetry group G is an assignment of a matrix $D(a)$ to each group element a in such a way that the representation of the concatenation of two elements a and b is nothing else than the product of the representations of the elements Challenge 293 e

$$D(a \circ b) = D(a)D(b) \quad . \tag{61}$$

For example, the matrix of equation (60), together with the corresponding matrices for all the other symmetry operations, have this property.*

For every symmetry group, the construction and classification of all possible representations is an important task. It corresponds to the classification of all possible multiplets a symmetric system can be made of. In this way, understanding the classification of all multiplets and parts which can appear in Figure 66 will teach us how to classify all possible parts of which an object or an example of motion can be composed of!

A representation is called *unitary* if all matrices D are unitary.** Almost all representations appearing in physics, with only a handful of exceptions, are unitary: this term is the

---

* There some obvious, but important side conditions for a representation: the matrices $D(a)$ must be invertible, or non-singular, and the identity operation of G must be mapped to the unit matrix. In even more compact language one says that a representation is a *homomorphism* from G into the group of non-singular or invertible matrices. A matrix D is invertible if its determinant $\det D$ is not zero.

In general, if there exists a mapping f from a group G to another G' such that

$$f(a \circ_G b) = f(a) \circ_{G'} f(b) \quad , \tag{62}$$

the two groups are called *homomorphic*, and the mapping f an *homomorphism*. A mapping which is also one-to-one is called a *isomorphism*.

A representation is called *faithful, true* or *proper* if it is also an isomorphism.

In the same way as groups, more complex mathematical structures such as rings, fields and associative algebras may also 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 $(A^T)_{\mathrm{ik}} = A_{\mathrm{ki}}$. The *complex conjugate* A^* of a matrix A is defined by $(A^*)_{\mathrm{ik}} = (A_{\mathrm{ik}})^*$. 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 $A^T = A^{-1}$, *hermitean* or *self-adjoint* (the two are synonymous in all physical applications) if $A^\dagger = A$ (hermitean matrices have real eigenvalues), and *unitary* if $A^\dagger = A^{-1}$. Unitary most restrictive, since it specifies that the corresponding transformations are one-to-one and invertible, which means that one observer never sees more or less than another. Obviously, if an observer can talk to a second one, the second one can also talk to the first.

The final important property of a multiplet or representation concerns is structure. If it can be seen as composed of sub-multiplets, it is called *reducible*, else *irreducible*; the same for representations. The irreducible representations obviously cannot be decomposed any further. For example, the symmetry group of Figure 66, commonly called D_4, has eight elements. It the general, faithful, unitary and irreducible matrixrepresentation Challenge 294 e

$$\begin{pmatrix} \cos n\pi/2 & -\sin n\pi/2 \\ \sin n\pi/2 & \cos n\pi/2 \end{pmatrix} \text{ for } n = 0..3, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}. \tag{63}$$

The representation is an *octet*. The complete list of possible irreducible representations of the group D_4 is given by *singlets*, *doublets*, and *quartets*. Are you able to find them all? Challenge 295 These representations allow the classification of all white and black ribbons that appear in the figure, as well as all coloured patches. The most symmetric elements are singlets, the least symmetric ones are members of the quartets. The complete system is always a singlet as well. Challenge 296 e

With these concepts we are ready to talk about motion with improved precision.

Table 20 Correspondences between the symmetries of an ornament, a flower and nature as a whole

Concept/system	Hispano-arabic pattern	Flower	Motion
Structure and components	set of ribbons and patches	set of petals, stem	motion path and observables
System symmetry	pattern symmetry	flower symmetry	symmetry of Lagrangian
Mathematical description of the symmetry group	D4	C5	in Galilean relativity: position, orientation, instant, and velocity changes
Invariants	number of multiplet elements	petal number	number of coordinates, magnitude of scalars, vectors and tensors
Representations of the components	multiplet types	multiplet types	tensors, including scalars and vectors
Most symmetric representation	singlet	part with circular symmetry	scalar
Simplest faithful representation	quartet	quintet	vector

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-hermitean* by $A^\dagger = -A$ and an *anti-unitary* 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 $\det A = 0$.

Concept/system	Hispano-arabic pattern	Flower	Motion
Least symmetric representation	quartet	quintet	no limit (tensor of infinite rank)

Symmetries, motion and Galilean physics

Every day we experience that we are able to talk to each other about motion. It must therefore be possible to find an *invariant* quantity describing it. We already know it: it is the *action*. Lighting a match is a change. It is the same whether it is lit here or there, in one direction or another, today or tomorrow. Indeed, the (Galilean) action is a number whose value is the same for each observer *at rest*, independent of his orientation or the instant he makes his observation.

In the case of the arabic pattern of Figure 66, the symmetry allowed to deduce the list of multiplets, or representations, that can be its building blocks. This approach must be possible for motion as well. We deduced the classification of the ribbons in the arabic pattern into singlets, doublets, etc. from the various possible observation viewpoints. For a moving system, the building blocks, corresponding to the ribbons, are the *observables*. Since we observe that nature is symmetric under many different changes of viewpoints, we can classify all possible observables. To do so, we need to take the list of all viewpoint transformations and deduce the list of all their representations.

Our everyday life shows that the world stays unchanged after changes in position, orientation and instant of observation. One also speaks of space translation invariance, rotation invariance and time translation invariance. These transformations are different from those of the arabic pattern in two respects: they are *continuous*, and they are *unbounded*. As a result, their representations will generally be concepts which can vary continuously and without bounds: they will be *quantities* or *magnitudes*. In other words, observables will be constructed with *numbers*. In this way we have deduced why numbers are *necessary* for any description of motion.[*]

Since observers can differ in orientation, most representations will be objects possessing a direction. To make a long story short, the change of observation position, orientation or instant leads to the result that all observables are either 'scalars', 'vectors' or higher-order 'tensors.'[**]

A *scalar* is an observable quantity which stays the same for all observers: it corresponds to a singlet. 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, and without direction. Other examples of scalars are the potential at a point and the temperature at a point. Velocity is obviously not a scalar, nor is the coordinate of a point. Can you find more examples and counterexamples? Challenge 297

---

[*] Only scalars, in contrast to vectors and higher order tensors, may also be quantities which only take a discrete set of values, such as $+1$ or -1 only. In short only scalars may be *discrete* observables.

[**] Later on, *spinors* will be added to this list, which will then be complete.

Energy is an interesting observable. It is a scalar if only changes of place, orientation, and instant of observation are considered. Energy is not a scalar if changes of observer speed are included. Nobody tried to find a generalization of energy which is a scalar also for moving observers, until Albert Einstein discovered it. More about this issue shortly.

Any quantity which has a magnitude, a direction, and which 'stays the same' with respect to the environment when changing viewpoint is a *vector*. For example, the arrow between two fixed points on the floor is a vector. Its length is the same for all observers; its direction changes from observer to observer, but not with respect to its environment. On the other hand, the arrow between a tree and the place where a rainbow touches the earth is *not* a vector, since that place does not stay fixed with respect to the environment, when the observer changes.

Mathematicians say that vectors are directed entities staying invariant under coordinate transformations. Velocities of objects, accelerations, and field strength are examples of vectors. (Can you confirm this?)The magnitude of a vector is a scalar; it is the same for any observer. By the way, a famous and baffling result of 19th-century experiments is that the velocity of light is *not* a vector for Galilean transformations. This mystery will be solved shortly.

Challenge 298 e

Tensors are generalized vectors. As an example, take the moment of inertia of an object. It specifies the dependence of the angular momentum on the angular velocity. For any object, doubling the magnitude of angular velocity doubles the magnitude of angular momentum; however, the two vectors are not parallel to each other if the object is not a sphere. If any two vector quantities are proportional like the two in the example, in the sense that doubling the magnitude of one vector doubles the magnitude of the other, but without the two vectors being parallel to each other, then the proportionality factor is a tensor. (Second order) *tensors* are thus quantities with a magnitude, a direction, and a *shape*.[*] Can you name another example?

See page 70

See page 85

Challenge 300

Let us get back to the description of motion. Table 20 shows that in physical systems we always have to distinguish between the symmetry of the whole Lagrangian – corresponding to the symmetry of the complete pattern – and the representation of the observables – corresponding (sloppily) to the symmetry of the ribbons. Since the action must be a scalar and since all observables must be tensors, Lagrangians contain sums and products of tensors only in combinations forming scalars. Lagrangians contain only scalar products or its

---

[*] Tensors (of rank 2) have magnitudes, directions, and connections between directions. 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. Tensors describe simple distributions in space. If vectors can be visualised as oriented arrows, tensors can be visualized as oriented ellipsoids.

A vector is described mathematically 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.

A nth-order tensor is the proportionality factor between a first order tensor, i.e. between a vector, and an $(n-2)$nd-order tensor. Tensors of higher orders correspond to more and more complex shapes. Vectors and scalars are first and zeroth order tensors. The order, by the way, also gives the number of indices an observable has. Can you showthis?

Challenge 299

generalizations. In short, Lagrangians always look like

$$L = \alpha a_{\mathrm{i}} b^{\mathrm{i}} + \beta c_{\mathrm{jk}} d^{\mathrm{jk}} + \gamma e_{\mathrm{lmn}} f^{\mathrm{lmn}} + \ldots \tag{64}$$

where the indices always come in matching pairs to be summed over. Therefore summation signs are usually simply left out.) The Greek letters represent constants. For example, the action of a free point particle in Galilean physics was given as

$$S = \int L\, dt = \frac{m}{2} \int v^2\, dt \tag{65}$$

which is indeed of the form just mentioned. We will encounter many other cases during our study of motion.*

Galileo already understood that motion is also invariant under change of viewpoints with different velocity. However, the action just given does not reflect this. It took another 250 years to find out the correct generalization: the theory of special relativity. Before studying it, we need to finish the present topic.

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, the symmetry under change of instant of time or time translation invariance, is also a case of viewpoint independence. (Can you find its irreducible representations?)The connection has several important consequences. Challenge 302 We have seen that symmetry implies invariance. It turns out that for *continuous* symmetries, such as time translation symmetry, this sentence can be made more precise: for

---

* 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 (65)? Surprisingly, Ref. 128 the answer is no. One of the first who 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 use as coordinates $(t, \mathbf{x})$ and (τ, ξ), the action is invariant under the Challenge 301 transformations

$$\xi = \frac{\mathbf{Rx} + \mathbf{x}_{\mathrm{o}} + \mathbf{v}t}{\gamma t + \delta} \quad \text{and} \quad \tau = \frac{\alpha t + \beta}{\gamma t + \delta} \quad \text{with} \quad \mathbf{R}^T \mathbf{R} = 1 \quad \text{and} \quad \alpha\delta - \beta\gamma = 1 \quad . \tag{66}$$

where $\mathbf{R}$ describes the rotation from the orientation of one observer to the other, $\mathbf{v}$ the velocity between the two observers, and $\mathbf{x}_{\mathrm{o}}$ the vector between the two origins at time zero.

The important special cases of these transformations are

$$\begin{aligned} &\text{The connected, static Galilei group} \quad \xi = \mathbf{Rx} + \mathbf{x}_{\mathrm{o}} + \mathbf{v}t \quad \text{and} \quad \tau = t \\ &\text{The transformation group SL(2,R)} \quad \xi = \frac{\mathbf{x}}{\gamma t + \delta} \quad \text{and} \quad \tau = \frac{\alpha t + \beta}{\gamma t + \delta} \end{aligned} \tag{67}$$

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 objects and systems. They will return to be of importance later on.

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.* She found the result in 1915 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. 129

Noether investigated continuous symmetries depending on a continuous parameter b. A viewpoint transformation is a symmetry if the action S does not depend on the value of b. For example, changing position as

$$x \mapsto x+b \tag{68}$$

leaves the action

$$S_o = \int T(v) - U(x) dt \tag{69}$$

invariant, as $S(b) = S_o$. This situation implies that

$$\frac{\partial T}{\partial v} = p = const \quad ; \tag{70}$$

in short, symmetry under change of position implies conservation of momentum. The converse is also true.

Challenge 303 In case of symmetry under shift of observation instant, we find

$$T+U = const \quad ; \tag{71}$$

time translation invariance implies constant energy. Again, the converse is also correct. One also says that energy and momentum are the *generators* of time and space translations.

The conserved quantity for a continuous symmetry is sometimes called the *Noether charge*, because the term *charge* is used in theoretical physics to designate conserved extensive observables. In other words, energy and momentum are Noether charges. 'Electric charge', 'gravitational charge' (i.e. mass), and 'topological charge' are other common examples. What is the conserved charge for rotation invariance? Challenge 304

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 total energy of any number of independent particles is constant. Finally, it means that the energy of a *system* of particles, i.e. including their interactions, is constant in time. Collisions are examples of the latter case. Noether's theorem makes all of these points at the same time, as you can verify using the corresponding Lagrangians. Challenge 305 e

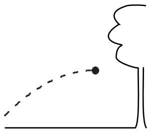
But Noether's theorem also makes, or better repeats, an even stronger statement: if energy were not conserved, time could not be defined. The whole description of nature requires the existence of conserved quantities, as we noticed when we introduced the concepts of object, state, and environment. For example, we defined objects as *permanent* entities, that is, as entities characterized by conserved quantities. We also saw that the introduction of time is possible only because in nature there are no surprises. Noether's theorem describes exactly

See page 33

See page 114

---

* Emmy Noether (Erlangen, 1882–Bryn Mayr, 1935), German mathematician. 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.



what such a surprise would have to be: the non-conservation of energy. It has never been observed.*

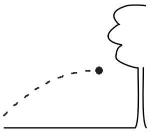
Since symmetries are so important for the description of nature, Table 21 gives an overview of all the symmetries of nature we will encounter. Their main properties are also listed. Except for those marked as 'approximate' or 'speculative', an experimental proof of incorrectness of any of them would be a big surprise indeed.

Table 21 The symmetries of relativity and quantum theory with their properties; at the same time, the complete list of logical *inductions* used in the two fields

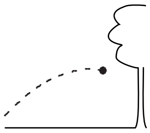
Symmetry	type [param. number]	space of action	group topology	possible representations	conserved quantity	vacuum/ matter is symmetric	main effect
Geometric or space-time, external, symmetries							
Time and space translation	$R \times R^3$ [4 par.]	space, time	not compact	scalars, vectors,	momentum and energy	yes/yes	allow everyday
Rotation	SO(3) [3 par.]	space	S^2	tensors	angular momentum	yes/yes	communica-tion
Galilei boost	R^3 [3 par.]	space, time	not compact	same	centre 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	
Dynamic, interaction-dependent symmetries: gravity							
$1/r^2$ gravity	SO(4) [6 par.]	config. space			perihelion direction	yes/yes	

---

* Quantum theory adds some details here, as we will find out when studying it.



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
Dynamic, classical and quantum mechanical motion symmetries							
Motion(‘time’) inversion T		Hilbert or phase space	discrete	even, odd	T-parity	yes/no	
Parity(‘spatial’) inversion P		Hilbert or phase space	discrete	even, odd	P-parity	yes/no	
Charge conjugation C	global, antilinear, anti-hermitean	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]
Dynamic, interaction-dependent, gauge symmetries							
Electromagnetic classical gauge inv.	[∞ par]	...	...	...	...	yes/yes	...
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							



Symmetry	type [param. number]	space of action	group topology	possible representations	conserved quantity	vacuum/ matter is symmetric	main effect
Particle exchange	discrete	Fock space and simil.	discrete	fermions and bosons	none	n.a./yes	Gibbs' paradox
Selected speculative symmetries of nature							
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 particle masses
Event symmetry		space-time				yes/no	

For details about the connection between symmetry and induction, see page 486. The explanation of the terms in the table will be completed in the rest of the walk.

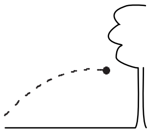
[a] Only approximate; 'massless' means that $m \ll m_{\text{Pl}}$, i.e. that $m \ll 22\,\mu\text{g}$.

[b] $N = 1$ supersymmetry, but not $N = 1$ supergravity, is probably a good approximation for nature at everyday energies.

[c] $i = 1..N$.

In summary, since 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 energy or linear and angular momentum. 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. To put it simply, our ability to talk about nature means that energy and momentum are conserved.

In general, 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, and Paul Dirac started off quantum electrodynamics. We will use the same method throughout our walk; in its third part we will uncover some symmetries which are even more mind-boggling than those of relativity. For the time being, we continue with the next method allowing a global description of motion.



Observation	frequency
Lowest vibration frequency of the earth Ref. 130	309 μHz
Wing beat of tiny fly	ca. 1000 Hz
Sound audible to young humans	20 Hz to 20 kHz
Sonar used by bats	over 100 kHz
Sonar used by dolphins	up to 120 kHz
Sound frequency used in ultrasound imaging	up to 15 MHz

Table 22 Some frequency values found in nature

Curiosities and challenges about motion symmetry

- What is the path followed by four turtles starting on the four angles of a square, if each of them continuously walks towards the next one? Challenge 306
- Challenge 307 What is the symmetry of a simple oscillation? And of a wave?
- Challenge 308 For what systems is motion reversal a symmetry transformation?

Simple motions of extended bodies – Oscillations and waves

We defined action as the integral of the Lagrangian, and the Lagrangian as the difference between kinetic and potential energy. For example, the Lagrangian of a mass attached to a spring is given by

$$L = mv^2/2 - kx \quad . \qquad (72)$$

Challenge 309 e Can you confirm it?The Lagrangian has a beautiful property: it describes the oscillation of the spring length. The motion is exactly the same as that of a long pendulum. It is called *harmonic motion*, because an object vibrating fast enough in this way produces a completely pure musical sound. (The musical instrument producing the purest harmonic waves is the transversal flute. It thus gives the best idea about how harmonic motion ‘sounds’.) The graph of a harmonic oscillation is called a *sine curve*; it can be seen as the basic building block of all oscillations. All other, non-harmonic oscillations in nature can be composed from it.

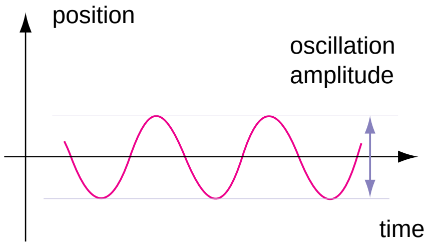
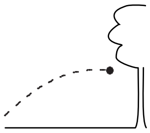


Figure 67 The simplest oscillation

Many oscillations diminish in time: they are damped. Systems with par *small* damping are useful for making precise clocks. Damping is measured by stating how many oscillations a system takes to reduce its amplitude to $1/e \approx 1/2.718$ times the original amount. This characteristic number is the so-called *Q-factor*. (Can you write down a simple Lagrangian for a damped oscillation?)In nature, damped oscillations do not Challenge 310



usually keep constant frequency; however, for the simple pendulum this remains the case to high accuracy. The reason is that for a pendulum, the frequency does not depend on the amplitude (as long as it is smaller than about 20 °). This is one reason that pendula appear as oscillators in mechanical clocks.

Obviously, for a good clock, the driving oscillation must not only show small damping, but also be temperature independent and insensitive to other external influences. An important development of the twentieth century was the introduction of quartz, a crystal of the size of a grain of sand that can made to oscillate with small temperature independence, a large Q-factor, and thus with very small energy, so that precise clocks can now run on small batteries.

All systems which oscillate also emit waves.

Waves

Waves are travelling oscillations. All waves are characterized by an oscillation frequency and by a propagation velocity. Water waves show this most clearly.

Waves appear in all *extended* bodies, be they solids, liquids, or gases. In fluid bodies, waves are *longitudinal*, meaning that the wave motion is in the same direction as the wave oscillation. Sound in air is an example of a longitudinal wave. In solid bodies waves can also be *transversal*; in that case the wave oscillation is perpendicular to the travelling direction.

Waves appear also on *interfaces* between bodies; water-air interfaces show waves. Even saltwater-freshwater interfaces, so-called dead water, show waves, and any aeroplane trip allows to study the regular cloud arrangements on the interface between warm and cold air layers in the atmosphere. Seismic waves travelling along the boundary between the sea floor and the sea water also exist. Surface waves usually are neither longitudinal nor transversal, but of a mixed type. Their speed often depends on the wavelength; for example, the speed of sea waves increases with the square root of the wavelength.

Waves can also exist in empty space. Both light and gravity waves are examples. The theory of electromagnetism and that of relativity will tell us more about their properties.

Any study of motion must also include the exploration of wave motion. Waves show six main effects that set them apart from the motion of bodies.

- Waves go around corners. This observation is called diffraction.
- Waves change direction when they change medium. This effect is called refraction.
- Waves can have a frequency dependent propagation speed. This effect, already mentioned above, is called dispersion.
- Waves can add up or cancel each other out; this effect is called interference. Interference and superposition is strongly related to the linearity of waves.
- Transverse waves can oscillate in different directions: they show polarization.
- Often, the wave amplitude decreases over time: waves show damping.

Material bodies do not behave this way when they move. The famous debate whether electrons or light are waves or particles thus requires to check whether these effects can be observed or not. This is the topic of quantum theory. Before we study it, can you give the conditions an observation needs Challenge 311 n to fulfil so that it surely cannot be a wave?

Figure yet to be included

Figure 68 Decomposing a general wave or signal into harmonic waves

As a result of having a frequency and a propagation velocity, all sine waves are characterized by the distance between two neighbouring wave crests; it is called the wavelength. All sine waves obey the basic relation

$$\lambda f = c \quad . \tag{73}$$

However, in many cases the wave velocity c depends on the wavelength of the wave. For example, this is the case for water waves. On the other hand, the speed of sound in air does not depend on the wavelength to a high accuracy.

Like a non-harmonic oscillation, also a non-harmonic wave can be decomposed into sine waves. Figure 68 gives examples. If the various sine waves contained in a disturbance propagate differently, the original wave will change in shape while it travels. That is why an echo sounds different from the original sound; for the same reason, a thunder sounds different for a nearby or a far away lightning.

All systems which oscillate also emit waves. Any radio or TV receiver contains oscillators. As a result, any such receiver is also a transmitter; indeed, in some countries people who listen to radio without permission are searched for by listening to the radio waves emitted by these devices. Also inside the human ear numerous tiny structures, the hair cells, oscillate. As a result, the ear must also emit sound. This prediction, made in 1948 by Tommy Gold, was confirmed only in 1979 by David Kemp. These so-called *otoacoustic emissions* can be detected with sensitive microphones; they are presently being studied in order to unravel the still unknown workings of the ear. Ref. 131

Figure to be included

Figure 69 The electrical signals measured in a nerve

Since any travelling disturbance can be decomposed into sine waves, term wave is used by physicists for all disturbances, whether they look like sine waves or not. The most important disturbances are the localized wave groups; an example, together with its decomposition in harmonic waves, is shown in Figure 68. Wave groups are extensively used as signals.

A famous example of signal is the equation found by Hodgkin and Huxley as realistic Ref. 132

approximation for the oscillations in nerve potential:

$$= a + b(V - V_{\mathrm{K}})n^4 + c(V - C)m^3 h + d(V - V_{\mathrm{L}}) \quad . \tag{74}$$

The voltage across a nerve is determined by the material constants V_{K}, V_{Na} for potassium and sodium, the leakage voltage V_{L} and the particle densities m and n. The voltage V then shows the characteristic spikes measured in nerves, a shown in Figure 69.

Signals

All signals are motion of energy. Signals can be either objects or waves. A thrown stone can be a signal, as a whistle can be. We see that waves are more practical because they do not require transport of matter. It is easier to use electricity in a telephone to transport a voice than to send a messenger. Indeed, most of modern technological advance can be traced to the separation between signal and matter transport. We do not need to transport an orchestra to enjoy music; sending radio signals is sufficient. Instead of sending paper letters we write e-mail messages. Instead of going to the library we browse the internet.

But the biggest progress in communication resulted from the use of signals to transport sizeable amounts of energy. That is what electric cables do: transporting energy without transporting any matter. We do not need to attach our kitchen machines to the power station: we can get the energy via a copper wire.

For all these reasons, the term ‘signal’ is often meant to imply waves. In practice, this is surely the case; voice, sound, electric signals, radio or light signals are the most common examples.

When signals are sent, their content can be lost. Each of the five characteristics of waves can lead to content degradation. Can you provide examples? Challenge 312

Signals connect causes and effects; studying signals is thus important in the theory of relativity, where people enjoy studying questions such as ‘Can I avoid the marriage of my great-great parents?’ We will come back to the issue.

Solitary waves

Figure to be inserted

Figure 70 A solitary wave

In August 1834, the English engineer John Scott Russel (1808–1882) recorded a strange observation in a water canal in his countryside near Edinburgh. When a boat pulled through the channel was suddenly stopped, a strange water wave departed from it. It consisted of a single crest, about 10 m long and 0.5 m high, moving with about 4 m/s. He followed that crest with his horse for several kilometres: the wave died out only very slowly. There also was no dispersion as usual in water waves. Russell then started producing such waves in his laboratory, and extensively studied their properties. He showed that the speed v, the Ref. 133 amplitude A and the width λ are related by

$$v = 2A - 4/\lambda^2 \quad . \tag{75}$$

Such stable waves with a single crest are today called solitary waves. They appear only in special cases, when the dispersion and the nonlinearity of the system exactly compensate each other. In some systems such solitary waves can cross each other unchanged, even when travelling in opposite directions; in that case (and only in that case) one speaks of *solitons*.

In 1895, Korteweg and de Vries found that the solitary waves had the shape

$$u(x,t) = A/\cosh^2[(x - vt)/\lambda] \quad . \tag{76}$$

and that the relation found by Russel was due to the wave equation

$$\frac{\partial u}{\partial t} = \frac{\partial^3 u}{\partial x^3} - 6u\frac{\partial u}{\partial x} \tag{77}$$

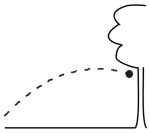
For many decades solitary waves were seen as mathematical or physical curiosities. Almost a hundred years later it became clear that this equation, now called *Korteweg-De Vries equation* is a universal model for weakly non-linear waves in the weak dispersion regime. In 1965, Kruskal and Zabusky also found that the solutions (76) actually are solitons. They interpenetrate each other without change of velocity or shape; a collision just gives a limited position shift.

Solitary waves also play a role in the flow of fluids, such as in ocean currents or as the red spot on Jupiter, and in light waves inside optical communication fibres, where they help to increase lossless signal transmission. Towards the end of the twentieth century a Ref. 133 second wave of interest into the mathematics of solitons arose, when quantum theorists got interested into them. A soliton is a middle thing between a particle and a wave; it has features of both concepts. For this reason, solitons are now an essential part of any description of elementary particles, as we will find out later on. See page 683

Curiosities and challenges about waves and extended bodies

One never knows enough about waves, oscillations and signals.

- What path is followed by a body moving in plane, but attached by a spring to a fixed point on the plane? Challenge 313 n
- Is a standing wave a wave? Challenge 314 n
- All waves are damped eventually. This effect is often frequency dependent. Can you provide a confirmation in the case of sound in air? Challenge 315 n
- Are water waves transversal or longitudinal? Challenge 316 n
- The speed of water waves limit the speeds of ships. A surface ship cannot travel (much) faster than about $v_{\text{crit}} = \sqrt{0.16gl}$, were $g = 9.8\,\text{m/s}^2$, l is its length, and 0.16 is a number



determined experimentally, called the critical Froude number. This relation is valid for large tankers – $l = 100\,\mathrm{m}$ gives $v_{\mathrm{crit}} = 13\,\mathrm{m/s}$ – down to ducks $l = 0.3\,\mathrm{m}$ gives $v_{\mathrm{crit}} = 0.7\,\mathrm{m/s}$. The critical speed is given by the wave speed of a wave with the same wavelength as the ship. In fact, moving at higher speeds than the critical value is possible, but takes a vastly higher amount of energy. Therefore all water animals and ships are faster when they swim under water – where the limit due to surface waves does not exist – than above water. Ducks can swim three times as fast under water than above water.

Challenge 317 n How much is the olympic swimming record away from the critical value?

▪ The group velocity of water waves (in deep water) is smaller than the velocity of the individual waves. As a result, when a group of wave crests travels, inside the group the crests move from back to front, appearing at the back, travelling forward, and then dying out at the front.

▪ On windy seas, the white wave crests have several important effects. The noise stems from tiny exploding and imploding water bubbles. The noise of waves on the open sea is thus the superposition of many small explosions. At the same time, white crests are the events where the seas absorb carbon dioxide from the atmosphere, and thus reduce global warming.

Challenge 318 n ▪ Why are there many small holes in the ceilings of many office buildings?

Ref. 1 ▪ Yakov Perelman lists the following problems in his delightful physics problem book. A stone falling into a lake produces circular waves. What is the shape of waves produced by a stone falling into a river? Challenge 319 n

▪ It is possible to build a lens for sound, in the same way as it is possible to build lenses for light. How would such a lens look like? Challenge 320

Challenge 321 ▪ What is the sound heard inside a shell?

▪ Light takes about eight minutes to arrive from the sun to the earth. What consequence does this have for a sunrise? Challenge 322 n

Challenge 323 ▪ Can you describe how a Rubik's cubes is built?And its generalizations to higher numbers of segments? Is there a limit to the number of segments? These puzzles are tougher than the search for a rearrangement. Similar puzzles can be found in many mechanisms, from robots to textile machines.

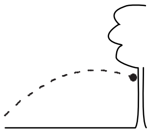
▪ Usual sound produces a pressure variation of 10^{-8} bar on the ear. How is this determined? Challenge 324

The ear is thus a sensitive device. It is now known that most cases in which sea mammals, like whales, swim onto the shore are due to ear problems; usually they became deaf after some military device (either sonar signals or explosions) destroyed their ear.

▪ *Infrasound*, inaudible sound below 20 Hz, is a modern topic of research. In nature, infrasound is emitted by earthquakes, volcanic eruptions, wind, thunder, waterfalls, falling meteorites, and the surf. Glacier motion, seaquakes, avalanches and geomagnetic storms also emit infrasound. Human sources are the start of missiles, traffic, fuel engines and air compressors. Ref. 134

It is known that high intensities of infrasound lead to vomit or disturbances of the sense of equilibrium (140 dB or more for 2 min), and even to death (170 dB for 10 min). The effects of lower intensities on human health are not known yet.

Infrasound can travel several times around the world before dying down, as the explosion of the Krakatoa volcano showed in 1883. With modern infrasound detectors, sea surf



can be detected hundreds of kilometres away. Infrasound detectors are even used to count meteorites at night. Very rarely, meteorites can also be heard.

- If you like engineering challenges, here is one that is still open. How can one make a robust and efficient system that transforms the energy of sea waves into electricity? Challenge325 h
- In our description of extended bodies, we assumed that each spot of a body can be followed throughout its motion. Is this assumption justified? What would happen if it were not? Challenge326 h

Do extended bodies exist?

We just studied the motion of extended bodies. Strangely enough, the question of their existence has been one of the most intensely discussed questions in physics. Over the centuries, it appeared again and again, at each improvement of the description of motion; the answer alternated between the affirmative and the negative. Many thinkers have been persecuted and many still are being persecuted by giving answers not politically correct! In fact, the issue already appears in everyday life.

Mountains, manifolds and fractals

Whenever we climb a mountain, we follow its outline. We usually describe this outline by a curved two-dimensional surface. In everyday life we find that this is a good approximation. But there are alternatives. The most popular is the idea that mountains are fractal surfaces. A *fractal* was defined by Benoit Mandelbrodt as a set which is self-similar under a countable but infinite number of magnification values. We encountered fractal lines earlier on. An example of an algorithm building a (random) fractal *surface* is shown on the left side of Figure 71. It produces shapes which look incredibly similar to real mountains. The results are so realistic that they are used in Hollywood movies. If this description would be correct, mountains would be extended, but not continuous. See page 47 Ref. 135

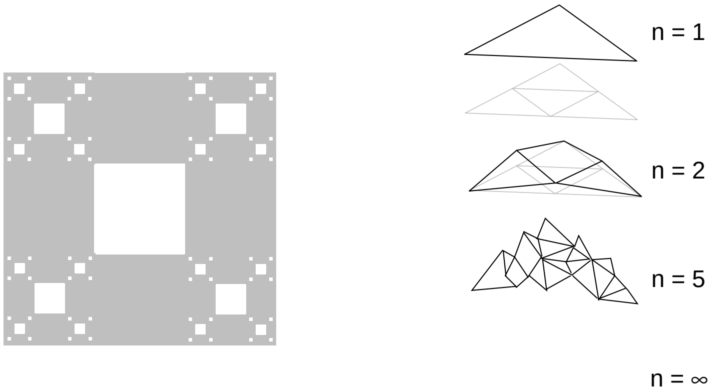


Figure 71 Floors and mountains as fractals

But mountains could also be fractals of a different sort, as shown in the right side of Figure 71. Floors could have an infinity of small and smaller holes. Can you devise an experiment to decide whether fractals or manifolds provide the correct description for mountain surfaces? Challenge 327

In fact, one could imagine that mountains have to be described by three dimensional versions of the right side of the figure. Mountains would then be some sort of mathematical swiss cheese. To settle the issue, a chocolate bar can help.

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 forever, namely by eating every day only half the remainder. However, this method only works if matter is scale-invariant. The method only works if matter is either *fractal*, as it then would be scale invariant for a discrete set of zoom factors, or *continuous*, in which case it would be scale invariant for any zoom factor. What case applies to nature?

See page 46 We have already encountered a fact making continuity a questionable assumption: continuity would allow us, as Banach and Tarski showed, to multiply food and any other matter by smart cutting and reassembling. Continuity would allow children to eat the *same* amount of chocolate every day, without ever buying a new bar. However, fractal chocolate is not ruled out in this way; an experiment can settle the question. Let us take fluid chocolate; or even simpler, let us take some oil – which is the main ingredient of chocolate anyway – and spread it out over an ever increasing surface. For example, one can spread a drop of oil onto a pond on a day without rain or wind; it is not difficult to observe which parts of the water are covered by the oil and which are not. Interestingly, a small droplet of oil cannot cover a surface larger than about – can you guess the value? Challenge 328 Trying to spread the film further inevitably rips it apart. The chocolate method thus does not work for ever; it comes to a sudden end. The oil experiment, first popularized by Kelvin, shows that there is a *minimum* thickness of oil films, with a value of about 2 nm.[*] This simple measurement can be conducted also in high school and shows that there is a smallest size in matter. The general conclusion is not a surprise, however. The existence of a smallest size – not its value – was already deduced by Galileo, when he studied some other, simple questions.[**]

---

[*] It is often claimed that Benjamin Franklin first conducted the experiment; that is wrong. Franklin did not measure the thickness and did not even ponder about minimal thickness; he poured oil on water, but missed the most important conclusion, which was taken only a century later by Kelvin. Also geniuses make mistakes.

[**] Galileo was brought to trial because of his ideas on atoms, not on the motion of the earth, as is often claimed. To get a clear view of the matters of dispute in the case of Galileo, especially of interest to 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. It is available also in 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 due to his views on the earth's motion was organized by his friend the pope to *protect* him from a sure condemnation to death about a different issue.

The reason for his arrest, as shown by the denunciation, 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 since matter is not

How high can animals jump?

Fleas can jump to heights a hundred times their size, humans only to heights about their own size. In fact, biological studies yield a simple observation: all animals, independently Ref. 136 of their size, achieve the same jumping height of $1.5 \pm 0.7\,\mathrm{m}$, whether they are humans, cats, grasshoppers, apes, horses, leopards, etc. Explaining this constancy takes only two lines. Are you able to do it? Challenge 329 n

The observation seems to be an example of scale invariance. But there are some interesting exceptions 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 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 to understanding the structure of matter. In fact, the materials of which we are made would not allow such change of scale, as the bones of giants would collapse under the weight they have to sustain. Bones have a finite strength because their constituents stick to each other with a finite attraction. Continuous matter could not break at all, and fractal matter would be infinitely fragile. Matter only breaks under finite loads because it is composed of smallest constituents.

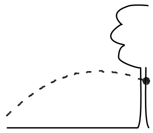
Felling trees

The lower, gentle slopes of Motion Mountain are covered by trees. Trees are fascinating structures. Take their size. Why do trees have limited size? Already in the 16th century, Galileo knew that increasing tree height is not possible without limits: at some point a tree would not have the strength to support its own weight. He estimated the maximum height to be around 90 m; the actual record, unknown to him at the time, is 152 m. But why does a limit exist at all? The answer is the same as for bones: wood has a finite strength because it is not scale invariant; and it is not scale invariant because it is made of small constituents, the atoms.

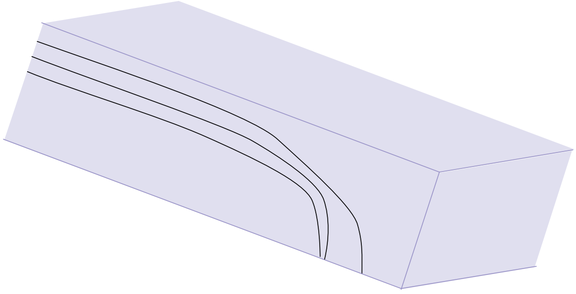
In fact, the origin for the precise value of the limit is more involved. Trees must not break under strong winds. Wind resistance limits the height-to-thickness ratio h/d to about 50 for standard-sized trees (for $0.2\,\mathrm{m} < d < 2\,\mathrm{m}$). Are you able to deduce this limit?Thinner trees Challenge 330

---

scale invariant, it is made of 'atoms' or, as he called them, 'piccolissimi quanti' – smallest quanta – which was and still is a heresy. A true catholic still is not allowed to believe in atoms. Indeed, atoms are not compatible with the change of bread and wine into human flesh and blood, called *transsubstantiation*, which is a central belief of the catholic faith. In Galileo's days, church tribunals punished heresy, i.e. deviating personal opinions, by the death sentence. Despite being condemned to prison in his trial, Galileo published his last book, written as an old man under house arrest, on the scaling issue. Today, the remainders of the catholic church continue to refuse to publish the proceedings and other documents of the trial. In addition, these remainders most carefully avoid the issue of atoms, as any catholic statement on the issue would start the biggest wave of humour on the planet. In fact, 'quantum' theory, after the term used by Galileo, has become the most precise description of nature ever.



are limited in height to less than 10 m by the requirement that they return to the vertical after being bent by the wind. Ref. 137



photograph to be included

Figure 72 Atomic steps in broken gallium arsenide crystals seen under a conventional light microscope

Such studies of natural constraints also answer the question of why trees are made from wood and not, for example, steel. You might want to check yourself that the maximum height of a column of a given mass is determined by the fraction E/ρ^2 between the elastic module and the square of the mass density. Challenge 331 n Wood is actually the material with the highest ratio. Ref. 138 Only recently a few selected engineering composites managed to achieve slightly better performance.

Why do materials break at all? All collected data yield the same answer and confirm Galileo's reasoning: because there is a smallest size in materials. 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; a crack could not start anywhere. If a body had a fractal swiss cheese structure, cracks would have places to start, but they would need only an infinitesimal effort for their propagation.

Experimental confirmation is not difficult. It is sufficient to break a thin single crystal, such as a gallium arsenide wafer, in two. The breaking surface is either completely flat or shows extremely small steps, as shown in Figure 72. These steps are visible in a normal light microscope. It turns out that all observed step heights are multiples of a smallest height; its value is about 1 nm. The smallest height, the height of an atom, contradicts all possibilities of scale invariance. Matter is not scale invariant.

Listening to silence

Climbing the slopes of Motion Mountain, we arrive in a region of the forest covered with deep snow. We stop one minute and look around. It is dark, all the animals are asleep, there is no wind, and there are no sources of sound. We stand still, without breathing, and listen to the silence. (You can have this experience also in a sound studio such as those used for musical recordings, or in a quiet bedroom at night, or putting wax in your ears.) In situations of complete silence, the ear automatically increases its sensitivity;[*] we then have a strange experience. We hear 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.

---

[*] The human ear can detect pressure variations at least as small as 20 μPa.

This and many similar experiments confirm that whatever we do, we can never eliminate noise from measurements. This unavoidable type of noise is called *shot noise* in physics. Measuring the properties of this type of noise, we find that they correspond precisely to what is expected if flows, instead of being motion of continuous matter, are transports of a large number of equal, small, and discrete entities. Indeed, simply listening to noise proves that electric current is made of electrons, that air and liquids are made of molecules, and that light is made of photons. In a sense, the sound of silence is the sound of atoms. Noise would not exist in continuous systems.

Little hard balls

> I prefer knowing the cause of a single thing to being king of Persia.
> Democritus

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 arguments on giants and trees, called them 'smallest quanta.' Today they are called 'atoms', in honour of a famous reasoning of the ancient Greeks.

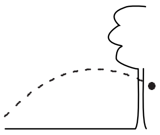
2500 years ago, this group of people asked the following question. If motion and matter are conserved, how can change and transformation exist? The philosophical school of Leucippos and Democritus of Abdera[*] studied in special detail two observations. He noted that salt dissolves in water. He also noted that fish can swim in water. He noted that the volume of water does not increase when the salt is dissolved. He also noted that when fish advance, they must put water aside. He deduced that there is only one possible solution which satisfies observations and which also reconciles conservation and transformation: 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, having a shape, and being indivisible, was imagined as being made of atoms. Atoms are particles which are hard, have a shape, but are indivisible. In other words, the Greek imagined nature as a big Lego set. Legos are first of all hard or *impenetrable,* i.e. repulsive at very small distances. Legos are *attractive* at small distance; they remain stuck together. Finally, Legos have *no interaction* at large distances. Atoms behave in the same way. (Actually, what the Greek called 'atoms' 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.)

---

[*] Leucippos of Elea (Leukippos) (ca. 490–ca. 430 BCE), Greek philosopher; Elea was a small town south of Naples/Napoli. It lies in Italy, but used to belong to the Magna Grecia. Democritus (Demokritos) of Abdera (ca. 460–ca. 356 or 370 BCE), also a Greek philosopher, arguably was the greatest philosopher who ever lived. Together with his teacher Leucippos, he was the founder of the atomic theory; Democritus was a much admired thinker, contemporary of Socrates. The vain Plato never even mentions him, as Democritus was a danger to his own fame. Democritus wrote many books which have been lost, as they were not copied during the middle ages due to his too scientific and materialistic world view, which was felt to be a danger by religious people.

[**] The story is told by Lucrece, or Titus Lucretius Carus, in his famous text *De natura rerum*, around 50 BCE. Especially if we imagine particles as little balls, we cannot avoid calling this a typically male idea.



Since atoms are so small, the experiments showing their existence took many years to convince everybody. In the nineteenth century, the idea of atoms was beautifully verified by the discovery of the 'laws' of chemistry and of gas behaviour. Later on, the noise effects were discovered.

no figure yet

Figure 73 The principle and the realization of an atomic force microscope

Nowadays, with the advances of technology, things are easier. Single atoms Ref. 139, 140 can be seen, photographed, hologrammed, counted, touched, moved, lifted, levitated, and thrown around. And indeed, like everyday matter, atoms have mass, size, shape, and colour. Single atoms have even been used Ref. 141 as light sources.

Several fields of modern physical research have fun playing with atoms in the same way that children do with Legos. Maybe the most beautiful example for these possibilities is provided by the many applications of the atomic force microscope. If you ever have the opportunity to see one, do not miss the occasion!* It is a simple device which follows the surface of an object with an atomically sharp needle; such needles, usually of tungsten, are Ref. 142 easily fabricated with a simple etching method. The height changes of the needle along its path over the surface are recorded with the 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 Ref. 143 on the surface. People can also scan a surface, pick up a given atom, and throw it towards a Ref. 144 mass spectrometer to determine what sort of atom it is.

As an aside, the construction of atomic force microscopes 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 scale invariant: in particular, it is neither smooth nor a fractal. Matter 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 its atoms produce the *shape* and the *extension* of objects, confirming the lego model of matter.** As a result, the description of motion of extended objects can

---

* A cheap version costs only a few thousand Euro, and will allow you to study the difference between consecrated and non-consecrated wafers.

** Studying matter in even more detail yields the now well-known picture that matter, at higher and higher magnifications, is made of molecules, atoms, nuclei, protons and neutrons, and finally, quarks. Atoms also contain electrons. A final type of matter, neutrinos, is observed coming from the sun and from certain types of radioactive materials. Even though the fundamental bricks have become smaller with time, the basic idea remains: matter is made of smallest entities, nowadays called elementary particles. The second part of our mountain ascent uncovers this connection in detail. Appendix C lists the measured properties of all elementary particles.

be reduced to the description of the motion of their atoms. Atomic motion will be a major theme in the following. One of its consequences is especially important: heat.

no image yet

Figure 74 The surface of a silicon crystal mapped with an atomic force microscope

Curiosities and challenges about fluids and solids

Before we continue, a few puzzles are due.

- What is the maximum length a vertically hanging wire can have? Can a wire be lowered from a suspended geostationary satellite down to the earth? This would allow to realize a space elevator. How would you build such a structure? What dangers would it face? Challenge 332 n
- Comic books have difficulties with the concept of atoms. Can Asterix throw Romans into the air using his fist? Do atoms allow the precise revolver shots shown by Lucky Luke? Can Spiderman's silk support him in his swings from building to building? Can the Road-runner stop running in three steps? Can the sun be made to stop in the sky by command? Can spaceships hover using fuel? Take any comic hero and ask yourself whether matter made of atoms would allow him the feats you read about him. Challenge 333
- When hydrogen and oxygen are combine to form water, the amount of hydrogen needed is exactly twice the amount of oxygen, if no gas is to be left over after the reaction. How does this confirm the existence of atoms? Challenge 334 n
- The most important component of air is nitrogen (about 78%). The second most important component is oxygen (about 21%). What is the third most common? Challenge 335 n
- A light bulb is placed, under water, in a stable steel cylinder with a diameter of 16 cm. A Fiat Cinquecento (500 kg) is placed on a piston pushing onto the water surface. Will the bulb resist? Challenge 336 n
- Which is most dense gas?The most dense vapour? Challenge 337
- Every year, the Institute of Maritime Systems of the University of Rostock organizes a contest. The challenge is to build a paper boat with the highest possible carrying capacity. The paper boat must weigh at most 10 g; the carrying capacity is measured by pouring lead small shot onto it, until the boat sinks. The 2002 record stands at 2.6 kg. Are you able to reach this value?(For more information, see Challenge 338

the www.paperboat.de web site.)

▪ A modern version of an old quiz – already posed by Daniel Colladon (1802–1893) – is the following. A ship of mass m in a river is pulled through ropes by horses walking along the river. If the river is of superfluid helium, meaning that there is no friction between ship and river, what energy is necessary to pull the ship along the river until a height h has been overcome? Challenge 339 n

▪ The Swiss professor Auguste Piccard (1884–1962) was a famous explorer of the stratosphere. He reached the height of 16 km in his *aerostat*.Inside the airtight cabin hanging under his balloon, he had normal air pressure. However, he needed to introduce several ropes attached at the balloon into the cabin, in order to be able to pull them, as they controlled his balloon. How did he get the ropes into the cabin while avoiding that the air leavesthe cabin? Challenge 340 Challenge 341

▪ A human cannot breathe under water, even if he has a tube going to the surface. At a few metres of depth, trying to do so is inevitably fatal. Even at a depth of 60 cm, the human body only allows breathing that way for a few minutes.Why? Challenge 342

▪ A human in air falls with a speed of about 180 km/h, depending on its clothing. How long does a fall take from 3000 m to 200 m? Challenge 343

▪ 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 344 n

▪ One or a few drops of tea usually flow along the pot opening (or falls onto the table). This phenomenon has even been simulated using supercomputer simulations of the Navier-Stokes equations describing the motion of liquids, by Kistler and Scriven. Tea pots are still shedding drops, though.

▪ The best giant soap bubbles can be made by mixing 1.5 l of water, 200 ml of corn syrup and 450 ml of dish cleaning liquid. Mix everything and then let it rest for four hours. You can then make the largest bubbles by dipping a metal ring of up to 100 mm diameter into the mixture.

Challenge 345 n But why do soap bubbles burst?

▪ Can humans start earthquakes? What would happen if all Chinese would jump at the same time from the kitchen table to the floor? Challenge 346

In fact, several strong earthquakes have been triggered by humans. this happens when water dams are filled, or when water is injected into drilling holes. It has been suggested that also the extraction of deep underground water causes earthquakes. If this proves to be correct, a large fraction of earthquakes could be human triggered.

Challenge 347 n ▪ How can a tip of a stalagtite be distinguished from a tip of a stalagmite?

Challenge 348 n ▪ How much more weight would a bath scale show, if you would stand on it in vacuum?

Challenge 349 n ▪ Why don't air molecules fall towards the bottom of the container and stay there?

Challenge 350 n ▪ Which of the two water funnels is emptied more rapidly? Applying energy conservation to the fluid (also called Bernoulli's 'law') provides the answer.

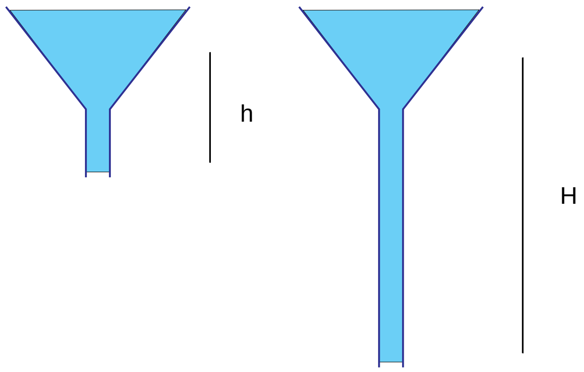


Figure 75 Which funnel is faster?

Why are objects warm?

We continue our short stroll through the field of global descriptions of motion with an overview of heat and its main concepts. For our adventure we only need to know a bit about heat. The main points that are taught in high school are almost sufficient: Ref. 146

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 principle of thermodynamics*.

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. The precise relation will be given shortly.

Heating implies flow of energy. Also friction heats up and slows down the moving bodies. In the old days, this 'creation' was even tested experimentally. It was shown that heat could be generated from friction, just by continuing rubbing, without any limit; this 'creation' implies that heat is not a material fluid extracted from the body – which in this case would be consumed after a certain time – but something else. Indeed, today we know that heat, even though it behaves like a fluid, is the disordered motion of particles.

To heat 1 kg of water by one degree, 4.2 kJ of mechanical energy

---

[*] This might change in future, when mass measurements improve in precision, thus allowing the detection of relativistic effects. See page 225

Observation	Temperature
Lowest, but unachievable temperature	0 K
In lasers, sometimes talking about negative temperature makes sense	
Temperature a perfect vacuum would have at earth's surface See page 638	40 zK
Lithium gas in certain laboratories – lowest value achieved by man, and possibly the coldest matter system in the universe	ca. 1 nK
Temperature of neutrino background in the universe	ca. 2 K
Temperature of photon gas background (or background radiation) in the universe	2.7 K
Liquid helium	4.2 K
Oxygen triple point	54.3584 K
Liquid nitrogen	77 K
Coldest weather measured (antarctic)	185 K=−88 °C
Average temperature of the earth's surface	287.2 K
Interior of human body	305.3 K
Hottest weather measured	331 K=58 °C
Boiling point of water at standard pressure	373.13 K or 99.975 °C
Liquid, pure iron	1810 K
Gold freezing point	1337.33K
Light bulb filament	2.9 kK
Earth's centre	4 kK
Sun's surface	5.8 kK
Space between earth and moon (no typo)	up to 1 MK
Sun's centre	20 MK
Inside the JET fusion tokamak	100 MK
Centre of hottest stars	1 GK
Universe when it was 1 s old	100 GK
Heavy ion collisions – highest man-made value	up to 3.6 TK
Planck temperature – nature's upper temperature limit	10^{32} K

Table 23 Some temperature measurements

need to be transformed through friction. The first to measure this with precision was, in 1842, the German physician Julius Robert Mayer (1814–1878). He performed this experiment as proof of the conservation of energy; indeed, he was the first to state energy conservation! It is one of the dark sides of modern physics that a medical doctor was the first to show the conservation of energy, and that furthermore, he was ridiculed by most physicists of his time. Worse, conservation of energy was accepted only when it was repeated many years later by two authorities: Hermann von Helmholtz (1821, Potsdam–1894) – himself also a physician turned physicist – and William Thomson (1824–1907) (later Lord Kelvin), who cited similar, but latter experiments by the English physicist James Prescott Joule (1818–1889).[*] All of them acknowledged Mayer's priority. Publicity by William Thomson eventually led to the naming of the unit of work after Joule.

In short, the sum of mechanical energy and of thermal energy is constant. This is usually called the *first principle of thermodynamics*. Equivalently, it is impossible to produce mechanical energy without paying with some other energy. This is an important statement, because among others it means that humanity will stop living one day. Indeed, we live mostly on energy from the sun; since the sun is of finite size, its energy content will be consumed one day. Can you estimate when this will happen? Challenge 351

There is also a second and a third principle of thermodynamics, to be mentioned later on. In fact, the study of these topics is called *thermostatics* if systems are at equilibrium, and *thermodynamics* if systems are away from equilibrium. In the latter case we distinguish situations *near* equilibrium, when equilibrium concepts such as temperature can still be used, from situations *far* from equilibrium, such as self-organization, where such concepts usually cannot be applied.

The study of heat tells many stories; two of them are most important: heat is due to particles and it is at the origin of the difference between past and future.

Why do balloons take up space?

---

[*] Joule is pronounced such that it rhymes with 'cool', as his descendants like to stress.

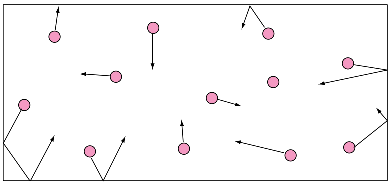


Figure 76 The basic idea of statistical mechanics about gases

Bernoulli was one of the first to think about atoms in the modern way. He focussed on studying gases.* Bernoulli reasoned that if atoms are small particles, with mass and momentum, he should be able to make quantitative predictions about the behaviour of gases, and check them with measurements. If the particles fly around in a gas, hitting each other, then the *pressure* of a gas in a container is produced by the steady flow of particles hitting the wall. It was then easy to conclude that if the particles are assumed to behave as tiny, hard and perfectly elastic balls, the quantities pressure p, volume V, and temperature T must be related by Challenge 352

$$pV = \frac{3}{2}NkT \tag{78}$$

where N is the number of particles contained in the gas. A gas made of particles with such textbook behaviour is called *ideal gas*. The relation is confirmed by experiment at room and higher temperatures, for all gases known. It thus provides a clear argument for the existence of atoms and their behaviour as normal, though small objects. (Can you imagine how N may be determined experimentally?) Challenge 353

The ideal gas relation (78) also allows an easy measurement of temperature itself. Indeed, temperature has been defined and measured for about a century in this way. Ref. 150 Most importantly of all, the ideal gas relation shows that there is a lowest temperature in nature, namely that temperature at which an ideal gas would have a vanishing volume. Sloppily speaking, this is the case when all particles are at rest. That happens, as is well known, at $T = 0\,\mathrm{K}$, i.e. at $-273.15\,°\mathrm{C}$.

The underlying approximation of hard constituents without any long-distance interactions is obviously not valid at very low temperatures. However, using improvements of the ideal gas relation (78), taking into account the deviations due to interactions between atoms or molecules, overcomes these limitations and is now standard practice, allowing to measure temperatures even at extremely low values. Ref. 151 The effects observed below 80 K, such as the solidification of air, frictionless transport of electrical current, or frictionless flow of liquids, form a fascinating world of their own; however, the beautiful domain of low temperature-physics will not be explored during this walk. Ref. 152, 153

But the ideal gas model helps to decide questions such as the one of Figure 77. Two *identical* rubber balloons, one filled up to a larger size than the other, are connected via a pipe and a valve. The valve is opened. Which one deflates? Challenge 354 n

Now you are able to take up the following challenge: how can you measure the weight of a car with a ruler only? Challenge 355

---

[*] By the way, the word '*gas*' is a modern construct. It was coined by the Brussels alchemist and physician Johan Baptista van Helmont (1579–1644), to sound similar to 'chaos'. It is one of the few words which have been invented by a particular person and then adopted all over the world.

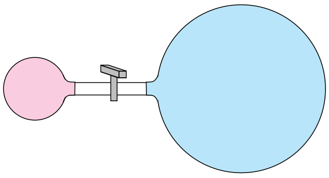


Figure 77 Which balloon wins?

Brownian motion

For many years, scientists had observed that small particles in a liquid never come to rest, when observed in a microscope. They keep executing a random zigzag movement. In 1827, the English botanist Robert Brown (1773–1858) then showed with a series of experiments that this observation is independent of the type of particle and of the type of liquid. In other words, Brown had discovered a fundamental noise in nature. Only in 1905 and 1906, Albert Einstein and independently, Marian von Smoluchowski, argued that this effect is due to the molecules of the liquid colliding with the pollen. He proposed an experiment to check this statement, even though at that time nobody was able to observe atoms directly. As expected, the experiment makes use of the properties of the noise.

Ref. 147

It had already been clear for a long time that if molecules, i.e. 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.[*]

Figure 76 shows that for monoatomic gases

Challenge 356

$$T_{\text{kin}} = \frac{3}{2} kT \qquad (79)$$

where T_{kin} is the kinetic energy per particle, and T is temperature. The so-called *Boltzmann constant* $k = 1.4 \cdot 10^{-23}\,\text{J/K}$ is the standard conversion factor between temperature and energy.[**]At a room temperature of 298 K, the kinetic energy is thus 6 zJ.

Using relation (79) to calculate the speed of air molecules at room temperature gives values of several hundred metres 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

Challenge 357

---

[*] 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. Atoms in a solid have six, whereas particles in monoatomic gases have only three. The number of degrees of freedom of molecules depends on their shape.

An excellent introduction into the physics of heat is the book by LINDA REICHL, *A Modern Course in Statistical Physics*, Wiley, 2nd edition, 1998.

[**] The important Austrian physicist Ludwig Boltzmann (1844, Wien–1906) is most famous for his work on thermodynamics, in which he explained all thermodynamic phenomena, inclusive entropy, as results of the behaviour of molecules. 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 then lead to quantum theory. It is said that Boltzmann committed suicide partly because of the resistance of the scientific establishment to his ideas, which are standard material nowadays.

slowed by collisions with air molecules, in the same way that pollen particles are hit in liquids.

It seems that the average distance the pollen particle has moved after n collisions should be zero, because all velocities are random. However, this is wrong, as experiment shows.

It is not known in which direction the particle moves, but it moves. An average *square* displacement, written $< d^2 >$, is observed. If the distance the particle moves after one collision is l, the average square displacement is given, as you should be able to
Challenge 358 show yourself,by

$$< d^2 >= nl^2 \quad . \qquad (80)$$

figure to be added

Figure 78 A typical path for a particle undergoing Brownian motion, its displacement distribution, and its average square displacement

For molecules with an average velocity v this gives

$$nl^2 = vlt \quad . \qquad (81)$$

In other words, the average square displacement increases proportionally with time. Repeatedly measuring the position of a particle should give the distribution shown in Figure 78 for the probability that the particle if found at a given distance from the starting point. This is called the *(Gaussian) normal distribution*. In 1908, the French physicist Jean Perrin (1870–
Ref. 148 1942) performed extensive experiments in order to test this prediction. He found complete correspondence of equation (81) 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 number of
Challenge 359 d molecules in a litre of water. Can you find out how?

Entropy

- It's irreversible.
- Like my raincoat!
Mel Brooks, *Spaceballs*, 1987.

Every domain of physics describes change with two quantities: energy and an extensive
Ref. 159 quantity characteristic of the domain. Table 24 provides an overview. An observable is called extensive if it increases with system size. Even though heat is related energy, the quantity physicists call heat is *not* an extensive quantity. Worse, what physicists call heat

---

Ref. 149 [*] In a delightful piece of research, Pierre Gaspard and his team showed in 1998 that Brownian motion is also chaotic.

Domain	extensive quantity	current	intensive quantity	energy flow	resistance to transport
	i.e. energy carrier	i.e. flow intensity	i.e. driving strength	i.e. power	i.e. intensity of entropy generation
Rivers	mass m	mass flow m/t	height difference gh	$P = ghm/t$	$R_\mathrm{m} = ght/m$ $[\mathrm{m}^2/\mathrm{s\,kg}]$
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	$P = UI$	$R = U/I$ $[\Omega]$
Thermo-dynamics	entropy$S$	entropy flow $I_S = dS/dt$	temperature T	$P = T I_S$	$R_S = Tt/S$ $[\mathrm{K}^2/\mathrm{W}]$
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 24 Extensive quantities in nature, i.e. quantities which *flow* and *accumulate*

is not the same as what we call heat in our everyday experience. The extensive quantity corresponding to what is called ‘heat’ in everyday language is called *entropy*,[*] in the same way as momentum is the extensive quantity describing motion. When two objects differing in temperature are brought into contact, an entropy flow takes place between them, like the flow of momentum taking place when two objects of different speed collide. Let us define the concept of entropy more precisely and explore its properties in some more detail.

Entropy is a quantity that represents the degree that energy is spread or shared among the particles of a system. Entropy measures the degree to which energy is mixed up in a system.

In other words, entropy adds up when identical systems are taken to be one. When two litres of water of the same temperature are poured together, entropy adds up.

Like any other extensive quantity, entropy can be accumulated in a body; it can flow in or out of bodies. When water is transformed into steam, the entropy added is indeed contained in the steam. In short, entropy is what is called ‘heat’ in everyday life.

In contrast to several other extensive quantities, entropy is not conserved. The sharing of energy in a system can be increased, for example by heating it. However, entropy is ‘half’ conserved: entropy does not decrease; mixing never undoes. When two liquids are poured together, they end up mixed, because mixing cannot be undone. In other words, what is

---

[*] 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.

Material	Typical entropy per particle
Monoatomic solids	0.3-10 k
Diamond	0.29 k
Graphite	0.68 k
Lead	7.79 k
Monoatomic gases	15-25 k
Helium	15.2 k
Radon	21.2 k
Diatomic gases	15-30 k
Polyatomic solids	10-60 k
Polyatomic liquids	10-80 k
Polyatomic gases	20-60 k
Icosane	112 k

Table 25 Some typical entropy values per particle at *standard* temperature and pressure as multiples of the Boltzmann constant

called equilibrium is simply the result of highest possible mixing. In short, the entropy in a closed system is increases until it reaches the maximum possible value.

When a piece of rock detaches from a mountain, it falls, tumbles into the valley, heating 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 motion that we have deduced so far.

Challenge 360

Rocks never fall upwards because mountains, valleys and rocks are made of many particles. Motions 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 tumbling rock. Irreversible processes are all those processes in which friction and its generalisations play a role. Irreversible processes are those which increase the sharing or mixing of energy. They are important: if there were no friction, shirt buttons and shoelaces would not stay fastened, we could not walk or run, coffee machines would not make coffee, and maybe most importantly of all, our memory would not work.

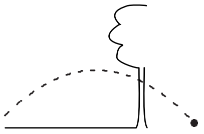
Ref. 160

See page 593

Irreversible processes transform macroscopic motion into the disorganized motion of all the small microscopic components involved: they increase the sharing and mixing of energy. It is therefore not impossible to reverse irreversible motion; it is only extremely improbable. We can say that entropy measures the amount of irreversibility; it measures the degree of mixing or decay a collective motion has undergone.

Entropy is not conserved. Entropy – ‘heat’ – can appear out of nowhere, since energy sharing or mixing can happen by itself. For example, when two different liquids at room temperature are mixed, the final temperature of the mix can differ, depending on the materials. Similarly, when electrical current flows through a room temperature body, the system can heat up or cool down, depending on the material.

Entropy is not conserved. The second principle 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, a *closed system* is a system which does not exchange energy or matter with its environment. Can you think of an example? Challenge 361

Entropy never decreases. Everyday life shows that in a closed system, the disorder increases with time, until it reaches some maximum. To reduce disorder, we need effort, i.e. work and energy. In other words, in order to reduce the disorder in a system, we need to connect the system to an energy source in some smart way. Refrigerators need electrical current precisely for this reason.

Entropy never decreases. As a consequence, *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 and white underwear, are so valued in our society. White objects defy decay.

Entropy allows to define the concept of equilibrium more precisely as the state of maximum entropy, or maximum energy sharing.

– CS – Some parts to be added. – CS –

Once it became clear that heat and temperature are due to the motion of microscopic particles, people 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) number W of possible microscopic states. A given macroscopic state can have many microscopic realizations. The logarithm of this number, multiplied by the Boltzmann constant k gives the entropy. Therefore the formula $S = k \ln W$ was inscribed by Max Planck on the tomb of Boltzmann.

In short, the higher the entropy, 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. If one prefers, entropy measures the *freedom* in the microstate that a system has. High entropy means high freedom fro the microstate. For example, 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} bits are abbreviated as $1\,\mathrm{J/K}$.*

To sum up, entropy is thus a specific measure for the characterization of disorder of thermal systems. Three points are 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. Ref. 162

Challenge 362 * This is only approximate; can you find the precise value?

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 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 $4\cdot 10^{-17}$ J/K. In short, what is usually called 'information' in everyday life is a negligible fraction of what a physicist calls information. Entropy is defined using the *physical* concept of information.

Finally, entropy is also *not* a measure for what in normal life is called the *complexity* of a situation. In fact, nobody has yet found a quantity describing this everyday experience. The task is surprisingly difficult. Have a try! Ref. 163 Challenge 363

In summary, if you hear the term entropy used with a different meaning than $S = k\ln W$, beware. Somebody is trying to get you, probably with some ideology.

We know from daily experience that transport of an extensive quantity always includes friction. Friction implies generation of entropy. In particular, 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

$$J = \kappa(T_\mathrm{i} - T_\mathrm{e}) \qquad (82)$$

where κ is a constant characterizing conduction ability of the wall. While conducting, the wall also produces entropy. The entropy flow σ is proportional to the difference of entropy flows between the interior and the exterior. In other words, one has

$$\sigma = \frac{J}{T_\mathrm{e}} - \frac{J}{T_\mathrm{i}} = \kappa\frac{(T_\mathrm{i} - T_\mathrm{e})^2}{T_\mathrm{i}T_\mathrm{e}} \quad . \qquad (83)$$

Note that we assumed in this calculation that everything is near equilibrium in each slice parallel to the wall, a reasonable assumption in everyday life. A typical case of a good wall has $\kappa = 1\,\mathrm{W/m^2K}$ in the range between 273 K and 293 K. One gets an entropy flow of

$$\sigma = 5\cdot 10^{-3}\,\mathrm{W/m^2K} \quad . \qquad (84)$$

Can you compare the amount of entropy produced in the flow with the amount transported?In comparison, a good goose feather duvet has $\kappa = 1.5\,\mathrm{W/m^2K}$, which in shops is also called 15 tog.* Challenge 364

There are two other ways, apart from heat conduction, to transport entropy: *convection*, used for heating houses, 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 PW/K. The entropy is (almost) the same that the earth receives from the sun. If more entropy had

---

* That unit is not yet as bad as the official (not a joke) BthU · h/sqft/cm/°F used in some remote provinces of our galaxy.

The insulation power 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 W/m · K for diamond, which is the best conductor of all, down to 0.1 W/m · K to 0.2 W/m · K for wood, a range between 0.015 W/m · K and 0.05 W/m · K for wools, cork and foams, and the small value of $5\cdot 10^{-3}$ W/m · K for krypton gas.

to be radiated away, the temperature of the surface of the earth would have to increase. This is called the *greenhouse effect*. Let's hope that it remains small in the near future.

Do isolated systems exist?

In all the discussions so far, we assumed that we could distinguish the system under investigation from the environment. In fact we assumed that at least in principle such *isolated* or *closed* systems, i.e. systems not interacting with their environment, actually exist. Probably our own human condition was the original model for the concept; we do experience having the possibility to act independently of the environment. Following this model, an isolated system is a system not exchanging any energy or matter with its environment. For many centuries experiments have shown no reason to question this definition.

The concept of an isolated system had to be refined somewhat with the advent of quantum mechanics. Nevertheless, the concept provides useful and precise descriptions of nature also in that domain. Only in the third part of our walk the situation will change drastically. There, the investigation of whether the universe is an isolated system will lead to surprising results. (What do youthink?)* We'll take the first steps towards the answer shortly. Challenge 365

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. But obviously, a difference exists, as we do not have the ability to remember the future. This is not 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? Such a device would have to solve a deep problem: how would it distinguish between the near and the far future?It does not take much to find out that any way to do this conflicts with the second principle of thermodynamics. Challenge 366 That is bad luck, as we would need precisely the same device to show that there is faster than light motion. Can you find the connection? Challenge 367

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 disappears. For few-particle systems, there is no difference between times gone 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 limits.

---

* A strange hint: your answer is most probably wrong.

Is everything made of particles?

> A physicist is the atom's way of knowing about atoms.
> George Wald

Ref. 154

Historically, the study of statistical mechanics has been of fundamental importance for physics. It was the first demonstration that physical objects are made of interacting particles. The story of the topic is in fact a long chain of arguments showing that all properties we ascribe to objects, such as size, stiffness, colour, mass density, magnetism, thermal or electrical conductivity, result from the interaction of the many particles they consist of. The discovery that *all objects are made of interacting particles* has often been called the main result of modern science.

See page 178 How was composition discovered? Table 24 lists the main extensive quantities used in physics. Extensive quantities are able to flow. It turns out that all flows in nature are composed of elementary processes. We saw that the flow of mass, volume, charge, entropy and substance are composed. Later, quantum theory will show the same for the flow of linear and angular momentum. *All flows are made of particles.*

This conceptual success has led many people to generalize it to the statement: 'Everything we observe is made of parts.' This approach has been applied with success to chemistry with molecules,[*] material science and geology with crystals, electricity with electrons, atoms with elementary particles, space with points, time with instants, light with photons, biology with cells, genetics with genes, neurology with neurons, mathematics with sets and relations, 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 368

See page 772 However, in the case of the *whole* of nature this idea is incorrect. It turns out to be a prejudice, and a prejudice so entrenched that for at least thirty years it has retarded further developments in physics. In particular, it does *not* apply to elementary particles and to space-time. Finding the correct description is the biggest challenge of our adventure, as it requires a complete change in thinking habits. There is a lot of fun ahead.

> 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

Curiosities and fun challenges about heat

Even though heat is disordered motion, it follows simple but surprising rules.

- If heat really is disordered motion of atoms, a big problem appears. When two atoms collide head-on, in the instant of smallest distance, none has velocity. Where did the kinetic

---

[*] A 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.

energy go? Obviously, 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 in order to show that heat cannot be atomic motion, but must be some sort of fluid. But since heat really is kinetic energy, atoms are indeed divisible, even though their name means 'indivisible'. We do not need any expensive experiment to show this.

▪ How does a usual, 1500 m^3 hot air balloon work? Challenge 369 n

▪ 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 *ice* at 0 °C and 1 kg of water at 100 °C? Challenge 370

▪ 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, Charles Blagden, together with a few friends, who remained in a room of that temperature for 45 minutes. Interestingly, the steak which he had taken with him was cooked 'well done' when he and his friends left the room. What condition had to be strictly followed in order to avoid cooking the people in the same way as the steak? Ref. 155 Challenge 371

▪ Why does water boil at 99.975 °C instead of 100 °C? Challenge 372 n

▪ Can you fill a bottle precisely with 1 ± 10^{-30} kg of water? Challenge 373 n

▪ If you do not like this text, here is a proposal. You can use the paper to make a cup, as shown in Figure 79, and boil water in it over an open flame. However, to succeed, you have to be a little careful. Can you find out in what way? Challenge 374 n

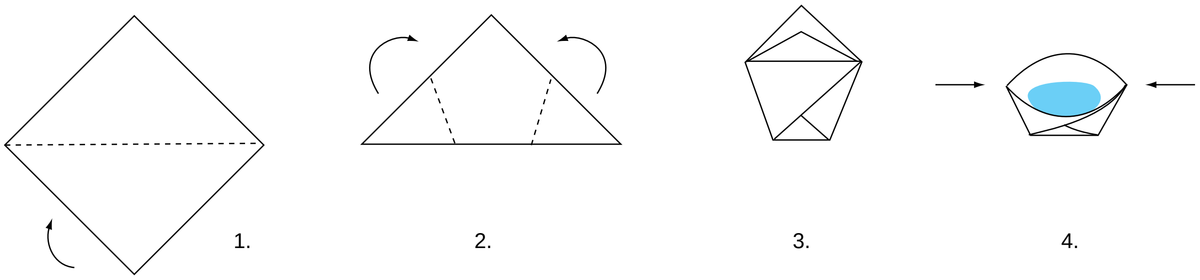


Figure 79 Are you able to boil water in this paper cup?

▪ One gram of fat, both of butter or human 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. Why are people or butter less dangerous than fuel? Challenge 375 n

▪ A famous exam question: How can you measure the height of a building with a barometer, a rope, and a ruler? Find at least six different ways. Challenge 376 n

▪ What is the probability that out of one million throws of a coin you get exactly 500 000 heads and as many tails?You may want to use Stirling's formula: $n! \approx \sqrt{2\pi n}\,(n/e)^n$ to calculate the result. Challenge 377

▪ By the way, does it make sense to say that the universe has an entropy? Challenge 378 n

▪ Can a helium balloon lift the tank which filled it? Challenge 379

▪ All friction processes, such as osmosis, diffusion, evaporation, or decay, are *slow*. They take a characteristic time. It turns out that any (macroscopic) process with a time-scale is irreversible. This is no real news: we know intuitively that undoing things always takes more time than doing them. That is the second principle of thermodynamics.

- It turns out that *storing* information is possible with negligible entropy generation. However, *erasing information* requires entropy. This is the prettiest result of the discussions on Ref. 156 irreversibility of macroscopic motion. This is the main reason that computers, as well as brains, require energy sources and cooling systems even if their mechanisms would need no energy at all.
- When mixing hot rum and cold water, how does the entropy increase due to the mixing compared to the increase due to the temperature difference? Challenge 380
- Why aren't there any small humans, e.g. 10 mm in size, as in many fairy tales? In fact, there are no warm-blooded animals of that size of any kind at all. Why? Challenge 381 n
- Shining light onto a body and repeatedly switching it on and off 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 photoacoustic *spectrum* for the material. This method allows the detection of gas concentrations in air of 1 part in 10^9. It is used among others to study the gases emitted by plants. Plants emit methane, alcohol and acetaldehyde in small quantities; the photoacoustic effect can detect these gases and help understanding the processes behind their emission.
- What is the rough probability that all oxygen molecules in the air move away from a given city for a few minutes, killing all inhabitants? Challenge 382
- If you pour a litre of water into the sea, stir thoroughly through all oceans and then take out a litre of the mixture, how many of the original atoms will you find? Challenge 383
- How long could you breathe in the room you are in if it were airtight? Challenge 384
- What happens if you put some ash onto a piece of sugar, and put fire to the whole?(Attention: this is dangerous and not for kids.) Challenge 385
- Entropy calculation is often surprising. For a system of N particles with two states each, there are $W_{\text{all}} = 2^N$ states. For its most probable configuration, with exactly half the particles in one state, and the other half in the other state, we have $W_{\text{max}} = N!/((N/2)!)^2$. Now, for a macroscopic amount of particles, we typically have $N = 10^{24}$. That gives $W_{\text{all}} \gg W_{\text{max}}$; indeed, the former is 10^{12} times larger than the latter. On the other hand, we find that $\ln W_{\text{all}}$ and $\ln W_{\text{max}}$ agree for the first 20 digits! Challenge 386 Even though the configuration with exactly half particles in each state is much more rare than the general case, where the number ratio is allowed to vary, the entropy turns out to be the same. Why? Challenge 387
- If heat is due to motion of atoms, our built-in senses of heat and cold simply are detectors of motion. How could they work? Challenge 388

By the way, the senses of smell and taste can also be seen as motion detectors, as they signal the presence of molecules flying around in air or in liquids. Do you agree? Challenge 389

- The moon has an atmosphere as well, although an extremely thin one, consisting of sodium (Na) and potassium (K). It has been detected up to nine moon radii from its surface. The atmosphere of the moon is generated from the surface by the ultraviolet radiation from the sun. Can you estimate the moon's atmospheric density? Challenge 390 n
- Does it make sense to add a line in Table 24 for the quantity of physical action? Why? Challenge 391
- Diffusion provides a length scale. For example, insects take up oxygen through their skin. As a result, the interior of their bodies cannot be much more distant from the skin than about a centimetre. Can you list other length scales produced by diffusion? Challenge 392

- Thermometers based on mercury can reach 750 °C. How is this possible, given that mercury boils at 357 °C? Challenge 393

- It is possible to build a power station by building a large chimney, so that air heated by the sun flows upward in it, driving a turbine doing so. It is also possible to realize a power station by building a long vertical tube, letting a gas such as ammonia rise into it which is then liquefied at the top by the low temperatures in the upper atmosphere; when it falls back down a second tube as a liquid – just like rain – it would drive a turbine. Why are such schemes, which are almost completely pollution free, not used yet? Challenge 394

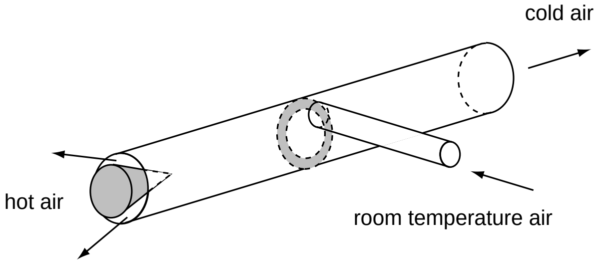


Figure 80 The Wirbelrohr or Ranque-Hilsch vortex tube

- One of the most surprising devices is the *Wirbelrohr* or Ranque-Hilsch vortex tube. By blowing compressed room temperature air into it at its midpoint, two flows of air are formed at its ends. One is extremely cold, easily as low as −50 °C, and one extremely hot, up to 200 °C. No moving parts and no heating devices are found inside. How does it work? Challenge 395 n

- A strange but fascinating application of heat are thermoacoustic engines, pumps and refrigerators. It is possible to use loud sound in closed metal chambers to move heat from a cold source to a hot one. These devices have few moving parts, and are being studied in a few places with the hope of future applications. Ref. 157

- What happens to entropy when gravitation is taken into account? We carefully left it out of the picture. In fact, many problems appear – just have a try to study the issue. Jakob Bekenstein stated that the state of highest entropy of matter is attained when the matter forms a black hole. Can you confirm this? Challenge 396

- Gerhard Müller has discovered a simple but beautiful way to observe selforganisation in solids. At the same time, his system provides a model for a famous geological process, the formation of hexagonal columns in basalt, such as the Devil's Staircase in Ireland. Similar formations are found in many other places of the earth. Just take some rice or corn starch, mix it with the about half the amount of water, put it into a pan, and dry it with a lamp. Ref. 158 Hexagonal columns form.The analogy works because the drying of starch and the cooling of lava are diffusive processes governed by the same equations, because the boundary conditions are the same, and because both materials react by a small volume reduction. Challenge 397 e

After this short trip into thermostatics, let us have an even shorter look at one aspect of thermodynamics.

Self-organization and chaos

> The study of non-linear physics is like
> the study of non-elephant biology.

Pattern	period	amplitude	origin
sand banks	2 to 10 km	2 to 20 m	tides
sand waves	100 to 800 m	5 m	tides
megaribbles	1 m	0.1 m	tides
ribbles	5 cm	5 mm	waves

Table 26 Sand patterns in the sea

Self-organization is the most general of all descriptions of motion. It studies the appearance of order. *Order* is the collective term for *shape*, such as the complex symmetry of Ref. 164 snowflakes, for *pattern*, such as the stripes of zebras, and for *cycle*, such as the creation of sound when singing. You might check that every example of what we call *beauty* is a Challenge 398 combination of shapes, patterns and cycles.Self-organization can thus simply be called the study of the origin of beauty.

Order appearance is found from the cell differentiation in an embryo inside a woman's body, the formation of colour patterns on tigers, tropical fish and butterflies, to the formation of the symmetrical arrangements of flower petals and the formation of biological rhythms.

Figures to be added

Figure 81 Examples of self-organization for sand

Fluids also provide a large number of phenomena where appearance and disappearance of order can be studied. 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 are examples.

All growth processes are self-organization phenomena. Have you ever pondered the growth of teeth, where a practically inorganic material forms shapes in the the upper and the lower rows fitting exactly into each other? Also the formation, before and after birth, of neural networks in the brain is a process of self-organization. Even the physical processes at the basis of thinking, with all its changing electrical signals, should at least partly be described along these lines.

Also biological evolution is a special case of growth. Wherever an aspect can be described quantitatively, the topic becomes fascinating. For example, take the evolution of animal shapes: it turns out that snake tongues are forked because that is the most efficient shape for following chemical trails left by prey and conspecifics. Also the fixed number of petals of flowers are consequences of self-organisation.

Ref. 165

See page 510

Many problems of self-organization are mechanical problems, such as the formation of mountains rows when continents move, the creation of earthquakes, or the creation of regular cloud arrangements in the sky. Pondering the mechanisms behind the formation of clouds you see from an aeroplane can transform a boring flight into a fascinating intellectual adventure.

Challenge 399

Figure(s) to be added

Figure 82 An oscillon formed by shaken bronze balls

Studies into the conditions required for order appearance or disappearance have shown that description requires only a few common concepts, independently of the physical system. This is best seen looking at a few examples.

All the richness of self-organization is shown by the study of plain sand. Why do sand dunes have ripples, as does the sand floor at the bottom of the sea? People also study how avalanches form on steep heaps and how sand behaves in hourglasses, in mixers, or in vibrating containers. Results are often surprising. For example, only recently Umbanhowar and Swinney have found that when a flat container with tiny bronze balls (less than a millimetre in diameter) is shaken up and down in vacuum at certain frequencies, the surface of this bronze ‘sand’ shows stable heaps. These heaps, so-called oscillons, also bob up and down. They can move and interact with other heaps. In fact, sand and dust is proving to be such a beautiful and fascinating topic that the prospect of each human becoming dust again does not look so grim at all.

Ref. 166

Figure to be included

Figure 83 Magic numbers of spheres swirled in a dish

A second, simple and beautiful example of self-organization is the effect discovered in 1999 by Karsten Kötter and his group. They found that the behaviour of a set of spheres swirled in a dish depends on the number of spheres used. Usually, all spheres get continuously mixed up. For certain ‘magical’ numbers, such as 21, stable ring patterns emerge, for which outside spheres remain outside, and inside ones remain inside. The rings are best visualized by colouring the spheres.

Ref. 167

These and many other studies of self-organizing systems have changed the description of nature in a number of ways. First of all, it was shown that patterns and shapes are similar to cycles: all are due to motion. Without motion, and thus without history, there is no order. there are neither patterns nor shapes. Every pattern has a history; every pattern is an example of motion.

Ref. 168

Secondly, patterns, shapes and cycles are due to the organized motion of large numbers of small constituents; systems which self-organize are always composite and *cooperative structures*.

Thirdly, all these systems show evolution equations which are *nonlinear* in the configuration variables. Linear systems do not self-organize. Many self-organizing systems also show *chaotic* motion.

Fourthly, the appearance and disappearance of order depends on the strength of a driving force, the so-called *order parameter*. Often, chaotic motion appears when the driving is increased beyond the value necessary for the appearance of order. An example of chaotic motion is turbulence, which appears when the order parameter, which is proportional to the speed of the fluid, is increased to high values.

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 self-organization without thermodynamics playing a role. Self-organizing 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.*

All self-organizing systems at the onset of order appearance can be described by equations of the general form

$$\frac{\partial A(t,x)}{\partial t} = \lambda A - \mu |A|^2 A + \kappa \Delta A + \text{higher orders} \quad . \qquad (85)$$

Here, the – possibly complex – observable A is the one which appears when order appears, such as the oscillation amplitude or the pattern amplitude. We note the driving term λA in which λ describes the strength of the driving, the nonlinearity in A, and the dissipative term ΔA. In cases that the dissipative term plays no role ($\kappa = 0$), when λ increases above zero, a *temporal* oscillation, i.e. a stable cycle with non-vanishing amplitude appears. Challenge 400

In case that the diffusive term does play a role, equation (85) describes how an amplitude for a *spatial* oscillation appears when the driving parameter λ becomes positive, as the solution $A = 0$ then becomes unstable. Challenge 401

In both cases, the onset of order is called a *bifurcation*, because at this critical value of the driving parameter λ 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 (85) and its numerous variations allow to describe many additional phenomena, such as spirals, waves, hexagonal patterns, topological defects, some forms of turbulence, etc. The main point is to distil the observable A and the parameters λ, μ and κ from the physical system under consideration. Ref. 169

---

* When you wish to describe the 'mystery' of human life, often terms like 'fire', 'river', or 'tree' are used as analogies. They all are examples of self-organized 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. We could even go further and speculate that pure beauty *is* pure self-organization. The lack of beauty indeed often results from a disturbed equilibrium between external breaking and external driving.

Self-organization is a vast field that is yielding new results almost by the week. In addition, to discover new topics of study, it is often sufficient to keep one's eye open; most effects are in the reach of high school students. Good hunting! Challenge 402

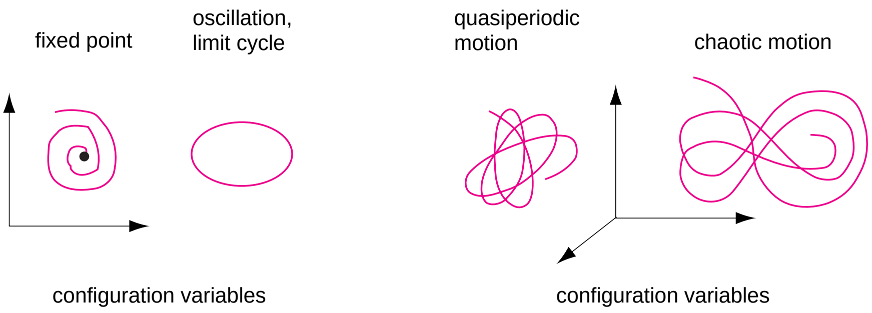


Figure 84 Examples of different types of motion in configuration space

When the driving parameter is increased as much as possible, order becomes more and more irregular, and at the end one usually finds chaos. For physicists, chaotic motion is the most irregular type of motion.* Chaos can be defined independently of self-organization, namely as that motion of systems for which small changes in initial conditions evolve into large changes of the motion, as shown in Figure 85. The weather is such a system, but it turns out that also dropping water-taps, many flows of liquids, the fall of dice, and many other systems are chaotic. For example, 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. 126

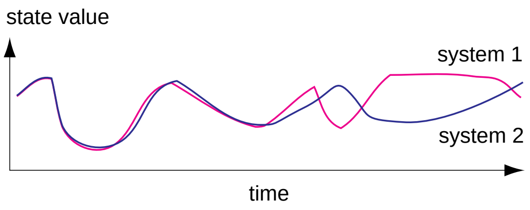


Figure 85 Sensitivity to initial conditions

As a note, can you show with simple words that the butterfly effect does not exist? The effect is an invention of newspapers; they claim that nonlinearity means that a small change in initial conditions can lead to large effects; thus a butterfly wing beat could lead to a tornado. This is wrong. Challenge 403 n

---

* 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, i.e. the earth, emerged from the chaos existing at the beginning. She then gave birth to the other gods, the animals, and the first humans.

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 many 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.

The steps from disorder to order to chaos, all examples of self-organization, are found in many fluid systems. Their study should lead, one day, to uncover the mysteries of turbulence. Ref. 170

Finally, self-organization is of interest also for a more general reason. Sometimes it is said 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 self-organization has definitely settled this debate. The properties of water molecules do allow to predict Niagara falls[*] and the diffusion of signal molecules do determine the development of a single cell into a full human being. In particular, cooperative phenomena determine the place where arms and legs 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 fur patterns on zebras and leopards, to cite only a few examples. Similarly, the mechanisms at the origin of the heart beat and many other cycles have been deciphered.

Self-organization 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. If you plan to enter science, evaluate taking this path. Challenge 405

These studies provide the last arguments to confirm what J. Offrey de la Mettrie wrote in 1748 in his famous book *L'homme machine*: humans are complex machines. Indeed, 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 self-organization is and will provide fascinating insights for many years to come, we now leave it and continue with our own adventure on the fundaments of motion.[**]

---

[*] Already small versions of Niagara fall, namely dripping water taps, show a large range of cooperative phenomena, including the chaotic, i.e. non-periodic, fall of water drops. Ref. 171 This happens when the water flow has the correct value, as you might want to test in your ownkitchen. Challenge 404

[**] An important case of self-organization 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 is also available in several expanded translations.

Curiosities and fun challenges about self-organisation

- When wine is made to swirl in a wine glass, after the wine has calmed down, the wine flowing down the glass walls forms little arcs. Can you explain in a few words what forms them? Challenge 406
- How does the average distance between cars parked along a street change over time, assuming a constant rate of cars leaving and coming? Challenge 407 d

5. Limits of Galilean physics – what is wrong with school physics

> I only know that I know nothing.
> Socrates (470–399 BCE), as cited by Plato.

Socrates' saying applies also to Galilean physics, despite its general success in engineering and in the description of everyday life.

Research topics in classical dynamics

Even though mechanics is now several hundred years old, research into its details is still not concluded.

- We mentioned already above the study of the stability of the solar system. The long-term future of the planets is unknown. In general, the behaviour of few body systems interacting through gravitation is still a research topic of mathematical physics. Answering the simple question on how long a given set of bodies gravitating around each other will stay together is a formidable challenge. This so-called *many-body problem* is one of the seemingly never-ending stories of theoretical physics. Interesting progress has been achieved, but the concluding answer is still missing. Ref. 172
- The challenges of self-organization, of nonlinear evolution equations, and of chaotic motion are still plenty and motivate numerous researchers in mathematics, physics, chemistry, biology, medicine and the other sciences.
- Perhaps the toughest of all problems in modern physics is the description of *turbulence*. When the young Werner Heisenberg was asked to continue research 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 despite the number and importance of its applications, only now, at the beginning of the twenty-first century, its secrets start to be unravelled. It is thought that the equations of motion describing fluids, the so-called *Navier-Stokes equations*, are sufficient to understand them.* But the mathematics behind them is mind boggling. There is even a one million dollar prize offered by the Fondation Clay at the Collège de France, for certain steps on the way of solving them. Ref. 173

What is contact?

> Democritus declared that there is a unique sort of motion: that ensuing from collision.

---

[*] They are named after Claude Navier and Georges Gabriel Stokes.

Ref. 174 Simplicius, Commentary on the Physics of Aristotle, 42, 10.

See page 61 We defined mass through the measurement of velocity changes during collisions. But why do objects change their motion in such instances? Why are collisions between two balls made of chewing gum different from those between two stainless steel balls? What happens during those moments of contact?

Contact is related to material properties, who in turn 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 chemistry and of cooking were related to the properties of motion only in the twentieth century, after having been independently pursued for thousands of years. Since material properties determine the essence of contact, we *need* knowledge about matter and about materials to understand the origin of mass, and thus of motion. The second part of the mountain ascent will uncover these 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. To make even this statement more precise, we distinguish between two 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.*

See Appendix B The overview of the most precise and accurate measurements at present shows that the record number of digits is 14. Why so few? Classical physics doesn't cover this issue. What is the maximum number of digits we can expect in measurements, what determines it, and how can we achieve it? These questions are still open at this point; they will be covered in the second part of our mountain ascent.

On the other hand, statements with false accuracy abound. What should we think of a car company – Ford – claiming that the drag coefficient c_w is 0.375? Or of the official claim that the world record in fuel consumption for cars is 2315.473 km/l? Or of the statement that 70.3% of all citizen share a certain opinion? One lesson we gain from the investigations into measurement errors is that we should never provide more digits for results than we can put our hand into fire for. Challenge 408 n

In our walk we aim for precision and accuracy, while avoiding false accuracy. Therefore, concepts have mainly to be *precise*, and descriptions have to be *accurate*. Any inaccuracy is a proof of lack of understanding. To put it bluntly, 'inaccurate' means *wrong*. Increasing the accuracy and precision of the description of nature implies leaving behind us all the mistakes we made so far. That is our aim in the following.

---

* For measurements, both precision and accuracy are best described by their *standard deviation*, as explained in Appendix B, on page 881.

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, we have 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.You always need a non-gravitational method to start and stop the stopwatch. Similarly, when you measure length, e.g. of a table, you have to hold a meter bar or some other device near it. The interaction necessary to line up the meter and the table cannot be gravitational. Challenge 409

A similar limitation applies even to mass measurements. Try to measure mass using gravitation alone.Any scale or balance needs other, usually mechanical, electromagnetic or optical interactions to achieve its function. Can you confirm that the same applies to speed and to angle measurements?In summary, whatever method we use, *in order to define velocity, length, time, and mass, interactions other than gravity are needed.* In short, our simple ability to measure shows that gravity is not all there is. Challenge 410 Challenge 411

A second fact hints that more is awaiting us. We found that not all observers agree on the measurement values of change; indeed, the value of the action of a system depends on the motion of the observer. So far, we have ignored this situation; for a complete description of motion we cannot do that though.

We need the concepts of space, time and mass to *talk* about motion, and we need to be able to talk to everybody. Since we cannot give the term 'motion' any meaning as long as we neglect the non-gravitational interactions in nature, we are forced to investigate these other interactions as well. To proceed as quickly as possible, we start studying an example of motion which we mentioned at the beginning but which we excluded from our investigations so far, even though it is used for the definition both of the meter and the second: the motion of light.

References

115 The beautiful story of the south pointing carriage is told in 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 reconstructed in 1947 by George Lancaster, a British engineer. Cited on page 130.

116 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 130.

117 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 130.

118 A historical account is given in WOLFGANG YOURGRAY & STANLEY MANDELSTAM, *Variational principles in dynamics and quantum theory*, Dover, 1968. Cited on pages 132 and 137.

119 MAX PÄSLER, *Prinzipe der Mechanik*, Walter de Gruyter & Co., 1968. Cited on page 135.

120 The relations between the possible Lagrangians is studied by ... Cited on page 136.

121 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 G.B. West, J.H. Brown & B.J. Enquist, *A general model for the structure, and allometry of plant vascular systems*, Nature **400**, pp. 664–667, 1999. Cited on page 142.

122 J.R. Banavar, A. Martin & A. Rinaldo, *Size and form in efficient transportation networks*, Nature pp. 130–132, 13 May 1999. Cited on page 142.

123 J.A. Moore, *An innovation in physics instruction for nonscience majors*, American Journal of Physics **46**, pp. 607–612, 1978. Cited on page 139.

124 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 141.

125 A good review article is by David W. Hughes, *Comets and Asteroids*, Contemporary Physics **35**, pp. 75–93, 1994. Cited on page 141.

126 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 190.

127 See e.g. B. Bower, *A child's theory of mind*, Science News **144**, pp. 40–41. Cited on page 145.

128 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*, http://arxiv.org/abs/math-ph/0102011. Cited on page 152.

129 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 153.

130 W. Z¨urn & R. Widmer-Schnidrig, *Globale Eigenschwingungen der Erde*, Physik Journal **1**, pp. 49–55, 2002. Cited on page 157.

131 An informative and modern summary of present research about the ear is http://www.physicsweb.org/article/world/15/5/8. Cited on page 160.

132 A.L. Hodgkin & A.F. Huxley, *A quantitative description of membrane current and its application to conduction and excitation in nerve*, Journal of Physiology **121**, p. 500, 1952. This famous paper of theoretical biology earned the authors a Nobel prize. Cited on page 160.

133 T. Filippov, *The Versatile Soliton*, Springer Verlag, 2000. See also J.S. Russell, *Report of the Fourteenth Meeting of the British Association for the Advancement of Science*, Murray, London, 1844, pp. 311–390. Cited on page 162.

134 See for example the article by G. Fritsch, *Infraschall*, Physik in unserer Zeit **13**, pp. 104–110, 1982. Cited on page 163.

135 M. Ausloos, Proceedings of the Royal Society in London A **400**, pp. 331–350, 1985. Cited on page 164.

136 T.A. McMahob & J. Tyler Bonner, *Form und Leben – Konstruktion vom Reißbrett der Natur*, Spektrum Verlag, 1985. Cited on page 166.

137 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. See also P. Nobel, *Plant Physiology*, 2nd Edition, Academic Press, 1999. Cited on page 167.

138 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 167.

139 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 & H. Dehmelt, *Localized visible Ba$^+$ mono-ion oscillator*, Physical Review A **22**, pp. 1137–1140, 1980. Cited on page 169.

140 Holograms of atoms have been produced first by Hans-Werner Fink & al., *Atomic resolution in lens-less low-energy electron holography*, Physical Review Letters **67**, pp. 1543–1546, 1991. Cited on page 169.

141 A single atom laser has been built in 1994 by K. An, J.J. Childs, R.R. Dasari & M.S. Feld, *Microlaser: a laser with one atom in an optical resonator*, Physical Review Letters **73**, p. 3375, 1994. Cited on page 169.

142 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. *L*925–*L*930, 1996. Cited on page 169.

143 Good pictures of atoms taken by atomic force microscopes can be found in Herman B. Elswijk, ..., Cited on page 169.

144 J. Spence & U. Weierstrass, ..., end 1997. Cited on page 169.

145 Some religions incredibly, still pretend that people believe the opposite. Cited on page .

146 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 172.

147 The famous paper on Brownian motion which so much contributed to Einstein's fame is A. Einstein, *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*, Annalen der Physik **17**, pp. 549–560, 1905. In the following years, Einstein wrote a series of subsequent papers elaborating on this topic. For example, he published his 1905 PhD thesis as A. Einstein, *Eine neue Bestimmung der Moleküldimensionen*, Annalen der Physik **19**, pp. 289–306, 1906, and he corrected a small mistake in A. Einstein, *Berichtigung zu meiner Arbeit: 'Eine neue Bestimmung der Moleküldimensionen'*, Annalen der Physik **34**, pp. 591–592, 1911, where, using also new data, he found the value $6.6 \cdot 10^{23}$. Cited on page 176.

148 The first tests of the prediction were performed by J. Perrin, Comptes Rendus de l'Académie des Sciences **147**, pp. 475–476, and pp. 530–532, 1908. He masterfully sums up the whole discussion in Jean Perrin, *Les atomes*, Librarie Félix Alcan, Paris, 1913. Cited on page 177.

149 Pierre Gaspard & al., *Experimental evidence for microscopic chaos*, Nature **394**, p. 865, 27 August 1998. Cited on page 177.

150 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 (in the range 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 887. Cited on page 175.

151 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 175.

152 The weight is given by the pressure of the tires times the contact area. Cited on page 175.

153 An overview of low temperature physics is given by ... Cited on page 175.

154 The quote is found in the introduction by George Wald to the text by Lawrence J. Henderson, *The Fitness of the Environment*, Macmillan, New York, 1913, reprinted 1958. Cited on page 183.

155 K. Schmidt-Nielsen, *Desert animals: physiological problems of heat and water*, Oxford University Press, 1964. Cited on page 184.

156 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. The summary: we should pay to throw the newspaper away, not to buy it. Cited on page 185.

157 See for example G. Swift, *Thermoacoustic engines and refrigerators*, Physics Today pp. 22–28, July 1995. Cited on page 186.

158 G. Mueller, J. Geophys. Research **103**, pp. 15239–15253, 1998. Cited on page 186.

159 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 177.

160 B. Polster, *What is the best way to lace your shoes?*, Nature 420 p. 476, 5 December 2002. Cited on page 179.

161 Ilya Prigogine, ... Cited on page .

162 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 180.

163 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 181.

164 About the shape of snowflakes, see the famous book by W.A. Bentlye & W.J. Humphreys, *Snow Crystals*, Dover Publications, New York, 1962. This second printing of the original from 1931 shows a large part of the result of Bentley's lifelong passion, namely several thousand photographs of snowflakes. Cited on page 187.

165 K. Schwenk, *Why snakes have forked tongues*, Science **263**, pp. 1573–1577, 1994. Cited on page 188.

166 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 188.

167 K. Kötter, E. Goles & M. Markus, *Shell structures with ‘magic numbers’ of spheres in a swirled disk*, Physical Review E **60**, pp. 7182–7185, 1999, See also the http://www.aip.org/physnews/graphics/html/magicball.html web site. Cited on page 188.

168 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 188.

169 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 189.

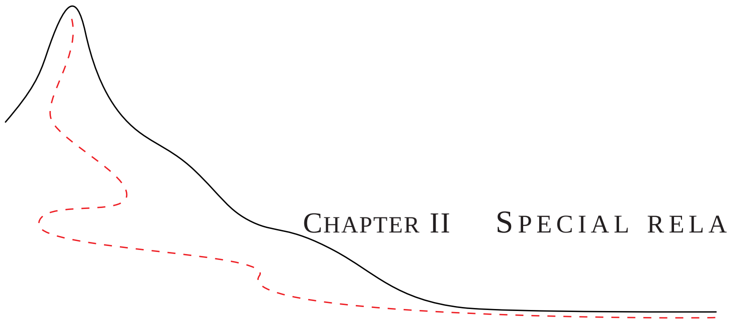
170 For a short, modern overview of turbulence, see L.P. Kadanoff, *A model of turbulence*, Physics Today **48**, pp. 11–13, Spetember 1995. Cited on page 191.

171 For a clear introduction, see T. Schmidt & M. Mahrl, *A simple mathematical model of a dripping tap*, European Journal of Physics **18**, pp. 377–383, 1997. Cited on page 191.

172 J.J. Lissauer, *Chaotic motion in the solar system*, Reviews of Modern Physics **71**, pp. 835–845, 1999. Cited on page 192.

173 The present status of the understanding of turbulence is given in ... Cited on page 192.

174 See Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 426, 1991. Cited on page 193.



CHAPTER II SPECIAL RELATIVITY

6. Light's speed and observer's rest

Typeset in January 2003

Light is important for describing motion precisely. We need it every day to read lines such as these; 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,[*] again using light. How do we measure length to high precision? With light. How do we measure time to high precision? With Ref. 175 light; once that from the sun was used, nowadays it is light from caesium atoms. In other words, light is important because it is the official standard for *undisturbed motion*. Physics would have evolved much more rapidly if, at some earlier time, light propagation had been recognized as the ideal example of motion.

But is light a moving phenomenon at all? It was already known in ancient Greece that this can be 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 Ref. 176 surface showing the shadow had already been reached by the Greek thinker Empedocles (ca. 490–ca. 430 BCE).

We can confirm this result with a different, but equally simple, argument. Speed can be measured. Therefore the *perfect* speed, which is used as the implicit measurement standard, Challenge 413 n must have a finite value. An infinite velocity standard would not allow measurements at all. Of course, we are implying here that the speed of light actually *is* the perfect speed. We will show this in a minute.

---

[*] Note that looking along the plane from all sides is not sufficient for this; a surface that a light beam touches right along its length in *all* directions does not need to be flat. Can you give an example? One needs other Challenge 412 n methods to check flatness with light. Can you specify one?

[**] Whenever a source produces shadows, the emitted entity is called *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 travelling 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.

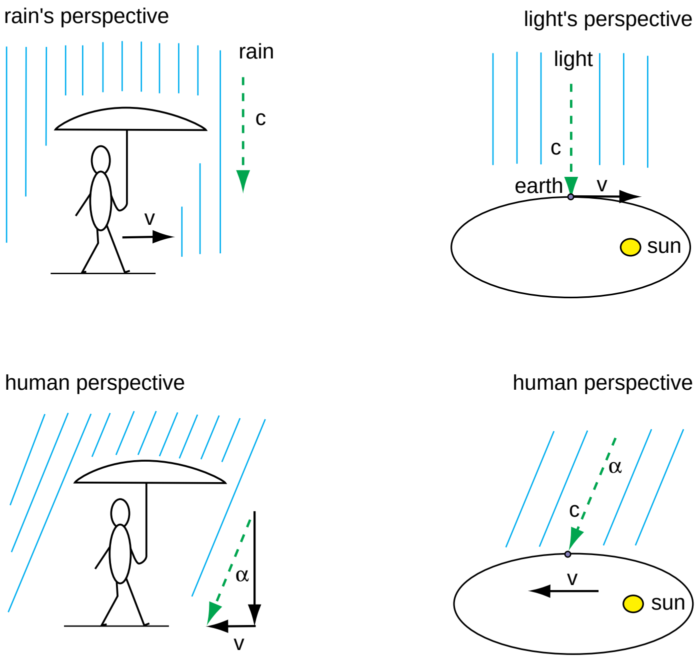


Figure 86 The rain method of measuring the speed of light

The speed of light is high; therefore it was not measured for the first time until 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 orbits of the moons of Jupiter. He obtained an incorrect value because he used the wrong value for their distance from earth. However, this was quickly corrected by his peers, including Newton himself. The result was then confirmed most beautifully by the next measurement, which was performed only fifty years later, in 1726, by the astronomer James Bradley (1693–1762). Being English, Bradley thought of the 'rain method' to measure the speed of light. Ref. 177

How can we measure the speed of falling rain? We walk rapidly with an umbrella, measure the angle α at which the rain appears to fall, and then measure our own velocity v. As shown in Figure 86, the velocity c of the rain is then given by

$$c = v/\tan\alpha \quad . \tag{86}$$

The same measurement can be made for light; we just need to measure the angle at which the light from a star above earth's orbit arrives at the earth. This effect is called the *aberration* of light; the angle is best found comparing measurements distant by six months. The value of the angle is $20.5''$; nowadays it can be measured with a precision of five decimal digits. Given that the velocity of the earth around the sun is $v = 2\pi R/T = 29.7\,\mathrm{km/s}$, the speed of light must therefore be $c = 3.00 \cdot 10^8$ 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 Mm/h = 14 km/s, with the growth of children, about 3 nm/s, or with the growth of stalagmites in caves, about 0.3 pm/s. We begin to realize why it took so long to measure the speed of light.** Table 165 gives a summary about what is known about the motion of light.

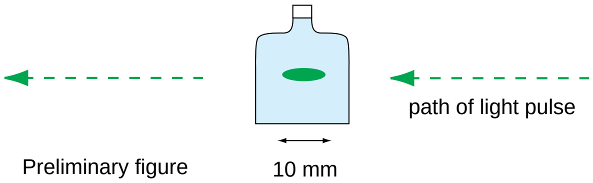


Figure 87 A photograph of a light pulse moving from right to left through milky water, taken by Dugay and Mattick

The speed of light is so high that it is even difficult to prove that it is *finite*. Perhaps the most beautiful way to prove this is to photograph a light pulse flying across one's field of view, in the same way as one takes the picture of a car driving by or of a bullet flying along. Ref. 179 Figure 87 shows the first such photograph, produced in 1971 with a standard off-the-shelf reflex camera, a very fast shutter invented by the photographers, and, most noteworthy, not a

---

[*] Umbrellas were not common in Britain in 1726; they became fashionable later, after being introduced from China. The umbrella part of the story is made up. In reality, Bradley first understood his unexpected result while sailing on the Thames, when he noted that on a moving ship the apparent wind has a different direction to that 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 was looking for, namely the star parallax.

Challenge 414 n By the way, it follows from special relativity that the correct formula is $c = v/\sin a$; 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). You measure 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 88) and the distance to the sun. The explanation is a puzzle left to the reader. Challenge 415 n

The angle in question is almost a right angle (which would yield an infinite distance), and good instruments are needed to measure it with precision, 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 °, 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 metres. Moon distance variations can even be measured down to the 1 centimetre range; can you guess how this is achieved? (Ref. 178; See page 884; Challenge 416 n)

Ref. 54 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 performed in 1849 by the French physicist Hippolyte L. Fizeau (1819–1896). His value was only 5% greater than the modern one. He sent a beam of light towards a distant mirror and measured the time the light took to come back. How far away does the mirror have to be? How do you think did he measure the time, without using any electric device? Challenge 417 n

Observations about light
light can move through vacuum;
the speed of light, its true signal speed, is the forerunner speed;
in vacuum its value is 299 792 458 m/s;
light transports energy;
light has momentum: it can hit bodies;
light has angular momentum: it can rotate bodies;
light moves across other light undisturbed;
light in vacuum always moves faster than any material body does;
the proper speed of light is infinite;
shadows can move with no speed limit;
light moves straight when far from matter;
light is a wave;
light beams are approximations;
in matter, the forerunner speed as well as the energy speed of light are lower than in vacuum;
in matter, the group velocity of light pulses can have any value, positive or negative, without limits.

Table 27 Properties of the motion of light

single piece of electronic equipment. (How fast does such a shutter have to be? How would you build such a shutter? And how would you open it at the right instant?) Challenge 418 n

In short, light is thus much faster than lightning, as you might like to check yourself. Challenge 419 n But once the velocity of light could be measured routinely, two surprising properties were discovered in the late nineteenth century. They form the basis of special relativity. Ref. 180

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

We all know that in order to throw a stone as far as possible, we run as we throw it; we know instinctively that in that case the stone's speed with respect to the ground is higher. However, to the initial astonishment of everybody, experiments show that light emitted from a moving lamp has the same speed as light emitted from a resting one. Many carefully and specially designed experiments confirmed this result to high precision; the speed of light can Ref. 181 be measured with a precision of better than 1 m/s, but even for lamp speeds of more than 290 000 000 m/s no differences have been found. (Can you guess what lamps were used?)In Challenge 420 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 with respect to the light source. The velocity of light is indeed the ideal, perfect measurement standard.*

---

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

The speed of light is also not far from the speed of neutrinos. This was shown most spectacularly by the observation of a supernova in 1987, when the flash and the neutrino pulse arrived spaced by a few hours. Can you deduce the maximal difference between the two speeds, knowing that the supernova was $1.7 \cdot 10^5$ light years away? Challenge 421

Ref. 183
See page 379

There is also a second set of experimental evidence for the constancy of the speed of light: every electromagnetic device, such as an electric toothbrush, shows that the speed of light 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, did the German–Swiss physicist Albert Einstein[*] show 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 detailed in the chapter on electrodynamics.[**] In simple words, if the speed of light were not constant, observers would be able to move at the speed of light. Since light is a wave, such observers would see a wave standing still. However, electromagnetism forbids the existence of such a phenomenon. Therefore, observers cannot reach the speed of light.

The constancy of the speed of light is in complete contrast with Galilean mechanics, and proves that the latter is *wrong* at high velocities. At low velocities the description remains good, because the error is small. But if we look for a description valid at *all* velocities, Galilean mechanics has to be discarded. For example, when we play tennis we use the fact that by hitting the ball in the right way, we can increase or decrease its speed. But with light this is impossible. Even if we take an aeroplane and fly after a light beam, it still moves away with the same speed. This is in contrast with cars. If we accelerate a car we are driving, the cars on the other side of the road pass by with higher and higher speeds as we drive 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 through positions $x(t)$ for O and through positions $\xi(\tau)$ (pronounced 'xi of tau') for Ω ('omega'). Since the speed of light is the same for both, we have

$$\frac{x}{t} = c = \frac{\xi}{\tau} \quad . \tag{87}$$

Ref. 186
Ref. 187

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

Ref. 184

[*] Albert Einstein (1879, Ulm–1955, Princeton); one of the greatest physicists of all time. He published three important papers in 1905, namely about Brownian motion, about special relativity and about the idea of light quanta. Each paper was worth a Nobel prize, but he was awarded the prize only for the last one. In 1905, he discovered the famous formula $E_o = mc^2$ (published early 1906). Although he was one of the founders of quantum theory, he later turned against it. His famous discussions with his friend Niels Bohr nevertheless helped to clarify the field in its most counter-intuitive aspects. 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 favourite target of attacks and discrimination by the establishment; in 1933 he emigrated to the USA. He was not only a great physicist, but also a great thinker; reading his collection of thoughts about topics outside physics is time well spent.

Ref. 185

[**] For information about the influences of relativity on machine design see the interesting textbook by Van Bladel.

Ref. 186

[***] Indeed, the presently possible measurement precision of $2 \cdot 10^{-13}$ does not allow to discern any changes of the speed of light with the speed of the observer.

However, in the situation described, we obviously have $x \neq \xi$. In other words, the constancy of speed of light implies that $t \neq \tau$, i.e. that *time is different for observers moving relative to each other*. Time is thus not unique.This surprising result, which in the mean time has been confirmed by many experiments, was first stated in detail in 1905 by Albert Einstein. Already in 1895, the discussion of this issue, especially in connection with viewpoint invariance, had been called 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 gravity the theory of *general relativity*. Both fields are full of fascinating and counter-intuitive results. In particular, they show that Galilean physics is wrong at high speeds.

Challenge 422 e

Ref. 188

Ref. 183

Obviously, many people tried to find arguments to avoid the strange conclusion that time differs from observer to observer. But all had to bow to the experimental results. Let us have a look at some of them.

Acceleration of light and the Doppler effect

Light *can* be accelerated. Every mirror does this! 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 methods only change the propagation direction; none has the power to change the speed of light in a vacuum. In short, light is an example of motion which cannot be stopped. Only a few other examples exist. Can you name one?

See page 398

Challenge 423 n

What would happen if we could accelerate light to higher speeds? It would mean that light is made of particles with non-vanishing mass. Physicists call such particles *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. The speed of light 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 the light, slowly, the following night, maybe after giving it an additional push to speed it up.[**]

Physicists have therefore studied 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. 189

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 respect to the receiver, the receiver will observe a *different colour* from that observed by the sender. This result is called the *Doppler effect*. Doppler studied the frequency shift in

---

[*] The most beautiful and simple introduction to 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. Only a century later there are books almost as beautiful, such as the text by Taylor and Wheeler.

Ref. 182

[**] 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.

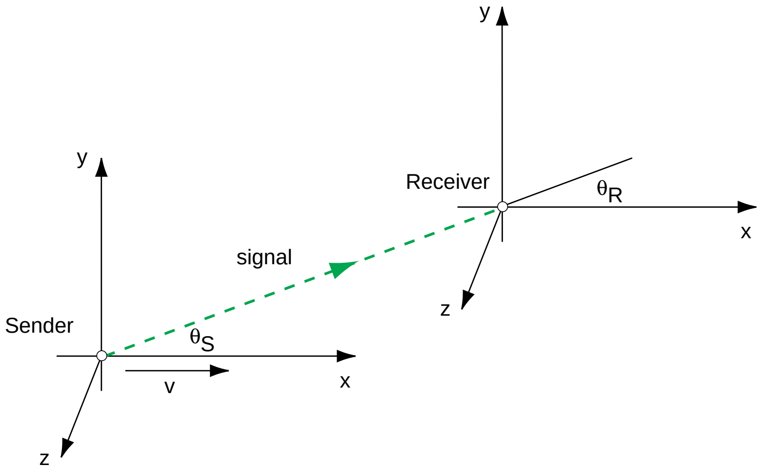


Figure 88 The set-up for the observation of the Doppler effect

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, a moving light source produces a colour at the receiver that is different from colour at the sending source. Simple geometry, starting from the fact that all wave maxima and minima emitted must also be received, leads to the result Challenge 424

$$\frac{\lambda_R}{\lambda_S} = \frac{1}{\sqrt{1-v^2/c^2}}\left(1-\frac{v}{c}\cos\theta_R\right) = \gamma\left(1-\frac{v}{c}\cos\theta_R\right) \quad . \tag{88}$$

Light from an approaching source is thus blue shifted, whereas light from a departing source is red shifted. The first observation of the Doppler effect for light was made by Johannes Stark in 1905 [**] by studying 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. Ref. 190 In contrast to sound waves, a colour effect is also found when the motion is *transverse* to the light signal. (How does the colour change in this case?) Challenge 425 n

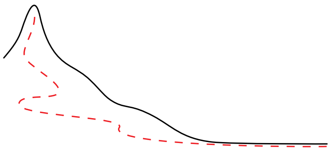
The colour shift is used in many applications. Almost all bodies are mirrors for radio waves. When one enters a building, often the doors open automatically. A little sensor above the door detects the approaching person. Usually, but not always, this is done by measuring the Doppler effect of radio waves emitted by the sensor and reflected by the approaching person. (We will see later that radio waves and light are two sides of the same phenomenon.) Police radar also works in this way.[***]

---

[*] Christian Doppler (1803, Salzburg–1853, Venezia), Austrian physicist.

[**] Johannes Stark (1874–1957), discovered in 1905 the optical Doppler effect in channel rays, and in 1913 the splitting of spectral lines in electrical fields, nowadays called the Stark effect. For both discoveries he received the 1919 Nobel prize for physics. He left his professorship in 1922 and later turned a full-blown national socialist. Member of the NSDAP since 1930, he became known for attacking statements about nature for ideological reasons only, and was thus as a person rightly despised by the academic community.

Challenge 426 n [***] At what speed does a red traffic light appear green?



Motion Mountain

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

The Doppler effect also makes it possible to measure the velocity of light sources. Indeed, it is commonly used to measure the speed of far away stars. In these cases, the Doppler shift is often characterized by the *red shift number z*, defined as

$$z = \frac{\Delta\lambda}{\lambda} = \frac{f_S}{f_R} - 1 = \sqrt{\frac{c+v}{c-v}} - 1 \quad . \tag{89}$$

Challenge 428 n Can you imagine how the number z is determined? Typical values for z found for light sources in the sky range from -0.1 to 3.5, but higher values, up to more than 5, have also been found. Can you determine the corresponding speeds? How can they be so high?

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 would *prove* that time is different for observers moving relative to each other. Why? Time is what we read from our watch. In order to determine whether another watch is synchronized with our own one, we look 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? Why is the conclusion about time differences not possible when the Doppler effect for sound is used? Challenge 429

Can one shoot faster than one's shadow?

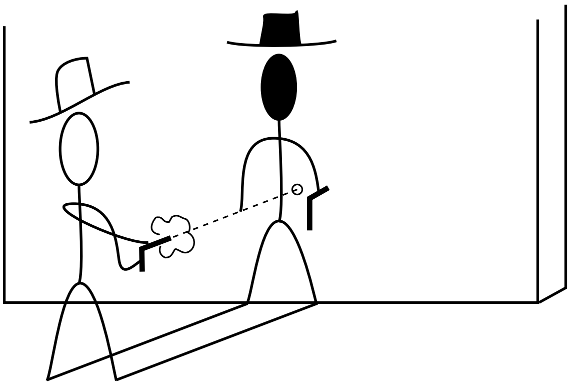


Figure 89 Lucky Luke

To realize what Lucky Luke does in Figure 89, both the bullet and the hand have to move faster than the speed of light. To achieve 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 the speed that can be achieved. This experiment is carried out daily in particle accelerators such as the Large Electron Positron ring, the LEP, of 27 km circumference located partly in France and partly in Switzerland, near Geneva. In that place, 40 MW of electrical power, the same amount used by a small city, Ref. 191 accelerates electrons and positrons to energies of over 16 nJ (104.5 GeV) each. The result is shown in Figure 90: even with these impressive means it is impossible to make 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.*

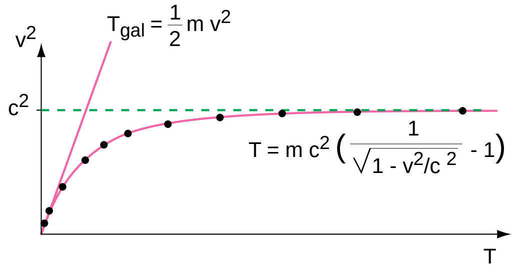


Figure 90 Experimental values (dots) for the electron velocity *v* as function of kinetic energy *T*

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

The 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 15 000 km/s or more, the expression $mv^2/2$ is *not* equal to the kinetic energy T 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. 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, as in Figure 90, it was found to be wrong. Challenge 430 n

The observation of speed of light as a *limit* speed for objects is easily seen to be a consequence of its *constancy*. Bodies that 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 more slowly than something else for *one* observer, it does so for all other observers as well. (Trying to imagine a world in which this would not be so is interesting: funny things would happen, such as things interpenetrating each other.) Challenge 431 n Therefore no object that 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.We also deduce that the maximum speed is the speed of *massless* entities. Challenge 432 Light, all the other types of electromagnetic waves and (probably) neutrinos are the only known examples. Ref. 193

A consequence of the existence of a limit velocity is important: velocities cannot simply be added or subtracted, as we are used to in everyday life. If a train is travelling 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

---

* There are still people who refuse 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 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 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. Ref. 192

velocity is given by

$$v_{se} = \frac{v_{st} + v_{te}}{1 + v_{st} v_{te}/c^2} \qquad (90)$$

Challenge 433 e Note that the result is never larger than c.We will deduce the expression in a moment, from reasoning alone.[*] It has been confirmed by literally all the millions of cases in which it has been checked so far. See page 379

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 used by the observer an *inertial frame of reference*. Examples of inertial observers (or frames) are – for *two* dimensions – those moving on a frictionless ice surface or on the floor inside a smoothly running train or ship; a full example – for all *three* spatial dimensions – is a cosmonaut in an Apollo capsule while travelling between the moon and the earth, as long as the engine is 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, as we will show below, the equivalent:[***]

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

$$v \leqslant c \qquad (91)$$

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

Ref. 196 This experimental statement was checked with high precision by Michelson and Morely[****] in the years from 1887 onwards. It has been confirmed in all subsequent experiments. Therefore the following conclusions can be drawn from it, using various (weak) implicit assumptions which will become clear during the rest of our ascent of Motion Mountain:

Ref. 195 [*] One can also deduce the Lorentz transformation from this expression.

[**] Note that the historical 'principle of relativity' is different from the one described here.

[***] This statement is due to the Dutch physicist Hendrik Antoon Lorentz (Arnhem, 1853– Haarlem, 1928). He 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 for the vacuum describe matter as well, as long as charged point particles are included. He also gave the correct description of the Lorentz force. Outside physics, he was active in the internationalization of scientific collaborations.

[****] 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, was Michelson's friend and long-time collaborator.

- 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 relative to each other. (Are you able to show this?) Challenge 434

Historically, it was the equivalence of all inertial observers which used to be called the principle of relativity. Nowadays this habit is changing, though slowly, mainly because the habit is connected to Poincaré and to Einstein himself. The *essence* of relativity however is the existence of a limit speed.

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 $\mathbf{v}$ relative to 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 of the speed of light in any direction for any two observers means that

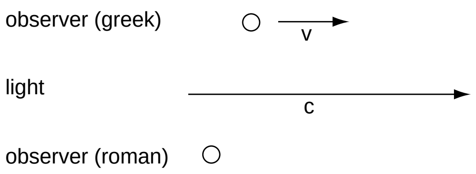


Figure 91 Two inertial observers, using coordinates (t,x) and (τ,ξ), and a beam of light

$$(c\mathrm{d}t)^2-(\mathrm{d}x)^2-(\mathrm{d}y)^2-(\mathrm{d}z)^2=(c\mathrm{d}\tau)^2-(\mathrm{d}\xi)^2-(\mathrm{d}\upsilon)^2-(\mathrm{d}\zeta)^2 \quad . \qquad (92)$$

Assume also that a flash lamp at rest for the Greek observer, thus with $\mathrm{d}\xi=0$, produces two flashes spaced by an interval $\mathrm{d}\tau$. For the Roman observer, the flash lamp moves, so that $\mathrm{d}x=v\mathrm{d}t$. Inserting this into the previous expression, and assuming linearity and speed direction independence for the general case, we find that intervals are relatedby Challenge 435 e

$$\begin{aligned}
\mathrm{d}t &= \gamma(\mathrm{d}\tau+v\mathrm{d}\xi/c^2)=\frac{\mathrm{d}\tau+v\mathrm{d}\xi/c^2}{\sqrt{1-v^2/c^2}} \qquad \text{with} \quad v=\mathrm{d}x/\mathrm{d}t\\
\mathrm{d}x &= \gamma(\mathrm{d}\xi+v\mathrm{d}\tau)=\frac{\mathrm{d}\xi+v\mathrm{d}\tau}{\sqrt{1-v^2/c^2}}\\
\mathrm{d}y &= \mathrm{d}\upsilon\\
\mathrm{d}z &= \mathrm{d}\zeta \quad . \qquad (93)
\end{aligned}$$

These expressions describe how length and time intervals measured by different observers are related. At relative speeds v that are small compared to the velocity of light, such as in everyday life, the time intervals are essentially equal; the *stretch factor* or *relativistic correction* or *relativistic contraction* γ is then equal to 1 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 92.

---

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

The expressions are also strange in another respect. When two observers look at each other, each of them claims to measure 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 on a bicycle. We explore this bizarre result in more detail shortly. Challenge 436

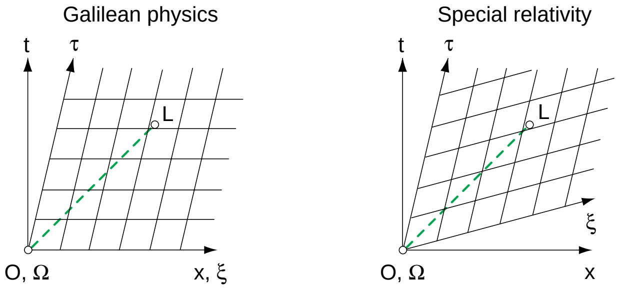


Figure 92 Space-time diagrams for light seen from two different observers

The stretch factor γ is equal to 1 in everyday life and for most practical purposes. The largest value humans have ever achieved is about $2 \cdot 10^5$; the largest observed value in nature is about 10^{12}. Can you imagine their occurrences? Challenge 437 n

Once we know how space and time *intervals* change, we can easily deduce how *coordinates* change. Figures 91 and 92 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, we have

$$\xi = \gamma(x - vt) \quad \text{and} \quad v = \frac{dx}{dt} \quad . \tag{94}$$

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

$$\tau = \gamma(t - xv/c^2) \quad . \tag{95}$$

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 these relations from the equations of electrodynamics, which had contained them, waiting to be discovered, since 1865.** In that year James Clerk Maxwell had published them in order to describe everything electric and magnetic. Ref. 197 See page 383

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

Many alternative formulas for boosts have been explored, such as expressions in which instead of the relative velocity also the relative acceleration of the two observers is included. Ref. 194

---

[*] About Hendrik Antoon Lorentz, see page 209.

[**] The Irishman George F. Fitzgerald had had already discovered the Lorentz transformations in 1889, but had, in contrast to Lorentz, not continued his research in the field.

However, all had to be discarded when compared to experimental results. But before we have a look at such experiments, we continue with a few logical deductions from the boost 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 four 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 manifold is characterized by a simple property: the *space-time interval* di between two nearby events, defined as Challenge 438 Ref. 198

$$\mathrm{d}i^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = c^2\mathrm{d}t^2(1-\frac{v^2}{c^2}) \quad , \tag{96}$$

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 (96) has a simple interpretation. It is the time measured by an observer moving from event (t,x) to event $(t+dt,x+dx)$, the so-called *proper time*, multiplied by c^2. We 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, i.e. free of curvature. In both cases, space is what is measured with a metre rule or with a light ray, and time is what is read from a clock. In both cases, space-time is fundamental; it is and remains the *background* and the *container* of things and events. We *live* in a Minkowski space-time, so to speak. Minkowski space-time exists independently of things. 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. Challenge 439

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 in Figure 92.

Relativity thus forces us to describe motion with space-time. That is interesting, because in space-time, *motion does not exist*. Motion exists only in space. In space-time, nothing moves. For each point particle, space-time contains a world-line. In other words, instead of asking why motion exists, we can equivalently ask why space-time is criss-crossed by world-lines. However, we are still far from answering either question.

Can we travel to the past? – Time and causality

Given that time is different for different observers, does time nevertheless order 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.

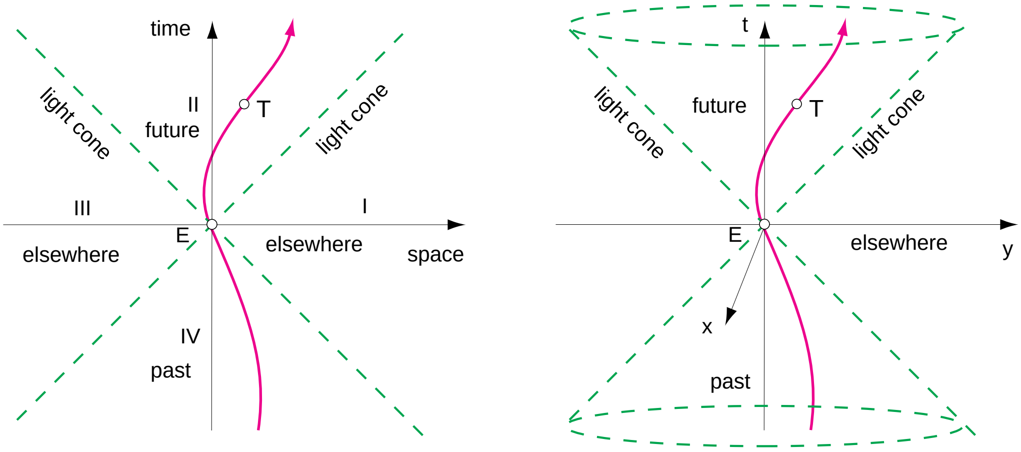


Figure 93 A space-time diagram of an object T seen from an inertial observer O in the case of one and of two spatial dimensions

Sequences of events can clearly be defined if one event is the cause of another. But this can only be the case if energy or signals travel from one event to another at speeds up to the speed of light. Figure 93 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 quadrant II (the *future light cone*). Events in 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 that *can* be ordered with respect to their origin – namely those inside the cones – and those that *cannot* – those outside the cones, happening elsewhere for all observers. 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!

In particular, the past light cone gives the complete set of events that 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 involves transport of energy, and thus cannot travel faster than 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 440

A vector inside the light cone is called *timelike*; one on the light cone is called *lightlike*, and one outside the cone is called *spacelike*. For example, the *world-line* of an observer, i.e. the set of all events that make up its history, consists of timelike events only. In fact, time is the fourth dimension; it expands space to space-time and thus ’completes’ space-time. There is not much more to know about the fourth dimension, or about thinking in four dimensions.

Special relativity thus teaches us that time can be defined *only* because light cones exist. If transport of energy at speeds faster than that of light did exist, time could not be defined. Causality, i.e. the possibility of (partially) ordering events for all observers, is 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 experience which tells that the speed of light is indeed maximal: our memory. If the future could influence the past, we would also be able to *remember* the future. To put it in another way, if the future could influence the past, the second principle of thermodynamics would not be valid, and then our memory would not work.* No other data from everyday life or from experiments provides any evidence that the future can influence the past. In other words, *time travel to the past is impossible*. How the situation changes in quantum theory will be revealed later on. Interestingly, time travel to the future *is* possible, as we will see shortly. Challenge 441 Ref. 199

Curiosities of special relativity

Faster than light: how far can we travel?

How far away from earth can we travel, given that the trip should not last more than a lifetime, say 80 years, and given that we are allowed to use a rocket whose speed can approach the speed of light as closely as desired? Given the time t we are prepared to spend in a rocket, given the speed v of the rocket and assuming optimistically that it can accelerate and decelerate in a negligible amount of time, the distance d we can move away is given by Challenge 442

$$d = \frac{vt}{\sqrt{1-v^2/c^2}} \quad . \tag{97}$$

The distance d is larger than ct already for $v > 0.71c$, and, if v is chosen large enough, it increases beyond all bounds! In other words, relativity itself does *not* limit the distance we can travel, not even that covered in a single second. We 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

$$w = d/t = \frac{v}{\sqrt{1-v^2/c^2}} = \gamma v \quad . \tag{98}$$

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 space-time 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 is made clear by Stephen Blau in a recent pedagogical paper. Ref. 200

[**] Using proper velocity, the relation given in (90) 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 443

$$w_{s\parallel} = \gamma_a \gamma_b (v_a + v_{b\parallel}) \quad \text{and} \quad w_{s\perp} = w_{b\perp} \quad , \tag{99}$$

where the signs $\parallel$ and $\perp$ mean the components in direction of motion and that perpendicular to $\mathbf{v}_a$, respectively. One can in fact write all of special relativity using ‘proper’ quantities, even though this is not done in this text. Ref. 201

Synchronization and aging: can a mother stay younger than her own daughter? – Time travel to the future

In the theory of special relativity time is different for different observers moving relative to each other. This implies that we have to be careful how to synchronize clocks that are far apart, even if they are at rest with respect to each other in an inertial reference frame. For example, if we have two identical watches showing the same time, and if we carry 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. Ref. 202, 203 The time difference for a person or a watch in a plane going around the earth once, at about 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

$$\frac{t}{t'} = \gamma \quad . \tag{100}$$

Also human bodies are clocks; they show the elapsed time, usually called *age*, by various changes in their shape, weight, hair colour, etc. If a person goes on a long and fast trip, on her return she will have aged *less* 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.

This can also be seen as a confirmation of the possibility of time travel to the future. With the help of a fast rocket that comes back to its starting point, we can arrive at local times that we would never have reached within our lifetime. Alas, as has just been said, we can *never* return to the past.*

In short, the question in the title of this section has a positive answer. Can you explain this to a friend, using a space-time diagram and the expression for proper time τ? Challenge 444 This famous result, usually called the *clock paradox* or the *twin paradox*, has also been confirmed in many experiments. Ref. 205 We give a simple example below.

We can also conclude that we cannot synchronize clocks simply by walking, clock in hand, from one place to the next. The correct way to do this is to exchange light signals. Can you describe how? Challenge 445

In summary, only with a clear definition of synchronization can we call two distant events simultaneous. In addition, special relativity shows that simultaneity depends on the observer. This is confirmed by all experiments performed so far.

Length contraction

The length of an object measured by an observer attached to the object is called its proper length. Special relativity makes a simple statement: the length measured by an inertial observer is always smaller or at best equal to the proper length. This result follows directly from the Lorentz transformations. Challenge 446 e

---

Ref. 204 * There are even special books on time travel, such as the well researched text by Nahin. Note that the concept of time travel has to be clearly defined; otherwise one gets into the situation of the clerk who called his office chair a time machine, as sitting on it allows him to get to the future.

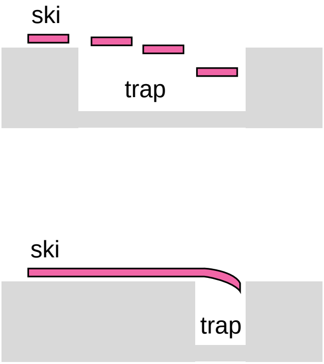


Figure 94 The observations of the trap digger and of the snowboarder

Can a rapid snowboarder fall into a hole that is a bit shorter than his board? Imagine him boarding so fast that the length contraction factor $\gamma = d/d'$ is 4.* For an observer on the ground, the snowboard is four times shorter, and when it passes over the hole, it will fall into it. However, for the boarder, it is the hole which is four times shorter; it seems that the snowboard cannot fall into it.

More careful analysis shows that, in contrast to the observation of hole digger, the snowboarder does not experience the board shape as fixed; while passing over the hole, the boarder observes that the board takes on a parabolic shape and falls into the hole, as shown in Figure 94. Can you confirm this?In other words, shape is *not* an observer invariant concept. (However, rigidity is such a concept, if defined properly; can you confirm this?) Ref. 206 Challenge 448 Challenge 449

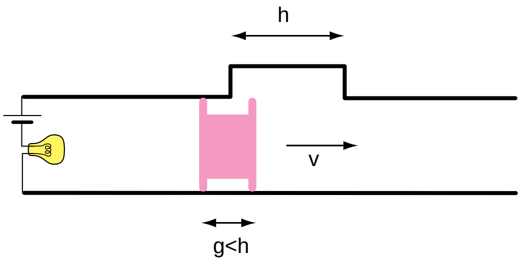


Figure 95 Does the conducting glider keep the lamp lit at large speeds?

The situation becomes more interesting in the case that the snowboard is replaced by a conductive bar and makes electrical contact between the two sides of the hole. As gravity is not needed, the whole arrangement is simplified by turning it on its side, as in Figure 95. 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. 207 Challenge 450

Figure 96 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 the two cars remain the same distance apart. 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. 208 Challenge 451

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 time $t = 0$, as described by an observer at rest Ref. 208

Challenge 447 n * Even the earth contracts in its direction of motion around the sun. Is the value measurable?

with respect to both of them. Both cars contain the same amount of fuel. (Let's forget the rope now.) We easily deduce 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 $t_\mathrm{f} = t_\mathrm{b}$. By using the Lorentz transformations you can deduce for the frame of the freely rolling twins the relation Challenge 452

$$t_\mathrm{b} = \gamma \Delta x v / c^2 + t_\mathrm{f} \quad , \tag{101}$$

which means that the front twin has aged *more* than the back twin. Therefore it seems that if we want to avoid grey hair as much as possible, we should always sit in the back of a bus or train when travelling. Is the conclusion correct?And is it correct to deduce that people on high mountains age faster than people invalleys? Challenge 453 Challenge 454

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 he could achieve in the Olympics. The ideal would be that each foot accelerates instantly to (almost) the speed of light. The highest walking speed can be achieved by taking the second foot off the ground at exactly the same instant at which the first is put down. In the beginning, by 'same instant' the student 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 97; it gives a limit speed for walking of half the speed of light. But then the student 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 is thus only one third of the speed of light. Ref. 209

Is the speed of shadow greater 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 constrains the motion of mass and energy. However, non-material points, non-energy transporting features and 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 214 Challenge 455 n

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 the cone-shaped shock wave around an aeroplane moving faster than the speed of sound. Cerenkov radiation is regularly observed; for example it is the cause of the blue glow of the water in nuclear reactors. Incidentally, the speed of light in matter can be quite low; in the centre of the sun, the speed of light is

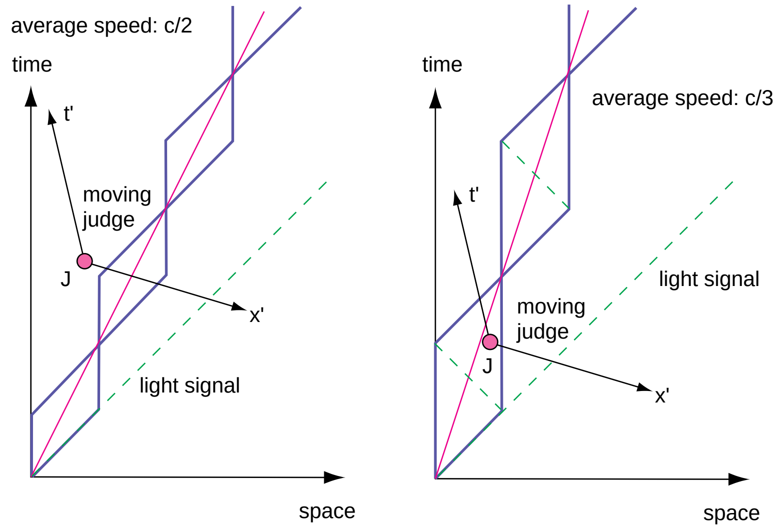


Figure 97 For the athlete on the left, the judge moving in the opposite direction sees both feet off the ground at certain times, but not for the athlete on the right

estimated to be only around 10 km/year, and in the laboratory, for some materials, it has been found to be 0.3 m/s. Ref. 210, 19

In contrast, the following examples show velocities that are genuinely faster than the externally measured velocity of light in vacuum. An example is the point marked X in Figure 98, the point at which scissors cut paper. If the scissors are closed rapidly enough, the point moves faster than light. Similar geometries can also be found in every window frame, and in fact in any device that has twisting parts.

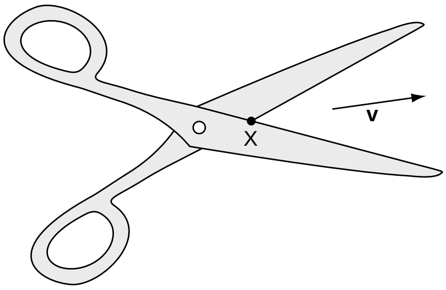


Figure 98 A simple example of motion that is faster than light

Another example of superluminal motion is the speed with which a music record – remember LPs? – disappears into its sleeve, as shown in Figure 99.

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

All these are typical examples of the *speed of shadows*, sometimes also called the *speed of darkness*. Both shadows and darkness can indeed move faster than light. In fact, there is no limit to their speed. Can you find another example? Challenge 456

In addition, there is an ever-increasing number of experimental set-ups in which the phase velocity or even the group velocity of light is higher than c. They regularly make headlines in the newspapers, usually of the type 'light moves faster than light'. This surprising result is discussed in more detail later on. See page 403

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

$$t = l/v - l/c \quad . \tag{102}$$

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

$$v_{\text{appr}} = \frac{l}{t} = \frac{vc}{(c-v)} \quad , \tag{103}$$

which is higher than c for any car velocity v higher than $c/2$. For cars this does not happen too often, but astronomers know a type of bright object in the sky called a *quasar* (a contraction of 'quasi-stellar'), which sometimes emits high-speed gas jets. If the emission is in or near the direction to the earth, the apparent speed is higher than c; such situations are now regularly observed with telescopes. Ref. 211

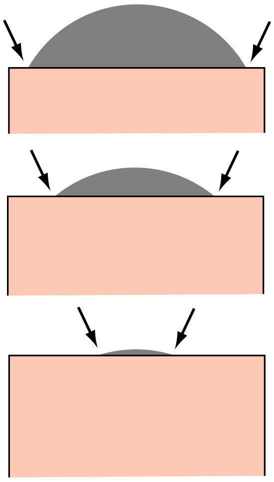


Figure 99 Another example of faster than light motion

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

$$v_{\text{leav}} = \frac{vc}{(c+v)} \quad , \tag{104}$$

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

The story has a final twist. We have just seen that motion faster than light can be observed in several ways. But could an *object* moving faster than light be observed at all? Surprisingly, 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, we can never see it approaching. If at all, tachyons can only be 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, could we notice it. We would first see a flash of light, corresponding to the bang of a plane passing with supersonic speed. Then we would see *two* images of the tachyon, appearing somewhere in space and departing in opposite directions, as can be deduced from Figure 100. Even if one of the two images were coming nearer, it would be getting fainter and smaller. This is, to say the least, rather unusual behaviour. Moreover, if you wanted to look at a tachyon at night, illuminating it with a torch, you would have to turn your head in the direction opposite to the arm with the torch! This requirement also follows from the space-time diagram; are you able to deduce this? Challenge 457 e Nobody has ever seen such phenomena; tachyons do not exist. Tachyons would be strange objects: they would have imaginary mass, they would accelerate when they lose energy, and a zero-energy tachyon would be infinitely fast. But no object with these properties has ever been observed. Worse, as we just saw, tachyons would seem to appear from nothing, defying laws Ref. 212

of conservation; 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 ascent of Motion Mountain. Tachyons therefore cannot be objects in the usual sense. In the second part of our adventure 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 discover.

But the best is still to come. Not only is it impossible to see approaching tachyons; it follows from equation (104) that departing ones seem to move with a velocity *lower* than the speed of light. In other words, we just found that if we ever see something move faster than light, it can be anything *but* a tachyon!

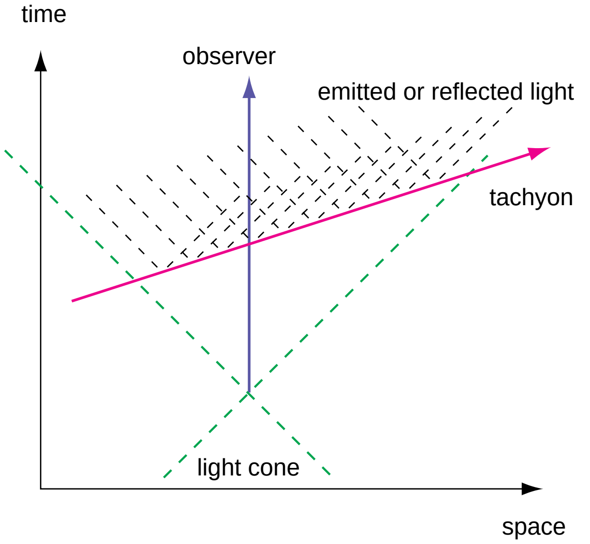


Figure 100 Hypothetical space-time diagram for tachyon observation

Parallel to parallel is not parallel – Thomas rotation

Relativity has strange consequences indeed. Even though any two observers can keep a stick parallel to the stick of another, even if they are moving with respect to each other, something strange results. A chain of sticks for which any two adjacent ones are parallel to each other will *not* ensure that the first and the last sticks are parallel. In particular, this is *never* the case if the motions of the various observers are in different directions, as is the case when the velocity vectors form a loop.

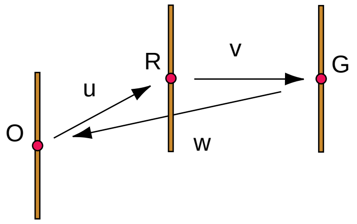


Figure 101 If O’s stick is parallel to R’s, and R’s is parallel to G’s, then O’s stick and G’s stick are not

This surprising result is purely relativistic, and thus occurs *only* in the case of speeds comparable to that of light. It results from the fact that in general, concatenations of pure boosts do not give a pure boost, but a boost and a rotation. Ref. 213

For example, if we walk with a stick in a fast circle, always keeping the stick parallel to the direction it had just before, at the end of the circle the stick will have an angle with respect to the original direction. Similarly, the axis of a rotating body circling a second body will *not* be pointing in the same direction after one turn, if the orbital velocity is comparable to that of light. This effect is called *Thomas precession*, after Llewellyn Thomas, who discovered it in 1925, a full 20 years after the birth of special relativity. It had escaped the attention of dozens of other famous physicists. Thomas precession is important in the inner working of atoms and we will return to it in that section of our adventure.

A never-ending story: temperature and relativity

Not everything is settled in special relativity. Do you want a problem to solve? Just deduce how temperature changes from one frame of reference to another, and publish the result. Challenge 458 e 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

$$T = T_o\sqrt{1-v^2/c^2} \quad , \tag{105}$$

thus always yielding lower values. But others maintain that T and T_o should be interchanged Ref. 214 in this expression. Even powers other than the simple square root have been proposed.

The origin of these discrepancies is simple: temperature is only defined for equilibrium situations, i.e. for baths. But a bath for one observer is usually not a bath for the other; for low speeds, a moving observer sees *almost* a bath; but at higher speeds the issue becomes tricky. The resulting temperature change may even depend on the energy range measured. So far, there do not seem to be any experimental observations that would allow the issue to be settled. Realizing such a measurement is a challenge for future experiments.

Relativistic mechanics

As the speed of light is constant and velocities do not add up, we need to rethink the definition of mass, as well as the definitions of momentum and energy.

Mass in relativity

See page 61 In Galilean physics, the mass ratio between two bodies was defined using collisions; it was given by the negative inverse of the velocity change ratio

$$\frac{m_2}{m_1} = -\frac{\Delta v_1}{\Delta v_2} \quad . \tag{106}$$

However, experiments show that the expression changes for speeds near that of light. In Challenge 459 fact, thinking alone can show this; are you able to do so?

The solution to this issue was found by Albert Einstein. He discovered that in collisions, the two Galilean conservation theorems $\sum_i m_i \mathbf{v}_i = \text{const}$ and $\sum_i m_i = \text{const}$ had to be changed into

$$\sum_i \gamma_i m_i \mathbf{v}_i = \text{const} \tag{107}$$

and

$$\sum_i \gamma_i m_i = \text{const} \quad . \tag{108}$$

These expressions, which are correct throughout the rest of our ascent of Motion Mountain, imply, among other things, that teleportation is *not* possible in nature. (Can you confirm

Challenge 460 n 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 we did not know the value of the relativistic correction, we could find it by a simple deduction from the collision shown in Figure 102. In the first frame of reference we have $\gamma_1 mv = \gamma_2 MV$ and $\gamma_1 m + m = \gamma_2 M$. From the observations of the second frame of reference we deduce that V composed with V gives v,in other words Challenge 461

$$v = \frac{2V}{1+V^2/c^2} \quad . \tag{109}$$

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

$$\gamma_v = \frac{1}{\sqrt{1-v^2/c^2}} \quad . \tag{110}$$

With this expression, and a generalization of the situation of Galilean physics, the *mass* ratio between two colliding particles is defined as the ratio

$$\frac{m_1}{m_2} = -\frac{\Delta(\gamma_2 v_2)}{\Delta(\gamma_1 v_1)} \quad . \tag{111}$$

(We do not give the generalized mass definition mentioned in Galilean mechanics and based on acceleration ratios, because it contains some subtleties that 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 may have.* In this way, the concept of mass remains a number characterizing the difficulty of accelerating a body, and it can still be used for *systems* of bodies as well.

Following the example of Galilean physics, we call the quantity

$$\mathbf{p} = \gamma m \mathbf{v} \tag{112}$$

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.

For low speeds, or $\gamma \approx 1$, the value is the same as that of Galilean physics. But for high speed, momentum increases faster than velocity, as it tends to infinity when approaching light speed. Momentum is thus not proportional to velocity any more.

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 billiards.

Challenge 462 e * The results below also show that $\gamma = 1 + T/mc^2$, where T is the kinetic energy ofa particle.

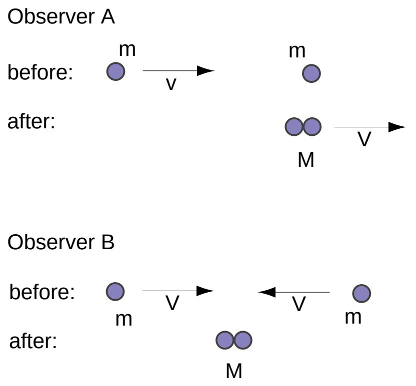


Figure 102 An inelastic collision of two identical particles seen from two different inertial frames of reference

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, you can find with a bit of dexterity that

Challenge 463

$$\tan\theta\tan\varphi = \frac{2}{\gamma+1} \quad . \tag{113}$$

In other words, the sum $\theta+\varphi$ is *smaller* than a right angle in the relativistic case. Relativistic speeds thus completely change the game of pool. Every accelerator physicist knows this, because 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 were wrong, most of these detectors would not work, as they would miss most of the particles after the collision, as shown in Figure 104.

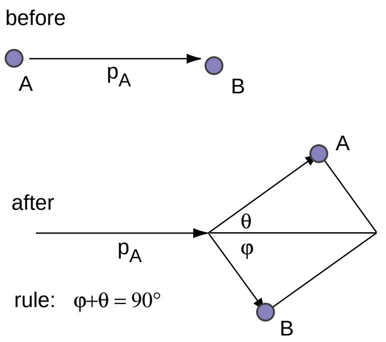


Figure 103 A useful rule for playing non-relativistic pool

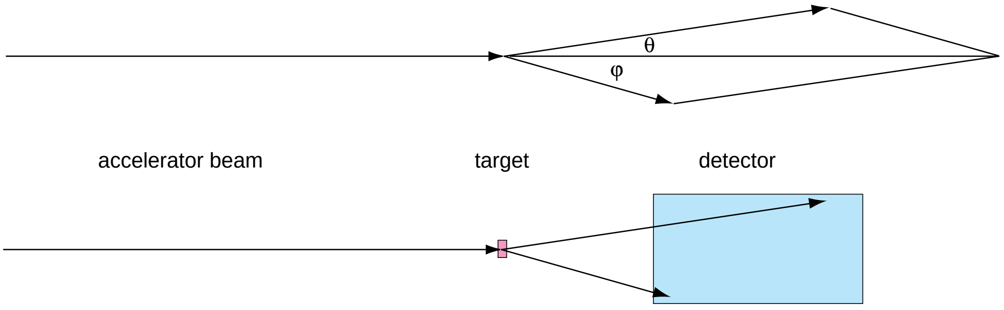


Figure 104 The dimensions of detectors in particle accelerators are based on the relativistic pool angle rule

Mass is concentrated energy

Let us go back to the simple collinear collision of Figure 102. What is the mass M of the final system? A somewhat boring calculation shows that Challenge 464

$$M/m = \sqrt{2(1+\gamma_V)} > 2 \quad . \tag{114}$$

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 $\sum_{\mathrm{i}} \gamma_{\mathrm{i}} m_{\mathrm{i}}$ 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

$$E = \gamma m c^2 \quad , \tag{115}$$

applying it both to the total system and to each component. The conservation of the corrected mass can then be read as the conservation of energy, simply without the factor c^2. In the example of the two identical masses sticking to each other, the two 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} of a body *at rest* and its mass m are related by

$$E_{\mathrm{o}} = mc^2 \tag{116}$$

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 kinetic energy T is then given by

$$T = \gamma m c^2 - mc^2 = \frac{1}{2} m v^2 + \frac{1 \cdot 3}{2 \cdot 4} m \frac{v^4}{c^2} + \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \frac{v^6}{c^4} + \ldots \tag{117}$$

which reduces to the Galilean value only for low speeds. In other words, special relativity says that every mass has energy, and that every form of energy in 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 are the same before and after the explosion, but the combined mass 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, we should be careful to 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 465 n

The mass-energy relation (115) means the death of many science fiction fantasies. It implies that there are *no* undiscovered sources of energy on or near earth. If such sources existed, they would be measurable through their mass. Many experiments have looked for, and are still looking for, such effects with a negative result. 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, because of the large conversion factor involved. Indeed, for any chemical reaction, bond energies are about 1 aJ (6 eV) per bond; this gives a weight change of the order of one part in 10^{10}, too small to be measured by weighing people or mass differences between food and excrement. 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. David Pritchard's group has developed Penning traps that allow masses to be determined from the measurement of frequencies; the attainable precision of these cyclotron resonance experiments is sufficient to confirm $\Delta E_o = \Delta mc^2$ for chemical bonds. In future, increased precision will even allow precise bond energies to be determined in this way. Since binding energy is often radiated as light, one can say that these modern techniques make it possible to *weigh* light. Ref. 215

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, we get again that the *rest energy* of the object is Challenge 466

$$E_o = mc^2 \quad . \tag{118}$$

In summary, collisions and any other physical processes need relativistic treatment whenever the energies involved are a sizeable fraction of the rest energies.

It is worth noting that 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 μJ. Which of the two systems is relativistic? Challenge 467

How are energy and momentum related? The definitions of momentum (112) and energy (115) lead to two basic relations. First of all, their magnitudes are related by Challenge 468

$$m^2c^4 = E^2 - p^2c^2 \tag{119}$$

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

$$\mathbf{p} = \frac{E}{c^2}\mathbf{v} \quad , \tag{120}$$

---

* For example, in the universe there may still be some extremely diluted, yet undiscovered, form of energy, called *dark matter*. It is predicted from (quite difficult) mass measurements. The issue has not been finally resolved.

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 our ascent of the Motion Challenge 469 Mountain, including the following situation.

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.

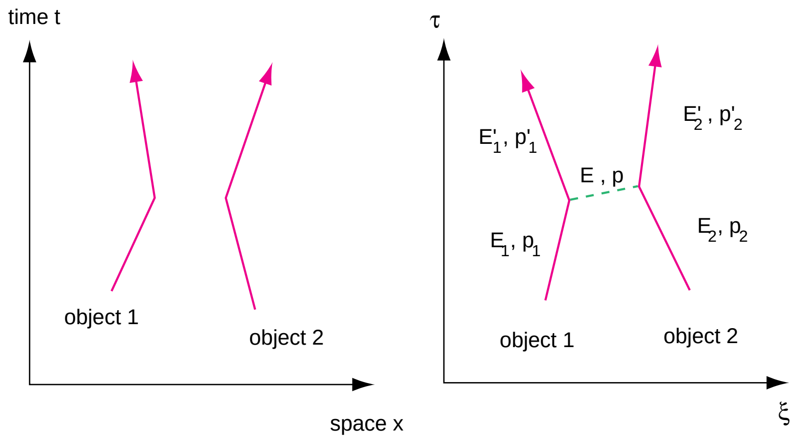


Figure 105 Space-time diagram of a collision for two observers

In everyday life an *impact*, i.e. a short distance interaction, is the event at which both objects change momentum. But the two colliding objects are located at *different* points when this happens. A collision is therefore described by a space-time diagram such as the one Ref. 216 in Figure 105, 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.

The right-hand side of Figure 105 shows the same process seen from another, Greek, frame of reference. The Greek observer says that the first object has changed its momentum *before* the second one. That would mean that there is a short interval when momentum and energy are *not* conserved!

The only way to 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 a prime after the collision, the unknown mass obeys Challenge 470

$$m^2c^4 = (E_1 - E_1')^2 - (p_1 - p_1')^2c^2 = 2m_1^2c^4 - 2E_1E_1'\left(\frac{v_1v_1'}{c^2} + 1\right) < 0 \quad . \qquad (121)$$

Here all quantities are taken to be positive numbers. This is a strange result, because a negative number means that the unknown 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 motion that is faster than light! We will see later that collisions are indeed the *only* processes where tachyons play a role in nature. Since the exchanged objects appear only during collisions, never on their own, they are called *virtual* objects, to distinguish them from the usual, *real* objects, which can move freely without restriction.* We will study their properties later on, in the part of the text on quantum theory. Only virtual objects may be tachyons. Real objects are always *bradyons*, or objects moving slower than light. Note that tachyons do not allow transport of energy faster than light, and 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 471

There is an additional secret hidden in collisions. In the right-hand side of Figure 105, 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 have encountered in our adventure. We will return to the topic in detail in the part of the text on quantum theory. Challenge 472 See page 550

Studying quantum theory 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 when a mountain keeps the trees on it in place, streams of virtual photons are continuously exchanged. This is one of the strange ways in which we will need to describe nature.

Systems of particles: no centre of mass

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

Figure 106 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 centre of mass. In other words, the centre of mass is not an observer-invariant concept. We can also deduce from the figure that the concept only makes sense for systems whose components move relative to each other with *small* velocities. For other cases, it is not uniquely definable. Will this hinder us in our ascent of the Motion Mountain? No. We are more interested in the motion of single particles than that of composite objects or systems. Ref. 217

Why is most motion so slow?

---

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

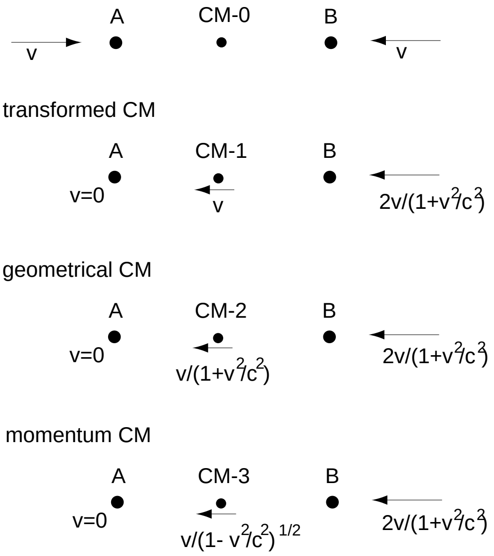


Figure 106 There is no way to define a centre of mass

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 per cent of the speed of light, is a difference noted. Most such situations are microscopic. We have already mentioned the electrons inside a television tube or inside accelerators. Another example is the particles making up cosmic radiation, which produced so many of the mutations that are the basis of evolution of animals and plants on this planet. Later we will discover that the particles involved in radioactivity are also relativistic.

But why don't we observe any rapid *macroscopic* bodies? Moving bodies with relativistic velocities, including observers, have a property not found in everyday life; when they are involved in a collision, part of their energy is converted into new matter via $E = \gamma mc^2$. In the history of the universe this has happened so many times that practically all the bodies still in relativistic motion are microscopic particles.

A second reason for the disappearance of rapid relative motion is radiation damping. Can you imagine what happens to charges during collisions or to charges in a bath oflight? Challenge 473

In short, almost all matter in the universe moves with small velocity relative to other matter. The few known counter-examples are either very old, such as the quasar jets mentioned above, or stop after a short time. The huge energies necessary for macroscopic relativistic motion are still found, e.g. in supernova explosions, but cease to exist after only a few weeks. Therefore the universe is mainly filled with slow motion because it is *old*. We will determine its age shortly. See page

The mass energy equivalence formula

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

$$E = \gamma mc^2 \tag{122}$$

which is often called the most famous formula of physics. Arguably, the formula could have been discovered thirty years earlier from the theory of electromagnetism. Einstein was thus lucky that nobody deduced the result before him. In fact, there is one exception. In 1903 and 1904, before Einstein's first relativity paper, a little known Italian engineer, Olinto

De Pretto, was the first to calculate, discuss and publish the energy value $E = mc^2$.* As an engineer, De Pretto did not pursue the topic further. On the other hand, it might even be that Einstein got the idea for the formula from De Pretto, possibly through his friend Michele Besso or other Italian-speaking friends he met when he visited his parents, who were living in Italy at the time. Of course, the merits of Einstein are not affected.

In the 1970s history repeated itself: a simple relation between the gravitational acceleration and the temperature of the vacuum was discovered, even though it could have been deduced 50 years earlier. Even a number of similar and preceding results were found in the
Challenge 474 n libraries. Could other simple relations be hidden in modern physics?

Four-vectors

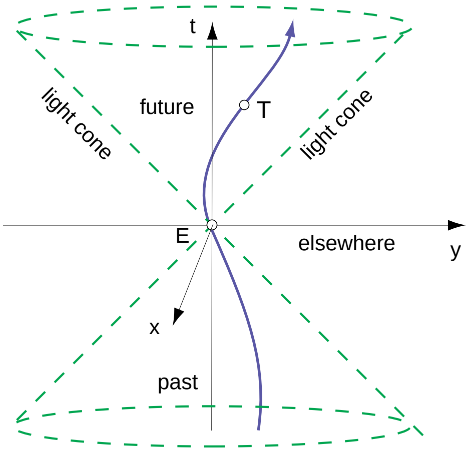


Figure 107 The space-time diagram of a moving object T

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. To describe events with precision, we use event coordinates, also called *4-coordinates*. These are written as

$$\mathbf{x} = (ct, \mathbf{r}) = (ct, x, y, z) = x^i \quad . \tag{123}$$

In this way, an event is a point in four-dimensional space-time, and is described by four coordinates. The coordinates are called the zeroth, namely time $x^0 = ct$, the first, usually called $x^1 = x$, the second, $x^2 = y$, and the third, $x^3 = z$. One can then define a *distance d* between events as the length of the difference vector. In fact, one usually uses the square of the length, to avoid all those nasty square roots. In special relativity, the magnitude ('squared length') of a vector is always defined through

$$\mathbf{xx} = {x_0}^2 - {x_1}^2 - {x_2}^2 - {x_3}^2 = ct^2 - x^2 - y^2 - z^2 = x_a x^a = \eta_{ab} x^a x^b = \eta^{ab} x_a x_b \quad . \tag{124}$$

---

* Umberto Bartocci, mathematics professor of the University of Perugia in Italy, published the details of this surprising story in several papers. The full account is found in his book U. BARTOCCI, *Albert Einstein e Olinto De Pretto: la vera storia della formula più famosa del mondo*, Ultreja, Padova, 1998.

In this equation we have introduced for the first time two notations that are useful in relativity. First of all, we sum over repeated indices. In other words, $x_a x^a$ means the sum over all products $x_a x^a$ for each index a, as just used above. Second, for every 4-vector **x** we distinguish two ways to write the coordinates, namely coordinates with superscripts and coordinates with subscripts. (In three dimensions, we only use subscripts.) They are related by the following general relation

$$x_a = \eta_{ab} x^b = (ct, -x, -y, -z) \quad , \tag{125}$$

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

$$\eta^{ab} = \eta_{ab} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad . \tag{126}$$

Don't panic; this is all, and it won't get more difficult! We now go back to physics.

The magnitude of a position or distance vector, also called the space-time *interval*, is essentially the proper time times c^2. The *proper time* is the time shown by a clock moving in a straight line and constant velocity from the starting point to the end point in space-time. The difference from the usual 3-vectors is that the magnitude of the interval can be positive, negative, or even zero. For example, if the start and end points in space-time require motion with the speed of light, the proper time is zero, as indeed is required for null vectors. If the motion is slower than the speed of light, the squared proper time is positive, and the distance is timelike. For negative intervals, and thus imaginary proper times, the distance is spacelike.**

Now we are ready to calculate and measure motion in four dimensions. The measurements are based on one central idea. 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 just mentioned *proper time* τ, which is defined as the time shown by a clock attached to the object. In relativity, motion and change are always measured with respect to clocks attached to the moving system. In particular, *relativistic velocity* or *4-velocity* **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

$$\mathbf{u} = \mathrm{d}\mathbf{r}/\mathrm{d}\tau \quad . \tag{127}$$

Using $\mathrm{d}t = \gamma\,\mathrm{d}\tau$ and thus

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma\frac{\mathrm{d}x}{\mathrm{d}t} \quad \text{where as usual} \quad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \quad , \tag{128}$$

---

* Note that 30% of all physics textbooks use the negative of η as metric, the so-called *spacelike convention*, and thus have negative signs in this definition. In this text, like in 70% of all physics texts, we use the *timelike convention*.

** In the latter case, the negative of the magnitude, which then is a positive number, is called the squared *proper distance*. The proper distance is the length measured by an odometer as the object moves along that distance.

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

$$u^0 = \gamma c \quad , \quad u^{\mathrm{i}} = \gamma v_{\mathrm{i}} \quad \text{or} \quad \mathbf{u} = (\gamma c, \gamma \mathbf{v}) \quad . \tag{129}$$

For small velocities we have $\gamma \approx 1$, and then the last three components of the 4-velocity are those of the usual, Galilean 3-velocity. For the magnitude of the 4-velocity $\mathbf{u}$ we find $\mathbf{uu} = u_a u^a = \eta_{ab} u^a u^b = c^2$, 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.*

Challenge 476 n Note that the magnitude of a 4-vector can be zero even though all components of such so-called *null* vectors are different from zero. Which motion has a null velocity vector?

Similarly, the *relativistic acceleration* or *4-acceleration* $\mathbf{b}$ of a body is defined as

$$\mathbf{b} = \mathrm{d}\mathbf{u}/\mathrm{d}\tau = \mathrm{d}\mathbf{x}^2/\mathrm{d}\tau^2 \quad . \tag{131}$$

Ref. 219 Using $\mathrm{d}\gamma/\mathrm{d}\tau = \gamma \mathrm{d}\gamma/\mathrm{d}t = \gamma^4 \mathbf{va}/c^2$, we get the following relations between the four components of $\mathbf{b}$ and the 3-acceleration $\mathbf{a} = \mathrm{d}\mathbf{v}/\mathrm{d}t$:

$$b^0 = \gamma^4 \frac{\mathbf{va}}{c} \quad , \quad b^i = \gamma^2 a_i + \gamma^4 \frac{(\mathbf{va}) v_i}{c^2} \quad . \tag{132}$$

Challenge 477 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 it does depend on the value of the 3-acceleration $\mathbf{a}$. (What is the connection between 4-acceleration and 3-acceleration for a an observer moving with the same speed as the object?)We note that 4-acceleration lies *outside* the light cone, i.e. that it is a spacelike 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 velocities involved are low, even the highest accelerations can be treated with Galilean methods.

---

* In general, a *4-vector* is defined as a quantity (h_0, h_1, h_2, h_3), which transforms as

$$\begin{aligned} h'_0 &= \gamma(h_0 - v h_1/c) \\ h'_1 &= \gamma(h_1 - v h_0/c) \\ h'_2 &= h_2 \\ h'_3 &= h_3 \end{aligned} \tag{130}$$

Challenge 475 when changing from one inertial observer to another moving with a relative velocity v in x direction; the corresponding generalization for the other coordinates are understood. Can you deduce the addition theorem (90) from this definition, applying it to 4-velocity?

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

$$\mathbf{J} = \mathrm{d}\mathbf{b}/\mathrm{d}\tau = \mathrm{d}^2\mathbf{u}/\mathrm{d}\tau^2 \quad . \tag{133}$$

Challenge 478 For the relation with the 3-jerk $\mathbf{j} = \mathrm{d}\mathbf{a}/\mathrm{d}t$ we then get

$$\mathbf{J} = (J^o, J^i) = \left(\frac{\gamma^5}{c} (\mathbf{jv} + a^2 + 4\gamma^2 \frac{(\mathbf{va})^2}{c^2}) \quad , \quad \gamma^3 j_i + \frac{\gamma^5}{c^2} ((\mathbf{jv}) v_i + a^2 v_i + 4\gamma^2 \frac{(\mathbf{va})^2 v_i}{c^2} + 3(\mathbf{va}) a_i) \right) \tag{134}$$

Challenge 479 which we will use later on. Surprisingly, $\mathbf{J}$ does not vanish when $\mathbf{j}$ vanishes.Why not?

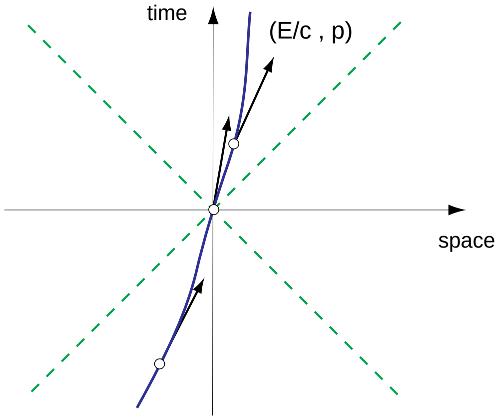


Figure 108 Energy-momentum is tangent to the world line

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

$$\mathbf{P} = m\mathbf{u} \tag{135}$$

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

$$\mathbf{P} = (\gamma mc, \gamma m\mathbf{v}) = (E/c, \mathbf{p}) \quad . \tag{136}$$

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, also called *momenergy*, like the 4-velocity, is *tangent* to the world line of a particle. This follows directly from the definition, since

$$(E/c, \mathbf{p}) = (\gamma mc, \gamma m\mathbf{v}) = m(\gamma c, \gamma \mathbf{v}) = m(\mathrm{d}t/\mathrm{d}\tau, \mathrm{d}\mathbf{x}/\mathrm{d}\tau) \quad . \tag{137}$$

The (square of the) length of momenergy, namely $\mathbf{PP} = \eta_{ab}P^aP^b$, is by definition the same for all inertial observers and found to be

$$E^2/c^2 - p^2 = m^2c^2 \quad , \tag{138}$$

thus confirming a result given above. We have already mentioned that energies or situations are called *relativistic* if the kinetic energy $T = E - E_\mathrm{o}$ is not negligible when compared to the rest energy $E_\mathrm{o} = mc^2$. 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. (What is their energy–momentum relation?) Challenge 480

Note that 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 of calling 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'. Ref. 218 This last statement is more at the intellectual level of the tabloid press, and not worthy of any motion expert.

We note that 4-force $\mathbf{K}$ is defined as

$$\mathbf{K} = \mathrm{d}\mathbf{P}/\mathrm{d}\tau = m\mathbf{b} \tag{139}$$

and that, therefore, contrary to an often heard statement, force remains mass times acceleration in relativity. From the definition of $\mathbf{K}$ we deduce the relation with 3-force Ref. 219, 220 $\mathbf{F} = \mathrm{d}\mathbf{p}/dt = m\mathrm{d}(\gamma\mathbf{v})/dt$, namely*

$$\mathbf{K} = (K^\mathrm{o}, K^\mathrm{i}) = (\gamma^4 m\mathbf{va}/c, \gamma^2 m a_\mathrm{i} + \gamma^4 v_\mathrm{i}\frac{m\mathbf{va}}{c^2}) = (\frac{\gamma}{c}\frac{\mathrm{d}E}{dt}, \gamma\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}) = (\gamma\frac{\mathbf{Fv}}{c}, \gamma\mathbf{F}) \quad . \tag{140}$$

∗ Some authors define 3-force as $\mathbf{F} = \mathrm{d}\mathbf{p}/\mathrm{d}\tau$; then $\mathbf{K}$ looks slightly different. In any case, 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 Challenge 481 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 proportional to 3-velocity, but parallel to it.

Challenge 482 Also the 4-force, like the 4-acceleration, is orthogonal to the 4-velocity.The meaning of the zeroth component of the 4-force can be easily recognized: it is the *power* required to accelerate the object. But, since force is not an important concept in physics, we now turn to a different topic. See page 107

Rotation in relativity

If at night we turn around our 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?

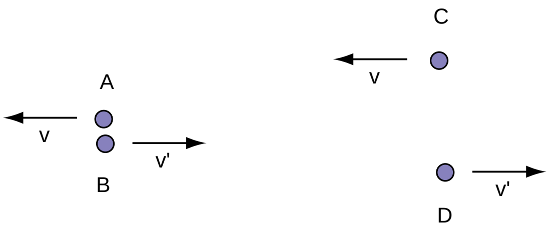


Figure 109 On the definition of relative velocity

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 that 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 higher than the speed of light. We encountered a first example above, when discussing the car in the tunnel, and we will encounter a few additional examples shortly. See page 219 See page 243

With this clarification, we can now have a short look at *rotation* in relativity. The first question is how lengths and times change in a rotating frame of reference. You may want to check that an observer in a rotating frame agrees with a non-rotating colleague on the radius of a rotating body; however, both find that the rotating body has a *different circumference* from before it started rotating.Sloppily speaking, the value of π *changes* for rotating observers: it increases with rotation speed. Challenge 483

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

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 484

We mention that 4-angular momentum is defined naturally as

$$l^{ab} = x^a p^b - x^b p^a \quad . \tag{141}$$

In other words, 4-angular momentum is a *tensor*, not a vector, as shown by its two indices. Angular momentum is also obviously conserved in special relativity, so that there are no surprises on this topic. Challenge 485

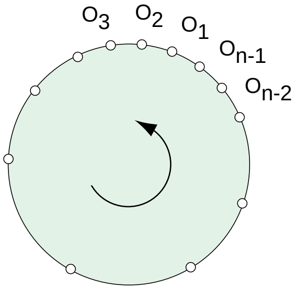


Figure 110 Observers on a rotating object

As usual, 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 the rest mass. You may want to calculate the fraction for the earth and the sun.It is not large. Challenge 486 By the way, how would you determine whether a small particle, too small to be seen, is rotating? Challenge 487

The action of a free particle

If we want to describe relativistic motion of a free particle with an extremal principle, we need a definition of the action. We already know that physical action measures the change occurring in a system. See page 132 For an inertially moving or free particle, the only change is the ticking of its proper clock. As a result, the action of a free particle will be proportional to the elapsed proper time. In order to get the standard unit of energy times time, or Js, for the action, the first guess for the action of a free particle is

$$S = -mc^2 \int_{\tau_1}^{\tau_2} \mathrm{d}\tau \quad , \tag{142}$$

where τ is the proper time along its path. This is indeed the correct expression. It shows that the proper time is maximal for straight-line motion with constant velocity. Can you confirm this? Challenge 488 All particles move in such a way that their proper time is maximal. In other words, we again find that in nature things change as little as possible. Nature is like a wise old man: its motions are as slow as possible. If you prefer, they are maximally effective.

The action can also be written in more complex ways, in order to frighten the hell out of readers. These other, equivalent ways to write it prepare for the future, in particular for general relativity:

$$S = \int L\,\mathrm{d}t = -mc^2 \int_{t_1}^{t_2} \frac{1}{\gamma}\,\mathrm{d}t = -mc \int_{\tau_1}^{\tau_2} \sqrt{u_a u^a}\,\mathrm{d}\tau = -mc \int_{s_1}^{s_2} \sqrt{\eta^{ab} \frac{\mathrm{d}x_a}{\mathrm{d}s} \frac{\mathrm{d}x_b}{\mathrm{d}s}}\,\mathrm{d}s \quad , \tag{143}$$

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

$$\eta^{ab} = \eta_{ab} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad . \tag{144}$$

You can easily confirm the form of the action by deducing the equation of motion with the usual procedure. Challenge 489

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 general principle is also valid for particles under the influence of gravity, as we will see in the section on general relativity, and under the influence of electric or magnetic interactions. In fact, it is valid in all cases. In nature, *proper time is always maximal*. Above, we saw that the action measures the change going on in a system. Minimizing proper time is the way that nature minimizes change. We thus again find that nature is the opposite of a Hollywood movie; nature changes in the most economical way possible. Speculating on the deeper meaning of this result is left to your personal preferences; enjoy it!

See page 132

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 do agree on the position, shape and orientation of the light cone at a point. The light cones at each point thus are the basic physical 'objects' with which space-time is described in the theory of relativity. Given the importance of light cones, we might ask if inertial observers are the only ones that observe the same light cones. Interestingly, it turns out that there are *other* such observers.

The first group of these additional observers is made up of those using different units of measurement, namely units in which all time and length intervals are multiplied by a *scale factor* λ. The transformations among these points of view are given by

$$x_a \mapsto \lambda x_a \tag{145}$$

and are called *dilations*.

A second type of additional observers are found by applying the so-called *special conformal transformations*. They are combinations of an *inversion*

$$x_a \mapsto \frac{x_a}{x^2} \tag{146}$$

with a *translation* by a vector b_a, namely

$$x_a \mapsto x_a + b_a \quad , \tag{147}$$

and a second inversion. This gives for the expression for the special conformal transformations

$$x_a \mapsto \frac{x_a + b_a x^2}{1 + 2b_a x^a + b^2 x^2} \quad \text{or} \quad \frac{x_a}{x^2} \mapsto \frac{x_a}{x^2} + b_a \quad . \tag{148}$$

These transformations are called *conformal* because they do not change angles of (infinitesimally) small shapes, as you may want to check.The transformations thus leave the *form* (of infinitesimally small objects) unchanged. For example, they transform infinitesimal circles

Challenge 491

Challenge 490

* For the massless neutrinos, the action does not work. Why? Can you find an alternative?

into infinitesimal circles. They are called *special* because the full *conformal group* includes the dilations and the inhomogeneous Lorentz transformations as well.*

The way in which special conformal transformations leave light cones invariant is rather 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, 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, as it does not fix a scale factor. As expected, matter is necessary to do so. Indeed, (special) conformal transformations are not symmetries of situations containing matter. Only 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. Challenge 493

As a final remark, conformal invariance includes inversion symmetry. Inversion symmetry means that the large and small scales of a vacuum are related. It thus seems as if the constancy of the speed of light is related to the existence of inversion symmetry. This mysterious connection gives us a glimpse of the adventures we will encounter in the third part of our ascent of Motion Mountain. Conformal invariance turns out to be an important property of will produce many incredible surprises.**

Challenge 492 * The set of all *special* conformal transformations forms a group with four parameters;adding dilations and the inhomogeneous Lorentz transformations one gets fifteen parameters for the *full* conformal group. The conformal group is locally isomorphic to SU(2,2) and to the simple group SO(4,2); these concepts are explained in Appendix D. Note that all this is true only for *four* space-time dimensions; in *two* dimensions, the other important case, especially in string theory, the conformal group is isomorphic to the group of arbitrary analytic coordinate transformations, and is (thus) infinite-dimensional. See page 904

** 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 494

However, since all particles observed up to now have masses that 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. The story gets even more interesting when one or both of the observers are accelerating.

One sometimes hears that special relativity cannot be used to describe accelerating observers. That is wrong: the argument would imply that also Galilean physics could not be used for accelerating observers, in contrast to everyday experience. Special relativity's only limitation is that it cannot be used in non-flat, i.e. curved, space-time. Accelerating bodies do exist in flat space-times, and therefore can be discussed in special relativity.

Ref. 221 As an appetizer, let us see what an accelerating, Greek, observer says about the clock of 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 are 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 Greek observer, and call them τ_1 and τ_2.* We see 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

$$\frac{\Delta\tau}{\Delta t} = \frac{\mathrm{d}\tau}{\mathrm{d}t} = \sqrt{1 - v^2/c^2} = \frac{1}{\gamma_v} \qquad (149)$$

Challenge 495 as we are now used to.We find again that moving clocks run slow.

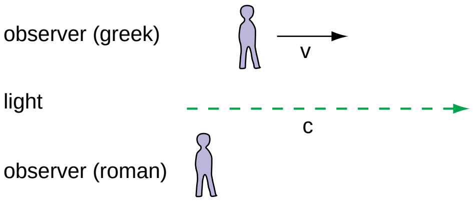


Figure 111 Simplified situation for an inertial and an accelerated observer

Ref. 221 For accelerated motions, the differential version of the reasoning is necessary. In other words, the Roman/Greek clock rate ratio is again $\mathrm{d}\tau/\mathrm{d}t$, and τ and $\tau + \mathrm{d}\tau$ are calculated in the same way as just defined from the times t and $t + \mathrm{d}t$. Assume again that the Greek observer moves along $\mathbf{x}(t)$, as measured by the Roman one. We find directly that

$$\tau = t - \mathbf{x}(t)\mathbf{v}(t) \qquad (150)$$

and thus

$$\tau + \mathrm{d}\tau = (t + \mathrm{d}t) - [\mathbf{x}(t) - \mathrm{d}t\,\mathbf{v}(t)][\mathbf{v}(t) + \mathrm{d}t\,\mathbf{a}(t)] \quad . \qquad (151)$$

Together, this yields

$$\text{‘}\mathrm{d}\tau/\mathrm{d}t\text{’} = \gamma_v(1 - \mathbf{v}\mathbf{v}/c^2 - \mathbf{x}\mathbf{a}/c^2) \quad , \qquad (152)$$

a result showing that accelerated clocks can run *fast* instead of slow, depending on their position **x** and the sign of their acceleration **a**. There are quotes in the expression because we see directly that the Greek observer notes

$$\text{‘}\mathrm{d}t/\mathrm{d}\tau\text{’} = \gamma_v \quad , \qquad (153)$$

---

* These sets form what mathematicians call *hypersurfaces.*

which is *not* the inverse of equation (152). This difference becomes most apparent in the simple case of two clocks with the same velocity, one of which is accelerated constantly towards the origin with magnitude g, whereas the other moves inertially. We then have

$$`\mathrm{d}\tau/\mathrm{d}t' = 1 + gx \tag{154}$$

and

$$`\mathrm{d}t/\mathrm{d}\tau' = 1 \quad . \tag{155}$$

We will encounter this situation shortly. But first we clarify the concept of acceleration.

Acceleration for inertial observers

Accelerations behave differently from velocities under change of viewpoint. Let us first take the simple case in which everything moves along the x-axis: the object and two inertial observers. If a Roman inertial observer measures an acceleration $a = \mathrm{d}v/\mathrm{d}t = \mathrm{d}^2x/\mathrm{d}t^2$, and the Greek observer, also *inertial* in this case, an acceleration $\alpha = \mathrm{d}\omega/\mathrm{d}\tau = \mathrm{d}^2\xi/\mathrm{d}\tau^2$, we get Ref. 222

$$\gamma_v^3 a = \gamma_\omega^3 \alpha \quad . \tag{156}$$

The relation shows that accelerations are *not* Lorentz invariant; they are so only if the velocities are small compared to the speed of light. This is in contrast to our everyday experience, where accelerations are independent of the observer. Note that expression (156) simplifies in the case that the accelerations are measured at a time t in which v vanishes – i.e. measured by the so-called comoving inertial observer. In that case the acceleration a is called proper acceleration, as its value describes what the Roman observer *feels*, e.g. the experience of being pushed into the accelerating seat.

The value of 3-acceleration **a** measured by a general observer is related to the value $\mathbf{a}_\mathrm{c}$ measured by the comoving observer by the expression Ref. 223

$$\mathbf{a} = \frac{1}{\gamma_v^2}\left(\mathbf{a}_\mathrm{c} - \frac{(1-\gamma_v)(\mathbf{v}\mathbf{a}_\mathrm{c})\mathbf{v}}{v^2} - \frac{\gamma_v(\mathbf{v}\mathbf{a}_\mathrm{c})\mathbf{v}}{c^2}\right) \quad . \quad . \tag{157}$$

A simple squaring gives

$$a^2 = \frac{1}{\gamma_v^4}\left(a_\mathrm{c}^2 - \frac{(\mathbf{a}_\mathrm{c}\mathbf{v})^2}{c^2}\right) \tag{158}$$

which shows that the comoving or proper acceleration is always larger than the acceleration measured by an outside observer. The faster the outside inertial observer is, the smaller the acceleration he observes.Acceleration is indeed not a relativistic invariant. Challenge 496 e

In summary, acceleration complicates many issues. In fact, it is such an interesting topic that it merits a deeper investigation. To keep matters simple, we only study *constant* accelerations. Interestingly, this case is also a good introduction to black holes and, as we will see shortly, to the whole universe. See page 338

Accelerating frames of reference

How do we check whether we live in an inertial frame of reference? An *inertial frame (of reference)* has two properties: first, the speed of light is constant. In other words, for any two observers in that frame the ratio c between twice the distance measured with a ruler and the time taken by light to travel from one point to another and back again is always the same. The ratio is independent of time and of the position of the observers. Second, lengths and distances measured with a ruler are described by Euclidean geometry. In other words, rulers behave as in daily life; in particular, distances found by counting how many rulers (rods) have to be laid down end to end, the so-called *rod distances*, behave as in everyday life. For example, they follow Pythagoras' theorem in the case of right-angled triangles.

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

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

A general *frame of reference* is a continuous set of observers remaining at rest with respect to each other. Here, 'at rest with respect to each other' means that the time for a light signal to go from one observer to another and back again is constant in time, or equivalently, 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. We therefore note that 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 we have chosen the coordinate system in such a way that all the rigidly connected observers have constant coordinate values, we speak of a *rigid coordinate system*. Obviously, these are the most useful to describe accelerating frames of reference.*

Ref. 224 Note that if two observers both move with a velocity $\mathbf{v}$, as measured in some *inertial* frame, they observe that they are at rest with respect to each other *only* if this velocity is
Challenge 497 *constant*.Again we 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 they accelerate together to (or decelerate from) relativistic speeds in precisely the same way. Relativistic acceleration requires careful thinking.

An observer who always *feels* 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 the observer is at rest *at that moment*, always has the same value $\mathbf{b}$. It is important to note that uniform acceleration

---

Ref. 225 * There are essentially only two other types of rigid coordinate frames, apart from the inertial frames:

- the frame $\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 - c^2\mathrm{d}t(1 + g_k x_k/c^2)^2$ 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 $\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + 2\omega(-y\,\mathrm{d}x + x\,\mathrm{d}y)\mathrm{d}t - (1 - r^2\omega^2/c^2)\mathrm{d}t$. Here the z-axis is the rotation axis, and $r^2 = x^2 + y^2$.

is *not* uniformly accelerating when always observed from the *same* inertial frame K. This is an important difference with respect to the Galilean case.

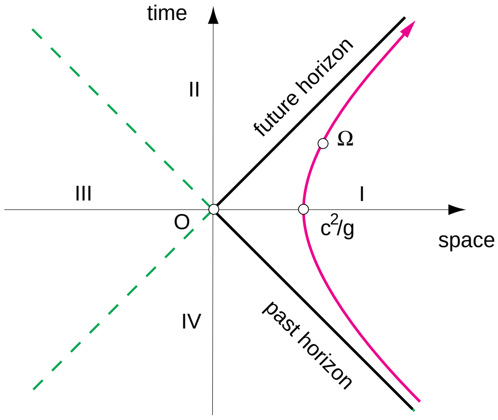


Figure 112 The hyperbolic motion of an rectilinearly, uniformly accelerating observer and its event horizons

For uniformly accelerated motion in the sense just defined, we need

$$\mathbf{b}\cdot\mathbf{b} = -g^2 \qquad (159)$$

where g is a constant independent of t. The simplest case is uniformly accelerating motion that is also *rectilinear*, i.e. for which the acceleration **a** is parallel to **v** at one instant of time and (therefore) for all other times as well. In this case we can write, using three-vectors, Ref. 226 Challenge 498

$$\gamma^3\mathbf{a} = \mathbf{g} \quad \text{or} \quad \frac{\mathrm{d}\gamma\mathbf{v}}{\mathrm{d}t} = \mathbf{g} \quad . \qquad (160)$$

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

$$v = \frac{gt}{\sqrt{1+\frac{g^2t^2}{c^2}}} \quad , \qquad (161)$$

where it was assumed that $v_o = 0$. We note that for small times we get $v = gt$ and for large times $v = c$, both as expected. The momentum of the Greek observer increases linearly with time, again as expected.Integrating, we find that the accelerated observer Ω moves along the path Challenge 499

$$x(t) = \frac{c^2}{g}\sqrt{1+\frac{g^2t^2}{c^2}} \quad , \qquad (162)$$

where it was assumed that $x_o = c^2/g$, in order to keep the expression simple. Because of this result, visualized in Figure 112, 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 $\mathrm{d}t = \gamma\mathrm{d}\tau$. Using the expression for the velocity $v(t)$ of equation (161) we get* Ref. 226, 228

$$t = \frac{c}{g}\sinh\frac{g\tau}{c} \quad \text{and} \quad x = \frac{c^2}{g}\cosh\frac{g\tau}{c} \qquad (163)$$

Ref. 229

[*] Use your favourite mathematical formula collection to deduce this. The abbreviations sinh $y = (e^y - e^{-y})/2$ and cosh $y = (e^y + e^{-y})/2$ imply that $\int \mathrm{d}y/\sqrt{y^2+a^2} = \operatorname{arsinh} y/a = \operatorname{Arsh} y/a = \ln(y+\sqrt{y^2+a^2})$.

for the relationship 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.

Does all this sound boring? Just imagine accelerating on a motor bike at $g = 10\,\mathrm{m/s}^2$ for the proper time τ of 25 years. That would bring you beyond the end of the known universe! Isn't that worth a try? Unfortunately, neither motor bikes nor missiles that accelerate like Challenge 500 this exist, as their fuel tank would have to be enormous.Can you confirm this even for the most optimistic case?

For uniform acceleration, the coordinates transform as

$$\begin{aligned} t &= \left(\frac{c}{g}+\frac{\xi}{c}\right)\sinh\frac{g\tau}{c} \\ x &= \left(\frac{c^2}{g}+\xi\right)\cosh\frac{g\tau}{c} \\ y &= \upsilon \\ z &= \zeta \quad , \end{aligned} \tag{164}$$

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

$$\mathrm{d}\sigma^2 = (1+g\xi/c^2)^2c^2\mathrm{d}\tau^2 - \mathrm{d}\xi^2 - \mathrm{d}\upsilon^2 - \mathrm{d}\zeta^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \tag{165}$$

and since for $\mathrm{d}\tau = 0$ distances are given by Pythagoras' theorem, the Greek reference frame Ref. 227 is indeed rigid.

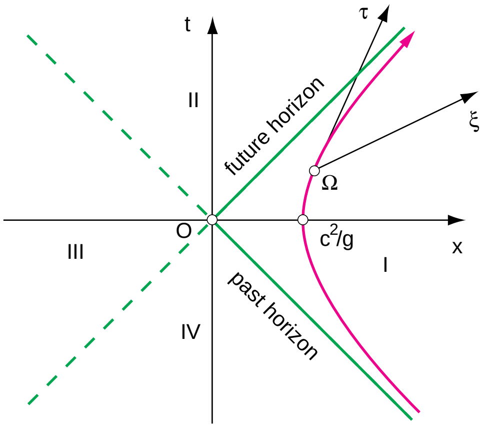


Figure 113 Do accelerated objects depart from inertial ones?

After this forest of formulae, 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 (162). 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 ob- Challenge 501 server 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 Ref. 230 $O\Omega$. Using expression (162) this turns out to be

$$d_{O\Omega} = \sqrt{\xi^2} = \sqrt{x^2 - c^2t^2} = c^2/g \quad , \tag{166}$$

which, surprisingly enough, is constant in time! In other words, the Greek observer will observe that he stays at a constant distance from the Roman one, in complete contrast to what the Roman observer says. Take your time to check this strange result in some other way. We will need it again later on, to explain why the earth does not explode. (Are you able to guess the relationship to this issue?) Challenge 502

Two pretty and challenging problems are first, to state how the acceleration ratio enters mass definition in special relativity, and second, to deduce the addition theorem for accelerations. See page 222 Challenge 503

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. Onegets the two relations* Challenge 504

$$\begin{aligned} \xi &= -\frac{c^2}{g}(1 - \operatorname{sech}\frac{g\tau}{c}) \\ d\xi/d\tau &= -c \operatorname{sech}\frac{g\tau}{c}\tanh\frac{g\tau}{c} \quad . \end{aligned} \tag{168}$$

These equations are strange. It is clear 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 woman standing on a long road. Seen from the car, the woman moves away; however, after a while, the only thing one notices is that she is slowly approaching the horizon. In Galilean physics, both the car driver and the woman on the road see the other person approaching each other's horizon; in special relativity, only the accelerated observer makes this observation.

Studying a graph of the situation confirms the result. In Figure 113 we can see that light emitted from any event in regions II and III cannot reach the Greek observer. Those events are hidden from him and cannot be observed. Strangely enough, however, light from the Greek 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 a uniformly accelerating frame of reference. Uniformly accelerating frames of reference

---

* The functions appearing above are defined using the expressions from the footnote on page 240:

$$\operatorname{sech} y = \frac{1}{\cosh y} \quad \text{and} \quad \tanh y = \frac{\sinh y}{\cosh y} \quad . \tag{167}$$

produce event horizons at a distance $-c^2/g$. For example, a person who is standing can never see further than this distance below his feet.

By the way, is it true that a light beam *cannot* catch up with an observer in hyperbolic motion, if the observer has a sufficient distance advantage at the start? Challenge 505 n

Acceleration changes colours

We saw above that a moving receiver sees different colours from 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. Ref. 226, 231 Indeed, if light is emitted in the direction of the acceleration, the expression for the space-time interval gives

$$c^2 \mathrm{d}\tau^2 = \left(1 + \frac{g_o x}{c^2}\right)^2 \mathrm{d}t^2 \qquad (169)$$

in which g_o is the proper acceleration of an observer located at $x = 0$. We can deduce in a straightforward waythat Challenge 506

$$\frac{f_R}{f_S} = 1 - \frac{g_R h}{c^2} = 1/\left(1 + \frac{g_S h}{c^2}\right) \qquad (170)$$

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 507 n

The formula usually given, namely

$$\frac{f_R}{f_S} = 1 - \frac{gh}{c^2} \quad , \qquad (171)$$

is only correct to first approximation, and not exactly what was just found. In accelerated frames of reference, we have to be careful with the meaning of every quantity used. For everyday accelerations, however, the differences between the two formulae are negligible. Are you able to confirm this? Challenge 508

Can light move faster than c?

What speed of light is measured by an accelerating observer? Using expression (171) above, an accelerated observer deduces that

$$v_{\text{light}} = c\,(1 + \frac{gh}{c}) \qquad (172)$$

which is higher than c in the case when light moves in front or 'above' him, and lower than c for light moving behind or 'below' him. This strange result concerning 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 also been confirmed by experiment. Ref. 232 In other words, the speed of light is only constant when it is defined as $c = \mathrm{d}x/\mathrm{d}t$, *and* if d$x$ and dt are measured with a ruler 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 ruler defining distances or the clock measuring times is 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.

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. Challenge 509 n

In fact, all these difficulties are only noticeable for distances l that 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, $9.5 \cdot 10^{12}$ km, in order to observe any sizable effects. In short, *c is the speed of light relative to* nearby *matter only.*

By the way, everyday gravity is equivalent to a constant acceleration. Why then don't distant objects, such as stars, move faster than light following expression (172)? Challenge 510 n

What is the speed of light?

We have seen that the speed of light, as usually defined, is given by c only if either the observer is inertial or the observer measures the speed of light passing nearby, 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.

An additional point is often forgotten. Usually, length is measured by the time it takes light to travel. In such a case the speed of light will obviously be constant. However, how does one check the constancy in the present case? One needs to eliminate length measurements. The simplest way to do this 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 Ref. 234

$$t_3 - t_1 = 2\,(t_2 - t_1) \quad . \tag{173}$$

Here it was assumed that the clocks were synchronized according to the prescription on page 215. If the factor were not exactly two, the speed of light would not be constant. In fact, all experiments so far have yielded 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 511 n

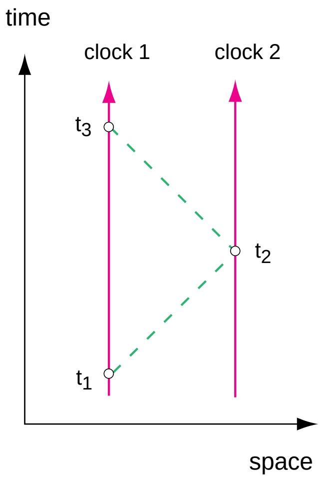


Figure 114 Clocks and the measurement of the speed of light as two-way velocity

Limits on the length of solid bodies

An everyday solid object breaks when some part of it moves with 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

$$\frac{v^2}{c^2} \geqslant \frac{2d}{l} \quad . \tag{174}$$

We see that we can avoid the breaking of fragile objects by packing them into foam rubber – which increases the stopping distance – of roughly the same thickness as the object's size. This may explain why boxes containing presents are usually so much larger than their contents!

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

$$la < c^2 \quad , \tag{175}$$

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

Ref. 235

Challenge 512 n

$$l\alpha < c^2/2 \quad , \tag{176}$$

where c is now the speed of light. The speed of light thus limits the size of solid bodies. For example, for $9.8\,\mathrm{m/s^2}$, the acceleration of a quality motor bike, this expression gives a length limit of 9.2 Pm, about a light year. Not a big restriction; most motor bikes are shorter.

However, 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 togetherin an accelerator?[**]

Challenge 513

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

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

Challenge 514

$$l < \frac{4\pi\varepsilon_o c^2 d^2 m}{e^2} \quad . \tag{177}$$

---

[*] 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, the speed of sound 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 $v^2/r \approx \hbar^2/m^2r^3 \approx 10^{31}\,\mathrm{m/s^2}$; this is one of the highest values found in nature.

The nearer electrons can get, the smaller they must be. The present experimental limit shows that the size is smaller than 10^{-19} m. Can electrons be point-like? We will come back to this issue during the study of quantum theory.

Special relativity in four sentences

This section of our ascent of Motion Mountain is rapidly summarized.

- All free floating observers find that there is a unique, perfect velocity in nature, namely a common maximum energy velocity, which is realized by massless radiation such as light or radio signals, 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 another, as described by the Lorentz transformations (94) and (95), 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 $E = \gamma mc^2$, as again confirmed by experiment.
- Applied to accelerated objects, these results lead to numerous counter-intuitive consequences, such as the twin paradox, the appearance of event horizons and the appearance of short-lived tachyons in collisions.

In summary, special relativity shows that motion is relative, defined using the propagation of light, conserved, reversible and deterministic.

Could the speed of light vary?

For massless light, the speed of light is the limit speed. Assuming that light is indeed exactly massless, is it possible at all that the speed of light changes from place to place or that it changes with time? This tricky question still makes a fool out of many physicists. On first sight, the answer is a loud ‘Yes, of course! Just have a look to what happens when the value of c is changed in formulas.’ However, this statement is wrong.

Since the speed of light enters our definition of time and space, and thus enters, even if we do not notice it, the construction of all rulers, all measurement standards and all measuring set-ups, there is *no way* to detect whether the value actually varies. No imaginable experiment could detect a variation of the limit speed.‘That is intellectual cruelty!’, you might say. ‘All experiments show that the speed of light is invariant; we had to swallow one counter-intuitive result after the other to accept the constancy of the speed of light, and now we are even supposed to admit that there is no other choice?’ Yes, we are. That is the irony of progress in physics. The observer invariance of the speed of light is counter-intuitive and astonishing when compared to the lack of observer invariance of everyday, Galilean speeds. But had we taken into account that every speed measurement always is, whether we like it or not, a comparison with the speed of light, we would not have been astonished by the constancy of the speed of light; we would have been astonished by the strange way small speeds behave. There is no way, in principle, to check the invariance of a standard. To put it in other words, the most counter-intuitive aspect of relativity is not the invariance of c; the most counter-intuitive aspect is the disappearance of c from the formulas of everyday motion. Challenge 515

During our exploration of Galilean physics, once we had defined the basic concepts of velocity, space and time, we turned our attention to gravitation. Since the invariance of the speed of light has forced us to change these basic concepts, we now return to study gravitation in the light of this invariance.

* * *

References

175 For more details on the definition of length, see Appendix B on physical measurement units, on page 872. Cited on page 199.

176 ARISTOTLE, *The sense*, VI, 446 a 26. Cited in JEAN-PAUL DUMONT, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 157, 1991. Cited on page 199.

177 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 200.

178 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 the http://almagest.as.utexas.edu/~rlr/mlrs.html and http://cddisa.gsfc.nasa.gov/cstg/slr_cstg.html web sites. Cited on page 201.

179 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 201 is also taken from there. Cited on page 201.

180 You can learn the basics of special relativity with the help of the web, without referring to any book, by using as a starting point the http://physics.syr.edu/research/relativity/RELATIVITY.html web page. This page mentions most English language relativity pages available on the web. Some links in other languages are mentioned on the Motion Mountain web site. Search engines can help as well. Cited on page 202.

181 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, Bulletin of the American Physical Society **45**, 2000. He assumed that both sides of the burster emit light. The large speed difference and the pulse sharpness then yield the limit. 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 completely watertight however; 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 202.

182 Edwin F. Taylor & John A. Wheeler, *Spacetime Physics – Introduction to Special Relativity*, second edition, Freeman, 1992. Cited on page 204.

183 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. It still well worth reading, and every physicist should have done so. It is now available in many languages. Cited on pages 203 and 204.

184 Albert Einstein, *Mein Weltbild*, edited by Carl Selig, Ullstein Verlag, 1998.

185 Jean van Bladel, *Relativity and Engineering*, Springer 1984. Cited on page 203.

186 See e.g. C. Will, *Theory and Experiment in Gravitational Physics*, Revised edition, Cambridge University Press, 1993. Cited on page 203.

187 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 203.

188 R.J. Kennedy & E.M. Thorndike, *Experimental establishment of the relativity of time*, Physical Review **42**, pp. 400–418, 1932. See also H.E. Ives & G.R. Stilwell, *An experimental study of the rate of a moving atomic clock*, Journal of the Optical Society of America **28**, pp. 215–226, 1938, and **31**, pp. 369–374, 1941, For a modern, high precision version, see C. Braxmeier, H. Müller, O. Pradl, J. Mlynek, A. Peters & S. Schiller, *New tests of relativity using a cryogenic optical resonator*, Physical Review Letters **88**, p. 010401, 2002. Cited on page 203.

Cited on page 204.

189 Rod S. Lakes, *Experimental limits on the photon mass and cosmic vector potential*, Physical Review Letters **80**, pp. 1826–1829, 1998. The speed of light is independent of frequency within a factor of $6 \cdot 10^{-21}$, as was shown from gamma ray studies by B.E. Schaefer, *Severe limits on variations of the speed of light with frequency*, Physical Review Letters **82**, pp. 4964–4966, 1999.

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 204.

190 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 205.

191 The beautiful graph is taken from Z.G.J. Guiragossian & al., ..., Physical Review Letters **34**, pp. 335–338, 1975. Cited on page 207.

192 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 208.

193 The speed of neutrinos is the same as that of light within 9 decimal digits, since neutrinos and light arrived together, within 12 seconds, 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, in ... The new results taken by the Canadian ... collaboration also confirm that neutrinos have a mass in the 1 eV region. Cited on page 208.

194 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 211.

195 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 209.

196 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 209.

197 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 211.

198 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 212.

199 If the future could influence the past ... Cited on page 214.

200 STEVEN K. BLAU, *Would a topology change allow Ms. Bright to travel backward in time?*, American Journal of Physics **66**, pp. 179–185, 1998. Cited on page 214.

201 On the ‘proper’ formulation of relativity, see for example ... Cited on page 214.

202 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 215.

203 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 215.

204 PAUL J. NAHIN, *Time machines – time travel in physics, metaphysics, and science fiction*, Springer Verlag and AIP Press, second edition, 1999. Cited on page 215.

205 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 215.

206 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 216.

207 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 216.

Challenge 516 n

208 S.P. BOUGHN, *The case of the identically accelerated twins*, American Journal of Physics **57**, pp. 791–793, 1989. Cited on page 216.

209 G.R. RYBICKI, *Speed limit on walking*, American Journal of Physics **59**, pp. 368–369, 1991. Cited on page 217.

210 The slowness of the speed of light inside stars is explained e.g. by ... Cited on page 218.

211 The first examples of such astronomical observations were provided by A.R. Whitney *&* al., *Quasars revisited: rapid time variations observed via very-long-baseline interferometry*, Science **173**, pp. 225–230, 1971, and by M.H. Cohen *&* al., *The small-scale structure of radio galaxies and quasi-stellar sources at 3.8 centimetres*, Astrophysical Journal **170**, pp. 207–217, 1971. See also T.J. Pearson, S.C. Unwin, M.H. Cohen, R.P. Linfield, A.C.S. Readhead, G.A. Seielstad, R.S. Simon & R.C. Walker, *Superluminal expansion of quasar 3C 273*, Nature **290**, pp. 365–368, 1981. An overview is given in J.A. Zensus & T.J. Pearson, editors, *Superluminal radio sources*, Cambridge University Press, 1987. 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 219.

212 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 219.

213 J.P. Costella, B.H.J. McKellar, A.A. Rawlinson & G.J. Stephenson, *The Thomas rotation*, American Journal of Physics **69**, pp. 837–847, 2001. Cited on page 220.

214 See S.S. Costa, G.E.A. Matsas, *Temperature and relativity*, http://arxiv.org/gr-qc/9505045. Cited on page 221.

215 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 225.

216 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 230. Cited on page 226.

217 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*, Phroceeding of the Royal Society in London, A **195**, pp. 62–81, 1948. Cited on page 227.

218 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 232.

219 A jewel among the textbooks on special relativity is the booklet by ULRICH E. SCHRÖDER, *Spezielle Relativitätstheorie*, Verlag Harri Deutsch, Thun, 1981. Cited on pages 231 and 232.

220 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 232.

221 R.J. LOW, *When moving clocks run fast*, European Journal of Physics **16**, pp. 228–229, 1995. Cited on page 237.

222 WOLFGANG RINDLER, *Relativity – Special, General and Cosmological*, Oxford University Press, 2001. A beautiful book by one of the masters of the field. Cited on page 238.

223 G. STEPHENSON & C.W. KILMISTER, *Special Relativity for Physicists*, Longmans, London, 1965. Cited on page 238.

224 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 239.

225 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 239.

226 R.H. GOOD, *Uniformly accelerated reference frame and twin paradox*, American Journal of Physics **50**, pp. 232–238, 1982. Cited on pages 240 and 243.

227 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 241.

228 J. DWAYNE HAMILTON, *The uniformly accelerated reference frame*, American Journal of Physics **46**, pp. 83–89, 1978. Cited on page 240.

229 The best and cheapest mathematical formula collection remains the one by K. ROTTMANN, *Mathematische Formelsammlung*, BI Hochschultaschenbücher, 1960. Cited on page 240.

230 See for example D.J. RAYMOND, *A Radically modern approach to introductory physics*, on the http://www.physics.nmt.edu/raymond/ph13xbook/index.html web site. Cited on pages 242 and 250.

231 EDWARD A. DESLOGE, *The gravitational red shift in a uniform field*, American Journal of Physics **58**, pp. 856–858, 1990. Cited on page 243.

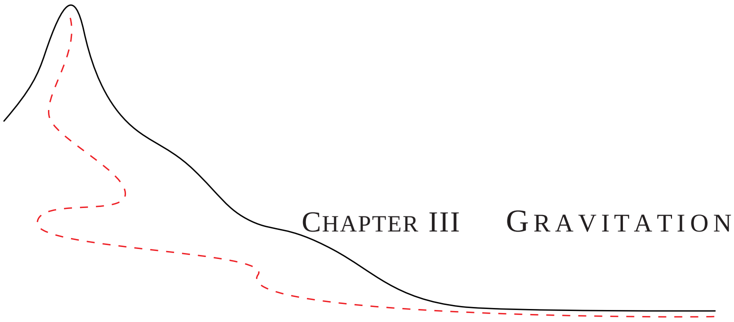
232 The propagation delays to be discussed in the chapter on general relativity can be seen as confirmations of this effect. Cited on page 243.

233 W. RINDLER & L. MISHRA, *The nonreciprocity of relative acceleration in relativity*, Physics Letters A **173**, 1993, pp. 105–108. Cited on page .

234 The subtleties of the one-way and two-way speed of light are explained in ... Cited on page 244.

235 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 245.

Typeset in
January 2003



CHAPTER III GRAVITATION AND RELATIVITY

Gravitational influences do transport energy.* In the description of motion, the next goal must therefore be to increase the precision in such a way that this transport happens at most with the speed of light. Henri Poincaré stated this requirement already in 1905. The results following from this principle will be fascinating; we will find that empty space can move, that the universe has a finite age, that objects can be in permanent free fall, that empty space can be bent, but that empty space is much stiffer than steel. Despite these strange statements, the theory and all its predictions have been confirmed by each one of the numerous experiments ever performed.

Describing motion due to gravity using the relation $a = GM/r^2$ not only allows speeds larger than light, e.g. in orbits; it is also unclear how the values of a and r depend on the observer. The expression thus 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. 236, 237

7. The new ideas on space, time, and gravity

Sapere aude.
Horace**

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. A body is at rest whenever it is not disturbed by other bodies. In the Galilean description of the world, rest is the *absence of velocity*. In special relativity, rest became *inertial motion*, since no inertially moving observer can distinguish its own motion from rest: nothing disturbs him. Both the rock in the waves and the rapid protons crossing the galaxy as cosmic rays are at rest. Including gravity leads us 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 Challenge 517 e

---

* The details of this statement are far from simple. They are discussed on page 274 and page 298.

** 'Venture to be wise.' Horatius, Quintus Flaccus, Ep. 1, 2, 40.

be. Nobody knows this better than Joseph Kittinger, the man who in August 1960 stepped out of a balloon capsule at the record height of 31.3 km. At that altitude, the air is so thin that during the first minute of his free fall he felt completely at rest, as if he were floating. Being an experienced parachutist, he was so surprised that he had to turn upwards in order to convince himself that he was indeed getting away from his balloon! Despite his lack of any sensation, he was falling at up to 988 km/h with respect to the earth's surface. He only started feeling something from that moment in which he encountered the first layers of air. That was when his free fall started to be disturbed. Later, after four and a half minutes of fall, his special parachute opened, and nine minutes later he landed safely in New Mexico. Ref. 238

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*. Rest is lack of disturbance; so 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. This means that we must describe motion in such a way that not only inertial but also freely falling observers can talk to each other. In addition, a full description of motion must be able to describe gravitation and the motion it produces, and it must be able to describe motion for any observer imaginable. This is the aim that general relativity realizes.

To pursue this aim, we put the original result in simple words: 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 not at rest. What motion are they undergoing? And if free fall is rest, what is weight? And what then is gravity anyway? Let us start with the last question. Challenge 518 n

What is gravity?

Ref. 249 As William Unruh likes to explain, the constancy of the speed of light for all observers implies the following conclusion: *gravity is the uneven running of clocks at different places*.* Of course, this seemingly absurd definition needs to be checked.The definition does not talk about a single situation seen by *different* observers, as we often did in special relativity. The definition states that neighbouring, identical clocks, fixed against each other, run differently in the presence of a gravitational field when watched by the *same* observer; moreover, this difference is 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. Challenge 520 e

An observer feels no difference between gravity and constant acceleration. Thus we can use a result 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, See page 216

---

[*] 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 seewhy? Challenge 519

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 effect, the frequency of the light arriving at the front has changed. Inserting,we get [*] Challenge 521 e

$$\frac{\Delta f}{f} = \frac{g\Delta h}{c^2} = \frac{\Delta \tau}{\tau} \quad . \tag{178}$$

(The meaning of τ will become clear shortly.) The sign of the frequency change depends on whether the light motion and the train acceleration are in the same or in opposite directions. For actual trains or buses, the frequency change is quite small.But before we discuss the consequences of the result, let us check it with a different experiment. Challenge 523

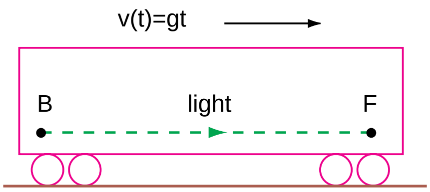


Figure 115 Inside an accelerating train or bus

To measure time and space, we use light. What happens to light when gravity is involved? The simplest experiment is to let light fall or rise. In order to deduce what must happen, we add a few details. Ref. 243 Imagine a conveyor belt carrying masses around two wheels, a low and a high one. The descending, grey masses are slightly larger. Whenever such a larger mass is near the bottom, some mechanism – not drawn – converts the mass surplus to light via $E = mc^2$ and sends the light up towards the top.[**] See Figure 116 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 grey masses on the descending side are always heavier, the belt would turn for ever, and this system could continuously *generate* energy. However, since energy conservation is at the basis of our definition of time, as we saw in the beginning of our walk, the whole process must be impossible. See page 152 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. Similarly, 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? The result deduced from the drawing isagain the one of formula (178). That is expected, as the two experiments describe equivalent situations, as you might want to check yourself. Challenge 525 e Challenge 526 The formula gives a relative change of frequency f of only $1.1 \cdot 10^{-16}$ /m on the surface of the

---

[*] The expression $v = gt$ is valid only for small speeds; nevertheless, the conclusion of the section is independent of this approximation. Challenge 522 e

[**] As in special relativity, here and in the rest of our mountain ascent, 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 528.

[****] How does this argument change if you include the illumination by thesun? Challenge 524

earth. For trees, this so-called *gravitational red shift* or *gravitational Doppler effect* is far too small to be observable, at least using normal light.

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. Ref. 256 See page 548 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, and independently Perot, showed that careful experiments indeed confirm the gravitational red shift. 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 red shift in their university tower using γ radiation. Ref. 244

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 makes old*. Can you confirm this conclusion? Challenge 527

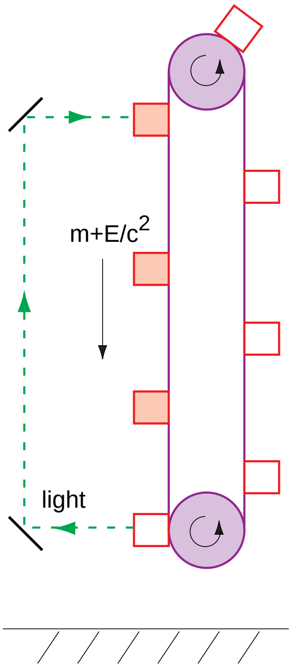


Figure 116 The necessity of blue and redshift of light: why trees are greener at the bottom

In 1972, by flying four precise clocks in an aeroplane, and keeping an identical one on the ground, Hafele and Keating found that clocks indeed run differently at different altitudes according to expression (178). Ref. 245 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. Ref. 246 And in 1977, Briatore and Leschiutta showed that a clock in Torino indeed ticks slower than one on the top of the Monte Rosa. Ref. 247 They confirmed the prediction that on earth, for every 100 m of height gained, people age more rapidly by about 1 ns per day. Challenge 528 In the mean time this effect has been confirmed for all systems for which experiments were performed, such as several other planets, the sun, and numerous other stars.

In summary, gravity is indeed the uneven running of clocks at different heights. Note that both an observer at the lower position and one at a higher position *agree* on the result; both find that the upper clock goes faster. In other words, when gravity is present, space-time is *not* described by the Minkowski space-time of special relativity, but by some more general space-time. To put it mathematically, whenever gravity is present, we have

$$ds^2 \neq c^2 dt^2 - dx^2 - dy^2 - dz^2 \quad . \qquad (179)$$

We will give the correct expression shortly.

Is this view of gravity as height-dependent really reasonable? No. It turns out that it is not yet strange enough. Since the speed of light is the same for all observers, we can say more.

If time changes with height, also length must do so! More precisely, if clocks run differently at different heights, also the length of meter bars changes with height. Can you confirm this for the case of horizontal bars at different heights? Challenge 529

If length changes with height, the circumference of a circle around the earth *cannot* be given by $2\pi r$. A similar discrepancy is also found by an ant measuring radius and circumference of a large circle traced on the surface of a basketball. Indeed, gravity implies that humans are in a similar position as ants on a basketball, with the only difference that the situation is translated from two to three dimensions. We conclude that wherever gravity plays a role, space is curved.

What tides tell about gravity

During his free fall, Kittinger was able to specify an inertial frame for himself. Indeed, he felt completely at rest. Does this mean that it is impossible to distinguish acceleration from gravitation? No; distinction *is* possible. We only have to compare *two* (or more) falling observers.

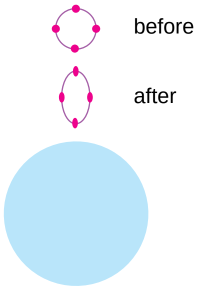


Figure 117 Tidal effects: what bodies feel when falling

Kittinger could not have found a frame which is also inertial for a colleague falling on the opposite side of the earth.In fact, such a common frame does not exist. In general, it is impossible to find a *single* inertial reference frame describing different observers freely falling near a mass. Challenge 530

The impossibility to find a common inertial frame applies even to *nearby* observers in a gravitational field. Two nearby observers observe that during 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.

That rings a bell. The squeezing of the body is the same effect that leads to the tides. Indeed, the bulging oceans can be seen as the squeezed earth in its fall towards the moon. Using this result of universal gravity we can state: the essence of gravity is the observation of tidal effects. See page 97 Ref. 241

In other words, gravity is simple only *locally*. Only locally it looks like acceleration. Only locally, a falling observer like Kittinger feels at rest. In fact, only a point-like observer does so! As soon as we take spatial extension into account, we find tidal effects. *Gravity is the presence of tidal effects.* The absence of tidal effects implies the absence of gravity. Tidal effects are the everyday consequence of height dependent time. Isn't this a beautiful conclusion?

In principle, Kittinger could have *felt* gravitation during his free fall, even with his eyes closed, had he paid attention to himself. Had he measured the distance change between his two hands, he would have found a tiny decrease which could have told him that he was falling, even with his eyes closed. This tiny decrease would have forced Kittinger to a strange conclusion. Two freely falling hands should move along two parallel lines, always keeping the same distance. Since the distance changes, in the space around him lines starting out in parallel do not remain so. Kittinger would have concluded that the space around him was similar to the surface of a sphere, where two lines starting out north, parallel to each other, also change distance, until they meet at the north pole. In other words, Kittinger would have concluded that he was in a *curved* space.

Studying the value of the distance decrease between his hands, Kittinger would even have concluded that the curvature changes with height. Physical space differs from a sphere, which has constant curvature; physical space is more involved. The effect is extremely small and cannot be felt by human senses. Kittinger had no chance to detect anything. Detection requires special high sensitivity apparatuses. However, the conclusion does not change. Space-time is *not* described by Minkowski space when gravity is present. Tidal effects imply space-time curvature.

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 with the title 'Revolution in science / new theory of the universe / Newtonian ideas overthrown'. The expedition had shown unequivocally – though not for the first time – that the theory of universal gravity, essentially given by $a = GM/r^2$, was wrong, and that instead space 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 Greek indices, in the middle of Time magazine. This did not happen to any other physicist before or afterwards.

Ref. 240 The expedition had performed an experiment proposed by Einstein himself. Apart from searching for the change of time with height, he had thought about a number of experiments to detect the curvature of space. In the one that eventually made him famous, Einstein proposed to take a picture of the stars near the sun, as is possible during a solar eclipse, and compare it with a picture of the same stars at night, when the sun is far away. Einstein predicted a change in position of 1.75' for star images at the border of the sun, a result *twice* as large as the effect predicted by universal gravity. (See page 99) The prediction, corresponding to about $1/40$ mm on the photographs, was confirmed in 1919, and thus universal gravity was ruled out.

Does this experiment *imply* that space is curved? No, it doesn't. 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. However, the eclipse is not the only data. We already 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. But these experiments also

---

* 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.

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.

If the speed of light is constant, but clocks and meter bars change with height, *space is curved near masses*.Many physicists in the twentieth century checked whether meter bars indeed behave differently in places where 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 observed. All results confirm the existence of curvature of space and space-time, and in addition confirm the predicted curvature values. In other words, meter bars near masses do indeed change their size from place to place, and even from orientation to orientation. Figure 118 gives an impression of the situation. Challenge 531

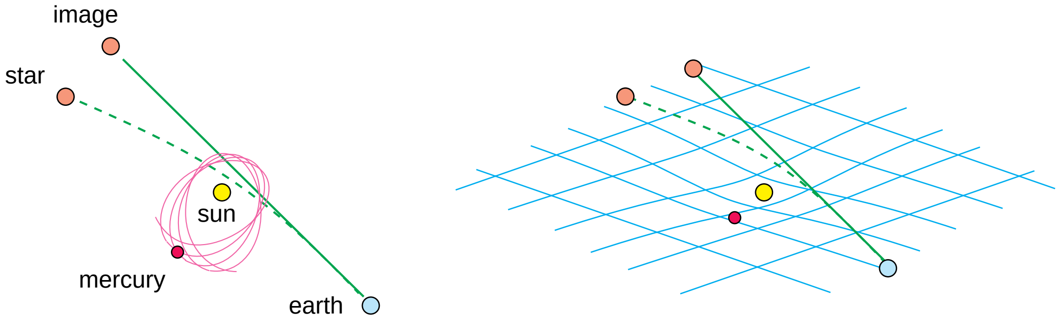


Figure 118 The mattress model of space: the path of a light beam and of a satellite near a spherical mass

But attention: the right hand figure, even though found in all textbooks, can be misleading. Ref. 239 It can be easily mistaken to show a *potential* around a body. Indeed, it is impossible to draw a graph showing curvature and potential separately. (Why?)We will see that for small curvatures, it is even possible to describe the meter bar change with a potential only. Challenge 532 Thus the figure does not really cheat, at least in the case of weak gravity. But for large and changing values of gravity, a potential cannot be defined, and thus there is indeed no way to avoid using curved space to describe gravity. In summary, if we imagine space as a sort of generalized mattress, on which masses produce deformations, we have a reasonable model of space-time. As masses move, the deformation follows them. Space thus behaves like a frictionless mattress with one additional dimension.

If gravity means curved space, we follow that any accelerated observer, such as a man in a departing car, must also observe that space is curved. But obviously, in everyday life we do not notice any such effect. How then would you devise a precision experiment to check the statement? Challenge 533

In fact, not only space, but also space-time is curved, even though Figure 118 shows the curvature of space only. We will shortly find out how to describe both the shape of space as well as the shape of space-time, and how to measure their curvature.

Let us have a first idea on how to describe nature when space-time is curved. In the case of Figure 118, the best description of events is the use of the time t shown by a clock located at spatial infinity; that avoids problems with the uneven running of clocks with different distances from the central mass. For the radial coordinate r the most practical choice to avoid problems with the curvature of space is to use the circumference of a circle around the central body divided by 2π. The curved shape of space-time is best described by the behaviour of the space-time distance ds, or by the wristwatch time $d\tau = ds/c$, between two neighbouring points with coordinates (t,r) and $(d+dt, r+dr)$. We found out above that gravity means that in spherical coordinates we have See page 212 See page 255

$$d\tau^2 = \frac{ds^2}{c^2} \neq dt^2 - dr^2/c^2 - r^2 d\varphi^2/c^2 \quad . \tag{180}$$

The inequality expresses that space-time is *curved*. The experiments on clock behaviour with height show that the space-time interval around a spherical mass is given by

$$d\tau^2 = \frac{ds^2}{c^2} = \left(1 - \frac{2GM}{rc^2}\right) dt^2 - \frac{dr^2}{c^2 - \frac{2GM}{r}} - \frac{r^2}{c^2} d\varphi^2 \quad . \tag{181}$$

This expression is called the *Schwarzschild metric* after one of its discoverers.* The metric (181) describes the curved shape 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 Challenge 534 n

$$h = \frac{2G}{c^2}\frac{M}{R} \quad . \tag{182}$$

This ratio expresses the gravitational strain with which lengths and the vacuum are deformed from the flat situation of special relativity, and thus also expresses the amount that clocks lag behind when gravity is present. On the surface of the earth the ratio h has the small value of $1.4 \cdot 10^{-9}$, and on the surface of the sun the somewhat larger value of $4.2 \cdot 10^{-6}$. Modern clocks can easily detect these changes. The consequences and uses will be discussed shortly.

We note that if a mass is highly concentrated, in particular when its radius gets *equal* to its so-called *Schwarzschild radius*

$$R_S = \frac{2GM}{c^2} \quad , \tag{183}$$

the Schwarzschild metric behaves strangely: *time disappears*. At the Schwarzschild radius, the wristwatch time stops. What happens precisely will be shown below. The situation is not common; the Schwarzschild radius for a mass like the earth is 8.8 mm and for the sun 3.0 km; you might want to check that the object size for all systems in everyday life is always larger than their Schwarzschild radius.Bodies who reach this limit are called *black* See page 340 Challenge 535 e

---

* 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 later on, directly from the field equations of general relativity.

Ref. 250 The other discoverer of the metric, unknown to Einstein, was the Dutch physicist J. Droste.

holes; we will study them in detail shortly. In fact, a principle of general relativity states 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 *mass generates curvature*. Special relativity then tells us that as a consequence, space should also be curved by the pressure of energy-momentum. For example, light should also curve space-time. Unfortunately, even the highest energy beams correspond to extremely small masses, and thus to unmeasurably small curvatures. Nevertheless it is still possible to show experimentally that energy also curves space, since in almost all atoms a sizeable fraction of the mass is due to the electrostatic energy among the positively charged protons. Indeed, in 1968 Kreuzer confirmed that energy curves space with a clever experiment using a floating mass. Ref. 251

It is straightforward to picture that the uneven running of clock is the temporal equivalent of spatial curvature.Taking the two together, we conclude that the complete statement is that in case of gravity, *space-time* is curved. Challenge 536

Let us sum up our chain of thoughts. Energy is equivalent to mass; mass produces gravity; gravity is equivalent to acceleration; acceleration is position dependent time. Since light speed is constant, we deduce that *energy-momentum tells space-time to curve.* This statement is the first half of general relativity.

We will soon find out how to measure curvature, how to calculate it, and how to compare the two results. In addition, different observers measure different curvatures. The set of transformations relating one viewpoint to another in general relativity, *diffeomorphism symmetry*, will also be object of our investigations.

Since matter moves, we can say even more. Not only is space-time curved near masses, it also bends back when a mass has passed by. In other words, general relativity states that space, as well as space-time, is *elastic*. However, it is rather stiff, and quite a lot stiffer than steel.* In fact, to curve a piece of space by 1% requires an energy density enormously larger than to curve a simple train rail by 1%.This and other fun consequences of space-time curvature and its elasticity will occupy us for a little while. Challenge 537

The speed of light and the constant of gravitation

Si morior, moror.**

All the experiments and knowledge about gravity can be summed up in just two general statements. The first principle states:

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

$$v \leqslant c \qquad (184)$$

for all observers, where c is the speed of light. Ref. 242

---

* A good book in popular style on the topic is DAVID BLAIR & GEOFF MCNAMARA, *Ripples on a cosmic sea*, Allen & Unwin, 1997.

** 'If I rest, I die.' Motto of the bird of paradise.

The description following from this first principle, *special* relativity, is extended to *general* relativity by adding a second principle, characterizing gravitation. A simple way is to use curvature:

▷ *The radius R and the surface A of a resting, spherical system of mass M and constant density are related by*

$$R-\sqrt{\frac{A}{4\pi}}=\frac{GM}{3c^2} \quad . \tag{185}$$

In other words, since mass bends space, the relation between the surface and the radius of a sphere is different from the expression $A=4\pi r^2$ we learned in school. This second principle introduces the gravitational constant G. We will see later on why the same constant also appears in the expression $a=GM/r^2$ for universal gravity.

But the relation between gravitation and curvature can also be expressed in a different way. In nature, there is a limit to the curvature of space-time.

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

$$\frac{L}{M}\geqslant\frac{4G}{c^2} \quad . \tag{186}$$

In other words, there is nothing more concentrated in nature than a black hole.

Yet another way to express the principle of gravitation is the following. Gravitation leads to attraction of masses. However, this attraction limits the acceleration a body can achieve.

▷ *For all observers, the acceleration a of a system of mass M is limited by*

$$Ma\leqslant\frac{c^4}{4G} \quad . \tag{187}$$

Every mass implies an acceleration limit, because acceleration requires 'fuel' to be left behind; however, for large amounts of fuel, the mass-energy left behind forms a black hole which prevents further acceleration. In short, there is a maximum force in nature.

We will discuss and motivate these new limits throughout this chapter, including the section on black holes. We will show that they are all equivalent to each other; no exception is known or possible. The limits reduce to the usual definition of gravity in the non-relativistic case. The limits tell *what* gravity is, namely curvature, and *how* exactly it behaves. Together, the first and any of the second principles imply all of general relativity.*

For example, are you able to show that the formula describing gravitational redshift complies with the general limit (186) on length to mass ratios? Challenge 538

We note that any formula that contains the speed of light c is based on special relativity, and if it contains the constant of gravitation G, it relates to universal gravity. If a formula contains *both* c and G, it is a statement of general relativity. The present chapter allows to regularly test this connection.

The mountain ascent so far has taught us that a precise description of motion requires the listing of all allowed viewpoints, their characteristics and their differences, as well as the

---

* This didactic approach is unconventional. Beware.

transformations from one viewpoint to the other. From now on, *all* viewpoints are allowed, without exception; anybody must be able to talk to anybody else. It makes no difference whether an observer feels gravity, is in free fall, is accelerated or is in inertial motion. Also people who exchange left and right, people who exchange up and down or people who say that the sun turns around the earth must be able to talk to each other. This gives a much larger set of viewpoint transformations than in the case of special relativity; it makes general relativity both difficult and fascinating. And since all viewpoints are allowed, the resulting description of motion is *complete*.*

Why does a stone thrown into the air fall back? – Geodesics

A genius is somebody who makes all possible mistakes in the shortest possible time.

In special relativity, we saw that inertial or free floating motion is that motion which connects two events that requires the *longest* proper time. In the absence of gravity, the motion fulfilling this requirement is *straight* motion. On the other hand, we are also used to think of straightness as the shape of light rays. Indeed, we all are used to check the straightness of an edge by looking along it. Whenever we draw the axes of a physical coordinate system, we imagine either drawing paths of light rays or drawing the motion of freely moving bodies.

See page 49

In the absence of gravity, object paths and light paths coincide. However, in the presence of gravity, objects do not move along light paths, as every thrown stone shows. In addition, in the presence of gravity both light and objects paths are bent, but by different amounts. Light does not define spatial straightness any more. In the presence of gravity, both light and matter paths are bent, though by *different* amounts. But the original statement remains: even when gravity is present, bodies follow paths of longest possible proper time. Such paths are called *(timelike) geodesics*. For light, the paths are called *(lightlike) geodesics*.

In other words, *stones fall because they follow geodesics*. Let us perform a few checks of this statement. We note that in space-time, geodesics are the curves with *maximal* length. This is in contrast with the case of pure space, such as the surface of a sphere, where geodesics are the curves of *minimal* length.

- Since stones move by maximizing proper time for inertial observers, they also must do so for freely falling observers, as Kittinger would argue. Then they do so for all observers.
- If fall is seen as a consequence of the earth's surface approaching – as we will argue later on – we can deduce directly that fall implies a proper time as long as possible.Free fall is motion along geodesics.

Challenge 539

- If gravitation is seen as the result of one of the splashy principles, the issue is also interesting. Take a vanishingly light object orbiting a mass M at distance R. It undergoes an acceleration a. Now the acceleration a of a component in a system of size $2R$ is limited, as we found out in special relativity, by $2aR \leqslant c^2$. This limit is compatible with the splashy size limit of systems*only* if the gravitational acceleration a is given by universal gravity's $a = GM/R^2$ at large distances. The size limit for physical systems is thus *equivalent* to universal gravity and to geodesic motion.

See page 245

Challenge 540

---

[*] Were it not for a small deviation called quantum theory.

▪ The acceleration limit given above can be seen as the combination of the size limit and the acceleration of universal gravity, as you may want to check.In other words, also the acceleration limit implies universal gravity. And we know that fall following universal gravity in turn implies longest proper time. Challenge 541

▪ If fall is seen as consequence of the curvature of space-time, the explanation is a bit more involved.

– CS – story to be filled in – CS –

In short, the straightest path in space-time for a stone thrown in the air and the path for the thrower himself cross again after a while: stones do fall back. Only if the velocity of the stone is too large, the stone does not fall back: it then leaves the attraction of the earth and makes its way through the sky.

▪ Let us turn to an experimental check. If fall is a consequence of curvature, then the path of *all* stones thrown on the surface of the earth must have the *same* curvature in space-time. Take a stone thrown horizontally, a stone thrown vertically, a stone thrown rapidly, or a stone thrown slowly: it takes only two lines to show that *in space-time* all their paths are approximated to high precision by circle segments,and that all paths have the *same* curvature radius r given by Challenge 542

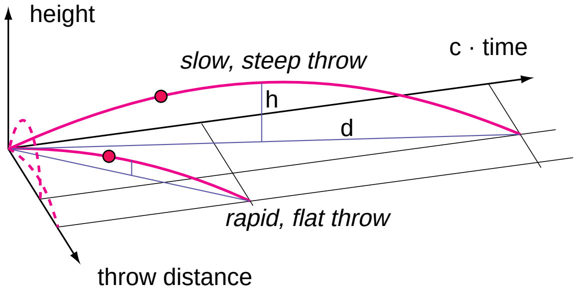


Figure 119 All paths of flying stones have the same curvature in space-time

$$r = \frac{c^2}{g} \approx 9.2 \cdot 10^{15}\,\mathrm{m} \quad . \tag{188}$$

The large value of the radius, corresponding to an extremely low curvature, explains why we do not notice it in everyday life. The parabola shape typical of the path of a stone in everyday life is just the projection of the more fundamental path in space-time. The important point is that the value of the curvature does *not* depend on the details of the throw. In fact, this simple result could have brought people onto the path of general relativity already a full century before Einstein; what was missing was the recognition of the importance of the speed of light. In any case, this simple calculation confirms that fall and curvature are connected. As expected and mentioned already above, the curvature diminishes at larger heights, until it vanishes at infinite distance from the earth.

In summary, the motion of any particle falling freely 'in a gravitational field' is described by the same variational principle as the motion of a free particle in special relativity: the path maximizes the proper time $\int d\tau$. We rephrase this by saying that any particle in free fall from point A to point B minimizes the action S given by

$$S = -mc^2 \int_A^B d\tau \quad . \tag{189}$$

That is all we need to know about the free fall of objects. As a consequence, any *deviation from free fall keeps young*. The larger the deviation, the younger one stays.

As we will see below, the minimum action description of free fall has been tested extremely precisely, and no difference from experiment has ever been observed. We will also find out that for free fall, the predictions of general relativity and of universal gravity differ substantially both particles near the speed of light and for central bodies of high density. All experiments show that whenever the two predictions differ, general relativity is right and universal gravity or other alternative descriptions are wrong. See page 351 Ref. 252

Free fall is 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 rest.

The next question is whether energy falls in the same way as mass. *Bound* energy does so, as is proven by comparing the fall of objects made of different materials. They have different percentages of bound energy. (Why?) For example, on the moon, where there is no air, cosmonauts dropped steel balls and feathers and found that they fell together, alongside each other. The independence on material composition has been checked over and over again, and no difference has ever been found. Challenge 543 n Ref. 253

Can light fall?

What about radiation? Light, like any radiation, is energy without rest mass. It moves like a stream of extremely fast and light objects. Therefore deviations from universal gravity become most apparent for light. How does light fall? Light cannot change speed. When light falls vertically, it only changes colour. But light can also change direction. 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 how he deflection angle depends on the mass of the body and the distance of passage. But nobody cared to check the result experimentally. Ref. 254 See page 99

Obviously, light has energy, and energy has weight; the deflection of light by itself is thus *not* a proof of the curvature of space. Also general relativity predicts a deflection angle for light passing masses, but of *twice* the classical value, because the curvature of space around large masses adds to the effect of universal gravity. The deflection of light thus only confirms the curvature of space if the *value* agrees with the one predicted by general relativity. And indeed, the observations coincide with the prediction. More details will be given shortly. See page 279

Can you show that the curvature of light near the earth is the same as the one of stones given by expression (188)? Challenge 544

In summary, all experiments show that not only mass, but also energy falls along geodesics, whatever its type, bound or free, and whatever the interaction, be it electromagnetic or nuclear. Moreover, the motion of radiation confirms that space-time is curved.

We thus find that *space-time tells matter, energy and radiation how to fall*. This statement is the second half of general relativity. It complements the first half, which states that energy tells space-time how to curve. To complete the description of macroscopic motion, we only need to add numbers to these statements, so that they become testable. As usual, we can proceed in two ways: we can deduce the equations of motion directly, or we can first deduce the Lagrangian and then deduce the equations of motion from it. But before we do that, 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

Curiosities about gravitation

General relativity is a beautiful topic with numerous interesting aspects.

- Take a plastic bottle and make some holes into it near the bottom. Fill the bottle with water, closing the holes with your fingers. If you let the bottle go, no water will leave the bottle during the fall. Can you explain how this experiment confirms the equivalence of rest and free fall? Challenge 545

- See page 216 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 earth? Challenge 546

- Acceleration of an object requires the use of power. There is a *maximum power* in nature, given by $P = c^5/4G$. It has the value $0.92 \cdot 10^{52}$ W. Power is flow of energy. If more power than the maximum value were used, a black hole would be created, and energy flow would be made impossible. Can you confirm that this maximum power in nature is consistent with the maximum accelereation mentioned above? Challenge 547

- A piece of wood floats on water. Does it stick out more or less in an elevator accelerating up? Challenge 548

- Challenge 549 How do cosmonauts weigh themselves to check whether they eat enough?

- 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} m/s^2 to 10^{-8} m/s^2, depending on the height of the orbit. When will an apple floating in space hit the wall of a space station? Challenge 550 n By the way, what is the magnitude of the tidal accelerations in this situation?

- See page 64 There is no negative mass in nature, as discussed in the beginning of our walk. This means that gravitation cannot be shielded, in contrast to 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 troubles this implies later on; for example, an *upper limit* for the entropy of physical systems will appear. See page 635

---

[*] 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.

- 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? Mathematically, this is possible; 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 statement that this does not happen in nature is also called the *weak energy condition*. Can you say whether it is included in the limit on length to mass ratios? Ref. 257 Challenge 551
- Like in special relativity, the length to mass limit $L/M \geqslant 4G/c^2$ is a challenge to devise experiments to overcome it. Can you explain what happens when an observer moves so rapidly past a mass that its length contraction reaches the limit? Challenge 552
- There is an important mathematical aspect 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 higher dimension. However, general relativity does not tell *why* space-time has three plus one dimensions. It is simply based on the fact. This deep and difficult question will be settled only in the third part of the mountain ascent.
- Henri Poincaré, who died in 1912, shortly before the general theory of relativity was finished, thought for a while that curved space was not a necessity, but only a possibility. He imagined that one could simply continue using Euclidean space and simply add that light follows curved paths. Can you explain why this project is impossible? Challenge 553
- Can two atoms hydrogen circle each other, in their respective gravitational field? What would the size of this ‘molecule’ be? Challenge 554 n
- Can two light pulses circle each other, in their respective gravitational field? Challenge 555 n
- The various motions of the earth mentioned in the section on Galilean physics, such as the rotation around its axis or the rotation around the sun lead to various types of time in physics and astronomy. The time defined by the best atomic clocks is called *terrestrial dynamical time*. By inserting leap seconds every now and then to compensate for the bad definition of the second (an earth rotation does not take 86 400, but 86 400.002 seconds) and, in minor ways, for the slowing of earths rotation, one gets the *universal time coordinate*. Then there is the time derived from this one by taking into account all leap seconds. One then has the – different – time which would be shown by a nonrotating clock in the centre of the earth. Finally, there is *barycentric dynamical time*, which is the time that would be shown by a clock in the centre of mass of the solar system. Only using this latter time satellites can be reliably steered through the solar system. In summary, relativity says goodbye to Greenwich mean time, as does British law, in one of a few cases were the law follows science. See page 82 See page 879 Ref. 258
- 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 side; 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 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 them. In addition, general relativity is essential for the so-called *global positioning system*, or GPS. This modern navigation tool** consists of 24 satellites with clocks flying around the world. Why does the system need general relativity to operate? Since both the satellites as well as any person on the surface of the earth travel in circles, we have $dr = 0$ and we can rewrite the Schwarzschild metric (181) as

Ref. 259

$$\left(\frac{di}{dt}\right)^2 = 1 - \frac{2GM}{rc^2} - r^2 \left(\frac{d\varphi}{dt}\right)^2 = 1 - \frac{2GM}{rc^2} - v^2 \quad . \tag{191}$$

Challenge 556 e

For the relation between satellite and earth time we then get

$$\left(\frac{dt_{\mathrm{sat}}}{dt_{\mathrm{earth}}}\right)^2 = \frac{1 - \frac{2GM}{r_{\mathrm{sat}}c^2} - \frac{v_{\mathrm{sat}}^2}{c^2}}{1 - \frac{2GM}{r_{\mathrm{earth}}c^2} - \frac{v_{\mathrm{earth}}^2}{c^2}} \quad . \tag{192}$$

Can you deduce how many microseconds a satellite clock runs fast every day, given that the GPS satellites turn around the earth every twelve hours? Since only three microseconds would give a position error of one kilometre after a single day, the clocks in the satellites are adjusted to run slow by the calculated amount. The necessary adjustments are monitored and confirm general relativity every single day within experimental errors.

Challenge 557 n

Ref. 260

- 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. 261

Challenge 558

- Could our impression that we live in 3 space dimensions be due to a limitation of our senses? How?

Challenge 559 n

- Can you estimate the effect of the tides on the colour of the light emitted by an atom?

Challenge 560

- The strongest possible gravitational field is the one of black small holes. 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 corresponding gravitational acceleration, assuming the neutron star has the same mass as the sun?

Ref. 262

Challenge 561

---

[*] 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

$$a = \frac{GM}{r^2} + f_2 \frac{GM}{r^2}\frac{v^2}{c^2} + f_4 \frac{GM}{r^2}\frac{v^4}{c^4} + f_5 \frac{Gm}{r^2}\frac{v^5}{c^5} + \cdots \tag{190}$$

where the numerical factors f_n are calculated from of order one. The first uneven terms are missing because of the reversibility of general relativistic motion, 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.

See page 278

For a *parametrized* post-Newtonian formalism, all factors f_n, including the uneven ones, are fitted through the data coming in; so far all these fits agree with the values predicted by general relativity.

[**] For more information, see the http://www.gpsworld.com web site.

- Light deflection changes the angular size δ of a mass M with radius r when observed at distance d. The effect leads to the pretty expression Ref. 263 Challenge 562 e

$$\delta = \arcsin\left(\frac{r\sqrt{1-R_S/d}}{d\sqrt{1-R_S/r}}\right) \quad \text{where} \quad R_S = \frac{2GM}{c^2} \quad . \tag{193}$$

What is the percentage of the surface of the sun an observer at infinity can see?We will come back to the issue in more detail shortly. Challenge 563 See page 347

- Much information about general relativity is available on the internet. As a good starting point for US-American material, see the http://math.ucr.edu/home/baez/relativity.html web site.

What is weight?

There is no way that a *single* (and point-like) observer can 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. For the same reason, objects in free fall have no weight.

Let us check the numbers. Obviously, an accelerating surface of the earth produces a weight for each body resting on it. The weight is proportional to the inertial mass. In other words, the inertial mass of a body is exactly identical to the gravitational mass. This is indeed observed, and to the highest precision achievable. Roland von Eötvös* performed many such high-precision experiments throughout his life, without finding any discrepancy. In these experiments, he used the fact that the inertial mass determines centrifugal effects and the gravitational mass determines free fall. Can you imagine how exactly he tested the equality? Ref. 248 Challenge 564

However, the mass equality is not a surprise. Remembering the definition of mass ratio as negative inverse acceleration ratio, independently of the origin of the acceleration, we are reminded 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 Galilean physics, and that the whole discussion is a red herring. See page 62 See page 100

The equality of acceleration and gravity allows us to tell the following story. Imagine to 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 we all feel like in a strongly accelerating car, pushed into direction opposite to the acceleration: for a short while, we feel heavier, until the elevator arrives at his destination.

---

[*] Roland von Eötvös (1848, Budapest–1919, Budapest), hungarian physicist. He performed many precision gravity experiments; among others, he discovered the effect named after him. The university of Budapest is named after him.

Why do apples fall?

> Vires acquirit eundo.
> Vergilius[*]

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. If you enjoy this way of seeing things, imagine an apple falling from a tress. In the moment it detaches, it stops being accelerated upwards by the branch. The apple can now enjoy the calmness of real rest. Our limited human perception calls this state of rest free fall. Unfortunately, the accelerating surface of the earth approaches mercilessly and, depending on the time the apple stayed at rest, the earth hits it with a corresponding velocity, leading to more or less severe shape deformation.

Falling apples also teach us not to be disturbed any more by the statement that gravity is the uneven running of clocks with height. In fact, this statement is *equivalent* to saying that the surface of the earth is accelerating upwards, as the discussion above showed.

Can this reasoning can be continued without limit? We can go on for quite a while; it is fun to show how the earth can be of constant radius even though its surface is accelerating upwards everywhere.However, the equivalence between acceleration and gravity ends as soon as *two* falling objects are studied. Any study of several bodies inevitably leads to the conclusion that *gravity is curved space-time*. Many aspects of this description can be understood with no or only a little mathematics. The next section will highlight some of the differences between universal gravity and general relativity, showing that only the latter description agrees with experiment. After that, a few concepts found the measurement of curvature are introduced and applied to the motion of objects and space-time. If the concepts get too involved for a first reading, skip these parts. Just continue with the sections the stars, cosmology and black holes, which again use little mathematics. Challenge 565

8. Motion in general relativity – bent light and wobbling vacuum

> I have the impression that Einstein understands relativity theory very well.
> Chaim Weitzmann, chemist, later first president of Israel

Before we tackle the details 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.

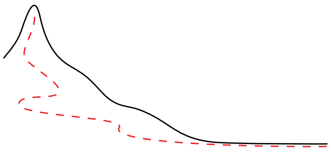
Weak fields

As mentioned above, one calls *strong* gravity all those situations for which the prediction by universal gravity strongly deviate from experiment. This happens when

$$\frac{2GM}{Rc^2} \approx 1 \qquad (194)$$

---

[*] 'Going it acquires strength.' Publius Vergilius Maro (Andes, 70 BCE–Brundisium, 19 BCE), *Aeneis* 4, 175.



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

and applies near black holes, as we will see below, or at extremely high energies, as we will discover in the third part of our mountain ascent. For most of nature, gravity is a *weak* effect, despite the violence of avalanches or of falling asteroids, and the number just mentioned much smaller than one. On the earth it is about 10^{-9}. 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; they allow to mention space-time curvature only in passing and to continue thinking about gravity as a source of acceleration. 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 gravity, but do appear in general relativity. Ref. 269

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 to the description of universal gravity. Are you able to deduce this effect from the figure? Challenge 567

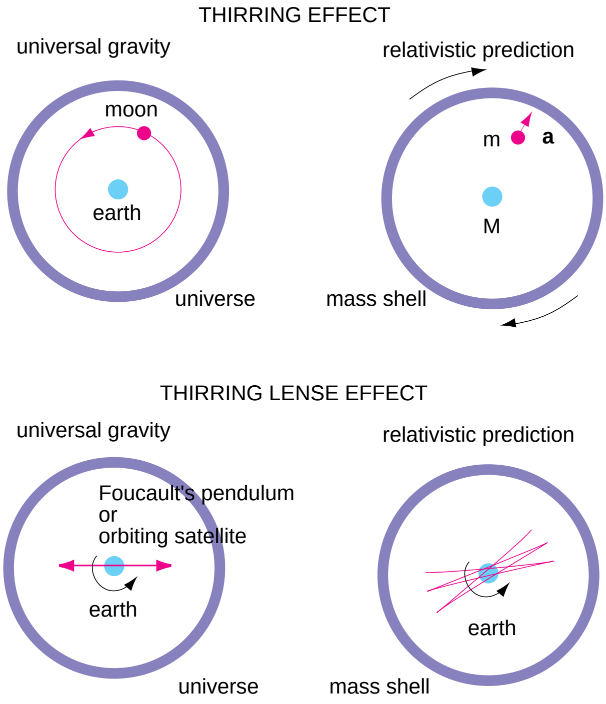


Figure 120 The Thirring and the Thirring-Lense effects

The *Thirring-Lense effect* is somewhat more complex. It predicts that an 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, appears because 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. For the same reason, the effect also moves the plane of an orbiting satellite.

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 beat by many years the efforts of much larger but much more sluggish groups.* The results confirm the tiny predictions by general relativity within about 25%. Ref. 270

Frame dragging effects have also been measured in binary star systems. This 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 that such subtle effects as frame dragging do take place. Ref. 271

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 nor 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; in addition, electromagnetism provides a good feeling on how the two fields behave. 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 See page 370

$$m\ddot{\mathbf{x}} = q\mathbf{E} - q\dot{\mathbf{x}} \times \mathbf{B} \quad . \tag{195}$$

In other words, change of speed is due to electric fields **E**, whereas magnetic fields **B** are those fields which give a velocity-dependent direction change of velocity, *without* changing the speed itself. The changes depend on the value of the charge q. In the case of gravity this expression, as we shall show below, becomes

$$m\ddot{\mathbf{x}} = m\mathbf{G} - m\dot{\mathbf{x}} \times \mathbf{H} \quad . \tag{196}$$

The role of charge is taken by mass. In this expression we already know the field **G**, given by

$$\mathbf{G} = \nabla\varphi = \nabla\frac{GM}{r} = -\frac{GM\mathbf{x}}{r^3} \quad . \tag{197}$$

---

[*] 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.

As usual, the quantity φ is the (scalar) potential. This is the field of universal gravity, as produced by every mass, and in this context is called the *gravitoelectric field*.

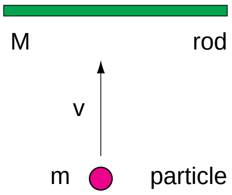


Figure 121 The reality of gravitomagnetism

In fact it is not hard to show that if gravito*electric* fields exist, gravito*magnetic* fields must exist as well, as the latter appear whenever one changes from an observer at rest to a moving one. Ref. 272 (The same argument is used in electrodynamics.) A particle falling perpendicularly towards an infinitely long rod already makes the point, as shown in Figure 121. 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 Challenge 568

$$m\mathbf{a} = -m\mathbf{v}\times\mathbf{H} \qquad (198)$$

where, as in electrodynamics, a static *gravitomagnetic field* obeys

$$\mathbf{H} = \nabla\times\mathbf{h} = 16\pi N\rho\mathbf{v} \qquad (199)$$

where ρ is mass density and N is a proportionality constant. The quantity **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 569 When the situation in Figure 121 is evaluated,one gets

$$N = \frac{G}{c^2} = 7.4\cdot 10^{-28}\,\mathrm{m/kg} \quad , \qquad (200)$$

an extremely small value. In addition, a second point renders the observation extremely difficult. In contrast to electromagnetism, in the case of gravity there is no way to observe Challenge 570 n *pure* gravitomagnetic fields (why?); they are always mixed with the usual, gravitoelectric ones. And as in the electrodynamic case, the gravitoelectric fields 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 any case, *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

$$\mathbf{H} = \nabla\times\mathbf{h} = \nabla\times\left(-2\frac{\mathbf{J}\times\mathbf{x}}{r^3}\right) \qquad (201)$$

exactly as in the electrodynamic case. In particular, like in electromagnetism, a spinning test particle with angular momentum **S** feels a *torque* if it is near a large spinning mass with angular momentum **J**. And obviously, this torque **T** is given by

$$\mathbf{T} = \frac{d\mathbf{S}}{dt} = \tfrac{1}{2}\mathbf{S}\times\mathbf{H} \quad . \qquad (202)$$

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 axis a and eccentricity e, Challenge 571

$$\dot{\mathbf{\Omega}} = -\frac{\mathbf{H}}{2} = -\frac{\mathbf{J}}{|\mathbf{x}|^3} + \frac{3(\mathbf{J}\mathbf{x})\mathbf{x}}{|\mathbf{x}|^5} = \frac{2\mathbf{J}}{a^3(1-e^2)^{3/2}} \qquad (203)$$

which is the prediction by Thirring and Lense.* 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.

The use of the split into gravitoelectric and gravitomagnetic effects corresponds to an approximation in which mainly clock effects of gravity are taken into account, neglecting spatial curvature. Despite this limitation, the approach is 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?To find the answer, thinking about the electromagnetic analogy can help. In addition, the split of the gravitational field into gravitoelectric and gravitomagnetic components also allows a simple description of gravity waves. Challenge 573

Gravitational waves

Table 28 The expected spectrum of gravitational waves

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

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

Jørgen Kalckar and Ole Ulfbeck have given a simple argument for the logical necessity of gravitational waves by studying two equal masses falling towards each other. They simply Ref. 273

Challenge 572 * A homogeneous spinning sphere has an angular momentum givenby $J = \frac{2}{5} M\omega R^2$.

imagined a spring between them. Such a spring will make them bounce towards each other again and again. The central spring stores the kinetic energy from the falling masses. That energy can be measured by determining, with a meter 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 gradually slow down the masses, until they again fall towards each other, thus starting the same cycle again.

However, the energy stored in the spring must get smaller with each cycle. At every bounce, the spring is compressed a little less. Whenever a sphere detaches from the spring, it obviously is decelerated by the other sphere due to the gravitational attraction. Now comes the point. The value of this deceleration depends on the distance to the other mass; but since there is a maximal propagation velocity, the effective deceleration is given by the distance the other mass *had* when its gravity reached the end of the spring. In short, while departing, the real deceleration is *larger* than the one calculated without taking the time delay into account.

Similarly, when a mass falls back down towards the other, it is accelerated by the other mass according to the distance it had when its gravity effect reached the other. Therefore, while going down, the acceleration is *smaller* than the one without time delay.

As a total effect, the masses arrive with a *smaller* energy than they departed with. The difference of these two energies is lost by each mass; it is taken away by space-time. The energy is radiated away as gravitational radiation. As we will see, this effect has already been measured, with the difference that the two masses, instead of being tied by a spring, where orbiting each other.

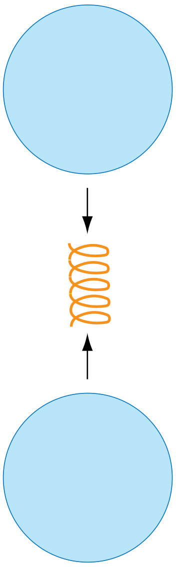


Figure 122 A Gedankenexperiment showing the necessity of gravity waves

The same story is told by mattresses. We remember that a mass deforms the space around it like a metal ball on a mattress deforms the surface around it. However, in contrast to mattresses, there is no friction between the ball and the mattress. If two metal balls continuously bang onto each other and then depart again, until they come back together, they will send out surface waves on the mattress.

A simple mathematical description of gravity waves appears with the split into gravito-Ref. 274 magnetic and gravitoelectric effects. It does not take much to extend gravitomagnetostatics and gravitoelectrostatics to *gravitodynamics*. Just as electrodynamics can be deduced from Coulomb's attraction when one switches to other inertial observers, gravitodynamics can be Challenge 574 deduced from universal gravity.One gets the four equations

$$\nabla \mathbf{G} = -4\pi G\rho \quad , \quad \nabla \times \mathbf{G} = -\frac{\partial \mathbf{H}}{\partial t}$$
$$\nabla \mathbf{H} = 0 \quad , \quad \nabla \times \mathbf{H} = -16\pi G\rho\mathbf{v} + \frac{N}{G}\frac{\partial \mathbf{G}}{\partial t} \qquad (204)$$

which except for a factor 16 instead of 4 in the last equation, are the same as those for electrodynamics.* You can easily deduce a *wave equation* for the gravitoelectricand the gravitomagnetic fields **G** and **H**. In other words, *gravity can behave like a wave; gravity can radiate*. All this follows from the expression of universal gravity when applied to moving observers, with the requirement that neither observers nor energy can move faster than c. The story with the spring and the mathematical story use the same assumptions and arrive to the same conclusion. Challenge 575

A few manipulations show that the speed of these waves is given by Challenge 576 e

$$c = \sqrt{\frac{G}{N}} \tag{205}$$

which corresponds to the famous electromagnetic expression See page 394

$$c = \frac{1}{\sqrt{\varepsilon_o \mu_o}} \quad . \tag{206}$$

The same letter has been used for the two speeds, as they are identical. Indeed, both influences travel with the speed common to all energy with vanishing rest mass.

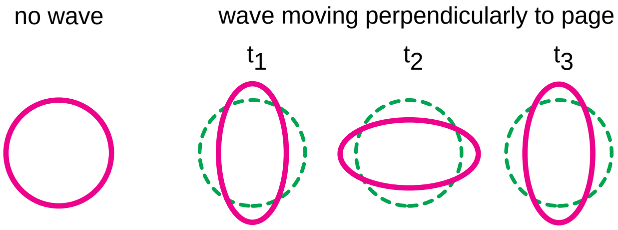


Figure 123 Effects on a circular or spherical body of a plane gravitational wave moving vertically to the page

How does one have to imagine these waves? We said that a gravity wave corresponds to a surface wave of the mattress. Gravitational waves are moving and oscillating deformations of space. As in the case of mattress waves, it turns out that gravity waves are *transverse*. Thus they can be polarized. (Mattress waves cannot, because they miss one dimension.) Gravity waves can be polarized in two independent ways. The effect of a gravitational wave in one polarization is shown in Figure 123. The effect of the other polarization is the same,rotated by 45 degrees.** Can you imagine what happens to the circular body if a circularly polarized wave hits it? Challenge 577 Challenge 578

---

* The additional factor reflects the property that the angular momentum to energy ratio (the ‘spin’) of gravity waves is different from that of electromagnetic waves. Gravity waves have spin 2, whereas electromagnetic waves have spin 1.

** A (small amplitude) plane gravity wave travelling in z-direction is described by a metric

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1+h_{xx} & h_{xy} & 0 \\ 0 & h_{xy} & -1+h_{xx} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{207}$$

How does one produce such waves? The conservation of energy does not allow changing mass monopoles. Also a spherical mass whose radius oscillates would not emit gravita- Challenge 579 tional waves. The conservation of momentum does not allow changing mass dipoles.As a result, only *changing quadrupoles* can emit waves. For example, two masses in orbit around each other will emit gravitational waves. Also any rotating object which is not cylindrically symmetric around the rotation axis will do so: rotating an arm leads to gravitational wave emission. Most of these relations also apply for masses on mattresses. Are you able to point Challenge 580 out the differences?

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:

$$h_{ab} = \frac{2G}{c^4}\frac{1}{r} d_{tt} Q^{\text{ret}}_{ab} = \frac{2G}{c^4}\frac{1}{r} d_{tt} Q_{ab}(t - r/c) \quad . \tag{211}$$

The expression shows that the amplitude of gravity waves *decreases only with* $1/r$, in contrast to naive expectations. However, this feature is the same as for electromagnetic waves. In addition, the small value of the prefactor, $1.6 \cdot 10^{-44}\,\mathrm{Wm/s}$, shows that truly gigantic systems are needed to produce quadrupole moment changes that yield any detectable length Challenge 581 variations in bodies. To be convinced, just insert a few numbers, keeping in mindthat the best present detectors are able to measure length changes down to $h = \delta l/l = 10^{-19}$. The production of sizeable gravitational waves by humans is (probably) impossible.

Gravitational waves, like all other waves, transport energy.* We specialize the general formula for the emitted power P to the case of two masses m_1 and m_2 in circular orbits Ref. 237 around each other at distance l and get

$$P = -\frac{dE}{dt} = \frac{G}{45c^5}\,\dddot{Q}^{\text{ret}}_{ab}\,\dddot{Q}^{\text{ret}}_{ab} = \frac{32}{5}\frac{G}{c^5}\left(\frac{m_1 m_2}{m_1 + m_2}\right)^2 l^4 \omega^6 \tag{212}$$

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

$$P = \frac{32}{5}\frac{G^4}{c^5}\frac{(m_1 m_2)^2 (m_1 + m_2)}{l^5} \tag{213}$$

---

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

$$h_{ab} = A_{ab} \sin(kz - \omega t + \varphi_{ab}) \tag{208}$$

as in all plane harmonic waves. The dispersion relation resulting from the wave equation is

$$\frac{\omega}{k} = c \tag{209}$$

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

In another gauge, a plane wave can be written as

$$g = \begin{pmatrix} c^2(1+2\varphi) & A_1 & A_2 & A_3 \\ A_1 & -1+2\varphi & h_{xy} & 0 \\ A_2 & h_{xy} & -1+h_{xx} & 0 \\ A_3 & 0 & 0 & -1 \end{pmatrix} \tag{210}$$

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 vec- Ref. 272 tor. It is as easy to define and use as in the electrodynamic case.

For elliptical orbits, the rate increases with the ellipticity.* Inserting the values in the case of the earth and the sun, we get a power of about 200 W, and a value of 400 W for the Jupiter-sun system. These values are so small that their effect cannot be detected at all.

For orbiting systems, the frequency of the waves is twice the orbital frequency, as you might want to check.These low frequencies make it difficult to detect them. Challenge 582

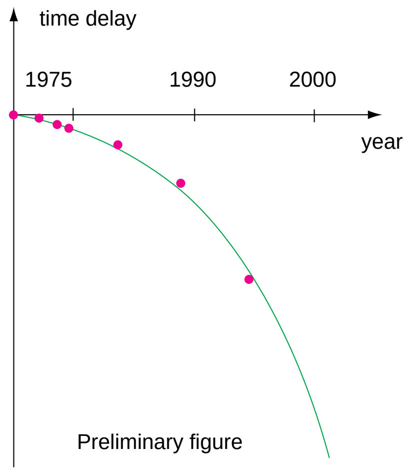


Figure 124 Comparison between measured time delay in the periastron of the binary pulsar PSR 1913+16 and the prediction due to energy loss by gravitational radiation

As a result, the only observation of effects of gravitational waves to date is in binary pulsars. Pulsars are extremely small stars; even with a mass equal to that of the sun, their diameter is only about 10 km; thus they can orbit each other at small distances and high speeds. Indeed, in the most famous binary pulsar system, PSR 1913+16, the two stars orbit each other in an amazing 7.8 h, even though their semimajor axis is about 700 Mm, just less than twice the earth-moon distance. Since their orbital speed is up to 400 km/s, the system is noticeably relativistic.

Pulsars have a useful property: due 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 system just mentioned. Ref. 276 Eliminating all other effects, and collecting data for 20 years, they found a slowing down of the orbital frequency shown in Figure 124. The slowdown is due to gravity wave emission. The results exactly fit the prediction by general relativity, *without any adjustable parameter*. (You might want to check that the effect must be quadratic in time.)This is the only case so far that general relativity has been Challenge 583 tested up to the $(v/c)^5$ precision. See page 267 To get an idea of the precision, this experiment detected a reduction of the orbit diameter of 3.1 mm per orbit, or 3.5 m per year! Ref. 276 The measurements were possible only because the two stars in this system are pulsars with small size, large velocities, and purely gravitational interactions. The pulsar rotation period around its axis, about 59 ms, is known with eleven digits of precision, the orbital time of 7.8 h is known to ten digits and the eccentricity of the orbit with 6 digits. Ref. 237

The *direct* detection of gravitational waves is one of the aims of experimental general relativity. The race is on since the 1990s. The basic idea is simple and taken from Figure 123: take four bodies for which the line connecting one pair is perpendicular to the line connecting the other pair. Then measure the distance changes of each pair. If a gravitational wave comes by, one pair will increase in distance and the other will decrease, at the *same* time.

Ref. 237 * See e.g. the explanation by Goenner.

** In 1993 he shared the Nobel prize in physics for his life's work.

Since gravitational waves cannot be produced in sufficient strength by humans, wave detection first of all requires a lot of time to wait for a strong enough wave to come by. Secondly, a system able to detect length changes of the order of 10^{-22} or better is needed – in other words, a lot of money. Any detection for sure will make the news in television.*

Figure to be included

Figure 125 Detection of gravitational waves

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 296 described here, the full field equations are used. More about the topic shortly.

By the way, if all change is due to motion of particles, as the Greeks maintained, how do
Challenge 585 gravity waves fit into the picture?

Light and radio wave bending

As we know 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 imagination that this effect will thus *increase* the bending of light near masses already deduced in 1801 by Soldner for universal gravity.

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
Ref. 278 value of the bending is very small. The angle of deflection α, to first order, is simply

$$\alpha = \int_{-\infty}^{\infty} \frac{\partial c}{\partial x} dy \tag{214}$$

Challenge 586 as you might want to confirm.The next step is to use the Schwarzschild metric

$$d\tau^2 = (1 - \frac{2GM}{rc^2})dt^2 - \frac{dr^2}{(c^2 - \frac{2GM}{r})} - \frac{r^2}{c^2}d\varphi^2 \tag{215}$$

Challenge 587 and transform it into (x, y) coordinates to first order. That gives

$$d\tau^2 = (1 - \frac{2GM}{rc^2})dt^2 - (1 + \frac{2GM}{rc^2})\frac{1}{c^2}(dx^2 + dy^2) \tag{216}$$

Ref. 277 * The topic of gravity waves is full of interesting sidelines. For example, can gravity waves be used to power a
Challenge 584 rocket? Yes, say Bonnor and Piper. You might ponder the possibilityyourself.

which again to first order leads to

$$\frac{\partial v}{\partial x} = (1 - \frac{2GM}{rc^2})c \quad . \tag{217}$$

It confirms what we know already, namely that far away observers see light *slowed down* when passing near a mass. Thus we can also speak of a height dependent index of refraction. In other words, constant *local* light speed leads to a *global* slowdown. This effect will be play a role again shortly.

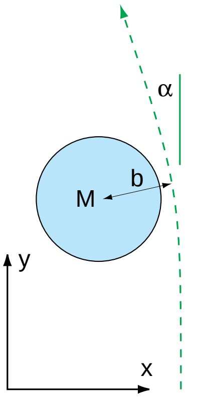


Figure 126 Calculating the bending of light by a mass

Inserting the last result in (214) and using a smart substitution, we get

Challenge 588

$$\alpha = \frac{4GM}{c^2}\frac{1}{b} \tag{218}$$

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, the result is the famous value of 1.75″, which was confirmed by the measurement of 1919. This experiment made Einstein famous, as it 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 two expeditions that took place before his correct calculations had failed.

See page 99

Ref. 279

See page 99

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. In the mean time, over a dozen independent experiments, using radio sources in the sky which lie on the path of the sun, confirmed general relativity's prediction within a few per cent.

Ref. 259, 236, 237

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?)We thus follow that if light would not be bent near a mass, it would go faster than c![*]

See page 286

Challenge 589

So far, bending of radiation has also been 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 yet, 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 is taking precision pictures of the night sky, is analysed properly in the future.

See page 325

---

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

Challenge 590 e

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, Ref. 280 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. 281 The first measurement was published in 1968, and directly confirmed the prediction by general relativity within experimental errors. All further tests up to the present have confirmed the prediction within experimental uncertainties, which nowadays are of the order of Ref. 282 one part in a thousand. The delay has now been measured in binary pulsars, as there are a few such systems for which the line of sight lies almost precisely in the orbital plane.

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 $a_{\mathrm{grav}} = a_{\mathrm{centri}}$, in other words, by setting $GM/r^2 = \omega^2 r$ and fixing energy and angular momentum. The mass of the particle does not appear explicitly.

In general relativity, the mass of the particle is made to disappear by rescaling energy Ref. 236, 237 and angular momentum as $e = E/mc^2$ and $j = J/m$. Next, the space curvature needs to See page 258 be included. We 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 Challenge 591 e time τ and time t at infinity:

$$\frac{dt}{d\tau} = \frac{e}{1-2GM/rc^2} \quad , \tag{219}$$

whereas the initial condition on the angular momentum J and its conservation implies that

$$\frac{d\varphi}{d\tau} = \frac{j}{r^2} \quad . \tag{220}$$

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

$$(\frac{dr}{cd\tau})^2 + V^2(j,r) = e^2 \tag{221}$$

where the effective potential V is given by

$$V^2(J,r) = (1-\frac{2GM}{rc^2})(1+\frac{j^2}{r^2c^2}) \quad . \tag{222}$$

Challenge 592 The expression differs slightly from the one in universal gravity,as you might want to check.

Challenge 593 e We now need to solve for $r(\varphi)$. For *circular* orbits we get *two* possibilities

$$r = \frac{6GM/c^2}{1 \pm \sqrt{1-12(\frac{GM}{cj})^2}} \tag{223}$$

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

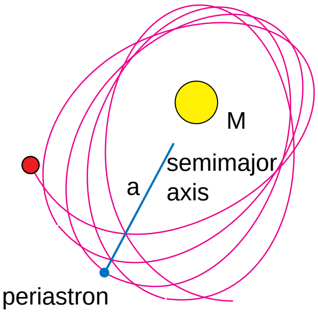


Figure 127 The orbit around a central body in general relativity

What is the situation for *elliptical* orbits? Setting $u = 1/r$ in (221) and differentiating, the equation for $u(\varphi)$ becomes

$$u' + u = \frac{GM}{j^2} + \frac{3GM}{c^2}u^2 \quad . \tag{224}$$

Without the nonlinear correction due to general relativity on the far right, the solutions are the famous *conic sections* Challenge 594 e

$$u_0(\varphi) = \frac{GM}{j^2}(1+\varepsilon\cos\varphi) \tag{225}$$

i.e. ellipses, parabolas, or hyperbolas, depending on the value of the parameter ε, the so-called *eccentricity*. We know their shapes from universal gravity. General relativity introduces the nonlinear term in (224). Thus the solutions are not conical sections any more; however, as the correction is small, a good approximation is given by Challenge 595 e

$$u_1(\varphi) = \frac{GM}{j^2}[1+\varepsilon\cos(\varphi - \frac{3G^2M^2}{j^2c^2}\varphi)] \quad . \tag{226}$$

The hyperbolas and parabolas of universal gravity are slightly deformed. Instead of elliptical orbits we get the famous rosetta path shown in Figure 127. Such a path is first of all characterized by a periastron shift. The *periastron*, or *perihelion* in the case of the sun, is the furthest point reached by an orbiting body. The periastron turns around the central body withan angle Challenge 596 e

$$\alpha \approx 6\pi\frac{GM}{a(1-\varepsilon^2)c^2} \tag{227}$$

for every orbit, where a is the *semimajor axis*. The angle has been measured for the orbits of Mercury, Icarus, Venus and Mars around the sun, as well as for several binary star systems. In all cases, expression (227) 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 his calculation lead him to exactly that value, he was overflowing with joy for many days.

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

In binary pulsars, the periastron shift can be as large as several degrees per year. Measurements have confirmed the predicted values to the full available precision. Ref. 282

However, even the rosetta orbit is not really stable, due to the emission of gravitational waves. But in the solar system, the power lost this way is completely negligible even over thousands of millions of years, as we saw above, so that the rosetta path is a good description.

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. This effect exists only in general relativity. The angle α describing the direction change is given by

$$\alpha = 2\pi\left(1-\sqrt{1-\frac{3Gm}{rc^2}}\right) \approx \frac{3\pi Gm}{rc^2} \quad . \qquad (228)$$

The angle change is called the *geodesic effect* – 'geodetic' in other languages. It is a further consequence of the split into gravitoelectric and gravitomagnetic fields, as you may want to show.Obviously, it does not exist in universal gravity. Challenge 597 e

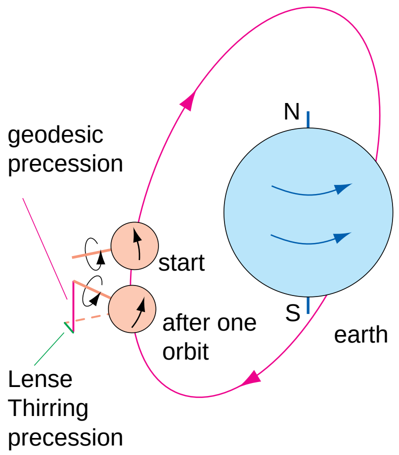


Figure 128 The geodesic effect

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

Ref. 283 The geodesic effect, or geodesic precession, was predicted by de Sitter in 1916; in particular, he proposed to detect that the earth moon-system would change its pointing direction in its fall around the sun. The effect is tiny; for the axis of the moon the precession angle is about 0.019 arcsec per year. The effect was first detected in 1987 by an Italian team for the earth-moon system, through a combination of radiointerferometry and lunar ranging, making use of the cat-eyes deposed by the cosmonauts on the moon. Experiments to detect it in artificial satellites are also under way. Ref. 284

At first sight, geodetic precession is similar to the Thomas precession found in special relativity. In both cases, a transport along a closed line results in the loss of the original direction. However, a careful investigation shows that Thomas precession can be *added* to geodesic precession by applying some additional, non-gravitational interaction, so that the analogy is shaky. See page 220

We now stop with the discussion of the weak gravity effects and return to the general case of relativistic motion. We now turn to strong gravity effects, where curvature cannot be neglected, and where there is more fun.

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 we roll it into a cone or a cylinder, it gets what is called *extrinsic curvature*; however, the sheet of paper still looks flat for any two-dimensional animal living on it – as approximated by an ant walking over it. In other words, the *intrinsic curvature* of the sheet of paper is zero even if the sheet as a whole is extrinsically curved. (Can a one-dimensional space have intrinsic curvature? Is a torus internally curved?) Challenge 598 n

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 we talk about curvature in general relativity, we always mean *intrinsic* curvature, since any observer in nature is by definition in the same situation as an ant on a surface: their experience, their actions and plans always only concern their closest neighbourhood in space and time.

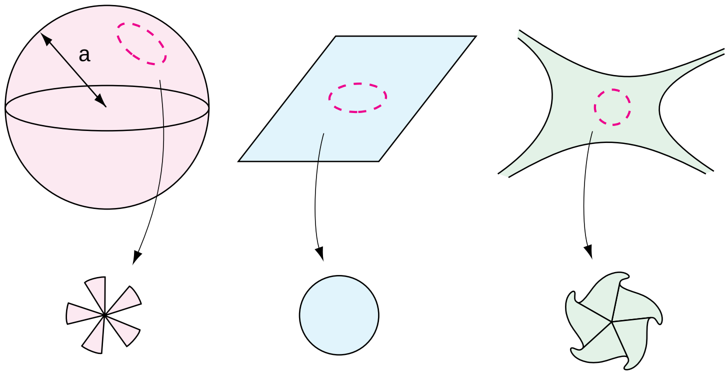


Figure 129 Positive, vanishing and negative curvature in two dimensions

But how precisely can an ant determine whether it lives on an intrinsically curved surface?* One way is shown in Figure 129. 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 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 two-dimensional surface is intrinsically curved whenever ironing is not sufficient to make a flat street map out of it.

This means that we 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

---

* Note that the answer to this question also tells how to 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.

a surface with diverging parallels, such as a saddle, is said to have *negative* curvature. In short, positive curvature means that we are more restricted in our movements, negative that we are not. Constant curvature even implies to be locked in. You might want to check this with Figure 129.

The third way to measure curvature uses triangles. On curved surfaces the sum of angles
Ref. 264 in a triangle is either larger or smaller than π. 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
Challenge 599 e circle on a sphere, the measured radius r, circumference C, and area A are relatedby

$$C = 2\pi r\left(1 - \frac{K}{6}r^2 + \ldots\right) \quad \text{and} \quad A = \pi r^2\left(1 - \frac{K}{12}r^2 + \ldots\right) \tag{229}$$

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:

$$K = 6\lim_{r\to 0}\left(1 - \frac{C}{2\pi r}\right)\frac{1}{r^2} \quad \text{or} \quad K = 12\lim_{r\to 0}\left(1 - \frac{A}{\pi r^2}\right)\frac{1}{r^2} \quad . \tag{230}$$

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, like the inside of any torus. Also a saddle is an example for the latter case, but, unlike the torus, 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;
Challenge 601 e one needs at least four dimensions, as you can find out if you try to imagine the situation.

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 in the 18th century. You might want to check this with a tea cup, or with a sculpture by Henry Moore, or with any other curved object from your surroundings, such as
Challenge 602 e aVolkswagen Beetle. The Gaussian curvature is in fact the product of the two corresponding inverse curvature radii. Thus, even though line curvature is not an intrinsic property, this special product is. Physicists are thus particularly interested in Gaussian curvature – and its higher-dimensional analogies.

For *three*-dimensional objects, the issue is a bit more involved. First of all, we have difficulties imagining the situation. But we can still visualize that the curvature of a small disk around a point will depend on its orientation. 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*. We can

---

[*] If the n-dimensional volume of a sphere is written as $V_n = C_n r^n$ and the n-dimensional surface as $O_n =$
Ref. 285 $nC_n r^{n-1}$, one can generalize the expressions to

$$K = 3(n+2)\lim_{r\to 0}\left(1 - \frac{V_n}{C_n r^n}\right)\frac{1}{r^2} K = 3n\lim_{r\to 0}\left(1 - \frac{O_n}{nC_n r^{n-1}}\right)\frac{1}{r^2} \quad \text{or} \tag{231}$$

Challenge 600 as shown by Vermeil. A famous riddle is to determine C_n.

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 603

$$K = 9 \lim_{r\to 0} (1 - \frac{A}{4\pi r^2}) \frac{1}{r^2} = 18 \lim_{r\to 0} \frac{r - \sqrt{A/4\pi}}{r^3} = 18 \lim_{r\to 0} \frac{r_{\text{excess}}}{r^3} \quad . \qquad (232)$$

Defining the *excess radius* as $r - \sqrt{A/4\pi}$, we get 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, we thus have to specify, as for any tensor in three dimensions, the main curvatures in three orthogonal directions.*

Challenge 604

What are the curvature values in practice? Already in 1827, the mathematician and physicist Friedrich Gauss checked whether the three angles between three mountain peaks near where he lived 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

$$\delta = \pi - (\alpha + \beta + \gamma) \approx A_{\text{triangle}} \frac{GM}{r^3 c^2} \quad . \qquad (233)$$

For the case of the earth and typical mountain distances, the angle is of the order of 10^{-14} rad. Gauss had no chance to detect any deviation, and in fact he didn't. Even with lasers and high precision set-ups, no deviation has been detected yet – on earth. The right-hand factor, which measures the curvature of space-time on the surface of the earth, is too small. But Gauss did not know, as we do today, that gravity and curvature go hand in hand.

Curvature and space-time

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

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 a central requirement in 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

---

* 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 $\det R/\det g$.

** 'Our head is round in order to allow our thougths to change direction.' Francis Picabia (1879, Paris – 1953, Paris) French dadaist and surrealist painter.

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.

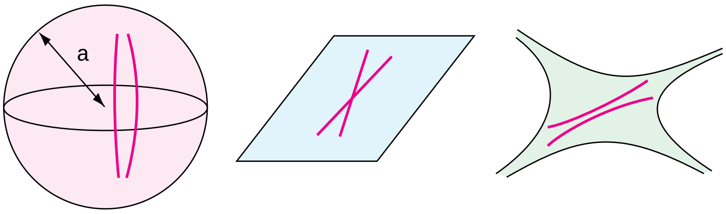


Figure 130 Curvature (in two dimensions) and geodesic behaviour

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 130 shows that the curvature K also describes how 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.

Challenge 605 e If a space is curved, the separation s will increase along the geodesics as

$$\frac{d^2 s}{dl^2} = -Ks + \text{higher orders} \tag{234}$$

where l measures the length along the geodesic, and as above K is the inverse square curvature radius. In space-time, this relation is extended by substituting proper length with proper time (times the speed of light). Thus separation and curvature are related by

$$\frac{d^2 s}{d\tau^2} = -Kc^2 s + \text{higher orders} \quad . \tag{235}$$

This turns out to be the definition of an acceleration. In other words, what in the purely spatial case is described by *curvature*, in the case of space-time becomes the *relative acceleration* of two particles freely falling from nearby points. But we encountered these accelerations already: they describe tidal effects. In short, space-time curvature and tidal effects are precisely the same. See page 97

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. The definition also shows that K is a tensor, so 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 so-called *sectional* curvature values does *not* depend on the observer. (This corresponds to the tensor trace.) Can you confirm this, by using the definition of the curvature just given?

Challenge 606

Ref. 265

Challenge 607

The sum of three sectional curvatures defined for mutually orthogonal planes $K_{(12)}$, $K_{(23)}$, and $K_{(31)}$, is related to the excess radius defined above. Can you find out how? Challenge 608

If a surface has *constant* (intrinsic) curvature, i.e. the same curvature at all locations, geometrical objects can be moved around without deforming them. Can you picture this? Challenge 609 e

In summary, curvature is not such a difficult concept. It describes the *deformation* of space-time. If we imagine 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 a hundred years it was a question of faith for every physicist to say that the vacuum is empty.) Picturing vacuum as a substance can help imagination in many ways in understanding general relativity.

Curvature and motion in general relativity

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

$$K_{(12)} + K_{(23)} + K_{(31)} = \frac{8\pi G}{c^4} W^{(0)} \tag{236}$$

where $W^{(0)}$ is the *proper* energy density at the point, *and* this statement is valid for *every* observer. This is all of general relativity in one paragraph.

An equivalent way to describe the expression is easily foundusing the excess radius defined above, and introducing the mass M by $M = VW^{(0)}/c^2$. We get Challenge 610 e

$$r_{\text{excess}} = r - \sqrt{A/4\pi} = \frac{G}{3c^2} M \quad . \tag{237}$$

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, both space curvature, and, as we will see shortly, space-time curvature *change* over space and time.

Ref. 266 * Another, equivalent way is to say that for small radii

$$A = 4\pi r^2 (1 + \frac{1}{9} r^2 R) \tag{238}$$

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

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

The quantities appearing in expression (236) are *independent* of the observer. But often people want to use observer-dependent quantities. The relation then gets more involved; the single equation (236) 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, the temporal curvatures (234) can be defined as the second spatial Challenge 611 e derivatives of a single scalar function φ via

$$K_{(0j)} = \frac{\partial^2\varphi}{\partial(x^j)^2} \quad . \tag{239}$$

Universal gravity is the description of general relativity for small speeds and small spatial curvature. Both limits imply, taking $W^{(0)} = \rho c^2$ and using $c \to \infty$, that

$$K_{(ij)} = 0 \quad \text{and} \quad \mathbf{K}_{(01)} + \mathbf{K}_{(02)} + \mathbf{K}_{(03)} = 4\pi G\rho \quad . \tag{240}$$

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 relation (236) between curvature and energy density indeed means Challenge 612 that time near a mass depends on the height, as stated in the beginning of this chapter?

The Schwarzschild metric

Ref. 265 What is the curvature of space-time near a spherical mass?

– CS – to be inserted – CS –

Challenge 613 The curvature of the Schwarzschild metric is givenby

$$\begin{aligned} K_{r\varphi} = K_{r\theta} = -\frac{G}{c^2}\frac{M}{r^3} \quad &\text{and} \quad K_{\theta\varphi} = 2\frac{G}{c^2}\frac{M}{r^3} \\ K_{t\varphi} = K_{t\theta} = \frac{G}{c^2}\frac{M}{r^3} \quad &\text{and} \quad K_{tr} = -2\frac{G}{c^2}\frac{M}{r^3} \end{aligned} \tag{241}$$

Ref. 265 everywhere. The dependence on $1/r^3$ follows from the general dependence of all tidal effects; See page 97 we had calculated them in the chapter on universal gravity. The factors c^2 are due to the unity of space-time, and only the numerical prefactors need to be calculated from Challenge 614 general relativity. The average curvature obviously vanishes.As expected, the values of the curvatures near the surface of the earth is exceedingly small.

Curiosities and fun challenges

- A fly has landed on the outside of a cylindrical glass, 1 cm below its rim. A drop of honey is located halfway around the glass, also on the outside, 2 cm below the rim. What is the shortest distance to the drop? Challenge 615 e

What is the shortest distance, if the drop is on the *inside* of the glass? Challenge 616 e

- Where are the points of highest and lowest Gaussian curvature on an egg? Challenge 617 e

– CS – more to come – CS –

All observers: heavier mathematics

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

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

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

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 of curvature is the one describing the distinction between the 4-volume calculated from a measured radius and the actual

---

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

[**] This section might be skipped at first reading. The section on cosmology, on page 313, then is the right point to continue.

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

4-volume. This is the *average curvature* at a space-time point and is described by the so-called *Ricci scalar R* defined as

$$R = -2K = -\frac{2}{r^2_{\text{curvature}}} \quad . \qquad (242)$$

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:

$$R = R^{\lambda}{}_{\lambda} = g^{\lambda\mu} R_{\lambda\mu} \quad . \qquad (243)$$

The Ricci scalar, describing the curvature averaged over space *and* time. In the image of a falling spherical cloud, the Ricci scalar describes the volume change of the cloud. The Ricci scalar always vanishes in vacuum. This result allows for example, on the surface of the earth, to relate the spatial curvatures and the changes of time with height. Challenge 618

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 (for vanishing cosmological constant) as

$$G_{ab} = R_{ab} - \frac{1}{2} g_{ab} R \quad . \qquad (244)$$

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:

$$G_{00} = K_{(12)} + K_{(23)} + K_{(31)} \quad . \qquad (245)$$

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, we have

$$G_{11} = K_{(02)} + K_{(03)} - K_{(23)} \quad . \qquad (246)$$

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 619

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

$$d\mathbf{p} = T\ d\mathbf{A} \quad . \tag{247}$$

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:

$$T = \left(\begin{array}{c|ccc} w & S_1 & S_2 & S_3 \\ \hline S_1 & t_{11} & t_{12} & t_{13} \\ S_2 & t_{21} & t_{22} & t_{23} \\ S_3 & t_{31} & t_{32} & t_{33} \end{array}\right) = \left(\begin{array}{c|c} \text{energy} & \text{energy flow density, or} \\ \text{density} & \text{momentum density} \\ \hline \text{energy flow or} & \text{momentum} \\ \text{momentum density} & \text{flow density} \end{array}\right) \tag{248}$$

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

The vanishing divergence, often written as

$$\partial_a T^{ab} = 0 \quad \text{or abbreviated} \quad T^{ab}{}_{,a} = 0 \tag{249}$$

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 620

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 $\mathbf{u}$, we have

$$T^{ab} = (\rho_o + p)u^a u^b - p g^{ab} \tag{250}$$

where 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.

---

* In the *comoving* frame we thus have

$$T^{ab} = \begin{pmatrix} \rho_o c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \quad . \tag{251}$$

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

$$p = p(\rho) \quad . \tag{252}$$

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

$$T^{ab} = \rho_o u^a u^b \quad . \tag{253}$$

Challenge 621 Can you explain the difference to the liquid case?

The divergence of the energy-momentum vanishes for all times and positions, as you may want to check. Challenge 622 This property is the same as for the Einstein tensor presented above. But before we elaborate on the issue, a short remark. Why don't we take account 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 symmetries of general relativity

– CS – To be written – CS –

By the way, Torre has shown that diffeomorphism symmetry and trivial scale symmetry are the *only* symmetries of the vacuum field equations.

Hilbert's action

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* could be described by an action integral, like all other physical systems.

Thus Hilbert set out to find the measure of change, as this is what an action describes, for motion due to gravity. Obviously, the measure must be observer invariant; in particular, it must include all possible changes of viewpoints, i.e. all the symmetries just described.

Motion due to gravity is determined by curvature. The only curvature measure independent of the observer is the Ricci scalar R and the cosmological constant Λ. It thus makes sense to expect that the change of space-time is described by an action

$$S = \frac{c^3}{16\pi G} \int (R + 2\Lambda)\, dV \quad . \tag{254}$$

---

* In certain special circumstances, such as weak fields, slow motion, or an asymptotically space-time, we *can* define the integral over the G^{oo} component of the Einstein tensor as negative gravitational energy. Gravitational energy is only defined *approximately*. Nevertheless, this approximation leads to the famous speculation that the total energy of the universe is zero. Do you agree? See page 329 Challenge 623

The cosmological constant Λ (added some years later) appears as a mathematical possibility to describe the most general diffeomorphism invariant action. We will see below that its value in nature, though small, seems to be different from zero.

The Hilbert action of a chunk of space-time is thus the integral of the Ricci scalar plus twice the cosmological constant over that chunk. The principle of least action states that space-time moves in such a way that this integral changes as little as possible.

– CS – to be finished – CS –

Einstein's field equations

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

> Do you believe in god? Prepaid reply 50 words.
> Subsequent telegram by a competing
> sorcerer 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 Einstein's famous field equations. They contain the full description of general relativity and are simply given by

$$G_{ab} = -\kappa\, T_{ab}$$
or
$$R_{ab} - \frac{1}{2} g_{ab} R = -\kappa\, T^{ab} - \Lambda g_{ab} \quad . \tag{255}$$

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

$$\kappa = \frac{8\pi G}{c^4} = 2.1 \cdot 10^{-43}\,/\mathrm{N} \tag{256}$$

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 or pressure Λ/κ. Its low value is quite hard to measure. The presently favoured value is

See page 318

$$\Lambda \approx 10^{-52}\,/\mathrm{m}^2 \quad \text{or} \quad \Lambda/\kappa \approx 0.5\,\mathrm{nJ/m^3} = 0.5\,\mathrm{nPa} \quad . \tag{257}$$

In summary, the field equations state that the curvature at a point is equal to the flow of energy-momentum through that point, taking into account the vacuum energy density. In short, *energy-momentum tells space-time how to curve.*[*]

---

[*] 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.

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 624 κ and setting $\varphi = (c^2/2)h_{oo}$ in $g_{ab} = \eta_{ab} + h_{ab}$, we find

$$\nabla^2\varphi = 4\pi\rho \quad \text{and} \quad \frac{d^2x}{dt^2} = -\nabla\varphi \tag{258}$$

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

$$a = G\frac{M}{r^2} , \tag{259}$$

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.*

- *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. 268 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 488 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 locally indistinguishable from gravitation; we used it to argue that space-time is semi-Riemannian, and that gravity is its curvature.

- *Mach's principle*: inertia is due to the interaction with the rest of the universe; this principle is correct, even though it is often maintained that it is not fulfilled in general relativity. In any case, it is not the essence of See page 331 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.

[*] 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

$$\nabla_e b_a = R_{ceda} v^c v^d \tag{260}$$

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

$$\nabla_e b^a = \nabla_e \nabla^a \varphi = R^a_{ced} v^c v^d \tag{261}$$

which implies that

$$\begin{aligned} \Delta\varphi = \nabla_a\nabla^a\varphi &= R^a_{cad} v^c v^d \\ &= R_{cd} v^c v^d \\ &= \kappa(T_{cd} v^c v^d - T/2) \end{aligned} \tag{262}$$

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

$$\Delta\varphi = 4\pi G\rho \tag{263}$$

as we wanted to show.

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 625 e

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

$$R = -\kappa T \quad . \tag{264}$$

This result also implies the relation between the excess radius and the mass inside a sphere. Challenge 626

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 specialized part of mathematical physics which is not studied here.*

Albert Einstein used to say that the general relativity only provides the understanding of one side of the field equations (255), but not of the other. Can you see which side he meant? Challenge 627

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

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 appear. 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, as stating 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

---

* For more mathematical details, see the famous three-women-book in two volumes by YVONNE CHOQUET-BRUHAT, CECILE DEWITT-MORETTE & MARGARET DILLARD-BLEICK, *Analysis, Manifolds, and Physics*, North-Holland, 1996 and 2001, even though the first edition of this classic appeared in 1977.

** See for example, the http://math1.uibk.ac.at/~werner/black-earth web site.

this is done, one discovers that not all curved space-times are possible, as some would lead to *negative* energy (or mass) densities. Such solutions would contradict the mentioned limit on 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. All orbits are geodesics, thus curves with the longest proper time. It is thus useful to be able to calculate these trajectories.* To start, one needs to know the *shape of space-time*, 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 neighbouring points through

$$ds^2 = dx_a\, dx^a = g_{ab}(x)\, dx^a\, dx^b \quad . \tag{265}$$

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 Challenge 628

$$\frac{d}{ds}\Big(g_{ad}\frac{dx^d}{ds}\Big) = \frac{1}{2}\frac{\partial g_{bc}}{\partial x^a}\frac{dx^b}{ds}\frac{dx^c}{ds} \quad , \tag{266}$$

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 air falls back, except if it is thrown with a speed larger than the escape velocity. Expression (266) 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. See page 262

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, using the details of the collision? Ref. 287 Challenge 629

---

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

[**] This is often written as

$$\frac{d^2x^a}{d^2s} + \Gamma^a_{bc}\frac{dx^b}{ds}\frac{dx^c}{ds} = 0 \tag{267}$$

where the condition

$$g_{ab}\frac{dx^a}{ds}\frac{dx^b}{ds} = 1 \tag{268}$$

must be fulfilled, thus simply requiring 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

$$\Gamma^a_{bc} = \left\{ \begin{matrix} a \\ bc \end{matrix} \right\} = \frac{1}{2}g^{ad}(\partial_b g_{dc} + \partial_c g_{db} - \partial_d g_{bc}) \quad , \tag{269}$$

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

For completion, we mention that light follows *lightlike* or *null geodesics*, an affine parameter u exists, and the geodesics follow

$$\frac{d^2x^a}{d^2u} + \Gamma^a_{bc}\frac{dx^b}{du}\frac{dx^c}{du} = 0 \qquad (270)$$

with the different condition

$$g_{ab}\frac{dx^a}{du}\frac{dx^b}{du} = 0 \quad . \qquad (271)$$

Given all these definitions of various types of geodesics, what are the lines drawn in Figure 118 on page 258? Challenge 630

Mass and ADM

The diffeomorphism invariance of general relativity makes life quite interesting. 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 since then is 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 631

Now that we can go on talking about mass without (too much) a bad conscience, we turn to the equations of motion. Ref. 333

Is gravity an interaction?

We tend to answer affirmatively, as in Galilean physics gravity was seen as an influence on the motion of bodies. In fact, we described it with by a potential, implying 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 an interaction, it is *not* due to a field, and there is *no* potential.

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 not an interaction. See page 488

However, that is going too far. An interaction transports energy between systems. We indeed found out that gravity can be said to transport energy only approximately. Gravitation is thus an interaction only approximately. But that is a sufficient reason to keep this characterization. In agreement with the strange conclusion, 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.

Another way to look at the issue is the following. Take a satellite orbiting Jupiter with energy-momentum $\mathbf{p} = m\mathbf{u}$. If we calculate the energy-momentum change along its path s, we get Challenge 632

$$\frac{d\mathbf{p}}{ds} = m\frac{d\mathbf{u}}{ds} = m(\mathbf{e}_a\frac{d\mathbf{u}^a}{ds} + \frac{d\mathbf{e}_a}{ds}\mathbf{u}^a) = m\mathbf{e}_a(\frac{d\mathbf{u}^a}{ds} + \Gamma^a_{bd}\mathbf{u}^b\mathbf{u}^c) = 0 \qquad (272)$$

where $\mathbf{e}$ describes the unit vector along a coordinate axis. The energy-momentum change vanishes along any geodesic, as you might check. Challenge 633 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 (272) the second term alone is the gravitational force. Ref. 286 But the term can be made to vanish identically along any given world line. Challenge 634 In short, nothing changes between two bodies in free fall around each other: gravity could be said not to be an interaction. The properties of energy confirm this argument. Challenge 635 n

Of course, the conclusion that gravity is not an interaction is somewhat academic, as it contradicts daily life. But we will need it for the full understanding of motion later on. The behaviour of radiation confirms the deduction. In vacuum, radiation is always moving freely. In a sense, we can say that radiation always is in free fall. Strangely, since we called free fall the same as rest, we should conclude that radiation always is at rest. This is not wrong! We already 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 for 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,** 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; Challenge 636 e let us call them $\mathbf{p}$ and $\mathbf{q}$. And instead of a disk, we take the

---

* Refraction, the slowdown of light inside matter, is not a counterexample. Strictly speaking, light inside matter is constantly being absorbed and reemitted. In between these processes, light still propagates with the speed of light in vacuum. The whole process only *looks* like a slowdown in the macroscopic limit. The same applies to diffraction and to reflection. A list of apparent ways to bend light can be found on page 398; details of the quantum mechanical processes at their basis can be found on page 537.

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

parallelogram spanned by **p** and **q**. There are several possible definitions.

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 **p** and **q**, with area A, through

$$K(\mathbf{p},\mathbf{q}) = \frac{R\,\mathbf{pqpq}}{A^2(\mathbf{p},\mathbf{q})} = \frac{R_{abcd}p^a q^b p^c q^d}{(g_{\alpha\delta}g_{\beta\gamma} - g_{\alpha\gamma}g_{\beta\delta})p^\alpha q^\beta p^\gamma q^\delta} \tag{273}$$

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* R or quantities derived from it.[**]

But we can forget the just mentioned definition of curvature. There is a second, more physical way to look at the Riemann tensor. We know that curvature means gravity. As 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 637 e

It turns out the the tensor R describes precisely this relative acceleration, i.e. what we called the *tidal effects* earlier on. Obviously, the relative acceleration **b** increases with the separation **d** and the square (why?)of the speed **u** of the two particles. Therefore we can also define R as a (generalized) proportionality factor among these quantities: Challenge 638

$$\mathbf{b} = R\,\mathbf{u}\,\mathbf{u}\,\mathbf{d} \quad \text{or, more clearly} \quad b^a = R^a{}_{bcd}\,u^b\,u^c\,d^d \quad . \tag{277}$$

The components of the Riemann curvature tensor have the dimension of an inverse square length. Since it contains all information about intrinsic curvature, we follow that if R vanishes in a region, space-time in that region is flat. This connection is easily deducedfrom Challenge 639

---

[*] Bernhard Riemann (1826, Breselenz–1866, Selasca), important German mathematician.

[**] 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 **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, we have

$$g_{ab} = 1/g^{ab} \quad \text{and} \quad g_a{}^b = g^a{}_b = \delta^a_b \quad . \tag{274}$$

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

$$R^a{}_{bcd} = \frac{\partial\Gamma^a{}_{bd}}{\partial x^c} - \frac{\partial\Gamma^a{}_{bc}}{\partial x^d} + \Gamma^a{}_{ec}\Gamma^e{}_{bd} - \Gamma^a{}_{fd}\Gamma^f{}_{bc} \quad . \tag{275}$$

The curvature tensor is built from the second derivatives of the metric. On the other hand, we can also determine the metric if the curvature is known, using

$$g = ...R... \tag{276}$$

In other words, either the Riemann tensor R or the metric g specify the whole situation of a space-time.

this second definition.[*]

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

$$\begin{aligned} g_{ab} = {} & \eta_{ab} + \frac{1}{3} R_{acbd} x^c x^d + O(x^3) \\ = {} & \frac{1}{2} (\partial_c \partial_d g_{ab}) x^c x^d + O(x^3) \quad . \end{aligned} \tag{279}$$

The curvature term thus describes the dependence of the space-time metric from flat space-time. The curvature tensor R is a large beast; it has $4^4 = 256$ components at each point of space-time; however, its symmetry properties reduce them to twenty independent numbers.[**] The 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

$$R_{bc} = R^a{}_{bac} \quad . \tag{282}$$

Its components, like those of the Riemann tensor, are inverse square lengths.

Challenge 642 Can you confirm that $R_{abcd} R^{abcd} = 48 m^2 / r^6$ for the Schwarzschild solution:?

9. Why can we see the stars? – Motion in the universe

Zwei Dinge erfüllen das Gemüt mit immer neuer und zunehmender Bewunderung
und Ehrfurcht, je öfter und anhaltender sich das Nachdenken damit beschäftigt:
der bestirnte Himmel über mir und das moralische Gesetz in mir.[***]
Ref. 301 Immanuel Kant (1724–1804)

On clear nights, between two and five thousand stars are visible with the naked eye. Several

---

[*] 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. (Ref. 267) 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

$$\delta \mathbf{w} = -\varepsilon^2 R\, \mathbf{u}\, \mathbf{v}\, \mathbf{w} + \quad \text{higher order terms} \quad . \tag{278}$$

See page 130 More about the geodesic deviation can be found out by studying the behaviour of the famous south-pointing carriage. This device, common in China before the compass was discovered, only works if the world is flat. Indeed, on a curved surface, after following a large closed path, it will show a different direction than at the start of the trip. Can you explainwhy? (Challenge 640)

[**] The second definition indeed shows that the Riemann tensor is symmetric in certain indices and antisymmetric in others: (Challenge 641)

$$R_{abcd} = R_{cdab} \quad , \quad R_{abcd} = -R_{bacd} = -R_{abdc} \tag{280}$$

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

$$R_{abcd} + R_{adbc} + R_{acdb} = 0 \quad . \tag{281}$$

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.

[***] 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.

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 the simple fact that we can *see* the stars is the basis for a story much more fantastic than all myths. It touches almost all aspects of modern physics.

Which stars do we see at all?

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

Ref. 303

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.

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. It is a whole galaxy like our own, as Immanuel Kant already had conjectured in 1755. Several extragalactic objects are visible with the naked eye in the southern hemisphere: the Tarantula Nebula, and the large and the small Magellanic cloud. The Magellanic clouds are neighbour galaxies to our own. Other, temporary exceptions are the rare *novae*, exploding stars which can be seen also if they appear in nearby galaxies, or the still rarer *supernovae*, which can often be seen even in faraway galaxies.

In fact, the visible stars are special 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 643

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.

Why did this happen so late? Well, people had the same difficulty as when the the shape of the earth had to be determined. They 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-

---

* About the myths around the stars and the constellations, see e.g. the text by G. Fasching, *Sternbilder und ihre Mythen*, Springer Verlag, 1993. On the internet there are also the beautiful http://www.astro.wisc.edu/~dolan/constellations/constellations.html and http://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 Heracles feed at Hera's breast in order to make him immortal; the young Heracles, in a sign showing his future strength, sucked so forcefully that the milk splashed all over the sky.

dimensional *shape*; it is shown in Figure 131. Our galaxy is a flat and circular structure, with a diameter of 100 000 light years; in the centre, it has a spherical bulge. As said before, Challenge 644 it rotates once in about 200 to 250 million years. (Can you guess how this is measured?)The rotation quite is slow: since the sun exists, it made only about 20 to 25 full turns around the centre.

Figure 132 The Andromeda nebula M31, our neighbour galaxy

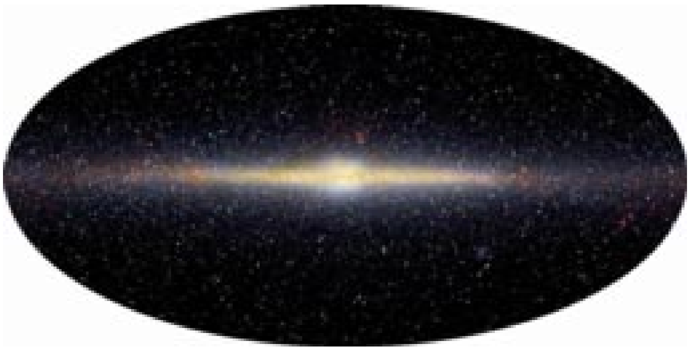


Figure 131 How our galaxy looks in the infrared

It is even possible to measure the *mass* of our galaxy. The trick is to use a binary pulsar on its outskirts. If it is observed for many years, one can deduce its acceleration around the galaxy centre, as the pulsar reacts with a frequency shift which can be measured on earth. However, many decades of observations are needed, and many spurious effects have to be eliminated. Nevertheless, such measurements are ongoing. Ref. 304

Figure 133 The elliptical galaxy NGC 205

Figure 134 The colliding galaxies M51 and M110

What do we see at night?

Astrophysics leads to a strange conclusion about matter, quite different from what we are used to think in classical physics: *the matter observed in the sky is found in clouds*. *Clouds* are systems in which the matter density diminishes with the distance from the centre, with no clear border, and with no clear size. It turns out that all astrophysical objects are best described by clouds.

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

A second aspect is that in the universe, *almost all of the clouds are plasma clouds*. A *plasma* is an ionized gas, such as fire, lightning, the inside of neon tubes, the sun etc. At least 99.9% of all matter in the universe is in the form of plasmas. Only an exceptionally small percentage exists in solid or liquid form, such as toasters, subways or their users.

A third aspect is the shape of the observed components. All clouds seen in the universe are rotating. Most clouds are therefore flattened. Finally, many clouds emit something along the rotation axis. This has been observed for stars, for pulsars, for our planetary system, for galaxies, for quasars, and for many other systems.

In summary, the universe is mostly made of rotating, flattened clouds emitting jets along their axes. A more detailed overview of the information collected by modern astronomy and astrophysics about various clouds in the universe is given in the following table.* Ref. 289

Table 29 Some observations about the universe

Aspect	main properties	value
Phenomena		
galaxy formation	observed by Hubble	several times
	trigger event	unknown
galactic collisions	momentum	$p \approx \ldots$
	star formation	
star formation	cloud collapse	
novae	new bright stars,	$L > \ldots$
	later surrounded by bubble	$R \approx t \cdot c/100$
supernovae	new bright star,	$L > \ldots$
	matter forms	
gamma ray bursts	luminosity	up to $3 \cdot 10^{47}$ W, almost equal to the whole visible universe
	energy	ca. 10^{46} J
	duration	ca. 0.015– 1000 s
	observed number	ca. 2 per day
hypernovae, 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 $4.6 \cdot 10^9$ a
Observed components		

---

* 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. The most recent discoveries are best followed on the http://hubble.nasa.gov web site.

Aspect	main properties	value
intergalactic space	mass density	...
quasars	redshift	up to 5.8
	luminosity	..., about the same as one galaxy
galaxy superclusters	number	ca. 10^8 inside horizon
our own local supercluster		with about 4000 galaxies
galaxy groups		100 Zm, with a dozen up to 1000 galaxies
our local group		with 30 galaxies
galaxies	size	0.5 to 2 Zm
	number	ca. 10^{11} inside horizon
	containing	10 to 400 globular clusters
	containing	typically 10^{11} stars
our galaxy	diameter	$1.0(0.1)$ Zm
	mass	10^{42} kg or $5 \cdot 10^{11}$ solar masses Ref. 302
	containing	100 globular clusters each with 1 million stars
	speed	$600\,\mathrm{km/s}$ towards Hydra-Centaurus
nebulae, clouds	composition	
our local 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	size	2 light years (Oort cloud)
our solar system	speed	$370\,\mathrm{km/s}$ from Acquarius towards Leo
stars		
giants and supergiants	large size	up to 10^{12} m
brown dwarfs	low temperature	below 2800 K Ref. 305
L dwarfs	low temperature	
T dwarfs	low temperature	
white dwarfs	high temperature	
neutron stars	nuclear mass density, small size	$\rho \approx 10^{-\cdot\cdot}\,\mathrm{kg/m^3}$ $r \approx 10\,\mathrm{km}$
pulsars	radio emission	
magnetars	high magnetic fields	
black holes	horizon radius	$r = 2GM/c^2$

General properties

Aspect	main properties	value
cosmic horizon	distance	ca. 10^{26} m=100 Ym
expansion	Hubble's constant	between $59\,\mathrm{km\,s^{-1}Mpc^{-1}}$ and $70\,\mathrm{km\,s^{-1}Mpc^{-1}}$, or ca. $2 \cdot 10{-}18\,/\mathrm{s}$
vacuum energy density	finite	$0.5\,\mathrm{nJ/m^3}$
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 $11 \cdot 10^{-27}\,\mathrm{kg/m^3}$ or 1 to 6 hydrogen atoms per cubic metre

Aspect	main properties	value
baryons	density	one sixth of the previous
other	density	five sixths unknown
photons	number density	4 to $5 \cdot 10^8$ $/\mathrm{m}^3$ $= 1.7$ to $2.1 \cdot 10^{-31}$ $\mathrm{kg/m}^3$
neutrinos	number density	not measured
average temperature	photons	2.7 K
	matter	ca. 0 K
	neutrinos	not measured, predicted is 2 K
original inhomogeneity	amplitude of radiation anisotropy	
	amplitude of matter clustering	

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

What is the universe?

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

The universe, as the name says, is what turns around us at night. For a physicist, 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. Many 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 adventure, 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 mountain ascent of Motion Mountain becomes impossible.

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

Challenge 645

Challenge 646

The colour and the motion of the stars

> 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 imagine how distance and velocity are determined?)This wealth of data can be summed up Challenge 647 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 nor homogeneous. But at large scales the differences average out. This large scale homogeneity Ref. 293 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 US-American astronomer Edwin Hubble[**] published the first measurement of the relation between speed and distance. Despite his use of incorrect length scales he found a relation

$$v = H\, d \quad , \tag{283}$$

where the proportionality constant H, so-called *Hubble constant*, is known today to have a value between 59 $\mathrm{km\,s^{-1}Mpc^{-1}}$and 70 $\mathrm{km\,s^{-1}Mpc^{-1}}$. (Hubble's own value was so far outside this range that it is never cited.) 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.

Figure yet to be included

Figure 135 The relation between star distance and velocity

---

[*] The Theogony, attributed to the probably mythical Hesiodos, was finalized around 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.

In fact, the discovery by Wirtz and Lundwerk implies that every galaxy moves away from all the others. (Why?)In other words, the matter in the universe is *expanding*. The scale of this expansion and the enormous dimensions involved are amazing. The motion of all the thousand millions galaxy groups in the sky are described by the single equation (283)! Of course, some deviations are observed for nearby galaxies, as mentioned above, and for far away galaxies, as we will see. Challenge 648

The cosmological principle and the expansion taken together 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, 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?)Despite various similar predictions by other authors, in one of the most famous cases of missed scientific communication, the radiation was found only much later, by two researchers completely unaware of all this work. A famous paper in 1964 by Doroshkevich and Novikov had even stated that the antenna used by the (unaware) later discoverers was the best device to search for the radiation! In any case, only in 1965, Arno Penzias and Robert Wilson discovered the radiation, in one of the most beautiful discoveries of physics, for which both later received the Nobel prize for physics. The radiation turns out to be described by the black body radiation for a body with a temperature of 2.7 K; it follows the black body dependence to the precision of about 1 part in 10^4. Ref. 290 Challenge 649 Ref. 291 Ref. 292

But apart from expansion and cooling, the past twelve thousand million years also produced a few other memorable events.

Do stars shine every night?

> Do you see the stars shine?
> I am the only in the world one who knows why.
> Eddington, in conversation.

Stars seems to be there for ever. In fact, every now and then a new star appears in the sky: a *nova*. The name is latin for ‘new’. 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.

It turns out that one can follow the age of a star in the so-called *Hertzsprung-Russel diagram*. The diagram, central to every book on astronomy, is a beautiful example of a

---

[*] George Gamow (1904, Odessa –1968), Russian-American physicist; he explained alpha decay as a tunnel effect and predicted the microwave background. He wrote the first successful popular science texts, such as *1, 2, 3, infinity* and the *Mr. Thompkins* series, which later were imitated by many others.

Figure yet to be included

Figure 136 The Hertzsprung-Russel diagram

standard method used by astrophysicists: collecting statistics over many examples of a type of object, one can deduce their life cycle, even though the lifetime is much longer than than that of a human. For example, it is possible, by clever use of the diagram, to estimate the age of stellar clusters, and thus provide a minimum age of the universe. The result is around twelve thousand million years.

One conclusion is essential: since stars shine, they also *die*. 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 observed at large distances, such as quasars, are not stars, but much more massive and bright systems. These mechanisms are still being studied by astrophysicists.

On the other hand, since the stars shine, they were also *formed* somehow. The fascinating details of these investigations are part of astrophysics and will not be explored here.

Yet we do not have the full answer to our question. Why do stars shine at all? Clearly, they shine because they are hot. They are hot because of nuclear reactions in their interior. We will discuss these processes in more detail in the chapter on the nucleus.

See page 672

A short history of the universe

> The soul is a spark of the substance of the stars.
> Heraclitos of Ephesos (ca. 540–ca. 480 BCE)

Ref. 288

The adventures the universe has experienced, or better, the adventures the matter and radiation inside it have experienced, are summarized in Table 30.* The steps not yet discussed will be studied in quantum theory. This history table has applications no physicist would have imagined. The sequence is so beautiful and impressive that nowadays it is used in certain psychotherapies to point out to people the story behind their existence and to remind them of their own worth. Enjoy.

---

* On the remote history of the universe, see the excellent texts by G. Börner, *The Early Universe – Facts & Fiction*, 3rd edition, Springer Verlag, 1993, or Barry Parker, *Creation – The Story of the Origin and the Evolution of the Universe*, Plenum Press, 1988. For an excellent popular text, see M. Longair, *Our Evolving Universe*, Cambridge University Press, 1996.

Table 30 A short history of the universe

Time from now[a]	Time from big bang[b]	Event	Temperature
$\approx 13 \cdot 10^9$ a	$\approx t_{\mathrm{Pl}}$ [b]	Time, space, matter, and initial conditions make no sense	10^{32} K $\approx T_{\mathrm{Pl}}$
$13 \cdot 10^9$ 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	75 000 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 black body 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 form; 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		**Palaeozoic** starts, after a gigantic ice age; animals appear; oldest fossils	
$400 \cdot 10^6$ a		Land plants appear	

Time from now[a]	Time from big bang[b]	Event	Temperature
$370 \cdot 10^6$ a		Wooden trees appear	
$220 \cdot 10^6$ a		**Mesozoic** starts: mammals appear, insects are exterminated	
$150 \cdot 10^6$ a		Continent Pangaea splits into Laurasia and Gondwana The star cluster of the Pleiades forms	
$150 \cdot 10^6$ a		Birds appear	
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	
500 000 a		Formation of youngest stars in galaxy	
300 000 a		Homo sapiens appears	
100 000 a		Beginning of last ice age	
90 000 a		Homo sapiens sapiens appears	
20– to 11 000 a		End of last ice age	
6 000 a		First written texts	
2 500 a		Physics starts	
500 a		Coffee and pencil usage spreads, modern physics starts	
200 a		Electricity usage begins	
100 a		Einstein publishes	
10 – 120 a		You were an unicellular being	
present	ca. $13 \cdot 10^9$ a	You are reading this	$T_\gamma = 2.73$ K, $T_\nu \approx 1.6$ K, $T_b \approx 0$ K
future		You enjoy life; for details and reasons, see page 443	

a. The time coordinate used here is the one given by the coordinate system defined by the microwave background radiation, as explained on page 315.

b. This quantity is not exactly defined since the big bang is not a space-time event. More on the issue on page 758.

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-centred.

Research in astrophysics is directed at discovering and understanding all phenomena observed in the skies. Here we skip most of this fascinating topic, since as usual, we focus on motion. Interestingly, general relativity allows to explain many of the general observations about motion in the universe.

The history of space-time

> A number of rabbits run away from a central point in various directions, all with the same speed. While running, one rabbit turns its head, and makes a startling observation. What does it see? Challenge 650 n

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

Ref. 294 The first to dare to take this conclusion was Aleksander Friedmann.* In 1922 he deduced the detailed evolution of the universe in the case of homogeneous, isotropic mass distribution. His calculation is a classic. For a universe which is homogeneous and isotropic for every point, the line element is given by Challenge 651

$$ds^2 = c^2 dt^2 - a^2(t)(dx^2 + dy^2 + dz^2) \tag{284}$$

and matter is described by a density ρ_M and a pressure p_M. Inserting all this into the field equations, we get two equations

$$\left(\frac{\dot a}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\rho_M + \frac{\Lambda}{3} \quad \text{and} \tag{285}$$

$$\ddot a = -\frac{4\pi G}{3}(\rho_M + 3p_M)\, a + \frac{\Lambda}{3}\, a \tag{286}$$

which imply

$$\dot\rho_M = -3\frac{\dot a}{a}(\rho_M + p_M) \quad . \tag{287}$$

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 all mass densities to the so-called *critical mass density* ρ_c given by Challenge 652

$$\rho_c = \frac{3H_o^2}{8\pi G} \approx 8 \pm 2 \cdot 10^{-27}\,\mathrm{kg/m^3} \tag{288}$$

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

---

* 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, Charleroi–1966, Leuven), 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 321, 322

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 used for this expression in all other cases as well. One thus speaks of dimensionless mass densities Ω_M defined as

$$\Omega_\mathrm{M} = \rho_\mathrm{o}/\rho_\mathrm{c} \quad . \tag{289}$$

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

$$\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_\mathrm{c}} = \frac{\Lambda c^2}{8\pi G \rho_\mathrm{c}} = \frac{\Lambda c^2}{3H_\mathrm{o}^2} \quad . \tag{290}$$

A third dimensionless parameter Ω_K describes the curvature of space. It is defined as

$$\Omega_\mathrm{K} = \frac{-k}{R_\mathrm{o}^2 H_\mathrm{o}^2} \tag{291}$$

and its sign is opposite to the one of the curvature; Ω_K vanishes for vanishing curvature. Note that a positively curved universe, when homogeneous and isotropic, is necessarily closed and of finite volume. A flat or negatively curved universe with the same matter distribution can be open, i.e. of infinite volume, but does not need to be so. It could 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 $H_\mathrm{o} = \dot{a}_\mathrm{o}/a_\mathrm{o}$. From equation (285) we then get:

Challenge 653

$$\Omega_\mathrm{M} + \Omega_\Lambda + \Omega_\mathrm{K} = 1 \quad . \tag{292}$$

In the past, when data was lacking, physicists were divided into two camps: the *claustrophobics* believing that $\Omega_\mathrm{K} > 0$ and the *agoraphobics* who believe that $\Omega_\mathrm{K} < 0$. More details about the measured values of these parameters will be given shortly. The diagram of Figure 137 shows the most interesting ranges of parameters together with the corresponding behaviour of the universe.

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

$$65 \pm 5\,\mathrm{km/sMpc} \approx 2 \cdot 10^{-18}\,/\mathrm{s} \tag{293}$$

which correspond to an age of the universe of 13.5 ± 1.5 thousand million years. In other words, the age deduced from the history of space-time corresponds with the age, given above, deduced from the history of stars.

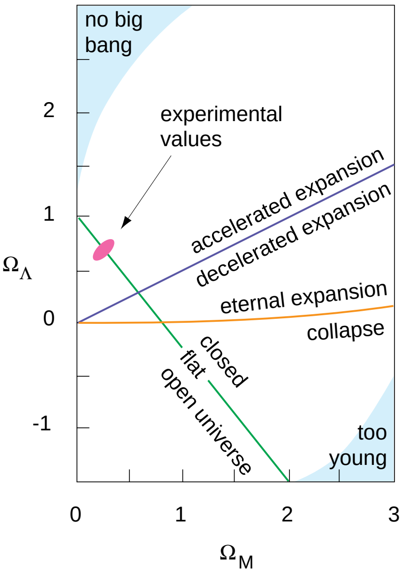


Figure 137 The ranges for the Ω parameters and their consequences

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

$$q_\mathrm{o} = -\frac{\ddot{a}_\mathrm{o}}{a_\mathrm{o} H_\mathrm{o}^2} = \frac{1}{2}\Omega_\mathrm{M} - \Omega_\Lambda \tag{294}$$

The parameter q_o is positive if the expansion is slowing down, and negative if 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 (285) using $\tau = t H_o$ and $x(\tau) = a(t)/a(t_o)$, yielding

$$\begin{aligned} \left(\frac{dx}{d\tau}\right)^2 + U(x) &= \Omega_K \\ \text{with } U(x) &= -\Omega_\Lambda x - \Omega_\Lambda x^2 \end{aligned} \qquad (295)$$

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

For vanishing Ω_Λ, the universe either expands for ever, or recollapses, depending on the value of the mass-energy density,

For non-vanishing (positive) Ω_Λ, the potential has exactly one maximum; if the particle has enough energy to get over the maximum, it will accelerate continuously. That is the situation the universe seems to be in today.

For a certain time range, the result is shown in Figure 138. 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 later on.

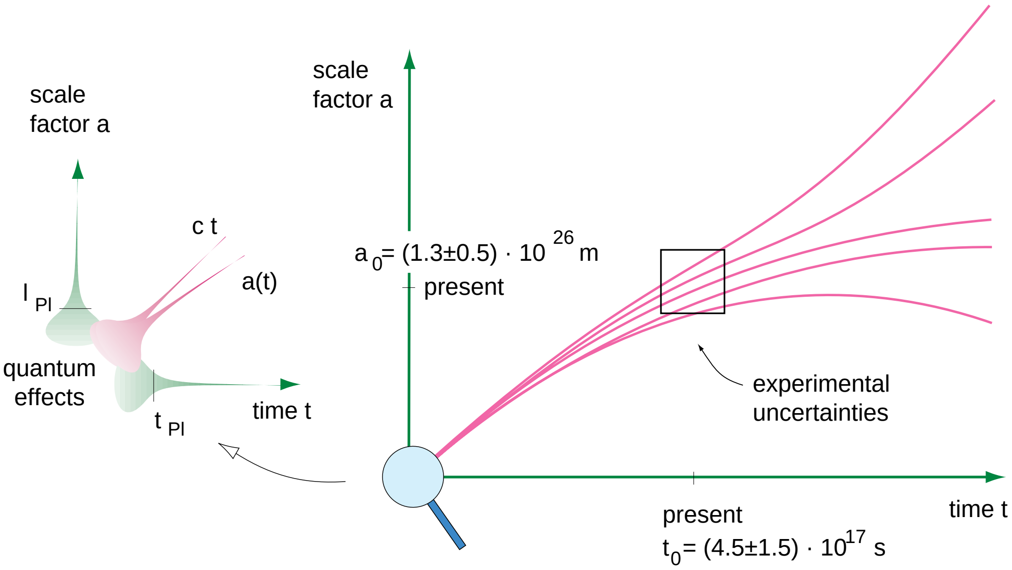


Figure 138 The evolution of the universe's scale R for different values of its mass density

The main result of Friedmann's work was that a homogeneous 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 30 and is the one meant when we talk about the *age* of the universe.

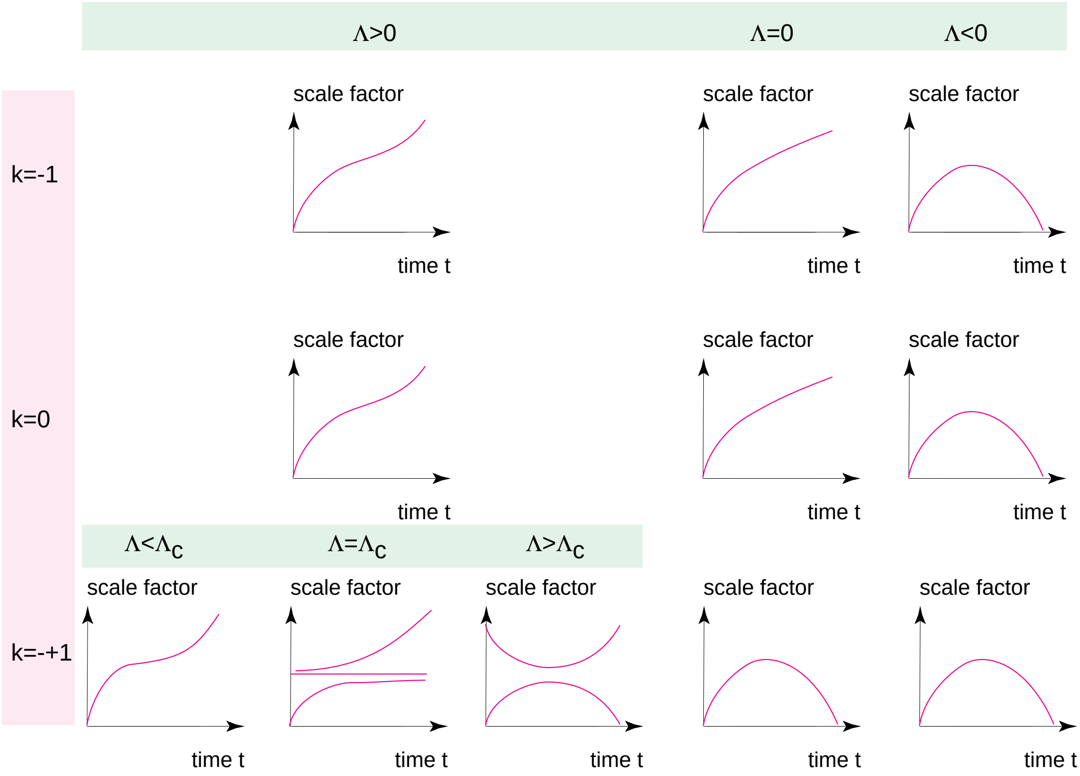


Figure 139 The long term evolution of the universe's scale factor *a* for various parameters

An overview of the possibilities for the long time evolution is given in Figure 139. 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 $U(x)$.

De Sitter had found, much to Einstein's personal dismay, that an empty universe with $\rho_M = p_M = 0$ and $k = 1$ is also possible. This type of universe expands for large times. Challenge 654

Lemaître had found expanding universes for positive mass, and his results were also contested by Einstein in the beginning. When later the first measurements confirmed the calculations, massive and expanding universes became popular. They were promoted to the standard model 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, as we will find out shortly. But first of all we settle an old issue.

Challenge 655

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

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 the colour of the sky during the day, is not due to the temperature of the sky, but to the light from the stars. In addition, if we look for temperature radiation, we do find some. Measurements show that even the empty 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. 308 The universe is indeed colder than the stars. But why is this so? If the universe were homogeneous on large scales and infinitely large, it would have an infinite number of stars. Given any direction to look at, we 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 656

In other words, we would effectively live inside an oven 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 1823 by the astronomer Wilhelm Olbers.[*] As he extensively discussed the question, it is also called 'Olber's paradox'. Thus the Today we know that even if all matter in the universe were converted into radiation, the universe would still not be as bright as just calculated. Equivalently, the lifetime of stars is way too short to produce the oven brightness just mentioned. So something is wrong. Ref. 309

In fact, two main effects have the power to avoid the contradiction with observations. First, since the universe is finite 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. Ref. 310 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. 308 In addition, the darkness of the sky is possible only because the speed of light is *finite*. Can you confirmthis? Challenge 658

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 Ref. 311 is red shifted due to the Doppler effect. Under reasonable assumptions, the temperature of

---

[*] Heinrich Wilhelm Matthias Olbers (1758, Arbergen – 1840, Bremen), astronomer. He discovered two planetoids, Pallas and Vesta, and five comets; he developed the method to calculate parabolic orbits for comets which is still in use today. Olbers also actively supported F.W. Bessel in his career choice. The paradox is named after him, though others had made similar points before, such as the Swiss astronomer de Cheseaux in 1744 and Johannes Kepler in 1610.

[**] 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 657

this radiation changes with the scale factor of the universe as

$$T \sim \frac{1}{R(t)} \quad . \qquad (296)$$

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

From the brightness of the sky at night, measured to be about $3 \cdot 10^{-13}$ times that of an average star like the sun, we can deduce something interesting: the density of stars in the universe must be much smaller than in our galaxy. The density of stars in the galaxy can be deduced by counting the stars we see at night. But the average star density in the galaxy would lead to much higher values for the night brightness if it were constant throughout the universe. We can thus deduce that the galaxy is much *smaller* than the universe simply by measuring the brightness of the night sky and by counting the stars in the sky! Can you make the explicit calculation? Ref. 309 Challenge 659

In summary, the sky is black at night because space-time is of finite, but old age. As a side issue, here is a quiz: is there an Olbers' paradox also for gravitation? Challenge 660

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.

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. This is presently being done with help of computerized searches of the sky, using the largest available telescopes. Ref. 296

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

A third method is the determination of the mass density using the gravitational lensing effect for the light of distant quasars bent around galaxies or galaxy clusters. See page 325

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

At present, these four completely independent sets of measurements provide the values Ref. 295

$$(\Omega_{\mathrm{M}}, \Omega_{\Lambda}, \Omega_{\mathrm{K}}) \approx (0.3, 0.7, 0.0) \qquad (297)$$

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 327

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

$$\Lambda = \Omega_\Lambda \frac{3H_o^2}{c^2} \approx 10^{-52}\,/\mathrm{m}^2 \qquad (298)$$

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

$$\rho_\Lambda c^2 = \frac{\Lambda c^4}{8\pi G} \approx 0.5\,\mathrm{nJ/m^3} \approx \frac{10^{-46}\,(GeV)^4}{(\hbar c)^3} \qquad (299)$$

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 294

$$\Delta\varphi = 4\pi G(\rho + 3p/c^2) \qquad (300)$$

Ref. 306 we get

$$\Delta\varphi = 4\pi G(\rho_\mathrm{M} - 2\rho_\Lambda) \quad . \qquad (301)$$

Challenge 661 Thus the gravitational acceleration follows

$$a = \frac{GM}{r^2} - \frac{\Lambda}{3}c^2 r = \frac{GM}{r^2} - \Omega_\Lambda H_o^2\, r \qquad (302)$$

which shows that a *positive* vacuum energy indeed leads to a *repulsive* gravitational effect. Inserting the mentioned value for the cosmological constant Λ we find 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 662

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

Why is the universe transparent?

Could the universe be filled with water, which is transparent, as maintained by some 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. 307 Challenge 663

---

* 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. This issue is called the *dark matter problem*; it is one of the important unsolved questions of cosmology.

The universe is thus transparent because it is mostly empty. But *why* is it so empty? First of all, in the times when the size of the universe was small, all antimatter annihilated with the corresponding amount of matter. Only a tiny fraction of matter, which originally was slightly more abundant than antimatter, was left over. This 10^{-9} fraction is the matter we see now. As a consequence, the number of photons in the universe is 10^9 larger than that of electrons or quarks. See page 672

In addition, 300 000 years after antimatter annihilation, all available nuclei and electrons recombined, forming atoms, and their aggregates, like stars and people. No free charges interacting with photons were lurking around any more, so that from that period onwards light could travel through space like it does today, being affected only when it hits some star or some dust particle.

If we remember that the average density of the universe is $10^{-26}\ \mathrm{kg/m^3}$ and that most of the matter is lumped by gravity in galaxies, we can imagine what an excellent vacuum lies in between. As a result, light can travel along large distances without noticeable hindrance.

But why is the vacuum transparent? That is a much deeper question. Like all material questions, it requires quantum theory; thus we reserve it for a later stage of our walk.

The big bang and its consequences

> Learn to die.
> Plato (427–347 BCE)

Above all, the big bang model, which is deduced from the colour of the stars and galaxies, states that about twelve thousand million years ago the whole universe was extremely small. This fact 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, we need to look for other, verifiable consequences. The central ones are the following: See page 502 Ref. 297

- 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 have been 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. 297 Ref. 298

---

* 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 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 our mountain ascent. The whole phenomenon cannot be called 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.

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; only 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 initial 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 value, since such values cannot exist in nature.* In any case, there is a general agreement that arguments based on *pure* general relativity alone cannot make correct statements on the big bang. Most newspaper article statements are of this sort.

See page 768

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. 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 the first chapter of the third part of the mountain ascent. 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. In fact, quantum theory shows that near the big bang, events can *neither* be ordered *nor* even be defined.

See page 729

---

* Many physicists are still wary to make such strong statements at this point. The first sections of the third part of the mountain ascent give the precise arguments leading to them.

See page 724

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 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 in the beginning of our walk.

Ref. 299

See page 41

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 112

Ref. 300

Challenge 664

In summary, the big bang does not contain a beginning nor does it imply one. We will uncover the correct way to think about it in the third part of our mountain ascent.

Does the big bang imply creation?

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

Creation, i.e. the appearance of something out of nothing, needs an existing concept of 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 495

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 ascent 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 mountain ascent.

Challenge 665

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. It appears ever more clearly at this point. But what then *is* the big bang? We'll find out in the third part. We now return to the discussion of what the stars can tell us about nature.

See page 44

Why can we see the sun?

First of all, the sun is visible 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 *incandescence*, as this effect is called, below. See page 428

Thirdly, we can see the sun because we and our environment and the sun's environment are *colder* than the sun. In fact, incandescent bodies can be distinguished from their background only if the background is colder. This is a consequence of the properties of incandescent light emission, usually called *black body radiation*. The radiation is material independent, so that for an environment with the same temperature as the body, nothing can be seen at all. Just have a look on the photograph of page 431 as a proof.

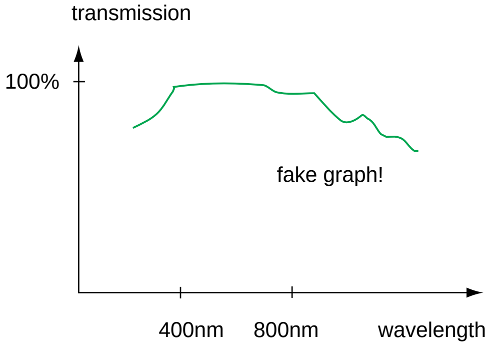


Figure 140 The absorption of the atmosphere

Finally, we can see the sun because it is not a black hole. If it were, it wouldn't emit (almost) any light. Obviously, each of these conditions applies for stars as well. For example, we can only see them, because the night sky is black. But then,

Why are the colours of the stars different?

Stars are visible because they emit visible light. We encountered several important effects which determine colours: 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 are good approximations of black bodies, so that the black body radiation See page 429 law sometimes is not an accurate description for their colour. However, most of the stars are reasonable approximations of black bodies. The temperature of a star depends mainly on its Ref. 312 size, its mass, its composition and its age, as the astrophysicists are happy to explain. Orion is a good example of a coloured constellation; each star has a different colour. Long term exposure photographs beautifully show this.[**]

---

[*] 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.

[**] See for example the book ...

Table 31 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	ϰ Orionis	blue-white
B	12 kK	Rigel	β Orionis	blue-white
B	kK	Alnilam	ε Orionis	blue-white
B	17(5) kK	Regulus	α Leonis	blue-white
A	9.9 kK	Sirius	α Canis Majoris	blue-white
A	8.6 kK	Megrez	δ Ursae Majoris	white
A	7.6(2) kK	Altair	α Aquilae	yellow-white
F	7.4(7) kK	Canopus	α Carinae	yellow-white
F	6.6 kK	Procyon	α Canis Minoris	yellow-white
G	5.8 kK	Sun	ecliptic	yellow
K	3.5(4) kK	Aldebaran	α Tauri	orange
M	2.8(5) kK	Betelgeuse	α Orionis	red

The basic colour determined by temperature is changed by two effects. The first, the *Doppler red shift*, depends on the speed v between source and observer following Challenge 666

$$z = \frac{\Delta\lambda}{\lambda} = \frac{f_S}{f_O} - 1 = \sqrt{\frac{c+v}{c-v}} - 1 \quad . \qquad (303)$$

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 5.0, corresponding to more than 94% of the speed of light.Note that in the universe, the red shift is also related to the scale factor R by Challenge 667

$$z = \frac{R(t_o)}{R(t_{\text{emission}})} - 1 \quad . \qquad (304)$$

Light at a red shift of 5.0 thus was emitted at an age 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

$$z = \frac{\Delta\lambda}{\lambda} = \frac{f_S}{f_O} - 1 = \frac{1}{\sqrt{1 - \frac{2GM}{c^2 R}}} - 1 \quad . \qquad (305)$$

It is usually quite a bit smaller than the Doppler shift. Can you confirm this? Challenge 668

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: See page 333

Are there dark stars?

It could be that some stars are not seen because they are dark. This possibility of dark matter, if widespread, would lead to incorrect matter density estimates for the universe, and thus to incorrect evolution predictions for its fate. This issue is therefore of great interest and hotly debated. It is known that objects more massive than Jupiter but less massive than the sun can exist in states which 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.

Another possibility for dark stars are black holes. They are discussed in detail in a separate section below.

Are all stars different? – Gravitational lenses

Per aspera ad astra.*

Are we sure that at night, two stars are really different? The answer is no. Recently, it was shown that two stars were actually two images of the same object. This was found by comparing the flicker of 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 the estonian astrophysicist Johannes Pelt and his research group while observing two quasar images of the system Q0957+561. Ref. 313

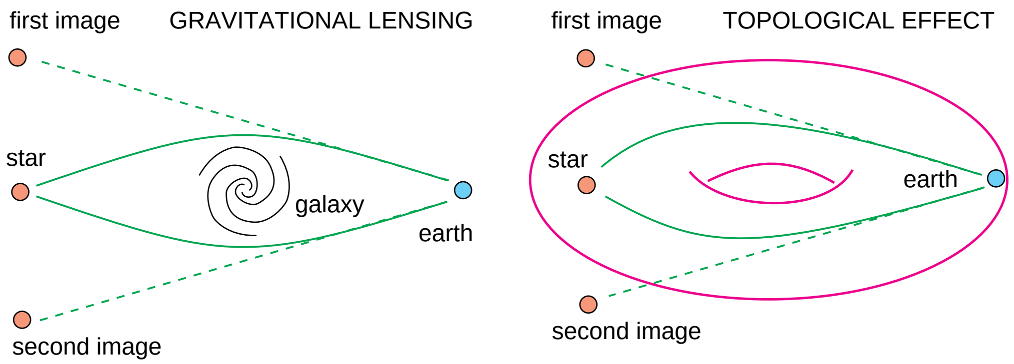


Figure 141 How one star can lead to several images

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 already considered by Einstein; however he did not believe that it was observable. The real father of gravitational lensing is Fritz Zwicky, who predicted in 1937 that the effect would be quite frequent and easy to observe, if lined-up galaxies instead of lined-up stars were considered, as indeed it turned out to be the case. Ref. 314

Interestingly, when the time delay is known, astronomers are able to determine the size of the universe from this observation. Can you imagine how? Challenge 669

In fact, if the two observed objects are lined up behind each other, the more distant one is seen as *ring* around the nearer one. Such rings have indeed been observed, and the object

---

[*] Through hardship to the stars. A famous Latin motto.

B1938+666 is one of the most beautiful ones. Using this method, some astronomers are even searching for earth-like planets around other stars.

Photographs to be included

Figure 142 The Zwicky-Einstein ring B1938+666

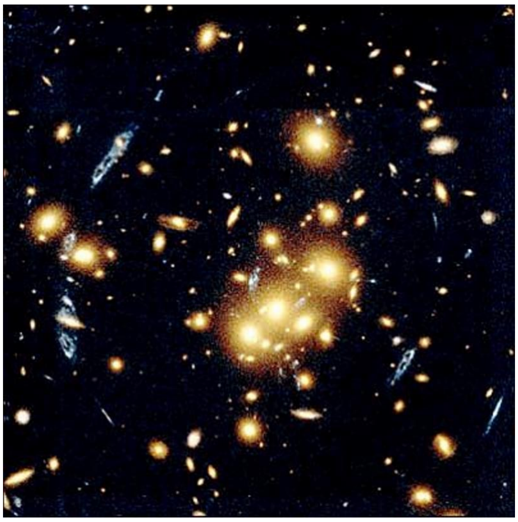


Figure 143 Multiple galaxy images around CL0024+1654

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, we start 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 these situations, 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 ever increasing, finite, but without boundary. This statement presupposes that the universe has the same topology, the same 'shape' as that of a sphere with an additional dimension.

Ref. 315 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, simply 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.* Maybe even the study of fluctuations of the cosmic background radiation can tell us something. All this research is still in its infancy.

Since little is known, we can ask about the range of possible answers. As just mentioned, in the standard model 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 $k = -1$ it is a hyperbolic manifold H^3.

See page 314 In addition, Figure 137 showed that depending on the value of the cosmological constant, space could be finite and bounded, or infinite and unbounded. In all Friedman-LeMaître calculations, simple connectedness is usually tacitly assumed, even though it is not at all required.

It could well be that space-time is *multiply* connected, like a higher-dimensional version of a torus. It could also have even more complex topologies.** In these cases, 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 range of possibilities is not limited to the simply and multiply connected cases suggested by classical physics. An additional and completely unexpected twist will appear in the third part of our walk, when quantum theory is included in the investigations. See page 767

What is behind the horizon?

The universe is a big place;
perhaps the biggest.

---

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

** 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. 316

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 the horizon moves away from us with a velocity Ref. 317 Challenge 670

$$v_{\text{horizon}} = 3c \quad . \tag{306}$$

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 671

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 question is still open. The most precise description is provided by the hypothesis of inflation.

Why are there stars all over the place? – Inflation

What were the initial conditions of matter? Obviously matter was distributed in a constant density over space expanding with great speed. How could this happen? The person to have explored this question most thoroughly is Alan Guth. So far, we based our studies of the night sky, cosmology, on two observational principles: the isotropy and the homogeneity of the universe. In addition, the universe is (almost) flat. Inflation is an attempt to understand the origin of these observations. Flatness at the present instant of time is strange: the flat state is an unstable solution of the Friedmann equations. Since the universe is still flat after twelve thousand million years, it must have been even more flat near the big bang.

Guth argued that the precise flatness, the homogeneity and the isotropy could follow if in the first second of its history, the universe had gone through a short phase of exponential size increase, which he called *inflation*. This exponential size increase, by a factor of about 10^{26}, would homogenize the universe. This extremely short evolution would be driven by a still unknown field, the *inflaton field*. Inflation also seems to describe correctly the growth of inhomogeneities in the cosmic background radiation. Ref. 318

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

Why are there so few stars? – The energy and entropy content of the universe

Die Energie der Welt ist constant. Die Entropie der Welt strebt einem Maximum zu.*

---

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

Rudolph Clausius

The matter-energy density of the universe is near the critical one. Inflation, described in the previous section, is the favourite explanation for this connection. This implies that the actual number of stars is given by the behaviour of matter at extremely high temperatures, and by the energy density left over at lower temperature. The precise connection is still the topic of intense research. But this issue also raises a question about the 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:

$$E = E_b + E_\gamma + E_\nu \approx \frac{c^2 M_o}{T_o} + ... + ... \approx \frac{c^2}{G} + ... \quad . \tag{307}$$

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

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

However, the discussion of *entropy* puts a strong question mark behind all these seemingly obvious statements. Many people try to give values for the entropy of the universe. Ref. 319 Some checked whether the relation

$$S = \frac{kc^3}{G\hbar}\frac{A}{4} = \frac{kG}{\hbar c}4\pi M^2 \quad , \tag{308}$$

Challenge 672 correct for black holes, also applies to the universe,hereby assuming that all the matter and all the radiation of the universe can be described by some average temperature. They argue that the entropy of the universe is obviously low, so that there must be some ordering principle behind it. Others even speculate where the entropy of the universe comes from, and whether the horizon is the source for it.

But let us be careful. Clausius assumes, without the slightest doubt, that the universe is a closed system, and thus deduces the above statement. Let us check this assumption. Entropy describes the maximum energy that can be extracted from a hot object. After the discovery of the particle structure of matter, it became clear that entropy is also given by the number of microstates that can make up a specific macrostate. But 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 See page 33 beginning, we defined the state as all those properties of a system which allow to distinguish

---

* Except for the case when pressure can be neglected.

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 not exist at all! Challenge 673

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

It turns out that homogeneous mass distributions are *unstable*. If for any reason the density fluctuates, regions of higher density will attract matter and increase in density, whereas regions of lower density will deplete. Can you confirm the instability, simply by assuming a space filled with dust and $a = GM/r^2$?

But how did the first inhomogeneities form? That is one of the big problems of modern physics and astrophysics, and there is no accepted answer yet. Several modern experiments 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 some information on the way to settle the issue. Ref. 320

Why are stars so small compared with the universe?

Given that the matter density is around the critical one, the size of stars, which contain most of the matter, is a result of the interaction of the elementary particles composing them. Below we will show that general relativity (alone) cannot explain any size appearing in nature. The discussion of this issue is a theme of quantum theory. See page 335

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

Can we distinguish between expanding space and galaxies moving apart? Yes, we can. Are you able to find an argument or to devise an experiment to do so? Challenge 674

The expansion of the universe does not apply to the space on the earth. The expansion is calculated for a homogeneous and isotropic mass distribution. Matter is not homogeneous nor isotropic inside the galaxy; the approximation of the cosmological principle is not valid down here. It has even been checked experimentally by studying atomic spectra in various places in the solar system that there is *no* Hubble expansion taking place around us. Ref. 321

Is there more than one universe?

'Several' universes might be an option when we study the question whether we see all the stars. But you can check that neither definition of universe given above, be it 'all matter-energy' or 'all matter-energy and all space-time', allows to answer the question positively. Challenge 675

There is no way to define a plural for universe: either the universe is everything, and then it is unique, or it is not everything, and then it is not the universe. We will discover that quantum theory does not change this conclusion, despite recurring reports of the contrary. See page 599

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

Ref. 322 Stars and arms prove that motion is obviously relative, not absolute.* This observation leads to two possible formulations of what Einstein called *Mach's principle*.

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

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

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

This formulation is more controversial. Many interpret this formulation as meaning that the *value of mass* of an object depends on the distribution of mass in the rest of the universe. That would mean that one needs to investigate whether mass is 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. Ref. 323 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. 324

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 676

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 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

---

* 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.

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. 324

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 677

At rest in the universe

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

In other words, the night sky is black because we move with almost no speed through background radiation. If the earth had a large velocity relative to the background radiation, the sky would be bright at night, at least in certain directions. Can you confirm this? Challenge 678

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 679

By the way, is the term 'universe' correct? Does the universe rotate, as its name implies? If by universe one means the whole of experience, the question does not make sense, because rotation is only defined for bodies, i.e. for parts of the universe. However, if by universe one only means 'all matter', the answer *can* be determined by experiments. It turns out that the rotation is extremely small, if there is any. In short, he who talks about the universe is actually lying. Ref. 325

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 if 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. 326 Challenge 680

Since light does not attract light moving along, light is not 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 681 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 M. Bronstein demonstrated in the 1930s that any light decay process would have a larger rate for smaller frequencies. So people checked the shift of radio waves, in particular the famous 21 cm line, and compared it with the shift of light from the same source. No difference was found for all galaxies tested. Ref. 327

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 thousands of millions of years. Ref. 328

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 682 See page 324

The visibility of the stars at night has indeed opened the door to numerous properties of nature. We now continue our mountain ascent with a more fundamental issue, nearer to our quest for the fundaments of motion.

10. 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? 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 viewpoint transformation

$$\begin{aligned} x'_0 &= x_0 + x_1 \\ x'_1 &= -x_0 + x_1 \\ x'_2 &= x_2 \\ x'_3 &= x_3 \end{aligned} \qquad (309)$$

is allowed in general relativity, and leaves the field equations invariant. You might want to search for other examples. Challenge 683

The consequence is clearly in sharp contrast with everyday life: diffeomorphism invariance makes it *impossible* to distinguish space from time *inside* general relativity. More explicitly, the coordinate x^0 cannot simply be identified with the physical time t, as implicitly

---

[*] 'Care about time.' Lucius Annaeus Seneca (ca. 4 BCE–65), *Epistolae* 88, 39

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.

How did we sweep this problem under the rug so far? We used several ways. The simplest 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. 311

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 this method is in fact a combination of the previous two.

But we are on a special quest 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 positions x_i and time t are connected, and how we can define invariant quantities. The question also prepares us for the moment when gravity is combined with quantum theory, as we will do in the third part of our mountain ascent.

A fundamental solution requires a description of clocks together with the system under consideration, and a deduction of how the reading t of the clock relates to the behaviour of the system in space-time. But we know that any description of a system requires measurements, e.g. in order to determine the initial conditions. And initial conditions require space

---

[*] 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 mountain ascent. Challenge 684

and time. We enter a vicious circle; that is precisely what we wanted to avoid in the first place.

A suspicion arises. Does a fundamental difference between space and time exist at all? Let us have a tour of the various ways to investigate the question.

Can space and time be measured?

In order to distinguish space and time in general relativity, we 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, we need *electrodynamics* to solve it. Only using the electromagnetic charge e can we form length scales, of which the simplest one is given by See page 194

$$l_{\text{scale}} = \frac{e}{\sqrt{4\pi\varepsilon_o}} \frac{\sqrt{G}}{c^2} \approx 1.4 \cdot 10^{-36}\,\text{m} \quad . \tag{310}$$

In fact, only *quantum mechanics* provides a real solution to this issue, as can be seen by rewriting the elementary charge e as the combination of nature's fundamental constant using

$$e = \sqrt{4\pi\varepsilon_o c\hbar\alpha} \quad . \tag{311}$$

This changes expression (310) into

$$l_{\text{scale}} = \sqrt{\frac{\alpha\hbar G}{c^3}} = \sqrt{\alpha}\, l_{\text{Pl}} \quad . \tag{312}$$

The expression shows that every length measurement is based on the electromagnetic coupling constant α and on the Planck length. Of course, the same is valid for time and mass measurements as well.There is no way to define or measure lengths, times and masses in general relativity alone.* Therefore, the answer to the section title being negative in general relativity, the next question is: Challenge 685

Are space and time necessary?

Ref. 330 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, which stands out. It is the only observable which allows to say that for any two observables a_1, a_2

Ref. 329 * In the past, John Wheeler used to state that his *geometrodynamic clock*, a device which measures time by bouncing back and forward between two parallel mirrors, was a counterexample; that is not correct, however. Can you confirmthis? Challenge 686 there is a third one a_3, for which

$$(a_3 - v) = (a_1 - v) + (a_2 - v) \quad . \tag{313}$$

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.

We conclude that it is possible to formulate general relativity without the use of space and time. Since both are unnecessary, it is unlikely that there is a 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 we have 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 space-time which are mutually compatible or exclusive. Ref. 298 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. Indeed, no candidate has ever been named.

That seems to point to a difference between space and time. But in fact, all these investigations do not help: they are based on the behaviour of matter. Thus these arguments imply a specific answer right from the start and do not allow to search for it. In short, also this topic cannot help to decide whether space and time differ. Let us look at the issue in another way.

Is general relativity local? – The hole argument

When Albert Einstein developed general relativity, he had quite some trouble with diffeomorphism invariance. Most startling is his famous *hole argument*, better called the *hole paradox*. Take the situation shown in Figure 144, 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 we change the curvature inside the hole while leaving the situation outside it unchanged, as shown in the inset of the picture? Ref. 331

On 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 we generalize this operation to the time domain, we get the biggest nightmare possible in physics: determinism is lost.

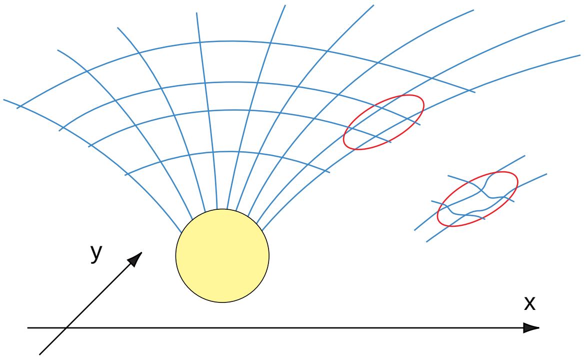


Figure 144 A ‘hole’ in space

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 favoured 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-time and gravitational field*. Space-time is a quality of the field, as Einstein put it, and not an entity with separate existence, as assumed in the graph. Coordinates have no physical meaning; only distances in space and time have one. In particular, diffeomorphism invariance proves that *there is no flow of time.* Time, like space, is only a relational entity: time and space are relative; they are not absolute.

This relativity also has practical consequences. For example, it turns out that many problems in general relativity are equivalent to the Schwarzschild situation, even though they appear completely different at first sight. As a result, researchers have ‘discovered’ the Schwarzschild solution (of course with different coordinate systems) over twenty times, often thinking that they had found a new, unknown solution. The topic has a startling consequence:

Is the earth hollow?

We live on the inside of a sphere,
as is proven by any pair of shoes.
Shoe soles are always used up at the ends,
and hardly at all in the middle.
Anonymous

The *hollow earth hypothesis*, i.e. the conjecture that we live on the *inside* of a sphere, was popular in paranormal circles around the year 1900, and still is among certain crackpots today, especially in Britain, Germany and the US. These views maintain that the solid earth encloses the sky, with the moon, the sun and the stars somewhere on the way to the centre. Obviously 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, it is said, and the hollow earth appears in all its glory.

See page 50

Interestingly, the reasoning is correct. There is *no way* to disprove this sort of description of the universe. In fact, as the great Austrian Physicist Roman Sexl used to explain, the diffeomorphism invariance of general relativity even proclaims the equivalence between the two views. The fun starts when either of the two camps wants to tell the other 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.Explaining the appearance of day and night, of the horizon and of the satellite images of the earth is possible. Explaining what happened during the flight to the moon is also fun. You can drive many bad physicists crazy in this way. The usual description and the hollow earth description are exactly equivalent. Ref. 332 Challenge 687 e

Are you able to confirm that even quantum theory, with its introduction of length scales into nature, does not change the situation? We see that diffeomorphism invariance is not an easy symmetry to swallow. Get used to it now, as the rest of our adventure will bring even more surprises. In the third part of our walk we will discover that there is an even larger symmetry in nature, quite similar to the change in viewpoint from the hollow earth view to the standard view. This symmetry is not only valid for distances measured from the centre of the earth, but for distance measured from any point in nature. Just be patient. Challenge 688 n See page 776

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 of our adventure 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. We will discover that that mass and space-time are on equal footing and that in a sense, particles and vacuum are made of the same substance. All distinctions between space and time turn out to be possible only at low, daily life energies. See page 742

In the beginning of our mountain ascent we found 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 found that space and time are required to define matter; now we found that we even need *flat* space-time to define it. See page 117

In summary, general relativity does not solve the circular reasoning we discovered in Galilean physics. General relativity even makes the issue less clear than before. Continuing the mountain ascent is really worth the effort. To increase our understanding, we now go to the limit case of gravitation.

11. Black holes – falling forever

Why study them?

Black holes are the most extreme case of gravity. Black holes realize nature's limit of length to mass ratios. Black holes produce the highest curvature possible. Therefore, they cannot be studied *without* general relativity. In addition, black holes are a central stepping stone towards unification and towards 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; Ref. 335 in addition, half a dozen smaller black holes have been identified elsewhere in our galaxy. It seems that the evolution of galaxies is strongly tied to the evolution of black holes. In addition, black holes are suspected at the heart of quasars and of gamma rays bursters. For this and many other reasons, black holes, the most impressive and the most relativistic systems in nature, are a fascinating subject of study.*

Horizons

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 Ref. 336 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 as nothing, not even light can leave them. In 1967, John Wheeler** coined the now standard term *black holes*.

It only takes a short calculation to show that light cannot escape from a mass whenever Challenge 689 the radius is smaller than a critical value given by

$$R_S = \frac{2GM}{c^2} \tag{314}$$

called the *Schwarzschild radius*. The formula is valid both in universal gravity and in general relativity, provided that in general relativity we take the radius as meaning the circumference divided by 2π. 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 ever observed one. In fact we will see that there is no way to observe such an object, in the same way that an objects faster than light cannot be observed.

Challenge 690 The surface gravity of a black hole is given by

$$g_{\text{surf}} = \frac{c^4}{4GM} = \frac{c^2}{2R_S} \quad . \tag{315}$$

A black hole thus swallows whatever falls into it, be it matter or radiation, without letting anything out. It acts like a cosmic trash can.

As it is impossible to send light from a black hole to the outside world, what happens Challenge 691 when a light beam is sent upwards from the horizon?And from slightly above the horizon?

---

* 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.

** 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 space-time*, Scientific American Library & Freeman, 1990, in which he explains general relativity with passion and in detail, but without any mathematics.

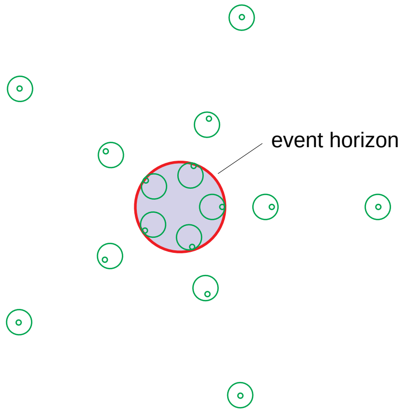


Figure 145 The light cones in the equatorial plane around a non-rotating black hole, seen from above

Black holes, when seen as astronomical objects, are thus different from planets. During the formation of planets, matter lumped together; as soon as it could not be compressed any further, an equilibrium was formed which determined 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 it overcomes all those interactions which make matter *impenetrable* in daily life. In Russian, black holes used to be called *collapsars*. A *black hole is matter in permanent free fall*. Note that despite this permanent free fall, its radius for an outside observer remains constant. Due to this permanent free fall, black holes are the only state of matter in thermodynamic equilibrium! All other states are metastable. Already in 1939, Robert Oppenheimer* and Hartland Snyder showed theoretically that a black hole forms whenever a star of sufficient mass stops burning. Ref. 338

The characterizing property of a black hole is its *horizon*. We have encountered horizons already in special relativity. Here the situation is similar; for an outside observer, the horizon 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 259

$$di^2 = (1-\frac{2GM}{rc^2})dt^2 - \frac{dr^2}{1-\frac{2GM}{rc^2}} - r^2 d\varphi^2/c^2 \quad . \qquad (316)$$

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 i of an observer at radius r is related to the time t of an observer at infinity through

$$di = \sqrt{1-\frac{2GM}{rc^2}}\,dt \quad , \qquad (317)$$

---

* Robert Oppenheimer (1904–1967), important US-American physicist. He can be called the father of theoretical physics in the USA. He worked on quantum theory and atomic physics. He then headed the development of the nuclear bomb during the second world war. He is also famous for being the most prominent as well as innocent victim of one of the greatest witch-hunts that were organized in his home country. See also the http://www.nap.edu/readingroom/books/biomems/joppenheimer.html web site.

we find that an observer at the horizon would have vanishing proper time. In other words, at the horizon the redshift is infinite.* Everything happening there goes on infinitely slowly for a far away observer. In other words, for a faraway observer nothing at all happens at the horizon.

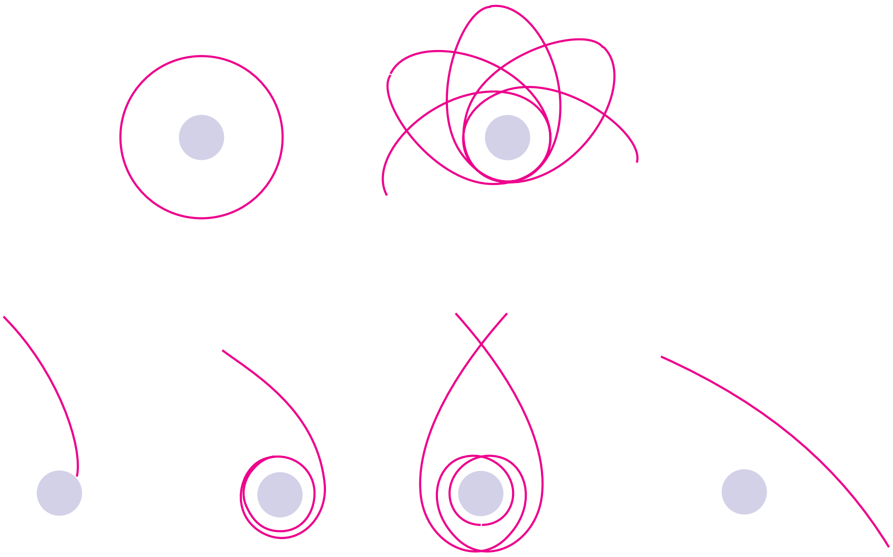


Figure 146 Motion of uncharged objects around a nonrotating black hole

Orbits

Ref. 327 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 in a plane only near *nonrotating* black holes.**

Around non-rotating black holes, also called *Schwarzschild black holes*, circular paths are impossible for radii less than $3R_S/2$ (can you show why?)and are unstable to perturbations from there up to a radius $3R_S$. Only at larger radii circular orbits are stable. Around black holes, there are no elliptic paths; the corresponding rosetta path is shown in Figure 146. Such a path shows the famous periastron shift in all its glory. Challenge 693

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, Challenge 694

---

* The surface of infinite redshift and the horizon coincide only for non-rotating black holes. For other black holes, such as rotating black holes, the two surfaces are distinct.

** For such paths, Kepler's rule connecting the average distance and the time of orbit

$$\frac{GMt^3}{(2\pi)^2} = r^3 \qquad (318)$$

still holds,provided the proper time and the radius measured by a far away observer is used. Challenge 692 that would be 42 km from its centre; at the distance of the earth, we would not be able to note any difference for the path of the earth around the sun.

Several times in our adventure we mentioned that gravitation is characterized by its tidal effects. Black holes show extreme properties also in this aspect. If a cloud of dust falls into a black hole, the size of the cloud increases when falling into it, until the cloud envelops the whole horizon. In fact, the result is valid for any extended body. This property of black holes will be of importance later on, when we will discuss the size of elementary particles.

For falling bodies coming from infinity, the situation near black holes is even more interesting. Of course there are no hyperbolic paths, only trajectories similar to hyperbolas for bodies passing far enough. 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 146. 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.In summary, relativity changes the motions due to gravity quite drastically. Challenge 695

Around *rotating* black holes, the orbits of point masses are even more complex than those shown in Figure 146; 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.

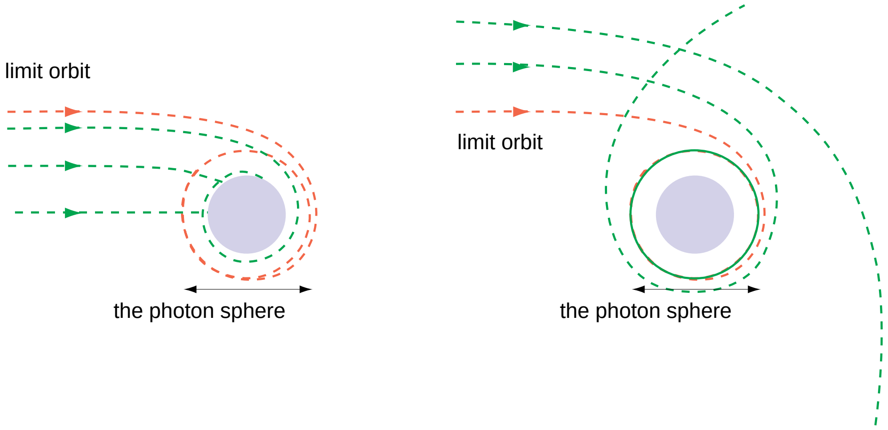


Figure 147 Motion of *light* passing near a non-rotating black hole

For *light* passing a black hole, the paths are equally interesting, as shown in Figure 147. There are no qualitative differences with the case of rapid particles, as relativity requires. For a non-rotating black hole, the path obviously lies in a single plane. Of course, if light passes sufficiently nearby there is strong bending of light, as well as capture. Again, light can also make one or several turns around the black hole before leaving or being captured. The limit between the two cases is the path in which light moves in a circle around a black hole, at $3R/2$. If we would be located on that orbit, we would see the back of our head by looking forward! However, this orbit is unstable.The space containing all orbits inside the circular one is called the *photon sphere*. The photon sphere thus divides paths leading to capture from those leading to infinity. As a note, there is no *stable* orbit for light around a black hole at all. Are there any rosetta paths for light around a black hole? Challenge 696 Challenge 697

For light 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 698

For *charged* black holes, the orbits for falling charged particles are even more complex. The electrical field lines need to be taken into account; several fascinating effects appear with no correspondence in usual electromagnetism, such as effects similar to electrical versions of the Meissner effect. The whole field is still partly unexplored and is one of today's research themes in general relativity.

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 height, following Wheeler's colourful language. Physicists also say that black holes 'have no hair,' meaning that (classical) black holes have no other degrees of freedom. This expression also was introduced by Wheeler.[***] 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. (However, the uniqueness theorem is not valid any more if the black holes carries nuclear quantum numbers, such as weak or strong charges.) Ref. 342 Ref. 343

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).

---

[*] There are other entities encountered so far with the same reduced number of characteristics: particles. More on the connection between black holes and particles will be uncovered in the third part of our mountain ascent.

[**] 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, who discovered the corresponding solution of Einstein's field equations in 1963. Electrically charged, but non-rotating black holes are often called *Reissner-Nordstrom black holes*, after the German physicist H. Reissner and the Danish physicist G. Nordström. The general case, charged and rotating, is sometimes named after Kerr and Newman. Ref. 339 Ref. 340

[***] It is not a secret that Wheeler was inspired by a clear anatomical image when he stated that 'black holes, in contrast to their surroundings, have no hair.'

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 many million solar masses. But for their angular momentum J and for their electric charge Q 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. Also for the charge there is a limit.The two limits are not independent; they are related by Challenge 699

$$\left(\frac{J}{cM}\right)^2 + \frac{GQ^2}{4\pi\varepsilon_o c^4} \leqslant \left(\frac{GM}{c^2}\right)^2 \quad . \tag{319}$$

The limit simply follows from the limit on length to mass ratios at the basis of general relativity.Rotating black holes realizing the limit (319) are called *extremal* black holes. The limit (319) also implies that the horizon radius of a general black hole is given by Challenge 700

$$r_\mathrm{h} = \frac{GM}{c^2}\left(1+\sqrt{1-\frac{J^2c^2}{M^4G^2}-\frac{Q^2}{4\pi\varepsilon_o GM^2}}\right) \tag{320}$$

For example, for a black hole with the mass and half the angular momentum of the sun, namely $2\cdot10^{30}$ kg and $0.45\cdot10^{42}$ kg m^2/s, the charge limit is about $1.4\cdot10^{20}$ C.

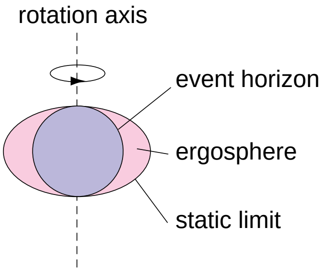


Figure 148 The ergosphere of a rotating black hole

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 (outer) horizon. The region in between, the *ergosphere*, as the name does not say, 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. Ref.

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 related to its mass by Challenge 701

$$A = \frac{16\pi G^2}{c^4}M^2 \quad . \tag{321}$$

---

[*] More about the still hypothetical magnetic charge later on. It enters like an additional type of charge into all expressions in which electric charge appears.

The surface-mass relation for a rotating and charged black hole is more complex; it is given by

$$A=\frac{8\pi G^2}{c^4}M^2\left(1+\sqrt{1-\frac{J^2c^2}{M^4G^2}-\frac{Q^2}{4\pi\varepsilon_o GM^2}}\right) \qquad (322)$$

where J is the angular momentum. In fact, the relation

$$A=\frac{8\pi G}{c^2}Mr_h \qquad (323)$$

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 which cannot have a magnetic field.

Black holes as energy sources

Can one extract energy from a black hole? Roger Penrose discovered that this is possible for Ref. 344 *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 stays constant, and only its rotation is slowed down.[**]

As a result, for a neutral black hole *rotating* with its maximum possible angular momentum, Challenge 703 $1-1/\sqrt{2}=29.3\%$ of its total energy can be extracted through the Penrose process.For black holes rotating more slowly, the percentage is obviously smaller.

For *charged* black holes, such irreversible energy extraction processes are also possible. Challenge 704 Can you think of a way?Using expression (319), we find that up to 50% of the mass of a Challenge 705 non-rotating black hole can be due to its charge.In fact, in the second part of the mountain See page 640 ascent we will encounter a process which nature seems to use quite frequently.

The Penrose process allows to determine how angular momentum and charge increase Ref. 345 the mass of a black hole. The result is the famous mass-energy relation

$$M^2=\frac{E^2}{c^4}=(m_{irr}+\frac{Q^2}{16\pi\varepsilon_o Gm_{irr}})^2+\frac{J^2}{4m_{irr}^2}\frac{c^2}{G^2}=(m_{irr}+\frac{Q^2}{8\pi\varepsilon_o\rho_{irr}})^2+\frac{J^2}{\rho_{irr}^2}\frac{1}{c^2} \qquad (324)$$

---

[*] And it would be much more dangerous, since any small object would hit such an against-the-stream satellite Challenge 702 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.

[**] 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_{irr} is the *irreducible mass* defined as

$$m_{\mathrm{irr}}^2 = \frac{A(M,0,0)}{16\pi}\frac{c^4}{G^2} = \left(\rho_{\mathrm{irr}}\frac{c^2}{2G}\right)^2 \tag{325}$$

and ρ_{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. 346

The black hole 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 706

$$S = \frac{A}{4}\frac{kc^3}{\hbar G} = \frac{Ak}{4\,l_{\mathrm{Pl}}^2} \quad . \tag{326}$$

This famous relation needs quantum theory for its deduction, as the absolute value of entropy, as for any other observable, is never fixed by classical physics alone. We will discuss the entropy expression later on in our mountain ascent. See page 642

If black holes have an entropy, they also must have a temperature. If they have a temperature, they must shine. Black holes thus cannot be black! The last conclusion was proven by Stephen Hawking in 1974 with extremely involved calculations. However, it could have been deduced already in the 1930s, with a simple Gedankenexperiment that we will present later on. See page 638 You might want to think about the issue, asking and investigating what strange consequences would appear if black holes had no entropy. Black hole radiation is a further, though tiny (quantum) mechanism for energy extraction, and is applicable even for non-rotating, uncharged black holes. See page 638 The interesting connections between black holes, thermodynamics, and quantum theory will be presented in the second part of our mountain ascent. Can you imagine other mechanisms that make black hole shine? Challenge 707

Paradoxes, curiosities, and challenges

> Tiens, les trous noirs. C'est troublant.*
> Anonyme

Black holes show many counterintuitive results. We have a look at the classical effects. The quantum effects are left for later on. . See page 644

---

[*] No translation possible. Traduttore, traditore.

- Following universal gravity, a black hole would allow that light climbs upwards from its surface and then falls back down. In general relativity, a black hole does not allow light to climb up at all; it can only fall. Can you confirm this? Challenge 708

- What happens to a person falling into a black hole? An outside observer gives a clear answer: the falling person *never* arrives there since she needs an infinite time to reach the horizon. Can you confirm this result?The falling observer however, reaches the horizon in a *finite* amount of his own time. Can you calculate it? Challenge 709 Challenge 710

This result is surprising, as it means that for an outside observer in a universe with *finite* age, black holes cannot have formed yet! At best, we can only observe systems busy forming black holes. In a sense, it might be correct to say that black holes do not exist. However, black holes could have existed right from the start in the fabric of space-time. On the other hand, we will find out later why this is impossible. In other words, it is important to keep in mind that the idea of black hole is a limit concept.

Independently of this last issue, we can confirm that in nature, the length to mass ratio always follows

$$\frac{L}{M} \geqslant \frac{4G}{c^2} \quad . \tag{327}$$

- Interestingly, the *size* of a person falling into a black hole is 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 711

This strange result will be of importance later on in our walk, and lead to important results for the size of point particles.

- An observer near a (non-rotating) black hole, or in fact near any object smaller than 7/4 times its gravitational radius, can even see the complete *back* side of the object, as shown in Figure 149. Can you imagine how the image looks?Note that in addition to the paths shown in Figure 149, 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 712 See page 267

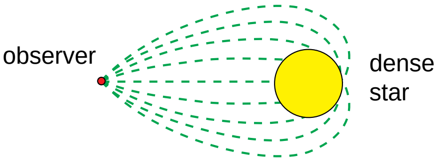


Figure 149 Motion of some light rays from a dense body to an observer

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; 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, Ref. 347

as you may want to check yourself.Only a rocket with engines switched on and pushing towards the sky can orbit a black hole at $3R/2$. Challenge 713

- By the way, how can gravity or an electrical field come out of a black hole, if no signal and no energy can leave it? Challenge 714 n
- Do *white holes* exist, i.e. time inverted black holes, in which everything flows out of instead of into some bounded region? Challenge 715
- In quantum theory, the *gyromagnetic ratio* is an important quantity for any rotating charged system. What is the gyromagnetic ratio for rotating black holes? Challenge 716
- A large black hole is, as the name implies, black. Still, it can be seen. If we would travel towards it in a space ship, we would note that the black hole is surrounded by a bright circle, like a thin halo. The ring at the distance of the photon sphere. It is due to those photons which coming from other stars towards the black hole, then circle the hole and finally, after one or several turns, end up in our eye. Can you confirm this result? Challenge 717 n
- Do moving black holes Lorentz-contract?Black holes do shine a little bit; it is true that the images they form are complex, as the light can turn around them a few times, before reaching the observer. In addition, the observer has to be far away, so that curvature has small effects. All these effects can be taken into account; nevertheless, the question remains subtle. The reason is that the concept of Lorentz contraction makes no sense in general relativity, as the comparison with the uncontracted situation is difficult to define precisely. Challenge 718
- Can you confirm that black holes provide a limit to power?Power is energy change over time. General relativity limits power to $P = c^5/4G$. In other words, no engine in nature can provide more than $0.92 \cdot 10^{52}$ W or $1.2 \cdot 10^{49}$ horsepower.[*] Challenge 719

Formation of and search for black holes

How might black holes form? At present, at least three mechanisms are distinguished; the question is still a hot subject of research. First of all, black holes could have formed during the early stages of the universe. These *primordial black holes* might grow through *accretion*, i.e. through the swallowing of nearby matter and radiation, or disappear through one of the mechanisms to be studied later on. Ref. 348 See page 640

Of the *observed* black holes, the so-called *supermassive* black holes are found at the centre of every galaxy studied so far. They have masses in the range from 10^6 to 10^9 solar masses. They are conjectured to exist at the centre of all galaxies and seem to be related to the formation of galaxies themselves. Supermassive black holes are supposed to have formed through the collapse of large dust collections, and to have grown through subsequent accretion of matter. The latest ideas imply that these black holes accrete a lot of matter in their early stage; the matter falling in emits lots of radiation, and thus would explain the brightness of quasars. Later on, the accretion calms down, and the less spectacular Seyfert galaxies form. Still later, these supermassive black holes almost get dormant, like the one in the centre of our own galaxy.

On the other hand, black holes can form when old massive stars *collapse*. It is estimated that when stars with at least three solar masses burn out their fuel, part of the matter will Ref. 349

---

[*] This statement is not yet found in the literature on general relativity. So beware.

collapse into a black hole. Such *stellar* black holes have a mass between one and a hundred solar masses; they can also continue growing through subsequent accretion. This situation provided the first 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 and formation conditions 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 Ref. 350 far is 30% of the theoretical maximum value. 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 centre of galaxies. Measuring their motion, one can deduce the mass of the body they orbit. The most favourite method to search for black holes is to look for extremely intense X-ray emission from point sources, using space-based satellites or balloon based detectors. If the distance to the object is known, its absolute brightness can be deduced; if it is above a certain limit, it must be a black hole, since normal matter cannot produce an unlimited amount of light. The method is being perfected with the aim to achieve the direct observation of energy disappearing into a horizon. This might have been Ref. 351 observed recently.

To sum up the experimental situation, 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 Ref. 335 masses. The central black hole of the galaxy M87 has 3 thousand million solar masses.

About a dozen stellar black holes between 4 and 20 solar masses are known in the rest Ref. 335 of our own galaxy, all discovered in the years after 1971, when Cygnus X-1 was found. In the year 2000, intermediate mass black holes have also been found. Astronomers are also studying how large numbers of black holes in star clusters behave, how often they collide and what sort of measurable gravitational waves these collisions produce. The list of discoveries and the related results are expected to expand dramatically in the coming years.

Singularities

Solving the equations of general relativity for various initial conditions, one finds that a cloud of dust usually collapses to a *singularity*, i.e. to a point of infinite density. The same conclusion appears when one follows the evolution of the universe backwards in time. In fact, Roger Penrose and Stephen Hawking have proven several mathematical theorems on the necessity of singularities for many classical matter distributions. These theorems assume Ref. 352 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 probably 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.

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, we can thus expect that every now and then, naked singularities emit tooth brushes. (Can you explain why dressed singularities are less dangerous?) Challenge 720

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 we encounter a prediction of an infinite value, we have extended our description of nature 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 does not allow this; quantum theory shows that it makes no sense to talk about 'singularities' nor about what happens 'inside' a black hole horizon, as time and space are not continuous at smallest distances.[**] See page 768

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

$$r_o \approx 3ct_o \approx 4 \cdot 10^{26}\,\mathrm{m} \tag{328}$$

and its matter content is about

$$m_o \approx \frac{4\pi}{3}\rho_o r_o^3 \quad \text{whence} \quad \frac{2Gm_o}{c^2} = 72\pi G\rho_o c t_o^3 = 6 \cdot 10^{26}\,\mathrm{m} \tag{329}$$

for a density of $3 \cdot 10^{-27}\ \mathrm{kg/m^3}$. Thus we have

$$r_o \approx \frac{2Gm_o}{c^2} \tag{330}$$

similar to the black hole relation $r_S = 2Gm/c^2$. Is this a coincidence?No, it is not; all systems with high curvature more or less obey the relation. But are we nevertheless falling into a large black hole? You can find out by yourself. Challenge 721 Challenge 722

12. General relativity in ten points – a summary for the layman

Sapientia felicitas.[***]

General relativity is the final description of *paths of motion*, or if one prefers, of *macroscopic motion*. General relativity describes how the observations of motion of *any* two observers

---

[*] There are also attempts to formulate such forbidding principles *inside* general relativity, called *cosmic censorship*, but we do not discuss them here. Ref. 353

[**] Many physicists are still wary to make such strong statements at this point, especially all those who claim that space and time are continuous even down to the smallest distances. The part on quantum theory and the first sections of the third part of the mountain ascent give the precise arguments leading to the opposite conclusion. See page 724

[***] 'Wisdom is happiness.' This is also the motto of Oxford university.

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. The preferred velocity is realized by massless radiation, such as light or radio signals.
- Any system of dimension L and mass M is bound by the limit

$$\frac{L}{M} \geqslant \frac{4G}{c^2} \tag{331}$$

which is realized only for black holes. From these two 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 or from tidal effects. We thus live in a pseudo-riemannian space-time. Measured times, lengths, and curvatures vary from observer to observer.
- Space-time and space 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 space curvature moves along with them. A built-in delay in this 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.
- Space is also *elastic*; it prefers being flat. Being 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 perform the distinction.

In addition, all matter and energy we observe in the sky 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.

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* Ref. 354 *moves*. Do objects really follow geodesics? As summarized in Table 32, all experiments agree with theory within measurement errors, i.e. at least within 1 part in 10^{12}. Ref. 355 In short, the way matter falls is indeed described by general relativity in all details.

The second set of measurements checks the dynamics of space-time itself. Does *space-time move* following the field equations of general relativity? In other words, is space-time

Measured effect	confirmation	type	reference
equivalence principle	10^{-12}	motion of matter	Ref. 248, 354
$1/r^2$ dependence (dimensionality of space-time)	10^{-10}	motion of matter	Ref. 356
time independence of G	10^{-19} /s	motion of matter	Ref. 354
redshift (light & microwaves on sun, earth, Sirius)	10^{-4}	space-time curvature	Ref. 244, 256, 354
perihelion shift (four planets, Icarus, pulsars)	10^{-3}	space-time curvature	Ref. 354
light deflection (light, radio waves around sun, stars, galaxies)	10^{-3}	space-time curvature	Ref. 354
time delay (radio signals near sun, near pulsars)	10^{-3}	space-time curvature	Ref. 354
gravitomagnetism (earth, pulsar)	10^{-1}	space-time curvature	Ref. 270
geodesic effect (moon, pulsars)	10^{-1}	space-time curvature	Ref. 284, 354
gravity wave emission delay (pulsars)	10^{-3}	space-time curvature	Ref. 354

Table 32 Present types of tests of general relativity

really bent by matter in the way the theory predicts? Many experiments have been performed, near and 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 Ref. 354, 355 digits. Note that even though numerous experiments have been performed, there are only few *types* of tests, as Table 32 shows; in the past, the discovery of a new type has always meant fame and riches.Most sought after, of course, is the direct detection of gravitational Challenge 723 waves.

Another comment on Table 32 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 us to reach order v^4/c^4. Soon a few effects of this order should also be detected even in- Ref. 354 side 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. See page 278

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. 355, 357

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- Ref. 354 gravity, and no unclear experimental data are known. All motion on earth and in the skies is described by general relativity. The importance of the achievement of Albert Einstein cannot be understated. We note that general relativity has not been tested for microscopic motion. In this context, *microscopic motion* is any example of motion for which the action is around the quantum of action, namely 10^{-34} Js. This issue is central to the third and last part of our adventure.

Research in general relativity and cosmology

Ref. 359 Despite all these successes, research in general relativity is more intense than ever.[*]

▪ The description of collisions and of many body problems, around the motion of stars, neutron stars and black holes, with its richness of behaviour, helps astrophysicists to improve their understanding of what they observe in their telescopes. Ref. 360

▪ 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 investigation. Ref. 363

▪ 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. 361

▪ Gathering data about galaxy formation is the main aim of many satellite systems and purpose-build telescopes. The main focus is the search for localized cosmic microwave background anisotropies due to protogalaxies. Ref. 362

▪ 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. 295

See page 640 ▪ Astrophysicists regularly discover new phenomena in the skies. For example, gamma ray bursts are still not completely understood. For a second or so, these explosion can be as bright as all other stars in the universe combined. Ref. 364

▪ 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. 365

▪ 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. 366

▪ Studying solutions with nontrivial topology, such as wormholes and particle-like solutions, is a fascinating field of enquiry, also related to string theory. Ref. 367

▪ Other formulations of general relativity, describing space-time with quantities other than the metric, are continuously developed, in the hope to clarify the relation to the quantum world. In fact, the unification of quantum physics and general relativity, the topic of the third part of this mountain ascent, will occupy researchers for many years to come. Ref. 368

▪ 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 more on the physics and less on the formalism. Ref. 369

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.

Could the constant of gravitation vary?

The constant of gravitation provides a limit for the density and the acceleration of objects, as well as for the power of engines. We based all our deductions on its invariance. Is it possible at all that the constant of gravitation G changes from place to place or that it changes with time? The question is tricky. On first sight, the answer is a loud 'Yes, of course! Just experience what happens when the value of G is changed in formulas.' However, this statement is wrong, as it was wrong for the speed of light c. See page 246

Since the constant of gravitation enters our definition of gravity and acceleration, and thus enters, even if we do not notice it, the construction of all rulers, all measurement standards and all measuring set-ups, there is *no way* to detect whether its value actually varies. No imaginable experiment could detect a variation.Every measurement of force is, whether we like it or not, a comparison with the limit force. There is no way, in principle, to check the invariance of a standard. This is even more astonishing because measurements of this types are regularly reported, and this chapter is no exception. But the result of any such experiment is easy to predict: no change will ever be found. Challenge 724 See page 352

The limits of general relativity

Even though successful, the description of motion presented so far is unsatisfactory; maybe you already have some gut feeling about certain unresolved issues.First of all, even though the speed of light is the starting point of the whole theory, we still do not know what light actually *is*. Finding out will be our next topic. Challenge 725

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? And where does mass come from anyway? What is mass? General relativity does not provide an answer; in fact, it does not describe matter *at all*. Einstein used to say that the left-hand side of the field equations, describing the curvature of space-time, was granite, the right-hand side, describing matter, was sand. Indeed, at this point we still do not know what mass is. As already remarked, to change the sand into rock we first need quantum theory and then, in a further step, its unification with relativity. This is also the program for the rest of our adventure.

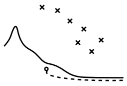
We also saw that matter is necessary to clearly distinguish space and time, and in particular, to understand the working of clocks, meter bars and balances. In particular, one question remains: why are there units of mass, length and time in nature *at all*? This deep question will also be addressed in the following.

Additionally, we found how little we know about the vacuum. We need to understand the magnitude of the cosmological constant and the number of space-time dimensions. Only then can we answer the simple question: Why is the sky so far away? General relativity does not help here. Worse, the smallness of the cosmological constant contradicts the simplest version of quantum theory; this is one of the reasons why we still have quite some height to escalate before we reach the top of Motion Mountain. See page 724

In short, to describe motion well, we realize that we need a more precise description of light, of matter and of the vacuum. In short, we need to know more about everything we know. Otherwise we cannot hope to answer questions about mountains, clocks and stars. See page 614

In a sense, it seems that we achieved quite little. Fortunately, this is not true. We learned so much that for the 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 mountain ascent. Despite the simplification to flat space-time, 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

> A man will turn over half a library to make one book.
> Samuel Johnson (1709–1784)

236 Albert Einstein, Sitzungsberichte der Preussischen Akademie der Wissenschaften II pp. 844–846, 1915, is a the first explanation of the general theory of relativity, in only three pages. The theory is then explained in detail in the famous article Albert Einstein, *Die Grundlage der allgemeinen Relativitätstheorie*, Annalen der Physik **49**, pp. 769–822, 1916.

A selection of English language textbooks for deeper study should include, in ascending order of depth and difficulty:

- Almost no formulas, but loads of insight are found in the enthusiastic text by John A. Wheeler, *A journey into gravity and space-time*, 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 beautiful, informative, and highly recommended text is Hans C. Ohanian & Remo Ruffini, *Gravitation and Spacetime*, W.W. Norton & Co., 1994.
- A well written and modern book, with emphasis on the theory, by one of the great masters of the field is Wolfgang Rindler, *Relativity – Special, General and Cosmological*, Oxford University Press, 2001.
- A classic is Steven Weinberg, *Gravitation and Cosmology*, Wiley, 1972.
- The 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.

For texts in other languages, see the next reference. Cited on pages 252, 280 and 281.

237 A beautiful German teaching text is the classic G. Falk & W. Ruppel, *Mechanik, Relativität, Gravitation – ein Lehrbuch*, Springer Verlag, Dritte Auflage, 1983.

A practical and elegant booklet is Ulrich E. Schröder, *Gravitation – Einführung in die allgemeine Relativitätstheorie*, Verlag Harri Deutsch, Frankfurt am Main, 2001.

A modern reference is Torsten Fliessbach, *Allgemeine Relativitätstheorie*, Akademischer Spektrum Verlag, 1998.

Excellent is Hubert Goenner, *Einführung in die spezielle und allgemeine Relativitätstheorie*, Akademischer Spektrum Verlag, 1996.

In Italian, there is the beautiful, informative, but expensive Hans C. Ohanian & Remo Ruffini, *Gravitazione e spazio-tempo*, Zanichelli, 1997. It is highly recommended. Cited on pages 252, 277, 278, 280 and 281.

238 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 253.

239 A good source for images of space-time is the text by G.F.R. Ellis & R. Williams, *Flat and curved space-times*, Clarendon Press, Oxford, 1988. Cited on page 258.

240 The expeditions had gone to two small islands, namely to Sobral, north of Brazil, and to Principe, in the gulf of Guinea. The results of the expedition appeared in the Times newspaper *before* they appeared in a scientific journal. Today this would be seen as a gross violation of scientific honesty. The results were published as F.W. Dyson, A.S. Eddington & C. Davidson, Philosophical Transactions of the Royal Society (London) **220A**, p. 291, 1920, and Memoirs of the Royal Astronomical Society **62**, p. 291, 1920. Cited on page 257.

241 More information about tides can be found in E.P. Clancy, *The Tides*, Doubleday, New York, 1969. Cited on page 256.

242 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 is observed in synchrotons and in star explosions. See ... Cited on page 260.

243 H. Bondi, *Gravitation*, European Journal of Physics **14**, pp. 1–6, 1993. Cited on page 254.

244 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 pages 255 and 352.

245 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 255.

246 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. 10 000 km. Cited on page 255.

247 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 255.

248 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, *The equivalence of inertial and passive* *gravitational mass*, Annals of Physics (NY) **26**, pp. 442–517, 1964, one of the most interesting and entertaining research articles in experimental physics, and by V.B. Braginsky & V.I. Panov, Soviet Physics - JETP **34**, pp. 463–466, 1971. Modern results, with errors less than one part in 10^{12}, are by Y. Su *&* al., *New tests of the universality of free fall*, Physical Review **D50**, pp. 3614–3636, 1994. Several experiments have been proposed to test the equality in space to less than one part in 10^{16}. Cited on pages 268 and 352.

249 This story is told e.g. by W.G. Unruh, *Time, gravity, and quantum mechanics*, gr-qc/9312027. Cited on page 253.

250 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 259.

251 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 260.

252 That bodies fall along geodesics has been checked by ... Cited on page 264.

253 So far, the experiments confirm that electrostatic and (strong) nuclear energy fall like matter to within one part in 10^8, and weak (nuclear) energy to within a few per cent. This is summarized in the reference 259. Cited on page 264.

254 J. Soldner, *Berliner Astronomisches Jahrbuch auf das Jahr 1804*, 1801, p. 161. Cited on page 264.

255 See the excellent explanation in N. Dragon, *Geometrie der Relativitätstheorie*, downloadable from the http://www.itp.uni-hannover.de/~dragon web site. Cited on page .

256 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 pages 255 and 352.

257 See for example K.D. Olum, *Superluminal travel requires negative energies*, Physical Review Letters **81**, pp. 3567–3570, 1998. Cited on page 266.

258 See the *Astronomical Almanac*, and its *Explanatory supplement*, H.M. Printing Office, London and U.S. Government Printing Office, Washington, 1992. For the information about various time coordinates used in the world, such as barycentric coordinate time, the time at the barycentre of the solar system, see also the http://tycho.usno.navy.mil/systime.html webpage. It also contains a good bibliography. Cited on page 266.

259 An overview is given in C. Will, *Theory and Experiment in Gravitational Physics*, chapter 14.3, revised edition, Cambridge University Press, 1993. (Despite being asstandard reference, the role of tides and the role of gravitational energy within the principle of equivalence, as discussed in the text, has also been criticized by other researchers.) See also C. Will, *Was Einstein right? – Putting general relativity to the test*, Oxford University Press, 1993. See also his paper http://arxiv.org/gr-qc/9811036. Cited on pages 267, 280 and 357.

260 The calculation omits several smaller effects, such as rotation of the earth, redshift and others. For the main effect, see Edwin F. Taylor, *'The boundaries of nature: special and general relativity and quantum mechanics, a second course in physics' – Edwin F. Taylor's acceptance speech for the 1998 Oersted Medal presented by the American Association of Physics Teachers, 6 January 1998*, American Journal of Physics **66**, pp. 369–376, 1998. Cited on page 267.

261 A.G. Lindh, *Did Popper solve Hume's problem?*, Nature **366**, pp. 105–106, 11 November 1993, Cited on page 267.

262 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 267.

263 R.J. Nemiroff, *Visual distortions near a black hole and a neutron star*, American Journal of Physics **61**, pp. 619–632, 1993. Cited on page 268.

264 Wolfgang Rindler, *Essential relativity*, Revised second edition, Springer, 1977. Cited on page 285.

265 M. Santander, L.M. Nieto & N.A. Cordero, *A curvature based derivation of the Schwarzschild metric*, American Journal of Physics **65**, pp. 1200–1209, 1997. Cited on pages 287 and 289.

266 R.P. Feynman, *Lectures on gravitation*, ... Cited on page 288.

267 The equivalence of the various definitions of the Riemann tensor is explained in ... Cited on page 301.

268 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 J.L. Anderson, *Principles of relativity physics*, chapter 4, Academic Press, 1967. Cited on page 295.

269 The Thirring effect was predicted in H. Thirring, Physikalische Zeitschrift *Über die Wirkung rotierender ferner Massen in der Einsteinschen Gravitationstheorie*, **19**, pp. 33–39, 1918, and also studied in H. Thirring, Physikalische Zeitschrift **22**, p. 29, 1921. The Thirring-Lense effect was predicted in H. Thirring & J. Lense, *Über den Einfluß der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie*, Physikalische Zeitschrift **19**, pp. 156–163, 1918. See also reference 283. Cited on page 271.

270 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 pages 272 and 352.

271 The detection of Thirring-Lense effects in binary pulsars is presented in ... Cited on page 272.

272 D. Bedford & P. Krumm, *On relativistic gravitation*, American Journal of Physics **53**, pp. 889–890, 1985. P. Krumm & D. Bedford, *The gravitational Poynting vector and energy transfer*, American Journal of Physics **55**, pp. 362–363, 1987. Cited on pages 273 and 277.

273 Jørgen Kalckar & Ole Ulfbeck, ... Cited on page 274.

274 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 275.

275 For an introduction to gravitational waves, see B.F. Schutz, American Journal of Physics **52**, p. 412, 1997, Cited on page .

276 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 278.

277 W.B. Bonnor & M.S. Piper, *The gravitational wave rocket*, gr-qc/9702005. Cited on page 279.

278 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 279.

279 A. Einstein, *Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes*, Annalen der Physik **35**, p. 898, 1911. Cited on page 280.

280 I.I. Shapiro & al., *Fourth test of general relativity*, Physical Review Letters **13**, pp. 789–792, 1964. Cited on page 281.

281 I.I. Shapiro & al., *Fourth test of general relativity: preliminary results*, Physical Review Letters **20**, pp. 1265–1269, 1968. Cited on page 281.

282 J.H. Taylor, *Pulsar timing and relativistic gravity*, Proceedings of the Royal Society, London A **341**, pp. 117–134, 1992. Cited on pages 281 and 283.

283 W. de Sitter, *On Einstein's theory of gravitation and its astronomical consequences*, Monthly Notes of the Royal Astrononmical Society **77**, pp. 155–184, p. 418E, 1916. For a discussion of de Sitter precession and Thirring-Lense precession, see also B.R. Holstein, *Gyroscope precession in general relativity*, American Journal of Physics **69**, pp. 1248–1256, 2001. Cited on pages 283 and 358.

284 B. Bertotti, I. Ciufolini, P.L. Bender, Physical Review Letters **58**, pp. 1062–1065, 1987. Later is was confirmed by I.I. Shapiro & al., *Measurement of the de Sitter precession of the moon: a relativistic three body effect*, Physical Review Letters **61**, pp. 2643–2646, 1988. Cited on pages 283 and 352.

285 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 285.

286 See for example the paper by K. Dalton, *Gravity, Geometry, and Equivalence*, http://www.arxiv.org/abs/gr-qc/9601004 and L. Landau & E. Lifshitz, *The Classical Theory of Fields*, Pergamon, 4th edition, 1975, p. 241. Cited on page 299.

287 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 297.

288 Macrobus, *Dream of Scipio*, XIV, 19. See Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 61, 1991. Cited on page 310.

289 See for example A.N. Cox, editor *Allen's Astrophysical Quantities*, AIP Press and Springer Verlag, 2000. Cited on page 305.

290 G. Gamow, *The origin of the elements and the separation of galaxies*, Physical Review **74**, p. 505, 1948. Cited on page 309.

291 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 309.

292 Arno A. Penzias & Rober W. Wilson, *A measurement of excess antenna temperature at 4080 Mcs*, Astrophysical Journal **142**, pp. 419–421, 1965. Cited on page 309.

293 An opposite idea is defended by ... Cited on page 308.

294 A. Friedmann, Zeitschrift für Physik **10**, p. 377, 1922, and Zeitschrift für Physik **21**, p. 326, 1924. Cited on page 313.

295 The experiments are discussed in detail in the excellent review by D. Giulini & N. Straumann, *Das Rätsel der kosmischen Vakuumenergiedichte und die beschleunigte Expansion des Universums*, Physikalische Blätter **556**, pp. 41–48, 2000. See also N.

STRAUMANN, *The mystery of the cosmic vacuum energy density and the accelerated expansion of the universe*, European Journal of Physics **20**, pp. 419–427, 1999. Cited on pages 318 and 353.

296 Supernova searches are being performed by ... Cited on page 318.

297 Up to his death, Fred Hoyle defended his belief that the universe is static, e.g. in G. BURBIDGE, F. HOYLE & J.V. NARLIKAR, *A different approach to cosmology*, Physics Today **52**, pp. 38–44, 1999. The team has also written a book with the same title, published in 2000 by Cambridge University Press. Cited on page 320.

298 STEPHEN W. HAWKING & G.F.R. ELLIS, *The large scale structure of space-time*, Cambridge University Press, Cambridge, 1973. Among others, this reference text discusses the singularities of space-time, and their necessity in the history of the universe. Cited on pages 320, 336 and 363.

299 Augustine ... Cited on page 322.

300 S. HAWKING, *A brief history of time – From the big bang to black holes*, 1988. Cited on page 322.

301 This famous quote is the first sentence of the final chapter, the 'Beschluß', of IMMANUEL KANT, *Kritik der praktischen Vernunft*, 1797. Cited on page 301.

302 P. JETZER, *Gravitational microlensing*, Naturwissenschaften **86**, pp. 201–211, 1999. Cited on page 306.

303 AETIUS, Opinions, III, I, 6. See JEAN-PAUL DUMONT, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 445, 1991. Cited on page 302.

304 Galaxy mass measurements are described by ... Cited on page 303.

305 G. BASRI, *The discovery of brown dwarfs*, Scientific American pp. 77–83, April 2001. Cited on page 306.

306 A. HARVEY & E. SCHUCKING, *Einstein's mistake and the cosmological contant*, American Journal of Physics **68**, pp. 723–727, 2000. Cited on page 319.

307 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 319.

308 H. KNUTSEN, *Darkness at night*, European Journal of Physics **18**, pp. 295–302, 1997. Cited on page 317.

309 P. PESIC, *Brightness at night*, American Journal of Physics **66**, pp. 1013–1015, 1998. Cited on pages 317 and 318.

310 PAUL WESSON, *Olber's paradox and the spectral intensity of extra-galctic background light*, Astrophysical Journal **367**, p. 399, 1991. Cited on page 317.

311 STEVEN WEINBERG, *Gravitation and cosmology*, John Wiley, 1972. An excellent book written with a strong personal touch, and stressing most of all the relation with experimental data. It does not develop a strong feeling for space-time curvature, and does not address the basic problems of space and time in general relativity. Excellent for learning how to actually calculate things, but less for the aims of our mountain ascent. Cited on pages 317 and 334.

312 Star details are explained in many texts. See for example ... Cited on page 323.

313 JOHANNES PELT, Astronomy and Astrophysics 1995. Cited on page 325.

314 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 even be found online, in the paper by J. WAMBSGANSS, Living Reviews in Relativity **1-12**, pp. 1–80, 1998, to be found 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 325.

315 M. Lachièze-Rey & J.-P. Luminet, *Cosmic topology*, Physics Reports **254**, pp. 135–214, 1995. See also B.F. Roukema, *Cosmic topology*, http://arxiv.org/abs/astro-ph/0010185 preprint. Cited on page 327.

316 Thanks to Steve Carlip for clarifying this point. Cited on page 327.

317 G.F.R. Ellis, T. Rothmann, *Lost horizons*, American Journal of Physics **61**, pp. 883–893, 1993. Cited on page 328.

318 A. Guth, *Die Geburt des Kosmos aus dem Nichts - Die Theorie des inflationären Universums*, Droemer Knaur, München 1999. Cited on page 328.

319 Entropy values for the universe have been discussed by Ilya Prigogine, ... For a different approach, see G.A. Mena Marugán & S. Carneiro, *Holography and the large number hypothesis*, http://arxiv.org/gr-qc/0111034. It also gives the often heard statement that the universe has an entropy that is much smaller than the theoretical maximum. The maximum is 'generally' estimated to be $k\,10^{120}$,whereas the actual value is 'estimated' to be $k\,10^{100}$. However, other authors give $k\,10^{84}$.

Challenge 726

In 1974, Roger Penrose also made statements about the entropy of the universe. Cited on page 329.

320 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 330.

321 The lack of expansion in the solar system is shown in ... Cited on page 330.

322 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 331.

323 The independence of inertia has been tested ... Cited on page 331.

324 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 pages 331 and 332.

In a related development, in 1953, Dennis Sciama published a frequently cited paper in which he argues that inertia of a particle is due to the gravitational attraction of all other matter in the universe. The paper is widely quoted, but makes no new statements on the issue. See D.W. Sciama, *On the origin of inertia*, **113**, pp. 34–42, ... 1953.

325 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, or E.F. Bunn, P.G. Fereira & J. Silk, *How anisotropic is the universe?*, Physical Review Letters **77**, pp. 2883–2886, 1996. Cited on page 332.

326 The issue has been discussed within linearized gravity by Richard Tolman, in his textbook *Relativity, thermodynamics, and cosmology*, Clarendon Press, 1934, on pp. 272–290. The exact problem has been solved by A. Peres, *Null electromagnetic fields in general relativity theory*, Physical Review **118**, pp. 1105–1110, 1960, and by W.B. Bonnor, *The gravitational field of light*, Commun. Math. Phys. **13**, pp. 163–174, 1969. See also N.V.

Mitskievic & K.K. Kumaradtya, *The gravitational field of a spinning pencil of light*, Journal of Mathematical Physics **30**, pp. 1095–1099, 198. Cited on page 332.

327 For a delightful popular account from a black hole expert, see Igor Novikov, *Black holes and the universe*, Cambridge University Press 1990. Cited on pages 333 and 341.

328 ... Carilli & al., *Limits on the change of fine structure constant ...*, Physical Review Letters 25 december 2000. Cited on page 333.

329 The geometrodynamic clock is discussed in D.E. Brahm & R.P. Gruber, *Limitations of the geometrodynamic clock*, General Relativity and Gravitation **24**, pp. 297–303, 1992. The clock itself was introduced by R.F. Marzke, in his Ph.D. thesis *The theory of measurement in general relativity*, 1959, performed with John Wheeler as thesis advisor. Cited on page 335.

330 R. Geroch, *Einstein Algebras*, Commun. Math. Phys. **26**, pp. 271–275, 1972. Cited on page 335.

331 A. Macdonald, *Einstein's hole argument*, American Journal of Physics **69**, pp. 223–225, 2001. Cited on page 336.

332 Roman U. Sexl, *Die Hohlwelttheorie*, Der mathematisch-naturwissenschaftliche Unterricht **368**, number 8, pp. 453–460, 1983. Roman U. Sexl, *Universal Conventionalism and Space-Time.*, General Relativity and Gravitation **1**, pp. 159–180, 1970. See also Roman U. Sexl, *Die Hohlwelttheorie*, pp. 241–258, in Arthur Scharmann & Herbert Schramm, editors, *Physik, Theorie, Experiment, Geschichte Didaktik, Festschrift für Wilfried Kuhn zum 60. Geburtstag am 6. Mai 1983*, Aulis Verlag, Deubner, Köln, 1984. Cited on page 338.

333 See for example, the definition of the ADM mass in ... Cited on page 298.

334 For a historical introduction, see the paper by R. Ruffini, *The physics of gravitationally collapsed objects*, ... Cited on page .

335 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 pages 339 and 349.

336 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 339.

337 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 .

338 The beautiful paper is R. Oppenheimer & H. Snyder, *On continued gravitational contraction*, Physical Review **56**, pp. 455–459, 1939. Cited on page 340.

339 R.P. Kerr, *Gravitanional field of a spinning mass as an example of algebraically special metrics*, Physical Review Letters **11**, pp. 237–238, 1963. Cited on page 343.

340 E.T. Newman, E. Couch, R. Chinnapared, A. Exton, A. Prakash & R. Torrence, *Metric of a rotating, charged mass*, Journal of Mathematical Physics **6**, pp. 918–919, 1965. Cited on page 343.

341 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 .

342 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 343.

343 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 343.

344 R. Penrose & R.M. Floyd, *Extraction of rotational energy from a black hole*, Nature **229**, pp. 177–179, 1971. Cited on page 345.

345 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 345.

346 J.D. Bekenstein, *Black holes and entropy*, Physical Review **D7**, pp. 2333–2346, 1973. Cited on page 346.

347 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 347.

348 About black holes in the early universe, see ... Cited on page 348.

349 About black holes formation via star collapse, see ... Cited on page 348.

350 Frederick Lamb, APS meeting 1998, binary star 4U1820-30, 20 000 lightyears from earth, Physics News Update, April 27, 1998. Cited on page 349.

351 The first evidence for the observation of matter falling into a black hole was publicized in early 2001. ... Cited on page 349.

352 For a readable summary of the Penrose-Hawking singularity theorems, see ... Details can be found in reference 298. Cited on page 349.

353 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 350.

354 T. Damour, *Experimental tests of relativistic gravity*, http://arxiv.org/gr-qc/9904057. It is the latest in a series of his papers on the topic; the first was T. Damour, *Was Einstein 100% right?*, http://arxiv.org/gr-qc/9412064. Cited on pages 351 and 352.

355 H. Dittus, F. Everitt, C. Lämmerzahl & G. Schäfer, *Die Gravitation im Test*, Physikalische Blätter **55**, pp. 39–46, 1999. Cited on pages 351 and 352.

356 The inverse square dependence has been checked down to 0.2 mm, as reported by ... Physics Today ...2000. Cited on page 352.

357 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. This effect, predicted by Kenneth Nordvedt, would appear if the gravitational energy in the earth-moon system would fall differently from the earth and the moon themselves. For a summary of the measurements, see J. Müller, M. Schneider, M. Soffel & H. Ruder, Astrophysical Journal Letters **382**, p. L101, 1991. Cited on page 352.

358 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.M. Will, *Gravitational radiation and the validity of general relativity*, Physics Today **52**, p. 38, October 1999. Cited on page .

359 Almost everything of importance 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 353.

360 Collisions and many body problems ... Cited on page 353.

361 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 353.

362 The ESA satellite Planck will measure the polarization of the cosmic microwave background. Cited on page 353.

363 Inflation and early universe ... Cited on page 353.

364 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 353.

365 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 353.

366 Torsion is presented in ... Cited on page 353.

367 Wormholes and nontrivial topologies ... A basic approach is the one by Mark Hadley, Classical and Quantum Gravity October 2000. Cited on page 353.

368 Other formulations of relativity ... Cited on page 353.

369 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 353.

370 A beautiful introduction into modern astronomy is Paolo Maffei, *I mostri del cielo*, Mondadori Editore, 1976. Cited on page .

371 Michael H. Soffel, *Relativity in astronomy, celestial mechanics and geodesy*, Springer Verlag, 1989. Cited on page .

372 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 .

373 The speed of light as function of height has been checked e.g. by ... The change is smaller than ... Cited on page .

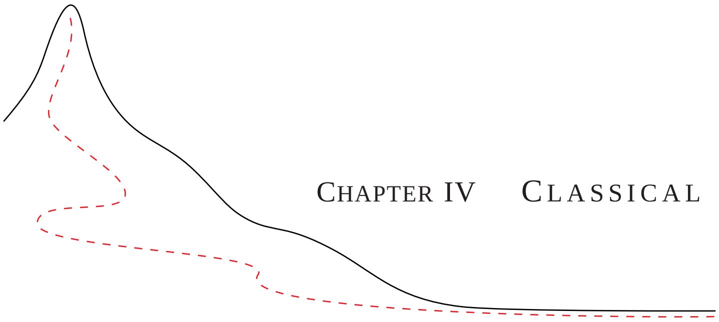
374 Steven Weinberg, *The cosmological constant problem*, Reviews of Modern Physics **61**, pp. 1–23, 1989. This influential review article on the views on cosmological constant problem at that time is frequently cited. Cited on page .

375 A. Ashtekar, 1986, *New variables for classical and quantum gravity*, Physical Review Letters **57**, pp. 2244–2247, Cited on page .

376 See the new book by R. Wald, ... Cited on page .

377 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
January 2003



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. We also learned that light is that moving entity which cannot be stopped; 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, such as the ways magicians levitate objects.

13. Liquid electricity and invisible fields

See page 110

Challenge 727

Revisiting the list of of motors found in this world, we remark that gravitation does not describe almost any of them. Neither the motion of sea waves, of fire, of 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 with a stethoscope?Without having done so, you cannot pretend to have experienced the mystery of motion. You have about 3000 million beats in your lifetime. Then they stop.

It was one of the most astonishing discoveries of science that heart beats, sea waves 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, the softness and the impenetrability of matter are due to internal electricity; also the emission of light is an electrical process. As these aspects are part of everyday life, we leave aside all complications due to gravity and curved space-time. The most productive way to study electrical motion is to start, like in the case of gravity, with those types of motion which are generated without any contact between the involved bodies.

Amber, lodestone, and mobile phones

> 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 a tree is cut, a viscous resin appears. With time it solidifies, and after millions of years it forms *amber*. When amber is rubbed with a cat fur, it acquires the ability to attract small 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 always surprised by the effect a rubbed comb has on a running water tap. Challenge 728 e

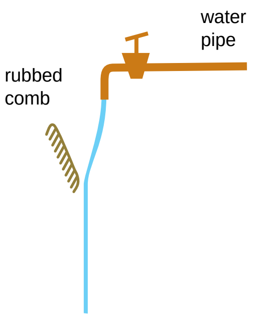


Figure 150 How to amaze kids

Another 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 two stones of this mineral are put 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 we also find various little objects in nature with more sophisticated properties. Some are able to switch on televisions, others unlock car doors, still others allow to talk with far away friends.

All these observations show that in nature there are situations where bodies exert influence on others *at a distance*. The space surrounding a body exerting such an influence is said to contain a field. A *(physical) field* is thus an entity which manifests itself by accelerating other bodies in that region of space. A field is some 'stuff' taking up space. Experiments show that fields have no mass. The field surrounding the mineral found in Magnesia is called a *magnetic field* and the stones themselves *magnets*.[*] The field around amber – called 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.[**]

The field around a mobile phone is called a *radio* field, or as we will see later, an *electromagnetic* field. In contrast to the previous fields, it oscillates over time. We will find out later that many other objects are surrounded by such fields, though often very weak. Objects which emit oscillating fields, such as mobile phones, are called radio transmitters or radio emitters.

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.

[**] The Kirlian effect, which allows to make so intriguingly beautiful photographs, is due to a time-varying electric field.

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. 379
Search in meteorites	none Ref. 379
Search in cosmic rays	none Ref. 379
Search with high energy accelerators	none Ref. 379

Table 33 Some searches for magnetic monopoles, i.e., for magnetic charges

Observation	Magnetic field
Lowest measured magnetic field	ca. 1 fT
Magnetic field produced by brain currents	ca. 0.1 pT to 3 pT
Intergalactic magnetic fields	1 pT to 10 pT
Magnetic field in the human chest, due to heart currents	ca. 100 pT
Magnetic field of our galaxy	0.5 nT
Magnetic field of earth	20 μT to 70 μT
Magnetic field below high voltage power line	ca. $10^{-\cdot\cdot}$ T
Magnetic field inside modern home	10^{-7} T to 10^{-4} T
Magnetic field near mobile phone	ca. $10^{-\cdot\cdot}$ T
Magnetic field in light beam	… T
Magnetic field near iron magnet	100 mT
Solar spots	ca. 1 T
Magnetic fields near high tech permanent magnet	max 1.3 T
Magnetic fields in particle accelerator	ca. 10 T
Maximum static magnetic field produced with superconducting coils	22 T
Highest long time static magnetic fields produced in laboratory using hybrid magnets	50 T
Highest *pulsed* magnetic fields produced without coil destruction	74 T
Pulsed magnetic fields produced, during about 1 μs, using imploding coils	ca. 1000 T
Field on neutron star	from 10^{6} T to 10^{11} T
Quantum critical magnetic field	$4.4 \cdot 10^{9}$ T
Highest field ever measured, on magnetar and soft gamma repeater SGR-1806-20	0.8 to $1 \cdot 10^{11}$ T
Field near nucleus	ca. 1 TT
Maximum (Planck) magnetic field	$2.2 \cdot 10^{53}$ T

Table 34 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.

Figure yet to be included

Figure 151 Lightning: a picture taken with a moving camera, showing the multiple strokes it consists of

How can one make lightning?

Everybody has seen a lightning or has observed 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 m/s. But *what* is moving? To find out, we have to find a way to make lightning by ourselves.

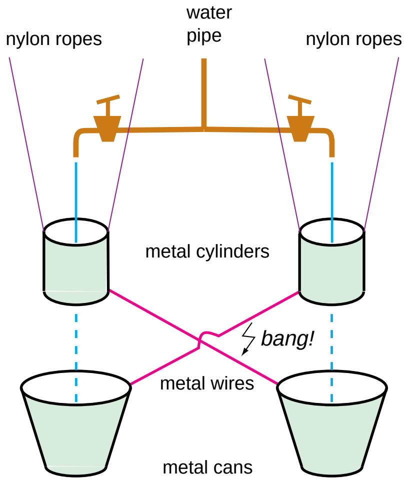


Figure 152 A simple Kelvin generator

In 1995, the car company General Motors accidentally rediscovered an old and simple method for achieving this. They had inadvertently built a spark generating mechanism into their cars; when filling the tank with fuel, sparks were generated which sometimes lead to the explosion of the fuel. They had to recall 2 million vehicles of its Opel brand. What had they done? Ref. 380

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, we would take a few water taps, four empty bean or coffee cans, of which two have been opened at both sides, some nylon rope and some metal wire. Ref. 381

Putting all together as shown in Figure 152 and letting the water flow, we find 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 did to repair the cars? Challenge 729 n

If we stop the water flow just before the next spark is due, we find that both buckets attract sawdust and pieces of paper. The generator thus does the same that rubbing amber does, just with more bang for the buck(et). Both buckets are surrounded by electric fields.

---

[*] William Thomson (1824–1907), important Irish unionist 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 strongly influenced the development of the theory of magnetism and electricity, the description of the aether, and 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 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.

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. We also find 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; charges of opposite sign flowing together cancel each other out.*

In summary, electric fields start at bodies, provided 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 a hundred years of use of electricity, the highest concentration of copper in the world is below the surface of Manhattan.

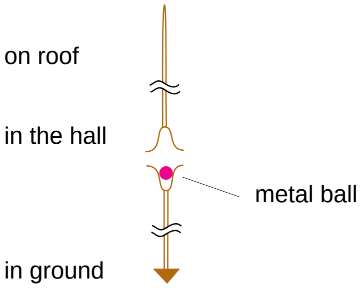


Figure 153 Franklin's personal lightning rod

Of course, one has to check whether real thunderstorm lightning actually is electrical in origin. In 1752, experiments performed in France, following a suggestion 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 had a lightning rod built through his own house, but of a somewhat unusual type, as shown in Figure 153. Can you guess what it did in his hall during bad weather, all parts being made of metal? Challenge 730 n

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 thus only flows when two differently charged bodies are brought into contact. Now, if charge can flow and accumulate, we 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

---

[*] In fact, there are many other ways to produces sparks or even *arcs*, i.e. sustained sparks; there is even a complete subculture of people who do this as a hobby at home. Those who have a larger budget do it professionally, in particle accelerators. See the http://www.mathematik.uni-marburg.de/~kronjaeg/hv/index.html web site.

[**] There is still research going on into the details of how lightning is generated and how it propagates. A little about this topic is said on page 411.

electric 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 35 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. $10^{62\pm2}$ C

Table 36 Values of electrical charge observed in nature

of mass m accelerated in a field, its unknown charge q is determined by the relation

$$\frac{q}{q_{\text{ref}}} = \frac{ma}{m_{\text{ref}}a_{\text{ref}}} \quad , \tag{332}$$

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 we need 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 we are forced to use for its description a scalar quantity q, which can take positive, vanishing, or negative values.

In everyday life these properties of electric charge, listed also in Table 35, 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 $\mathbf{x}$ 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

$$q\mathbf{E}(\mathbf{x}) = m\mathbf{a}(\mathbf{x}) \tag{333}$$

Observation	Electric field
Cosmic noise	ca. 10 μV/m
Field 1 m away from an electron	...
Field of a 100 W FM radio transmitter at 100 km distance	0.5 mV/m
Field in solar wind	...
Field in clouds	...
Field inside conductors, such as copper wire	0.1 V/m
Field inside a typical home	1 to 10 V/m
Field of a 100 W bulb at 1 m distance	50 V/m
Ground field in earth's atmosphere	100 to 300 V/m
Maximum electric field in air before sparks appear	1 to 3 MV/m =$1to3$ kV/mm
Electric fields in biological membranes	10 MV/m
Electric fields inside capacitors	up to 1 GV/m
Electric fields in most intense laser beams	100 TV/m
Electric fields in U^{91+} ions, at nucleus	1 EV/m
Maximum electric field in vacuum, limited by pair production	1.3 EV/m
Planck electric field	$6.5 \cdot 10^{61}$ V/m

Table 37 Some observed electric fields

taken at every point in space **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 unit V/m. Challenge 732 e

To describe motion due to electricity completely, we also need 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

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_o}\frac{Q}{r^2}\frac{\mathbf{r}}{r} \quad \text{where} \quad \frac{1}{4\pi\varepsilon_o} = 8.9\,\text{GVm/C} \quad . \tag{334}$$

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 origin of this dependence? Challenge 733 n

---

Challenge 731 * Does the definition of electric field given here assume a speed of the charge much smaller than that of light?

** 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. The SI system 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, The two previous equations allow to write the interaction between two charged bodies as

$$\frac{\mathrm{d}\mathbf{p}_1}{dt} = \frac{1}{4\pi\varepsilon_o}\frac{q_1 q_2}{r^2}\frac{\mathbf{r}}{r} = -\frac{\mathrm{d}\mathbf{p}_2}{\mathrm{d}t} \qquad (335)$$

where d**p** is the momentum change, and **r** is the vector connecting the two centres of mass. This famous expression for electrostatic attraction and repulsion, 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, all atoms are kept together by electrostatic attraction. As a final example to 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 734

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. People 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.

Do uncharged bodies attract each other? In first approximation they do not. But when the question is investigated more precisely, one finds that they can attract each other. Can you find the conditions for this to happen? In fact, the conditions are quite important, as our own bodies, which are made of neutral molecules, hold together in this way. Challenge 736 n

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. We have to be careful when using it. In fact the vocabulary issue hides a deeper question, which was still unanswered at the beginning of the twenty-first century: what is the nature of electric charge? Since charge flows, we can start by asking:

Can we feel the inertia of electricity?

If electric charge really is something *flowing* through metals, we should be able to observe the effects shown in Figure 154. Already Maxwell predicted most of them: electric charge should fall, have inertia, and be separable from matter. And indeed, each of these effects has been observed.[**] For example, when a long metal rod is kept vertically, we can measure an electrical potential difference, a voltage, between the top and the bottom. In other words,

---

Challenge 735 n [*] By the way, are batteries sources of charges?

[**] Maxwell also performed experiments to detect these effects (apart from the last one, which he did not predict), but his apparatuses where not sensitive enough.

we can measure the *weight* of electricity in this way. Similarly, we can measure potential Ref. 382 differences between the ends of an accelerated rod. In particular, we can measure a potential difference between the centre 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

$$q/m = 1.8 \cdot 10^{11}\,\mathrm{C/kg} \qquad (336)$$

for all metals, with small variations. In short, electrical current has mass. Therefore, whenever we switch on an electrical current, we get a *recoil*. This simple effect can easily be Ref. 383 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. 384 the beam producing the picture. It works best when metal objects have a sharp, pointed tip. The rays created this way – we could say that they are 'free' electricity – are called *cathode rays*. Within a few per cent, they show the same mass to charge ratio as expression (336). This correspondence thus shows that charges in metals move almost as freely as in air; that is the reason metals are such good conductors.

If electric charge falls *inside* vertical metal rods, we 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 exact compensation of the acceleration by the electrical field generated by the displaced electricity in the tube and the acceleration of gravity. Challenge 737 e Thus electrons should notbe able to fall through long thin cylinders. The experiment has indeed Ref. 385 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?

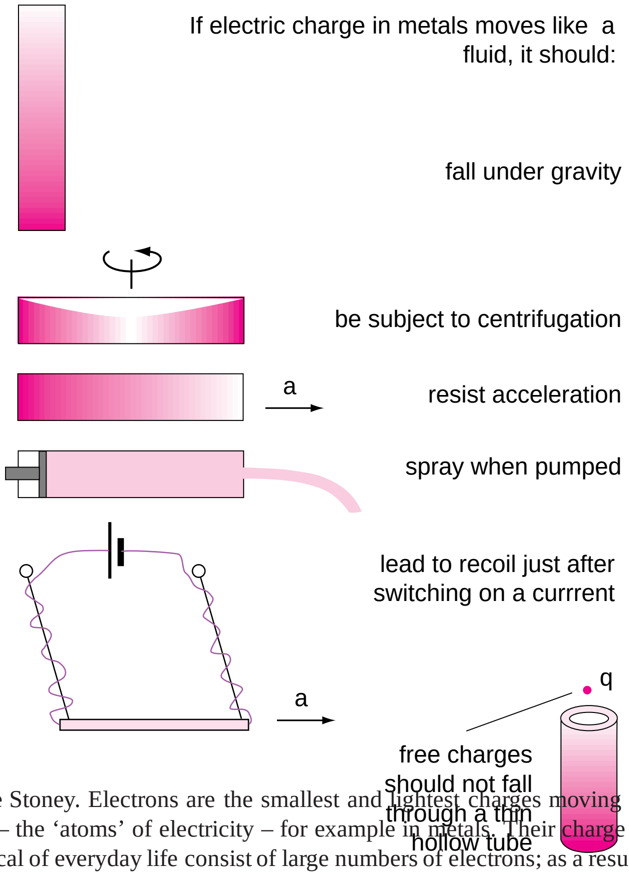


Figure 154 Consequences of the flow of electricity

---

[*] The name 'electron' is due to Johnstone Stoney. Electrons are the smallest and lightest charges moving in metals; they are, usually – but not always – the 'atoms' of electricity – for example in metals. Their charge is small, 0.16 aC, so that flows of charge typical of everyday life consist of large numbers of electrons; as a result, charge behaves like a continuous fluid.

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 738 n

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 $0.5c$.) This is due to the ability of metals to easily take in arriving charges and to let depart others. But when the speed of charges inside metals is measured, the *electrons*, 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 739 n

Inside liquids, charges move with different speed than inside metals, and their charge to mass ratio is also different. We all 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 metres long. A similar speed is observed inside semiconductors and inside batteries. In all these systems, moving charge is transported by *ions*; they are charged atoms. Ions, like atoms, are large and composed entities, in contrast to the tiny electrons.

In other systems, charges move both as electrons and as ions. Examples are neon lamps, fire, plasmas, or the sun itself. Inside 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 atoms and their structure will be told in the second part of our mountain ascent.

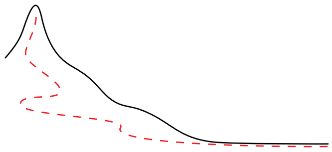
Magnets

Parallel to the study of electricity, the study of magnetism had progressed independently across the world. Towards the end of the 12th century, the compass came into use in Europe. There was a heated debate on whether it pointed to the north or the south. Already in 1269, the French military engineer Pierre de Maricourt (1219–1292) published his study of magnetic materials. He found that every magnet has two points of high magnetization, Ref. 386 and called them poles. He found that even by cutting a magnet, the resulting pieces always retain two poles, pointing to the north and the south when the stone is left free to rotate. There are no magnetic monopoles.

How can one make a motor?

> Communism is soviets plus electricity.
> Lenin (1870, Simbirsk-1924, Gorki)

The reason for Lenin's famous statement were two discoveries. One was made in 1820 by the Danish physicist Hans Christian Oersted (1777–1851) and the other in 1831 by the



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

English physicist Michael Faraday.[*] The consequences of these experiments changed the world completely in less than one century.

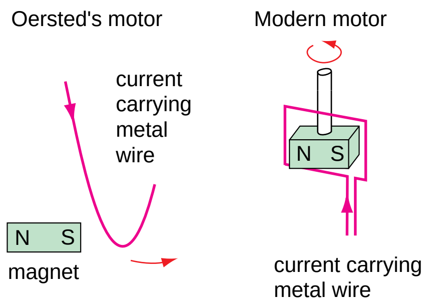


Figure 155 An ancient and a modern version of an electric motor

On the 21st of July of 1821, Oersted published a leaflet, in Latin, which took Europe by storm. Oersted had found that when a current is sent through a wire, a nearby magnet is put into motion. In other words, he found that the flow of electricity can move bodies.

Further 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 other experiments show that wires in which electricity flows behave like magnets.[**] In other words, Oersted had found that electricity could be turned into magnetism.

Shortly afterwards, Ampère[***] found that *coils* increase these effects dramatically. Coils behave like little magnets. In particular, coils, like 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.

Experiments show that the magnetic field turns out to always have a given direction in space, and to have a magnitude common to all (resting) observers. We are 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 now? Challenge 741 e

---

[*] 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 any mathematical training, later in 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 the idea of (magnetic) fields and field lines through all his experimental discoveries, such as induction, paramagnetism, diamagnetism, electrochemistry and the Faraday effect. Fields were later described mathematically by Maxwell, who at his time was the only person in Europe who took over the concept.

[**] In fact, if one imagines tiny currents moving in circles inside magnets, one gets a unique description for all magnetic fields observed in nature.

[***] 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: electrical current can deviate magnetic needles. Ampère worked for years on the problem, and in 1826 published the summary of his findings, which lead Maxwell to call him the 'Newton of electricity'. Ampère named and developed many parts of electrodynamics. The unit of electrical current is named after him.

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*

$$\mathrm{B} = \begin{pmatrix} 0 & -B_z & B_y \\ B_z & 0 & -B_x \\ -B_y & B_x & 0 \end{pmatrix} \quad , \tag{337}$$

called an *antisymmetric tensor*. (It is also called a *pseudovector*; note that also angular momentum and torque are examples of such quantities.) In summary, *magnetic fields* are defined by the acceleration they impart on moving charges. This acceleration turns out to follow

$$\mathbf{a} = \frac{e}{m}\mathbf{v}\mathrm{B} = \frac{e}{m}\mathbf{v} \times \mathbf{B} \tag{338}$$

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.** The Lorentz acceleration is the effect at the basis of any electric motor. An electric motor is a device using magnetic fields as efficiently as possible to accelerate charges flowing in a wire. Through their motion the wire is then moved as well. Electricity is thus transformed into magnetism and then into motion.

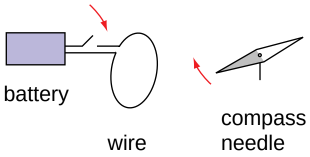


Figure 156 An electrical current always produces a magnetic field

Like in the electric case, we need to know how the *strength* of magnetic fields is determined. Experiments like Oersted's show that the magnetic field is due to moving charges, and that a charge moving with velocity **v** produces a field given by

$$\mathrm{B}(\mathbf{r}) = \frac{\mu_o}{4\pi} q \frac{\mathbf{v} \times \mathbf{r}}{r^3} \quad \text{where} \quad \frac{\mu_o}{4\pi} = 10^{-7}\,\mathrm{N/A^2} \ . \tag{339}$$

Again, the strange factor $\mu_o/4\pi$ is due to the historical way the electrical units were defined. 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.*** Challenge 743 e

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. Magnetism can thus also be turned into electricity. This important discovery allowed the production of electrical current flow with generators, so-called *dynamos*, using water power, wind power or steam power. They started the modern use of electricity in our world. Behind every electrical plug there is a dynamo somewhere.

---

[*] 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 **H** will not appear in this walk, but is important for the description of magnetism in materials.

[**] Does the definition of magnetic field given here assume a speed of the charge much smaller than that of light? Challenge 742

[***] Equation (339) is valid only for small velocities and accelerations. Can you find the general one? Challenge 744

Additional experiments show that magnetic fields also lead to electric fields when one changes to a moving viewpoint You might check this on any of the examples of the Figures 155 to 160. *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. Magnetic and electrical fields thus behave like space and time, which are also mixed up when changing from one inertial frame to the other. The theory of special relativity thus 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*

$$\mathrm{F}^{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix} \quad \text{or} \quad \mathrm{F}_{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix} \tag{340}$$

Obviously, the electromagnetic field, and thus every component of these matrices, depends on space and time. The matrices show that electricity and magnetism are two faces of the same effect.* In addition, since electric fields appear only in the topmost row and leftmost column, the expressions show that in everyday life, for small speeds, electricity and magnetism *can* be separated.

The total influence of electric and magnetic fields on fixed or moving charges is then given by the following expression:

$$\begin{aligned} & m\mathrm{b} = \mathrm{F}u \quad \text{or} \\ & m\,\frac{du^\mu}{d\tau} = q\mathrm{F}^\mu{}_\nu u^\nu \quad \text{or} \\ & m\frac{d}{d\tau}\begin{pmatrix} \gamma c \\ \gamma v_x \\ \gamma v_y \\ \gamma v_z \end{pmatrix} = q \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix} \begin{pmatrix} \gamma c \\ \gamma v_x \\ \gamma v_y \\ \gamma v_z \end{pmatrix} \quad \text{or} \\ & W = q\mathbf{E}\mathbf{v} \quad \text{and} \quad d\mathbf{p}/dt = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \end{aligned} \tag{341}$$

All four expressions describe the same content; the simplicity of the first one is the reason for the involved matrices (340) of the electromagnetic field. In fact, the extended *Lorentz relation* (341) 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 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 travelling of electrical signals in cables and in the nerves of the body. Ref. 387, 388

---

* Actually, the expression for the field contains everywhere the expression $1/\sqrt{\mu_0 \varepsilon_0}$ instead of the speed of light c. We will explain the reason for this substitution shortly.

The electromagnetic field tensor F is an *antisymmetric* 4-tensor. (Can you write down Challenge 745 the relation between $F^{\mu\nu}$, $F_{\mu\nu}$, and $F^{\mu}{}_{\nu}$?)Like any such such tensor, it has two invariants, i.e., two deduced properties which are the same for every observer: the expression $B^2 - E^2/c^2 =$ Challenge 746 $\frac{1}{2}\mathrm{tr}\,F^2$ and the product $4\mathbf{EB} = -c\,\mathrm{tr}\,F^{*}F$. (Can you confirm this?)

The first expression turns out to be the Lagrangian of the electromagnetic field. It is a scalar and implies that if E is larger, smaller, or equal cB for one observer, it also is for all other observers. The second invariant, a pseudoscalar, describes whether the angle between the electric and the magnetic field is acute or obtuse for all observers.*

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 useful fact that charges can flow in metals, that electromagnetic energy can be transformed into mechanical energy (sound, motors), into light (lamps), into heat and coldness (ovens, refrigerators), that electromagnetic fields can be sent across the air (radio and television, remote controls), and that electric or magnetic fields can be used to store information (computers).

How motors prove relativity right

> The only mathematical operation I performed in my life
> was to turn the handle of a calculator.
> Michael Faraday

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 Ref. 390 maximal speed impossible.

The argument is beautifully simple. We change the original experiment and imagine two long, electrically charged rods of mass m, moving in the same direction with velocity v and separation d. An observer moving with the rods would see an electrostatic repulsion Challenge 748 e between the rods given by

$$ma_e = -\frac{1}{4\pi\varepsilon_o}\frac{2\lambda^2}{d} \qquad (343)$$

---

* 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:

$$\begin{aligned}\kappa_3 &= \frac{1}{2}A_\mu A^\mu F_{\rho\nu}F^{\nu\rho} - 2A_\rho F^{\rho\nu}F_{\nu\mu}A^\mu \\ &= (\mathbf{AE})^2 + (\mathbf{AB})^2 - |\mathbf{A}\times\mathbf{E}|^2 - |\mathbf{A}\times\mathbf{B}|^2 + 4A^4(\mathbf{AE}\times\mathbf{B}) - A^8(E^2+B^2) \qquad (342)\end{aligned}$$

Ref. 389 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 Lorentz Challenge 747 invariants. (Why?)The latter, the frame independence 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*.

where λ is the charge per length of the rods. A second, resting observer sees two effects: the electrostatic repulsion and the attraction discovered by Ampère. He therefore observes Challenge 749 e

$$ma_{em} = -\frac{1}{4\pi\varepsilon_o}\frac{2\lambda^2}{d} + \frac{\mu_o}{2\pi}\frac{\lambda^2 v^2}{d} \quad . \tag{344}$$

It is easy to check that the second observer sees a repulsion, as the first one does, only if

$$v^2 < \frac{1}{\varepsilon_o \mu_o} \quad . \tag{345}$$

This maximum speed, with a value of 0.3 GM/s, is thus valid for any object carrying charges. But all everyday objects contain charges: there is thus a maximum speed for matter. Are you able to expand the argument to neutral particles as well?More on this limit velocity, which we know already, will be found out below. Challenge 750

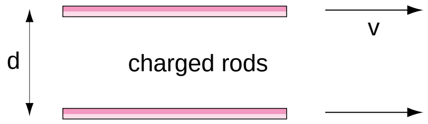


Figure 157 The relativistic aspect of magnetism

In summary, electric effects are due to flow of electric charges and to electric fields. Magnetism is due to *moving* electric charges. It is *not* due to magnetic charges.* The strength of magnetism, used in any running electric motor, proves relativity right: there is a maximum speed in nature. Both electric and magnetic fields carry energy and momentum. They are two faces of the same coin. However, our description of electromagnetism is not complete yet; we need the final description of the way charges *produce* the electromagnetic field.

Curiosities and fun challenges about things electric and magnetic

Et facta mirari et intellectua assequi.
Augustinus

Before we study the motion of electromagnetic field in detail, we have some fun with electricity.

▪ How can you distinguish a magnet from an unmagnetized metal bar of the same size and material, using no other means? Challenge 752 n

▪ How do you wire a light bulb, the mains, and three switches so that the light can be switched on at any of the switches and off at any other switch? And in case of four switches? Challenge 753 n Nobody will take a physicist serious who can write Maxwell's equations but cannot solve this little problem.

▪ The first appliances to generate electric currents where large rubbing machines. Then the Italian scientist Alessandro Volta (1745–1827) constructed a new device to generate electricity and called it a *battery*. Batteries are based on chemical processes, provide much

---

* 'Electrons move in metal with a speed of about 1 mm; thus if I walk with the same speed along a cable carrying a constant current, I should not be able to sense any magnetic field.' What is wrong with theargument? Challenge 751

more current, work in all weathers, are smaller and are easier to handle that electrostatic machines. The invention of batteries changed the investigation of electricity so much that the unit of electrical tension was deduced from the name of Volta. A 'battery' is a large number of cells; the term was taken over from a previous, almost purely military usage. An apple with an inserted copper and an inserted zinc piece of metal is one of the simplest possible cells. It provides about 1 V of electrical tension. Batteries in mobile phones are just a number of elaborated apple replacements.

Challenge 754 n ▪ Can you make a mirror that does not exchange left and right? In two different ways?

▪ A concave mirror shows an inverted image, if the mirror is bent along the horizontal line. What happens if this mirror is turned around the line of sight? Challenge 755 n

▪ A scotch tape roll is a dangerous device. Pulling the roll quickly leads to light emission (through triboluminescence) and to small sparks. It is suspected that several explosions in mines were triggered when such sparks ignited a combustible gas mixture.

▪ Take an envelope, wet it and close it. After letting it dry for a day or more, open it in the dark. At the place where the two papers are being separated from each other, the envelope glows with a blue colour. Why? Is it possible to speed up the test using a hair dryer? Challenge 756 n

▪ Electricity produced by friction and by flows of liquids is a small effect. However, in the 1990s, several oil tankers disappeared suddenly, because they had washed their oil tanks by pointing a hose spraying sea water on their walls. The spraying led to charging; with the oil fumes in the tank this led to an explosion and the tankers sank. Similar accidents also happen regularly when chemicals are filled from one tank to another.

▪ When a ship sinks, survivors usually end up in small boats drifting on the sea. Often they are saved by a rope hanging from a helicopter. It is essential a survivor only touches the rope *after* the rope has been in the water, as he can die of heart attack otherwise: the helicopter can be heavily charged.

▪ The names anode and cathode were suggested by William Whewell and popularized by Michael Faraday. Whewell formed them from the greek; they literally mean 'upward street' and 'descending street'.

▪ The shortest light pulse produced so far had a length of 100 as. How many wavelengths of green light would that correspond to? Challenge 757

▪ Why do we often see shadows of houses, shadows of trees, but never shadows of the electrical cables hanging over streets? Challenge 758 n

▪ How would you measure the speed of the tip of a lightning bolt? What range do you expect? Challenge 759

Challenge 760 ▪ How would you show that electrical charge comes in smallest chunks?

Ref. 432 ▪ One of the simplest possible electric motors was discovered by Faraday in 1831. A magnet suspended into mercury starts to turn around its axis if a current flows through it. In addition, if the magnet is made to turn from outside, the device (in other geometries also called Barlow's wheel) also works as a current generator, and people even tried to generate domestic current with such a system! Can you explain how it works? Challenge 761

▪ Cosmic radiation consists of charged particles hitting the earth. (We will discuss it in more detail later.) See page 663 Astrophysicists explain that these particles are accelerated by the magnetic fields around the galaxy. Ref. 424 However, the expression of the Lorentz acceleration shows that magnetic fields can only change the direction of the velocity of charges, not its magnitude. How can nature get acceleration nevertheless? Challenge 762

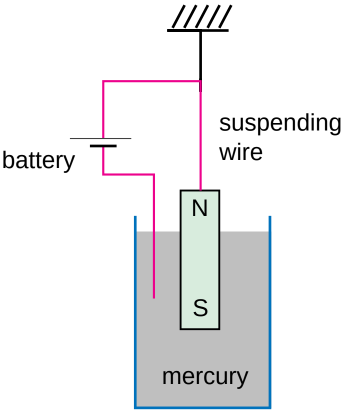


Figure 158 Unipolar motor

- The magnetic field of the earth, much higher than that of other planets because of the moon, with a dipole strength of $7.8 \cdot 10^{22}\,\mathrm{Am}^2$, shields us from lethal solar wind and cosmic radiation particles. We owe it our life. Ref. 442
- The ionosphere around the earth has a resonant frequency of 7 Hz; for this reason any apparatus measuring low frequencies always gets a strong signal at this value. Can you give an explanation of the frequency? Challenge 763 n
- What would be the potential of the earth if we could take all the electrons of a drop of water away? Challenge 764
- The sun is visible to the naked eye only up to a distance of 50 light years. True? Challenge 765
- At home, electricity is mostly used as alternating current. In other words, no electron actually flows through cables; as the speed of metal electrons is about 1 mm/s, electrons just move back and forward by 20 μm. Nothing is flowing in or out of the cables! Why do the electricity companies require an actual flow of money in return, instead of being satisfied with a back and forth motion of money? Challenge 766 e
- Do electrons and protons have the same charge? Experiments show that they are equal to within at least twenty digits. How would you check this?
- Charge is also velocity-independent. How would you check this?

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 159 and 160. Twenty years later, Heaviside and Hertz extracted the main points of Maxwell ideas and called the summary *Maxwell's theory of the electromagnetic field*. It consists of two equations (four in the nonrelativistic case). See page 910

---

* 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 colours 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.

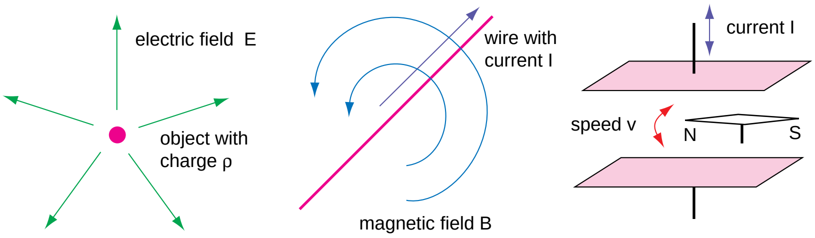


Figure 159 The first of Maxwell's equations

The first result is the precise statement that electromagnetic fields *originate at charges*, and nowhere else. The corresponding equation is variously written*

$$
\begin{aligned}
&d\mathrm{F} = j\sqrt{\frac{\mu_o}{\varepsilon_o}} \quad \text{or}\\
&d^{\nu}\mathrm{F}_{\mu\nu} = j^{\mu}\sqrt{\frac{\mu_o}{\varepsilon_o}} \quad \text{or} \qquad\qquad (346)\\
&(\partial_t/c, -\partial_x, -\partial_y, -\partial_z)\begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix} = \sqrt{\frac{\mu_o}{\varepsilon_o}}(\rho, j_x/c, j_y/c, j_z/c) \quad \text{or}\\
&\nabla\mathbf{E} = \frac{\rho}{\varepsilon_o} \quad \text{and} \quad \nabla\times\mathbf{B} - \frac{1}{c^2}\frac{\partial\mathbf{E}}{\partial t} = \mu_o\mathbf{j} \quad ,
\end{aligned}
$$

putting into many signs a simple statement: *electrical charge carries the electromagnetic field*. This statement, including its equations, are equivalent to the three basic observations of Figure 159. It describes Coulomb's relation, Ampère's relation, and the way changing
Challenge 767 currents induce magnetic effects, as you may want to check.

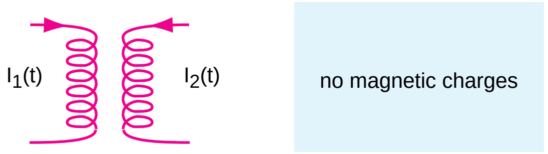


Figure 160 The second of Maxwell's equations

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 mountain ascent, the apparently special case of vacuum in fact describes all of nature.

results are described by the the relation variously written

$$\begin{gathered}
d\,{}^*\mathrm{F}=0 \quad \text{with} \quad {}^*\mathrm{F}^{\rho\sigma}=\tfrac{1}{2}\varepsilon^{\rho\sigma\mu\nu}\mathrm{F}_{\mu\nu} \quad \text{or} \\
\varepsilon_{\mu\nu\rho}\partial_\mu \mathrm{F}_{\nu\rho}=\partial_\mu \mathrm{F}_{\nu\rho}+\partial_\nu \mathrm{F}_{\rho\mu}+\partial_\rho \mathrm{F}_{\mu\nu}=0 \quad \text{or} \\
\begin{pmatrix} \gamma\frac{1}{c}\partial_t \\ \gamma\partial_x \\ \gamma\partial_y \\ \gamma\partial_z \end{pmatrix}
\begin{pmatrix} 0 & B_x & B_y & B_z \\ -B_x & 0 & -E_z/c & E_y/c \\ -B_y & E_z/c & 0 & -E_x/c \\ -B_z & -E_y/c & E_x/c & 0 \end{pmatrix}
=\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{or} \\
\nabla\mathbf{B}=0 \quad \text{and} \quad \nabla\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t} \quad .
\end{gathered} \tag{347}$$

The relation expresses the *lack of sources for the dual field tensor*, usually written ${}^*\mathrm{F}$. There are no magnetic charges, i.e. no magnetic monopoles in nature. In practice, this equation is always needed together with the previous one. Can you see why? Challenge 768

We now have a system as organized as the expression $a=GM/r$ or as Einstein's field equations for gravitation. Together with Lorentz' evolution equation (341), which describes how charges move given the motion of the fields, Maxwell's evolution equations (347) and (348) describe *all* electromagnetic phenomena at everyday scales, from portable phones, car batteries, to personal computers, lasers, lightning, holograms, and rainbows.

We will not study many applications of the equations; we continue directly towards our aim to understand the connection to everyday motion and to motion of light. In fact, the electromagnetic field has an important property which we mentioned already right at the beginning: it itself can also move.

Colliding charged particles

A simple experiment clarifies the just defined properties of electromagnetic fields. When two charged particles collide, their total momentum is *not* conserved.

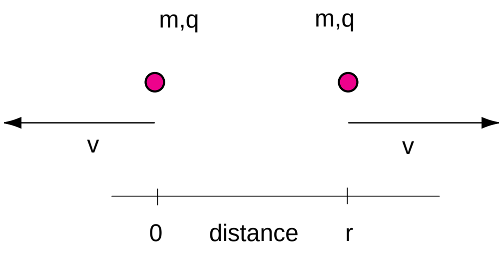


Figure 161 Charged particles after a collision

Imagine two particles of identical mass and identical charge just after a collision, when they move away 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* Challenge 769

$$E=\frac{q(1-v^2/c^2)}{4\pi e_o r^2} \quad . \tag{348}$$

In other words, the total system has a vanishing total momentum.

Take a second observer, moving with respect to the first with velocity v, so that 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. 393 Challenge 770 n

This at first surprising effect has even been put in form of a theorem, by Van Dam and Ref. 394 Wigner. They showed that for a system of particles interacting at a distance the total energy-momentum cannot remain constant in all inertial frames.

The total momentum of the system is conserved only because the electromagnetic field itself also carries momentum. For a moving observer, the electromagnetic field carries the missing momentum. If electromagnetic fields have momentum, they are able to *hit* objects and to be hit by them. As we will show below, also light is an electromagnetic field. Thus we should be able to move objects by shining light onto them. We 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.

We conclude that any sort of field leading to particle interactions must carry energy and momentum, as the argument applies to all such cases. In particular, it applies to the nuclear interactions. Indeed, in the second part of our mountain ascent we will even find an additional result: all fields are themselves composed of particles. The energy and momentum of fields then become an obvious state of affairs.

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 mountain ascent 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 of 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*. They can impart it to particles. The experiments with motors have shown that objects can add energy and momentum to fields. We therefore have to define a *state function* which allows us 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 momentum per charge the field Ref. 391 provides:

$$\mathbf{A} = \frac{\mathbf{p}}{q} \quad . \tag{349}$$

When a charged particle moves through a magnetic potential $\mathbf{A}(\mathbf{x})$, its momentum changes by $q\Delta\mathbf{A}$; it changes by the difference between the potential values at the start and end points, multiplied by its charge. Due to this definition, the vector potential has the property that

$$\mathrm{B} = \nabla \times \mathbf{A} = \mathrm{curl}\,\mathbf{A} \tag{350}$$

Challenge 771 n * What is the relation, for static fields, between field lines and (equi-) potential surfaces? Can a field line cross a potential surface twice?

For more details on topics such as these, see the *free* textbook by BO THIDÉ, *Electromagnetic Field Theory*, on his http://www.plasma.uu.se/CED/Book web site. And of course, in English, have a look at the texts by Ref. 388 Schwinger and by Jackson.

i.e. that the magnetic field is the curl of the magnetic potential.* For example, the vector potential for a long straight current carrying wire is parallel to the wire, and has the magnitudes Challenge 772 e

$$A(r) = -\frac{\mu_o I}{2\pi}\ln(r) \tag{351}$$

which depend 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 we find the vector potential Challenge 773

$$\mathbf{A}(\mathrm{r}) = -\frac{1}{2}\mathbf{B}\times\mathbf{r} \quad . \tag{352}$$

However, there is a catch. The magnetic potential is *not* defined uniquely. If $\mathbf{A}(\mathbf{x})$ is a vector potential, then

$$\mathbf{A}'(\mathbf{x}) = \mathbf{A}(\mathbf{x}) + \operatorname{grad}\Lambda \tag{353}$$

where $\Lambda(t,\mathbf{x})$ is some scalar function, is also a vector potential for the same situation. The magnetic field B stays the same, though. Can you confirm that the corresponding (absolute) momentum values also change? Challenge 774

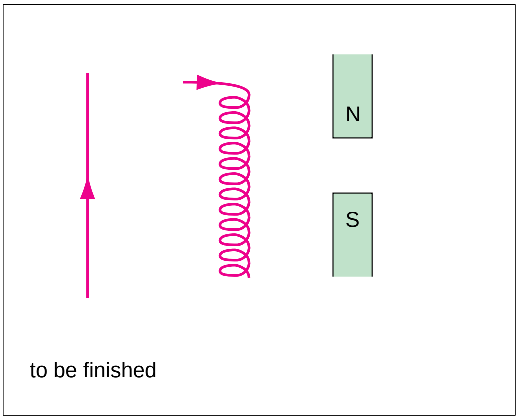


Figure 162 Vector potentials for selected situations

Not only momentum, also the energy of the electromagnetic field is defined ambiguously. Indeed, the second step in the specification of a state for the electromagnetic field is definition of the *electric potential* as the energy U per charge: Ref. 391

$$\varphi = \frac{U}{q} \tag{354}$$

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 corrected by the time dependence of momentum, i.e.

$$\mathbf{E} = -\nabla\varphi - \frac{\partial}{\partial t}\mathbf{A} \quad , \tag{355}$$

Obviously, there is a freedom in the choice of the definition of the potential. If $\varphi(\mathbf{x})$ is a possible potential, then

$$\varphi'(\mathbf{x}) = \varphi(\mathbf{x}) - \frac{\partial}{\partial t}\blacksquare \tag{356}$$

---

* The curl is called the *rotation* and abbreviated rot in most languages.

is also a potential function for the same situation. This freedom is the generalization of the freedom to define energy up to a constant. Nevertheless, the electric field **E** remains the same for all potentials.

In relativistic 4-vector notation, the energy and the momentum of the field appear together. The state function of the electromagnetic field becomes

$$A^{\mu} = (\varphi/c, \mathbf{A}) \quad . \tag{357}$$

It is easy to see that the description of the field is complete, since we have

$$\mathrm{F} = dA \quad \text{or} \quad \mathrm{F}^{\mu\nu} = \partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu} \tag{358}$$

which means that the electromagnetic field is completely specified by the 4-potential A. But as just said, the 4-potential itself is *not* uniquely defined. Indeed, any other gauge field A' related to A by the *gauge transformation*

$$A'^{\,\mu} = A^{\mu} + \partial^{\mu}\Lambda \tag{359}$$

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 we have to check 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. We add that many theoretical physicists use the term 'electromagnetic field' indifferently for the quantities $F^{\mu\nu}$ or A_{μ}.

There is a simple image, due to Maxwell, to help overcoming the conceptual difficulties of the vector potential. It turns out that the closed line integral over A_{μ} is gauge invariant, because Challenge 775 e

$$\oint A_{\mu} dx^{\mu} = \oint (A_{\mu} + \partial_{\mu}\Lambda)\, dx^{\mu} = \oint A'_{\mu} dx^{\mu} \quad . \tag{360}$$

In other words, if we picture the vector potential as a quantity allowing to associate a number to a tiny ring at each point in space, we get a good, gauge invariant picture of the vector potential.*

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.

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 definedby** Challenge 776

Ref. 392 * 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.

$$S_{\mathrm{CED}} = -mc^2\int d\tau - \tfrac{1}{4\mu_o}\int \mathrm{F}\wedge *\,\mathrm{F} - \int j\wedge A \qquad (361)$$

which in index notation becomes

$$S_{\mathrm{CED}} = -mc\int_{-\infty}^{\infty}\sqrt{\eta_{\mu\nu}\frac{dx_n^{\mu}(s)}{ds}\frac{dx_n^{\nu}(s)}{ds}}\,ds - \int_{\mathbf{M}}\left(\tfrac{1}{4\mu_o}\mathrm{F}_{\mu\nu}\mathrm{F}^{\mu\nu} + j_{\mu}A^{\mu}\right)d^4x \quad .$$

In other words, the least action principle, as usual, states that the change of a system is always as small as possible. New is the measure of the change produced by the electromagnetic field. Its internal change is given by the term F^*F, and the change due to interaction with matter is given by the term jA.

The action S_{CED} leads to the evolution equations by requiring that it be stationary under variations δ and δ' of the positions and of the fields which vanish at infinity. In other terms, the principle of least action requires that

$$\begin{aligned} \delta S = 0 \quad & \text{when} \quad x_\mu = x_\mu + \delta_\mu \quad \text{and} \quad A_\mu = A_\mu + \delta'_\mu \quad , \\ & \text{provided} \quad \delta x_\mu(\theta) \to 0 \quad \text{for} \quad |\theta| \to \infty \\ & \text{and} \quad \delta A_\mu(x_\nu) \to 0 \quad \text{for} \quad |x_\nu| \to \infty \quad . \end{aligned} \qquad (362)$$

In the same way as in the case of mechanics, using the variational method for the two variables A and x, we recover the evolution equations for particle and fields See page 135 Challenge 777

$$b^{\mu} = \frac{q}{m}\mathrm{F}^{\mu}_{\nu}u^{\nu} \quad , \quad \partial_{\mu}\mathrm{F}^{\mu\nu} = j^{\nu}\sqrt{\frac{\mu_o}{\varepsilon_o}} \quad , \quad \text{and} \quad \varepsilon^{\mu\nu\rho\sigma}\partial_{\nu}\mathrm{F}_{\rho\sigma} = 0 \qquad (363)$$

which we know already. Obviously, they are equivalent to the variational principle based on S_{CED}. Both descriptions have to be completed by specifying *initial conditions* for the particles and the fields, as well as *boundary conditions* for the latter. We need the first and zeroth derivatives of the position of the particles, and the zeroth derivative for the electromagnetic field.

Are you able to specify the Lagrangian of the pure electrodynamic field using the fields **E** and **B** instead of F and $^*F$? Challenge 778

The form of the Lagrangian 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 breaking of bodies, electric light bulbs, etc. Can you explain how this fits together? Challenge 779

In summary, with Lagrangian (361) all of classical electrodynamics is described and understood. For the rest of this chapter, we look at some specific topics from this vast field.

∗∗ The product described by the symbol ∧, 'wedge', has a precise mathematical meaning, defined in the next equation for this case. 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.

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 located, it describes the energy and momentum.

Since the electromagnetic field is not a localized entity like a point particle, but an extended entity, we need to know the *flow* of energy and momentum at every point in space, separately *for each direction*. This makes a description with a *tensor* necessary.

$$\mathrm{T}^{\mu\nu} = \left(\begin{array}{c|c} u & \mathrm{S}/c = c\mathrm{p} \\ c\mathrm{p} & T \end{array} \right) = \left(\begin{array}{c|c} 0 & \mathrm{E}\times\mathrm{B}/\mu_o c \\ \hline \mathrm{E}\times\mathrm{B} & -\varepsilon_o E_i E_j - B_i B_j/\mu_o \\ \mu_o c & 1/2\delta_{ij}(\varepsilon_o E^2 + B^2/\mu_o) \end{array} \right) \qquad (364)$$

– CS – to be continued – CS –

In summary, electrodynamic motion, like all other examples of motion encountered so far, is deterministic, conserved and reversible. That is no big news. But two special symmetries of electromagnetism deserve special mention.

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.

Electrodynamics give a second answer: a mirror is a device that switches magnetic north and south poles. Can you confirm this? Challenge 780

But is it always possible to distinguish left from right? This seems easy: this text is rather different from a bɘɿoɿɿim 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 163 is the original?

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 such as hop or the winding direction of snail shells, have been extensively studied.* Can you name a few more? See page 672 Challenge 781 n

---

* 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 in the region called *node*, determine the right-left asymmetry. The deep origin of this asymmetry is not yet elucidated, however. Ref. 395

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. But many people have more than one.

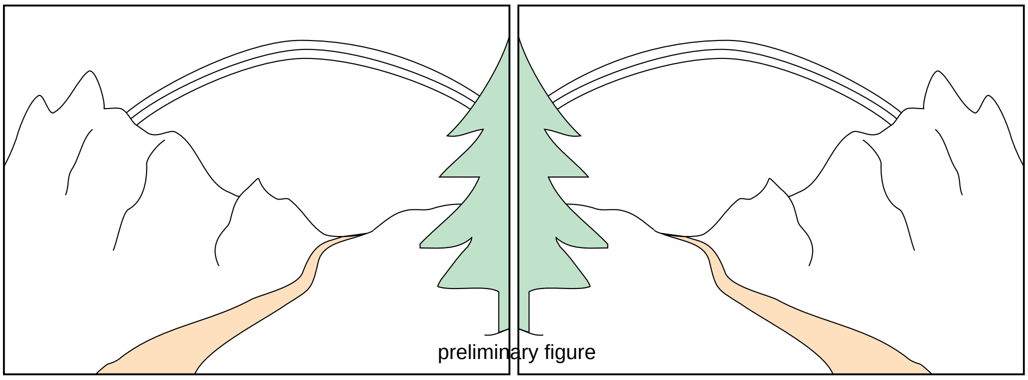


Figure 163 Which one is the original landscape?

The left-right symmetry of nature appears because everyday nature is described by gravitation and, as we will see, by electromagnetism. Both interactions share an important property: substituting all coordinates in their equations by the negative of their values leaves the equations 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 782 n

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. 396 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.

To sum up this digression, classical electrodynamics is left-right symmetric, or parity invariant. Can you show this using its Lagrangian? Challenge 783

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. See page 676 If found, the action (361) would have to be modified by the addition of a fourth term, namely the magnetic current density. However, no such particle has yet been detected, despite intensive search efforts.

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

$$\begin{aligned} \mathbf{E} &\to c\mathbf{B} \\ \mathbf{B} &\to -\mathbf{E}/c \end{aligned} \tag{365}$$

called (electromagnetic) *duality* transformation, transforms each vacuum Maxwell equation into the other. (In fact, there are even more such transformations; can you spot them?) Alternatively, the duality transformation transforms F into *F. In other words, in vacuum we *cannot* distinguish electric from magnetic fields.

Challenge 784 n

Matter would be symmetric under duality only if magnetic charges, also called magnetic monopoles, would exist. In that case the transformation (365) could be extended to

$$c\rho_{\mathrm{e}} \to \rho_{\mathrm{m}} \quad , \quad \rho_{\mathrm{m}} \to -c\rho_{\mathrm{e}} \quad . \tag{366}$$

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 somewhat difficult question: extending this duality to quantum theory, can you deduce what transformation is found for the fine structure constant, and why it is so interesting?)

See page 688

Challenge 785

Duality, by the way, is a symmetry that works *only* in Minkowski space-time, i.e. in space-times of 3 plus 1 dimensions. Mathematically, it is closely related to the existence of quaternions, to the possibility of interpreting Lorentz boosts as rotations in 3+1 dimensions, and last but not least, to the possibility to define other smooth mathematical structures than the standard one on the space R^4. These mathematical connections are still mysterious at the time being; they somehow point to the special role that four space-time dimensions play in nature. More details will become clear in the third part of our mountain ascent.

14. 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

$$\Box\,\mathbf{A} = 0 \quad \text{or} \quad \varepsilon_o \mu_o \frac{\partial^2 \varphi}{dt^2} + \frac{\partial^2 A_x}{dx^2} + \frac{\partial^2 A_y}{dy^2} + \frac{\partial^2 A_z}{dz^2} = 0 \quad . \tag{367}$$

Challenge 786 e

This is called a *wave equation*, because it admits solutions of the type

$$\mathbf{A}(t,x) = \mathbf{A}_0 \sin(\omega t - \mathbf{kx} + \delta) \tag{368}$$

which are commonly called *(plane) waves*. Such a wave satisfies equation (367) for any value of the *amplitude* A_0, of the *phase* δ, and of the *angular frequency* ω, provided the *wave vector* **k** satisfies the relation

$$\omega(\mathbf{k}) = \frac{1}{\sqrt{\varepsilon_0\mu_0}}k \quad \text{or} \quad \omega(\mathbf{k}) = \frac{1}{\sqrt{\varepsilon_0\mu_0}}\sqrt{\mathbf{k}^2} \qquad (369)$$

The relation $\omega(\mathbf{k})$ between the angular 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 (369) specifically characterizes electromagnetic waves in vacuum, and distinguishes them from all other types of waves.*

Equation (367) 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 cross each other without disturbing each other, and that waves can travel across 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 (368).

After Maxwell had predicted the existence of electromagnetic waves, in the years between 1885 and 1889 Heinrich Hertz** discovered and studied them. He fabricated 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 conversation. In this half second, the signal goes up to the geostationary satellite, down again and back. This gives a speed of $c \approx 4 \cdot 36\,000\,\mathrm{km}/0.5\,\mathrm{s} \approx 3 \cdot 10^5\,\mathrm{km/s}$, which is close to the precise value.

But Maxwell did more. He strengthened earlier predictions that *light* itself is a solution of equation (368) and therefore an electromagnetic wave, albeit with a much higher frequency. Let us see how we can check this.

It is easy to confirm the *wave* properties of light; 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). You can confirm this fact with your own fingers. Simply put your hand one or two centimetres in front of 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

---

* 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 mountain ascent.

** 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, in the explanation of Maxwell’s theory, and in the unfolding of radio communication technology. More about him on page 112.

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 788

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 we look carefully at a rainbow, below the main reed yellow green blue violet bow, we observe weaker, additional green blue and violet bows. Ref. 398 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.[**] See page 402 (More about the rainbow below.) It seems that in those times scientists either did not trust their own fingers, or did not have any.

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. Ref. 399 It turns out that the transmitted intensity depends on the width of the slit. Their surprising result is shown in Figure 164. Can you explain the origin of the unexpected steps in the curve? Challenge 789

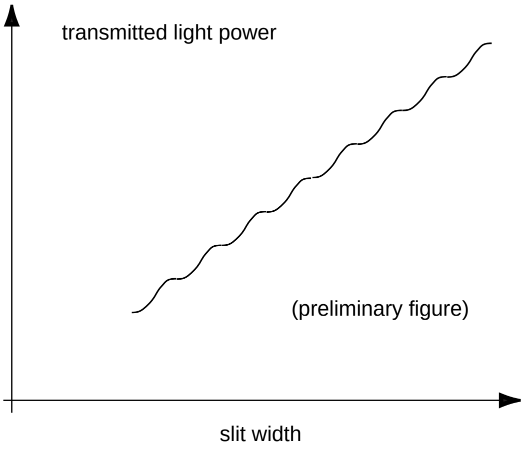


Figure 164 The light power transmitted through a slit as function of its width

Numerous other experiments on the creation, detection and measurement of electromagnetic waves have been performed in the nineteenth and twentieth century. For example, in 1800, William Herschel discovered *infrared light* using a prism and a thermometer. (Can you guess how?) Challenge 790 n A bit later, Johann Wilhelm Ritter, a colourful figure of natural Romanticism, discovered *ultraviolet light*, using silver chloride, AgCl, and gain a prism. The result of all these experiments is that electromagnetic waves can be distinguished first of all by their wavelength or frequency. See page 395 The main categories are listed in Table 38. For visible light, the wavelength lies between 0.4 μm (pure violet) and 0.8 μm (pure red).

---

[*] Where does the energy go in interference patterns? Challenge 787 n

[**] 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 opposing views. Young collaborated with Fraunhofer and Fresnel; at last, his results were made known by Fresnel and Helmholtz.

At the end of the twentieth century the final confirmation has become possible. Using quite sophisticated experiments researchers measured 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 (369) of light has finally been confirmed in all completeness. Ref. 400

We are left with one additional question about light. If light oscillates, in which direction does this happen? The answer is hidden in the parameter $\mathbf{A}_0$ in expression (368). Electromagnetic waves oscillate in directions *perpendicular* to their motion. Therefore, even for identical frequency and phase, waves can still differ: they can have different *polarization* directions. For example, the polarization of radio transmitters determine whether radio antennas of receivers have to be kept horizontal or vertical. Also for light, polarization is easily achieved, e.g. by shining it through stretched plastic films. When the polarization of light was discovered in 1808 by the French physicist Louis Malus, it definitively established its wave nature. Malus discovered it when he looked at the strange double images produced by feldspar, a transparent crystal found in many minerals. Feldspar ($KAlSi_3O_8$) splits light beams into two – it is *birefringent* – and polarizes them differently. That is the reason that feldspar is part of every crystal collection. Calcite ($CaCO_3$) shows the same effect. If you ever see a piece of feldspar or transparent calcite, have a look through it onto some written text.

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 it is hidden behind clouds, and many insects use polarization to distinguish water surfaces from mirages. Can you find out how?On the other hand, both the cornea and the lens of the human eye are birefringent. Ref. 401 Challenge 791 Ref. 402

Note that all possible polarizations of light form a continuous set. However, a general wave can be seen as a superposition of two orthogonal, linearly polarized waves with different amplitudes and different phases. Most books show pictures of plane, linearized electrodynamic waves. Essentially, the electric fields look like water waves generalized to three dimensions, the same for magnetic fields, and the two are perpendicular to each other. Can you confirm this? Challenge 792

Interestingly, a generally polarized plane wave can also be seen as the superposition of right and left *circularly polarized waves*. However, no figures of such waves are found in any textbook. Can you explain why? Challenge 793

So far it is clear that light is a wave. To confirm that light waves are indeed *electromagnetic* is more difficult. The first argument was by Riemann in 1858; he deduced that any electromagnetic wave must propagate with a speed

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \quad . \tag{370}$$

Already ten years before him, in 1848, Kirchoff had noted that the measured values on both sides agreed within measurement errors. A few years later, Maxwell gave a beautiful confirmation of the expression by deducing it from equation (369). You should be able to repeat the feat.Note that the right hand side contains electric and magnetic quantities, and the left hand side is an optical entity. Riemann's expression thus unifies electromagnetism with optics. Challenge 794

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 electromagnetic fields and accelerated electric charges. This implies that the influence of matter on light could be understood from its internal electromagnetic fields, and in particular, that subjecting matter to *external* electromagnetic fields should change the light it emits, the way it interacts with light, or generally, the material properties as a whole.

Challenge 795 e

For example, it is indeed found that electric fields can influence the light transmission of oil, an effect discovered by Kerr. With time, many more influences of matter in fields on light were found, and an extensive list is given in the table on page 423. It turns out that with a few exceptions the effects can *all* be described by the electromagnetic Lagrangian S_{CED} (361), or equivalently, by Maxwell's equations (363). In summary, classical electrodynamics indeed unifies the description of electricity, magnetism and optics; all phenomena in these fields, from the rainbow to radio, from lightning to electric motors, are found to be different aspects of the evolution of the electromagnetic field.

Table 38 The electromagnetic spectrum

Frequency	Wave-length	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 http://www.vlf.it
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	radio transmissions, radio amateurs, spying
15 -150 MHz	20 m-2 m	VHF	allow battery operated transmitters		remote controls, closed networks, tv, radio amateurs, radio navigation, military, police, taxi

Frequency	Wavelength	Name	Main properties	Appearance	Use
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, astronomy
3 -100 THz	1000 - 3 μm	IRC or far infrared		sunlight, living beings	seeing through clothes
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
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	idem; best background for reading black text
	580 nm	standard yellow			
530 -606 THz	566-495 nm	green	maximum eye sensitivity	algae and plants	highest brightness per light energy to the human eye

Frequency	Wave-length	Name	Main properties	Appearance	Use
	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 and 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, silicon lithography
		X-rays	penetrate materials	emitted by stars, plasmas, and 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
$>$ 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 were realized and relativity was developed.

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 determine which one he meant? Challenge 796 n

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. 403

In sugar syrup, light beams curve, as shown in Figure 165. In fact, light bends at any material interface. This effect, called *refraction*, is the same that makes aquaria seem less deep than they actually are. Refraction is a consequence of the change of light speed from material to material. Are you able to explain refraction, and thus explain the syrup effect? Challenge 797 n

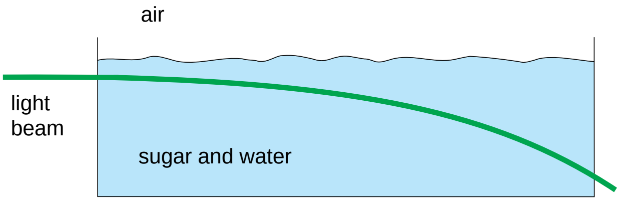


Figure 165 Sugar water bends light

Refraction in water droplets is also the basis of the rainbow, as shown on page 402, 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. 404

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 cosmonauts. 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. 405 Challenge 798 n

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 you can distinguish the two headlights of a car. Can you estimate it? Challenge 799

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 morning or 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 cosmonauts who went to the moon performed careful searches and confirmed that the claim is absurd. The story is one of the most tenacious Ref. 406

Challenge 800 urban legends. (Is it possible to see the wall from the space shuttle?)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 166 Light beams can spiral around each other

Diffraction also means that behind a small disk illuminated along its axis, the centre 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. 407

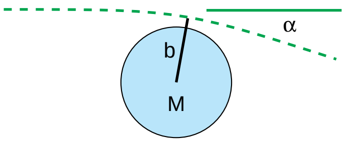


Figure 167 Masses bend light

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 332

In summary, light travels straight only if it travels *far from other matter*. In everyday life, 'far' simply means more than a few millimetres, because all gravitational and electromagnetic effects are negligible at these distances, mainly due to lights' truly supersonic speed.

Can one touch light?

If a little glass bead is put on top of a powerful laser, the bead remains suspended in mid air, as shown in Figure 168. That means that light has momentum. Therefore, contrary to what we said in the beginning of our mountain ascent, 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.

Like each type of electromagnetic field, and like every kind of wave, light carries energy;

Challenge 801 e the energy flow per surface and time is

$$\mathbf{P} = \frac{1}{\mu_o}\mathbf{E}\times\mathbf{B} \quad \text{giving an average} \quad < P > = \frac{1}{2\mu_o} E_{\max} B_{\max} \quad . \qquad (371)$$

---

[*] François Arago (1786–1853), French physicist. Augustin Jean Fresnel (1788–1827), engineer and part time physicist; he published 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 commission responsible for lighthouses, for which he developed the well-known Fresnel lens. He died prematurely, partly of exhaustion due to overwork.

Obviously, light also has momentum. It is related to the energy by

$$p = \frac{E}{c} \quad . \tag{372}$$

As a result, the pressure p exerted by light onto a body is given by Challenge 802

$$p = \frac{P}{c}(1+r) \tag{373}$$

where for black bodies we have $r = 0$ 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 803 n

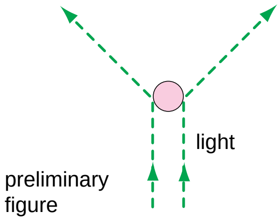


Figure 168 Levitating a small glass bead with a laser

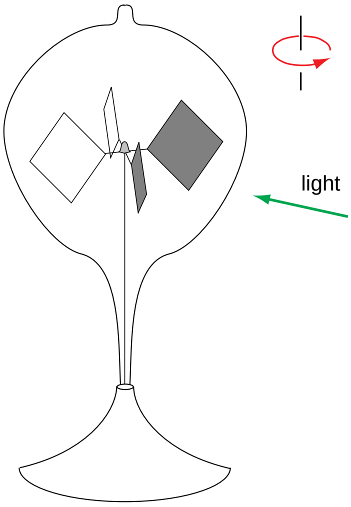


Figure 169 A commercial light mill turns *against* the light

In fact, rather delicate equipment is needed to detect the momentum of light, in other words, its radiation pressure. 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 169. 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, namely with the shiny side towards the light! (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 mountain ascent. Only in 1901, with the advent of much better pumps, it became possible to create a sufficiently good vacuum that allowed to measure the light pressure with such an improved, true radiometer. In fact, it is now possible to build such tiny propellers that the light shining on them makes them turn, like the wind makes a windmill turn.

Ref. 408 Challenge 804 n Ref. 409 Ref. 410 Ref. 411

---

* William Crookes (1832, London–1919, London), English chemist and physicist, president of the Royal Society, discoverer of Thallium.

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 that reason, the tail always points *away* from the sun, a well-known fact that you might want to check at the next opportunity. Challenge 805 e

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, to suspend, and to move small transparent spheres of 1 to 20 μm diameter using laser beams. It is possible to do this through a microscope, so that one can also observe at the same time what is happening. 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 fibres, by chemically attaching their ends to glass or teflon spheres and then pulling them apart with such an optical tweezer. Ref. 412

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. 412

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

$$L = \frac{E_{\text{nergy}}}{\omega} \quad . \tag{374}$$

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 Figure 170. Of course, the whole thing works even better with a laser beam. In the 1960s, a beautiful demonstration was performed with microwaves. A circularly polarized microwave beam from a maser – the microwave equivalent of a laser – 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 mountain ascent. Challenge 806 e Ref. 413 Challenge 807

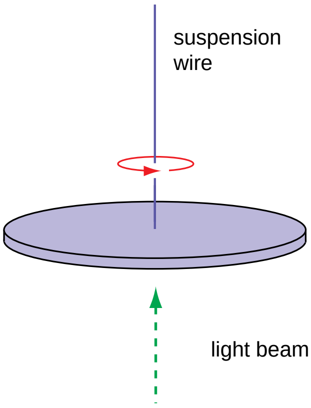


Figure 170 Light can rotate objects

We note that not for all waves angular momentum is energy per angular frequency. This is only the case for waves made of what in quantum theory will be called spin 1 particles. For example, for gravity waves the angular momentum is *twice* this value, and they are therefore expected to be made of spin 2 particles.

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 808 n

War, light, and lies

From the tiny effects of the equation (373) for light pressure we deduce that light is not an efficient tool for hitting objects. On the other hand, light is able to heat up objects, as we 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, and again in 2001, a group of people who read too many 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 809

Other people tried to persuade NASA to study the possibility to propel a rocket using emitted light instead of ejected gas. Are you able to estimate whether this is feasible? Challenge 810

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 811 n 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 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, laser beams and the rainbow. You can check this yourself, using an umbrella or a compact disk: they decompose light mixtures, but not pure colours. Challenge 812 e

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 171 so near to your eye that you cannot focus the stripes any more. The unsharp borders of the white stripes have a pink or a green shade. These colours are due to the imperfections of the lens in the human eye, its 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 813

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.Adding a few drops of milk to the water makes Challenge 814 e the lamp 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 and if the air is 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 famous 'rayon vert' described by Jules Verne in his novel of the same title. It is often Ref. 414

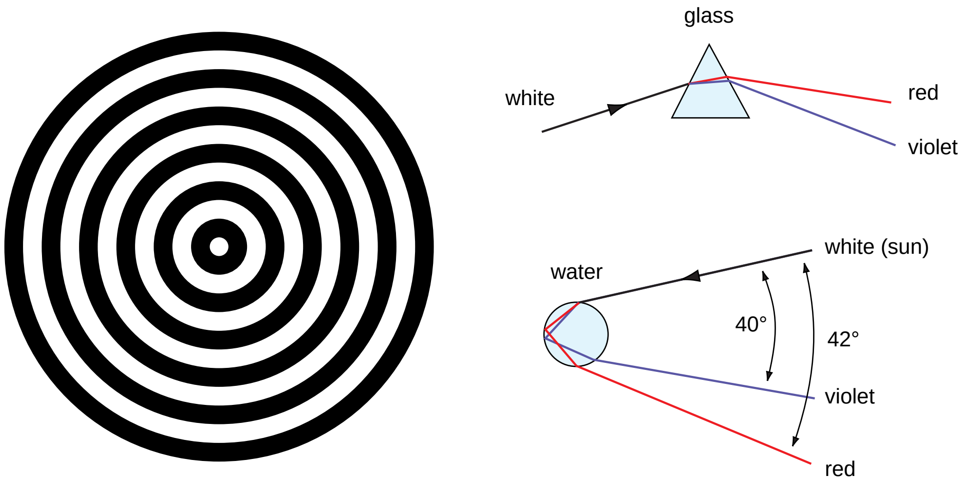


Figure 171 Proving that white light is a mixture of colours

seen on islands, for example in Hawaii.[*]

To clarify the difference between colours in physics and colour in human perception and 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 world, from the sea, the desert or the mountains, in the following order: 1st black and white, 2nd red, 3rd green and yellow, 4th blue, 5th brown; 6th come mauve, pink, orange, grey 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. The result also implies that people use these basic colour classes even if their language does *not* have a word for each of them. These strong statements have been confirmed for over 100 languages. Ref. 415

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. Indeed, 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.

---

[*] 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 60.

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

$$v_{\mathrm{ph}} = \frac{\omega}{k} \quad . \tag{375}$$

For example, the phase velocity determines interference phenomena. Light in vacuum has the same phase velocity $v_{\mathrm{ph}} = c$ for all frequencies. Are you able to imagine an experiment to test this to high precision? Challenge 815 n

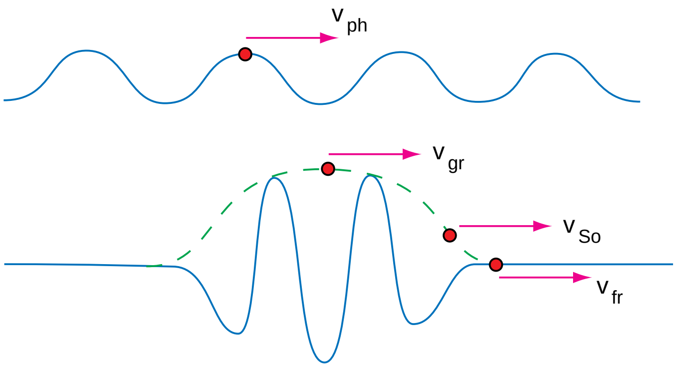


Figure 172 The definition of important velocities in wave phenomena

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 when at the same time the frequency is near to an absorption maximum. In these cases, experiments show that the phase velocity is *not* the signal velocity. For such situations, a better approximation to the signal speed is the *group velocity*, i.e. the velocity at which a group maximum will travel. This velocity is given by Ref. 416

$$v_{\mathrm{gr}} = \left.\frac{d\omega}{dk}\right|_{k_0} \tag{376}$$

where k_0 is the central wavelength of the wave packet. We observe that $\omega = c(k)k = 2\pi v_{\mathrm{ph}}/\lambda$ implies the relation

$$v_{\mathrm{gr}} = \left.\frac{d\omega}{dk}\right|_{k_0} = v_{\mathrm{ph}} + \lambda \frac{dv_{\mathrm{ph}}}{d\lambda} \quad . \tag{377}$$

This means that the sign of the last term determines whether the group velocity is larger or smaller than the phase velocity. For a travelling group, as shown by the dotted line in Figure 172, 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 $v_{\mathrm{gr}} = c$ for all values of the wave vector k.

You should be warned that still many publications propagate the false statement 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 evennegative; this happens when the light pulse is very narrow, i.e. when it includes a wide range of frequencies, or again when the frequency 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. The refractive index then is smaller than 1.[*] However, in all these cases the group velocity is *not* the same as the signal speed.[**] Challenge 816 Ref. 417

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 172. The definition cannot be summarized in a formula, but it does have 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, it is found that for no material Sommerfeld's signal velocity is larger than the speed of light in vacuum. Ref. 416

Sometimes 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 **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

$$\mathbf{v}_{\mathrm{en}} = \frac{\mathrm{Re}(\mathbf{P})}{W} = \frac{2c^2\ \mathbf{E}\times\mathbf{B}}{\mathbf{E}^2 + c^2\mathbf{B}^2} \quad . \tag{378}$$

However, like in the case of the front velocity, also in the case of the energy velocity we have to specify if we mean the energy transported by the main pulse or by the front. In vacuum, neither is ever larger than the speed of light.[****] (In general, the energy velocity in matter has a value slightly different from Sommerfeld's signal velocity.) Ref. 416

In recent years, the progress in light detector technology, allowing to detect even the tiniest energies, has forced people to take the fastest of all these energy velocities to describe signal velocity. Using detectors with the highest possible sensitivity we can use as signal

---

Challenge 817 n [*] Some people (incorrectly) pretend to have found $n < o$ for certain microwaves. Can you imagine what this would mean?

[**] 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 mountain ascent.

[***] 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.'

[****] 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, of course). Ref. 418

Note that the negative group velocity implies energy transport against the propagation velocity of light. This is possible only in *energy loaded* materials. Ref. 419

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 velocity*. It is simply given by Challenge 818

$$v_{\mathrm{fr}} = \lim_{\omega\to\infty} \frac{\omega}{k} \quad . \tag{379}$$

The forerunner velocity is *never* larger than the speed of light in vacuum, even in materials. 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. The forerunner velocity is the true signal velocity or the *true velocity of light*. Using it, all discussions on light speed become clear and unambiguous.

Recently, the issue reappeared in another way. A discussion started in 1960 with the 'prediction' by the soviet physicist Victor Veselago that the index of refraction could have *negative* values. In 2000, an experimental 'confirmation' for microwaves was published. But Ref. 420 in 2002 it was shown that negative refraction indices, which imply speeds larger than unity, are only possible for either phase velocity or even group velocity, but not for the energy or true signal velocity. The problems arise because in some physical systems the refraction angle for phase motion and for energy motion differ. If the term 'index of refraction' is consistently used to characterize the motion of energy, it cannot have negative values.

To finish this section, here are two challenges. Which of all the velocities of light is Challenge 819 n measured in experiments determining the velocity of light, e.g. when light is sent to the moon and reflected back? And now a more difficult one: why is the signal speed of light slower inside matter, as all experiments show? Challenge 820 n

Signals and predictions

When somebody reads a text through the phone to a neighbour who listens to it and maybe repeats it, we speak 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 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 Ref. 421 music – all music is predictable for short time scales – through a 'faster-than-light' system. To distinguish between the two situations, we note 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.

> 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 disclose their full importance only when we ask 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 odd dimensions.

In spaces of even dimensions, it is impossible to 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 –

We 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 requiring that the evolving 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. Checking this for a general number of dimensions, it turns out that the delta function is a solution of the wave equation 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 more than one, odd 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.

- You would have no mirror image, like a vampire.
- Light would not be oscillating, 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. You 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 821 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

---

* He took the question from a book on the sciences by Aaron Bernstein which he read at that time.

‘medium’ in which this happens the *aether*. The properties of the aether found in experiments are listed in Table 39.

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
spatial dimensionality	3
topology	R^3
mass and energy content	not detectable
friction on moving bodies	not detectable
own motion	not detectable

Table 39 Experimental properties of the aether and of flat vacuum

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. Ref. 422 Challenge 822 n

Sometimes it is heard 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 we use the change of curvature as definition for motion of the vacuum, vacuum *can* move, as we will find out in the section on general relativity; but aether still remains indistinguishable from it.*

Later 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: 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 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 39 will require some amendments later on. Ref. 423

15. Levitation and other fun challenges and curiosities

Electromagnetism and light are almost endless topics. Some aspects are too beautiful to be missed.

Ref. 423 * 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 mountain ascent.

- Since light is a wave, something must happen if it is directed to a hole smaller than its wavelength. What happens? Challenge 823

- Electrodynamics shows that light beams always push; they never pull. Can you confirm that 'tractor beams' are impossible in nature? Challenge 824 e

- It is well known that the glowing material in light bulbs is tungsten wire in an inert gas. This was the result of a series of experiments which started by the grandmother of all lamps, namely the cucumber. The older generation knows that a pickled cucumber, when attached to the 230 V of the mains, glows in bright green light. (Be careful; the experiment is dirty and dangerous)

- If you calculate 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 825

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. 425

- 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, even the electrical resistance of 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 conductive than gold atoms. Can you imagine why? Ref. 426 Challenge 826

- 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? Ref. 427 Challenge 827

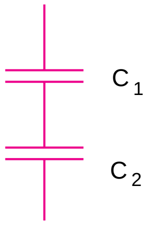


Figure 173 Capacitors in series

- Does it make sense to write Maxwell's equations in vacuum? Both electrical and magnetic fields require charges in order to be measured. But in vacuum there are no charges at all. In fact, only quantum theory solves this apparent contradiction. Are you able to imagine how? Challenge 828

- Grass is usually greener on the other side of the fence. Can you give an explanation based on observations for this statement? Ref. 428 Challenge 829

- Inside a conductor there is no electric field. Thus there is no danger if a lightning hits an aeroplane, as long the plane is made of metal. Aeroplanes are so-called *Faraday cages*. More generally speaking, a field or a charge on the metal surface of a body does not influence fields and charges inside it. Can you give an explanation? Challenge 830 n

The explanation will allow you to answer the following question. Are there Faraday cages for gravity as well? Why?

Cars also are good approximations of Faraday cages. If your car is hit by lightning in dry weather, you should wait a few minutes before leaving it, though. Can you imagine why?

---

* 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*, the proportionality factor between voltage and current, was named after him.

Faraday cages also work the other way round. Electric fields changing inside a Faraday cage are not felt outside. For this reason, radios, mobile phones and computers are surrounded by boxes made of metal or metal-sprayed plastics. The metal keeps the so-called *electromagnetic smog* inside buildings to a minimum.

For purely magnetic fields, the situation is more complex. It is quite difficult to shield the inside of a machine from outside magnetic fields. How would you do it? In practice one often uses layers of so-called *mu-metal*; can you guess what this material does? Challenge 831

- The *electric polarizability* is the property of matter responsible for the deviation of water flowing from a faucet by a charged comb. 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, how precisely combs get charged when rubbed, a phenomenon called *electrification*, is still one of the mysteries of modern science. See page 366
- A pure magnetic field cannot be transformed into a pure electric field by change of observations frame. The best that can be achieved is a state similar to an equal mixture of magnetic and electric fields. Can you provide an argument elucidating this relation? Challenge 832
- 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 would you make use of this effect in this case? Ref. 429 Challenge 833
- A team of camera men in the middle of the Sahara were using battery driven electrical equipment to make sound recordings. Whenever the microphone cable was a few tens of metres long, they also heard a 50 Hz power supply noise, even though the next power supply was thousands of kilometres away. An investigation found that the high voltage lines in Europe lose a considerable amount of power by irradiation; those 50 Hz waves are reflected by the ionosphere around the earth and thus can disturb recording in the middle of the desert. Can you estimate whether this observation implies that living directly near a high voltage line is dangerous? Challenge 834

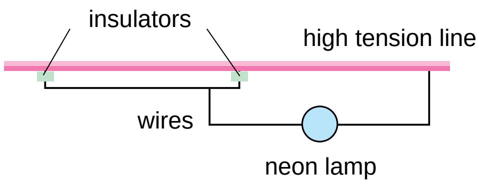


Figure 174 Small neon lamps on a high voltage cable

- On certain high voltage cables leading across the land scape, small neon lamps shine when the current flows. How is that possible? Challenge 835
- When two laser beams cross at a small angle, one can form light pulses which seem to move faster than light. Does this contradict special relativity? Ref. 433 Challenge 836
- 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 kilometres, 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. 430

Challenge 837 ▪ Can you explain to a non-physicist how amicroscope works?* Heisenberg almost missed his PhD exam because he could not.

▪ Is it really possible to see stars from the bottom of a deep pit or of a well during daytime, Challenge 838 as often stated in print?

▪ If the electric field is described as a sum of components of different frequencies, its Ref. 431 so-called Fourier components, the amplitudes are given by

$$\hat{\mathbf{E}}(k,t) = \frac{1}{(2\pi)^3/2} \int \mathbf{E}(x,t) e^{-i\mathbf{kx}} d^3x \qquad (380)$$

and similarly for the magnetic field. It then turns out that a Lorentz invariant quantity N, describing the energy per circular frequency ω, can be defined:

$$N = \frac{1}{8\pi} \int \frac{|\mathbf{E}(k,t)|^2 + |\mathbf{B}(k,t)|^2}{c|\mathbf{k}|} d^3k \qquad (381)$$

Challenge 839 n Can you guess what N is physically? (Hint: think about quantum theory.)

▪ Faraday discovered how to change magnetism into electricity, knowing that electricity could be transformed into magnetism. He also found how to transform electricity into light and into chemistry. He then tried to change gravitation into electricity. But he was not Challenge 840 successful. Why not?

▪ Take an envelope, wet it and close it. After letting it dry for a day or more, open it in the dark. At the place where the two papers are being separated from each other, the envelope Challenge 841 glows with a blue colour. Why?

▪ At high altitudes above the earth, gases are completely ionized; no atom is neutral. One speaks of the *ionosphere*, as space is full of positive ions and free electrons. Even though both charges appear in exactly the same number, a satellite moving through the ionosphere Challenge 842 n acquires a negative charge. Why? How does the charging stop?

▪ A capacitor of capacity C is charged with a voltage U. The stored electrostatic energy is $E = CU^2/2$. The capacitor is then detached from the power supply and branched onto an empty capacitor of the same capacity. After a while, the voltage obviously drops to $U/2$. However, the stored energy now is $C(U/2)^2$, which is half the original value. What Challenge 843 n happened?

▪ Perfectly spherical electromagnetic waves are impossible in nature. Can you show this Challenge 844 n using Maxwell's equation of electromagnetism, or even without them?

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. 434 the small 'hail' ice crystallites rising due to thermal upwinds. Since the collision takes part

---

* If not, read the beautiful text by ELIZABETH M. SLATER & HENRY S. SLATER, *Light and electron microscopy*, Cambridge University Press, 1993.

** From Latin 'cumulus,' meaning heap, and 'nimbus', meaning big cloud. The various types of clouds all have Latin names.

in an electric field, charges are separated in a way similar to the mechanism in the Kelvin generator. Discharge takes place when the electric field becomes too 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 lightning.[*] 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. See page 368

As a note, you might know 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 became known. Airline pilots and passengers 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 mountain ascent. 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 life? And why despite its value it cannot be used to extract large amounts of energy?) 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, lightning do not discharge the ground, they actually charge it up![***] Of course, lightning does discharge the cloud to ground potential difference, but by doing so, it actually sends negative charge down to the earth. Challenge 845 n

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 847

Ref. 435 [*] There is no ball lightning even though there is a Physics Report about them. Ball lightning is 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 846 [***] 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.

Why are sparks and lightning 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, usually air. This excitation is due to the temperature of 30 kK inside the channel of a typical lightning stroke. For everyday sparks, the temperature is 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 are dangerous. Also electromagnetic fields can be. In 1997, with beautiful calm 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 modern gas balloons the gondola is suspended by high quality nylon ropes. To avoid damage by lightning and in order to avoid electrostatic charging problems 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.

Electrical nerves

In 1789 the Italian medical doctor Luigi Galvani (1737–1798) discovered that electrical current makes muscles contract. Subsequent investigations confirmed that nerves make use of electrical signals. The details were clarified only in the 20th century. Nerve signals propagate using the motion of ions in the cell membrane of the nerve. The resulting signal speed is between 0.5 m/s and 120 m/s, depending on the nerve type.

How to prove you're holy

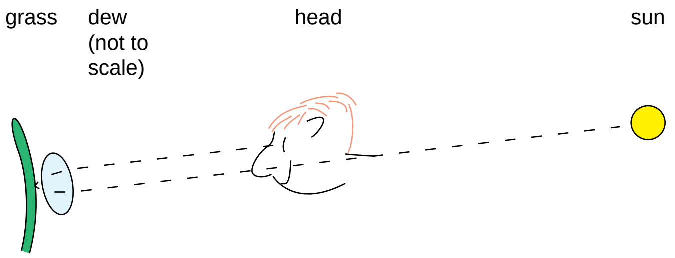


Figure 175 The path of light for dew on grass responsible for the aureole

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 observe it around your own head. It is sufficient to get up early in the 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 you do this in a group, you see the aureole only around *your own* head. Ref. 436

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 your 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 you look at the cat with a light source in your back. By the way, does a cat-eye work like a cat's eye? Ref. 437 Challenge 848

Do we see what exists?

Sometimes we see *less* than there is. Close the left eye, look at the white spot in Figure 176, 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 849

Figure 176 A limitation of the eye

On the other hand, sometimes we see *more* than there is, as the next two figures show.

Our eyes also see 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 centimetres 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

---

* 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 propagate the idea of conservation of energy. His other important book, *Die Lehre von den Tonempfindungen*, published in 1863, describes the basis of acoustics, and like the handbook, is still worth to be read.

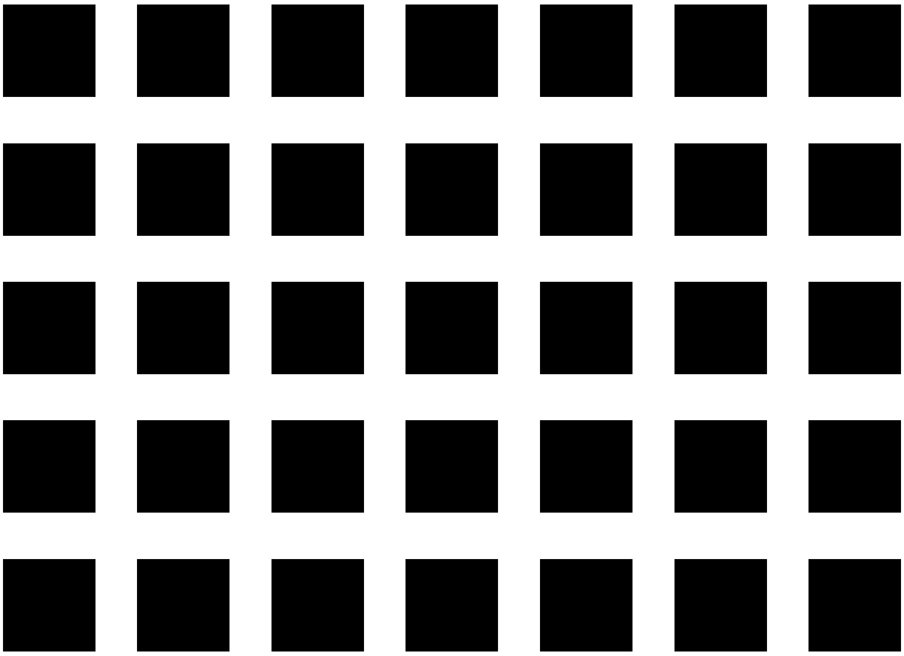


Figure 177 What is the shade of the crossings?

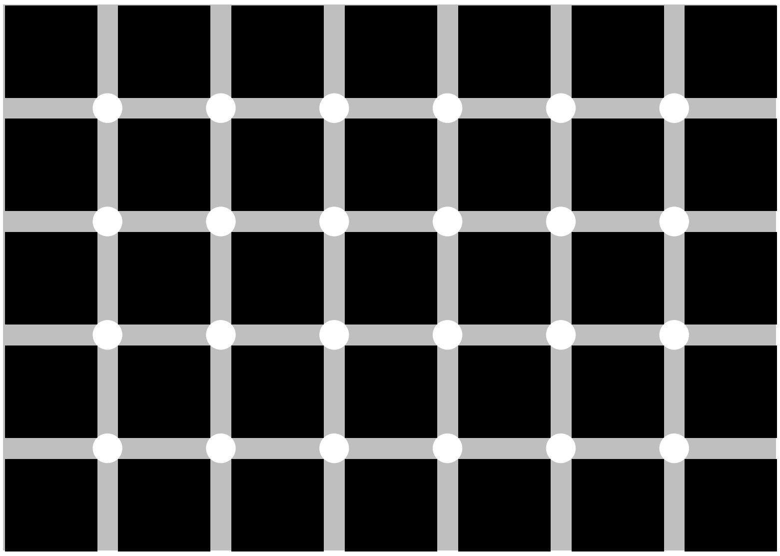


Figure 178 Do you see white, grey, or black dots?

disappear, and vice versa. This shows that the image inside the eye, on the retina, is inverted. Challenge 850 Are you able to complete the proof?

Another reason that we do not see a complete image of nature is that the eye has a limited sensitivity. This sensitivity peaks around 560 nm; outside the red and the violet, the eye does not detect radiation. We thus see only part of nature. Infrared photographs of nature are often so interesting because they show us something which usually remains hidden. Every expert of motion should also know that the sensitivity of the eye does *not* correspond to the brightest part of sunlight. This myth is spread around the world by numerous textbooks copying Ref. 438 from each other. Depending on whether frequency or wavelength or wavelength logarithm is used, the solar spectrum peaks at 500 nm, 880 nm or 720 nm. They eye's spectral sensitivity, like the completely different sensitivity of birds or frogs, is due to the chemicals used for detection; in short, the human eye can only be understood by a careful analysis of its particular evolution history.

Figure to be drawn

Figure 179 Eyes see inverted images

In summary, 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.

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 180, were made by the group of David Williams at the University at Rochester in New York. They used adaptive optics, Ref. 439 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.

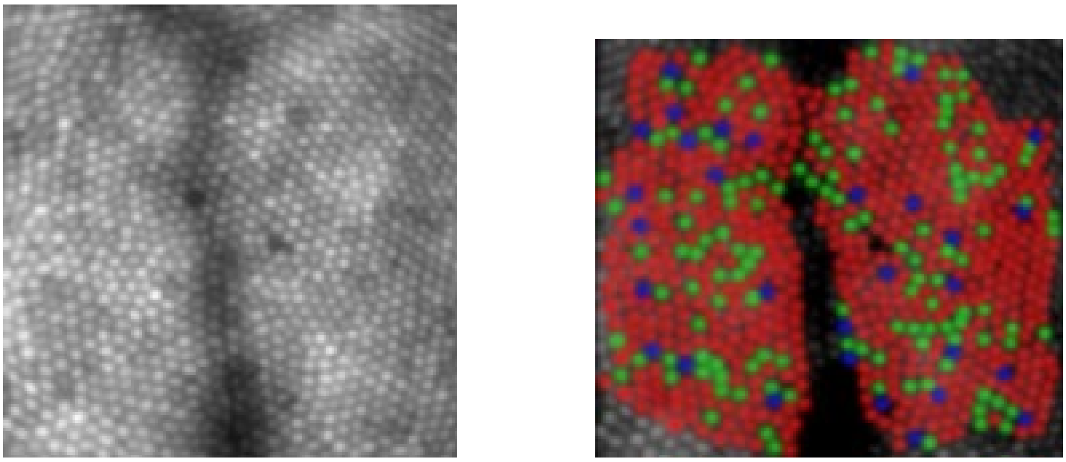


Figure 180 A high quality photograph of a live human retina, including a measured (false colour) indication of the sensitivity of each cone cell

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, See page 402 e.g. yellow, by a pure yellow laser beam, or by the mixture of red and green light.

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!

By the way, Figure 180 is quite astounding. In the human eye, the blood vessels are located in front of the light sensitive cones. Why don't they appear in the picture? And why don't they disturb us in everyday life? Challenge 851 n

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. A number of birds also have a much better eye resolution than humans.

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. 440 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 precise instant this started to be the case.

– CS – to be filled in – CS –

How does one make holograms and other 3-d images?

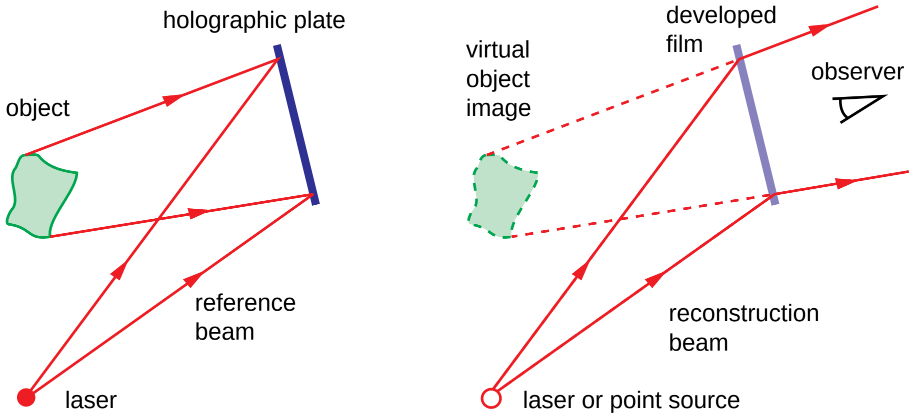


Figure 181 The recording and the observation of a hologram

Our sense of sight gives us the impression of depth mainly due to three effects. First of all, the two eyes see different images. Secondly, the images formed in the eyes are position dependent. Thirdly, our eye needs to focus differently for different distances.

A simple photograph does not capture any of the three effects. A photograph corresponds to the picture taken by one eye, from one particular spot, and at one particular focus.In fact, all photographic cameras are essentially copies of a single and static eye. Challenge 852 e

Any system wanting to produce the perception of depth must include at least one of the three effects just mentioned. In all systems so far, the third and weakest effect, varying focus with distance, is never used, as it is too weak. Stereo photography and virtual reality systems extensively use the first effect by sending two different images to the eyes. Also certain post cards and computer screens are covered by thin cylindrical lenses which allow to send two different images to the two eyes, thus generating the same impression of depth.

But obviously the most spectacular effect is obtained whenever position dependent images can be created. Some virtual reality systems mimic this effect by attaching a sensor to the head, and creating computer-generated images which depend on this position. However, such systems are not able to reproduce actual situations and thus pale when compared to the impression produced by holograms.

Holograms reproduce all what is seen from any point of a region of space. A *hologram* is thus a stored set of position dependent pictures of an object. It is produced by storing amplitude *and phase* of the light emitted by an object. To achieve this, the object is illuminated by a *coherent* light source, such as a laser, and the interference pattern is stored. Illuminating the developed photographic film by a coherent light source then allows to see a full three-dimensional image. In particular, due to the reproduction of the situation, the image floats in free space.

Is it possible to make *moving* holograms? Yes; however, the technical set-ups are still extremely expensive. So far, they only exist in a few laboratories. By the way, can you describe how you would distinguish a moving hologram from a real body, if you ever met one, without touching it? There is no way that holograms of people can walk around and frighten real people. They look too similar to the ghosts from moving pictures. Challenge 853 n

Research questions

The classical description of electrodynamics is coherent and complete; nevertheless there are still many subjects of research. Here are a few of them.

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 starts below the earth's crust. The *crust* is typically 30 to 40 km thick, though with variations at plate faults and near volcanoes. 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 Ref. 442

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.

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. 443

Challenge 854

$$E_{\text{nergy}} = \frac{1}{2}\varepsilon_o \int E^2_{\text{lectric field}}\, dV = \frac{1}{2}\varepsilon_o \int_d^\infty \left(\frac{1}{4\pi\varepsilon_o}\frac{q}{r^2}\right)^2 4\pi r^2 dr = \frac{q^2}{8\pi\varepsilon_o}\frac{1}{d} > 1.2\,\mu\text{J} \qquad (382)$$

On the other hand, the *mass* of an electron, usually given as 511 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 considerable difficulty describing electrons. In fact, a consistent description of charged point particles within classical electrodynamics is not possible. This pretty topic receives only a rare – but then often passionate – interest nowadays, because the puzzle is solved in a different way in the upcoming, second part of our mountain ascent.

Ref. 444

See page **??**

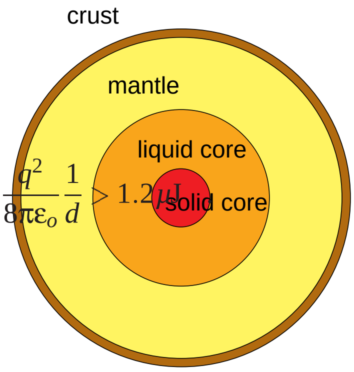


Figure 182 The structure of our planet

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 40. 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 423

Electrodynamics and general relativity interact in many ways. Only a few cases have been studied up to now. They are important for black holes and for empty space. For example, it seems that magnetic fields increase the stiffness of empty space. Many such topics will appear in the future.

Ref. 441

But maybe the biggest challenge imaginable in classical electrodynamics is to decode the currents inside the brain. Is it possible to read our thoughts with an apparatus placed outside the head? One could start with a simple challenge: is it possible to distinguish the thought 'yes' from the though 'no' by measuring electrical or magnetic fields around the head? In other words, is mind-reading possible? Maybe the twenty-first century will give us a positive answer. If so, the team performing the feat will be instantly famous.

Challenge855 h

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. Is it 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 into two groups: those which consume energy and those who do not. Among the methods consuming energy is 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. Levitation of liquids or solids by strong ultrasound waves is presently becoming popular in laboratories. All 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 Shanghai by a German consortium 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. Ref. 445 Ref. 446 Ref. 447

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, we cannot produce a local minimum of potential energy in the middle of a box using electric or gravitational fields. This impossibility is called *Earnshaw'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 92.) The theorem can also be proved by noting that given a potential minimum in free space, Gauss' theorem for a sphere around that minimum requires that a source of the field be present inside, which is in contradiction with the original assumption. Ref. 448

We can deduce 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

$$\frac{U}{V} = -\frac{1}{2}(\varepsilon - \varepsilon_o)E^2 \quad . \tag{383}$$

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.[**] Challenge 856

---

[*] 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 64. And even though gravity is not really due to a field, the result still holds in general.

Ref. 449 [**] 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 the gas. Can you find a liquid-gas combination where bubbles fall instead of rising; can you find one? Challenge 857

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.

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 858

$$\frac{U}{V} = -\frac{1}{2}\left(\frac{1}{\mu} - \frac{1}{\mu_o}\right)B^2 \tag{384}$$

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 859

There are thus two ways to get magnetic levitation: levitating a diamagnet or using a time dependent field. Ref. 450 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. Ref. 447 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. 451

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, liquid helium droplets as large as 2 cm, grasshoppers, fish, and frogs (all alive and without any harm) in this way. They are, like humans, all made of diamagnetic material. Ref. 452 Humans themselves have not yet been levitated, but the feat is being planned and worked on.

Diamagnets levitate if $\nabla B^2 > 2\mu_o \rho g/\chi$, where ρ is the mass density of the object and Challenge 860 $\chi = 1 - \mu/\mu_o$ its magnetic susceptibility.Since χ is typically about 10^{-5} and ρ of order $1000\,\mathrm{kg/m^3}$, field gradients of about $1000\,\mathrm{T^2/m}$ are needed. In other words, levitation requires fields changes of 10 T over 10 cm, nowadays common for high field laboratory magnets.

Finally, *time dependent* electrical or magnetic fields, e.g. periodic fields, can lead to levitation Ref. 445 in many different ways without any consumption of energy. This is one of the methods used in the magnetic bearings of turbomolecular vacuum pumps. Also single charged particles, such as ions and electrons, are now regularly levitated with Paul traps and Penning traps. Ref. 445 The mechanical analogy is shown in Figure 184.

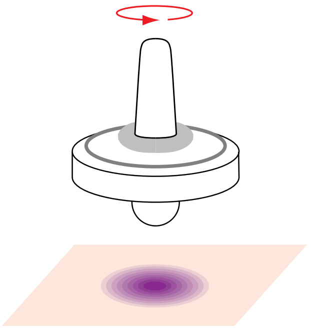


Figure 183 Floating 'magic' nowadays available in toy shops

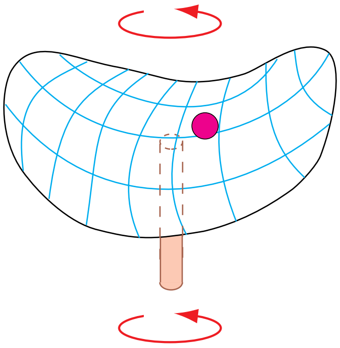


Figure 184 Trapping a metal sphere using a variable speed drill and a plastic saddle

Figure 183 shows a toy allowing to let one personally levitate a spinning top in mid air above a ring magnet, a quite impressive demonstration of levitation for anybody looking at it. It is not hard building such a device oneself. Ref. 453 Ref. 454

Even free electrons can be levitated, letting them float above the surface of fluid helium. 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. 455

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; we could thus say that it is indeed levitated by nuclear interactions. See page 672

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 we want to be precise, we 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 we look at the critical points with a microscope. More about this in the second part of our walk. Ref. 456

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 614

Solidity is due to electricity. 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.* Can you find or imagine a new one?For example, can electric charge change the colour of objects? Challenge 861

Table 40 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, fluid and business physics

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. 457
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
magnetoresistance	iron, metal multilayers	resistance changes with applied magnetic field Ref. 458
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

---

* 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
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. 459
Interactions with magnetic fields		
Hall effect	silicon; used for magnetic field measurements	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 by applied magnetic field
magnetoelastic effect	Fe, Ni	change of magnetization by tension or pressure
acoustomagnetic effect	metal alloys, anti-theft etiquettes	excitation of mechanical oscillations through magnetic field
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; essentially the same as the previous one
photoelectromagnetic effect	InSb	current flow due to light irradiation of semiconductor in a magnetic field
Voigt effect	vapours	birefringence induced by applied magnetic field
Cotton-Mouton effect	liquids	birefringence induced by applied magnetic field

Name of property	example	definition
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
Ettinghausen-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. 460
magnetocaloric effect	gadolinium, GdSiGe alloys	material cools when magnetic field is switched off Ref. 461
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 magnetic 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	$CsNO_3$, tourmaline, 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. 462
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	nematic liquid crystals	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 (which ones?)Challenge 862
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, photoeffect	Cs	current flows into vacuum due to light irradiation
internal photoelectric effect	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	low reflection, low absorption, low scattering
reflectivity	metals	light bounces on surface
polarization	pulled polymer sheets	light transmission depending on polarization angle
optical activity	sugar dissolved in water, quartz	rotation of polarization
birefringence	feldspar,cornea	refraction index depends on polarization direction, light beams are split into two beams
dichroism	feldspar, andalusite	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

Name of property	example	definition
luminescence: general term for the opposite of incandescence	GaAs, television	cold light emission
fluorescence	CaF_2, X ray production, light tubes, cathode ray tubes	light emission during and after light 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, feldspar	light emission during heating, used e.g. for archaeological dating of pottery Ref. 463
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
scattering	gases, liquids	light changes direction
Mie scattering	dust in gases	light changes direction
Raleigh scattering	sky	light changes direction, sky is blue
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; it does not exist; why? Challenge 863	
switchable mirror	LaH	voltage controlled change from reflection to transparency Ref. 464
radiometer effect	bi-coloured windmills	mill turn due to irradiation (see page 400)
luminous pressure	idem	opposite of the preceding one
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

Name of property	example	definition
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	Se, CdS	change of resistivity with light irradiation
optoacoustic affect, photoacoustic effect	gases, solids	creation of sound due to absorption of pulsed light
optogalvanic effect	plasmas	change of discharge current due to light irradiation
optical nonlinear effects: parametric amplification, frequency mixing, saturable absorption, n-th harmonic generation, optical Kerr effect, etc.		
phase conjugated mirror activity	gases	reflection of light with opposite phase
Material properties		
solidity, impenetrability	floors, columns, ropes, buckets	at most one object per place at a given time
Interactions with vacuum		
Casimir effect	metals	attraction of uncharged, conducting bodies

All matter properties in the list can be influenced by electric or magnetic fields or directly depend on them. This shows that the nature of all these material properties is electromagnetic. 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 the motion of charges inside each material must be complex indeed. Most effects are the topic of solid state physics,* fluid and plasma physics.

Solid state physics is by far the most important part of physics, when measured by the impact it had on society. Almost all effects have applications in technical products, and give work to many people. Can you name a product or business application for any randomly choseneffect from the table? Challenge 864 e

In our mountain ascent 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 working of light bulbs thus proves that metals are made of charged particles. *Incandescence*, as it is called, requires charges. Actually,

---

* 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.

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. 465

$$I(T) = fT^4 \frac{2\pi^5 k^4}{15c^2h^3} \quad \text{or} \quad I(T) = f\sigma T^4 \quad \text{with} \quad \sigma = 56.7\,\mathrm{nW/K^4m^2} \tag{385}$$

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 an ideal absorption coefficient and thus appear black. (Can you see why?)The heat radiation they emit is called *black body radiation*. Challenge 865

By the way, which object radiates more energy: a human body or an average piece of the sun of the same mass? Guess first! Ref. 466 Challenge 866 n

Why can we see each other?

Physicists have a strange use of the term ‘black’. Most bodies at temperatures at which they are red hot or even hotter are excellent 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 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 141. Ref. 467

Let us have a quick summary of black body radiation. Black body radiation has two important properties; first, the emitted light power increases with the fourth power of the temperature. With this 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 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 14.0 °C is due to the sun’s irradiation.) Challenge 867 Ref. 468

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:

$$u(\nu,T) = \frac{8\pi h}{c^3} \frac{\nu^3}{e^{h\nu/kT} - 1} \quad . \tag{386}$$

He made this important discovery, which we will discuss in more detail in the second part of our mountain ascent, simply by comparing this curve with experiment.** The new constant h, *quantum of action* or *Planck’s constant*, turns out to have the value $6.6 \cdot 10^{-34}$ Js, and is central to all quantum theory, as we will see. The other constant Planck introduced, the See page 516

---

* Most bodies are not black, because colour is not only determined by emission, but also by absorption of light.

** Max Planck (1858–1947), professor of physics in Berlin, was a central figure in thermostatics. He discovered and named *Boltzmann’s constant k* and the *quantum of action h*, often called Planck’s constant. His introduction

Boltzmann constant k, appears as a prefactor of temperature all over thermodynamics, as it acts as a conversion unit from temperature to energy.

The radiation law gives for the total emitted energy density the expression Challenge 868

$$u(T) = T^4 \frac{8\pi^5 k^4}{15 c^3 h^3} \tag{387}$$

from which equation (385) is deduced using $I = uc/4$. (Why?) Challenge 869

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 is deduced from equation (386) to be Challenge 870

$$\lambda_{\max} = \frac{hc}{4.956\,k}\,\frac{1}{T} = 2.9\,\mathrm{mm\,K}/T \quad \text{but} \quad \hbar\nu_{\max} = 2.82\,kT = 3.9\cdot 10^{-23}\,\mathrm{J/K}\cdot T \quad .(388)$$

Either of these expressions is called *Wien's colour displacement* after its discoverer.* For 37 °C, human body temperature, it gives a peak wavelength of 9.3 μm or 115 THz, which is thus the colour of the bulk of the radiation emitted by every human being. (The peak wavelength does not correspond to the peak frequency. Why?)On the other hand, following Challenge 871 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.

Note that a black body or a star can be blue, white, yellow, orange or red. It is never green. Can you explain why? Challenge 872

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 predictedradiation? Challenge 873

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. 469

An arrangement in which the walls and the objects inside are at the same temperature is called an *oven*. It turns out that it is *impossible* to 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 874

---

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 Adolf 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.

* Wilhelm Wien (1864, Gaffken–1824, München), East-Prussian physicist; he received the Nobel prize for physics in 1911 for the discovery of this relation.

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 because we do *not* live in thermal equilibrium with our environment.

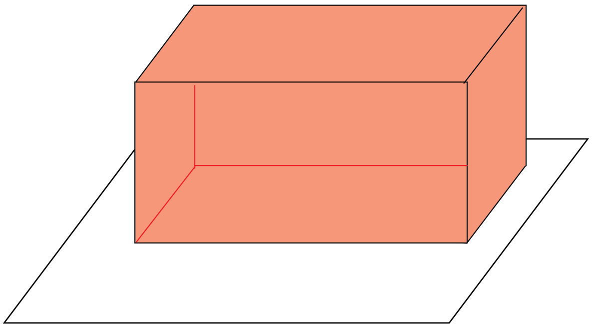


Figure 185 Hot bodies in a hot oven

Could electrodynamics be different?

Any interaction like Coulomb's rule (334) 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, we then recover electrodynamics. Ref. 470

In fact, also the requirements 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 small non-vanishing mass for the photon would change electrodynamics somewhat. 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 of 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;
- The field sources are the (moving) charges and the field evolution is 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;
- The field behaves like a continuous quantity, a distribution of little arrows, and propagates as a wave, as shown e.g. by radios and mobile phones;
- The field can exist and move also in empty space, as shown e.g. by the stars.

---

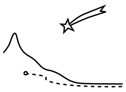
* This can be deduced from the special relativity in various ways, e.g. from the reasoning of page 379, or the formula in the footnote of page 232.

However, there is quite some fun waiting; even though this description is correct in everyday life, during the rest of our mountain ascent we will find that *each* of these ideas is in fact wrong. A simple example shows this.

At a temperature of zero Kelvin, when matter does not radiate thermally, we have the paradoxical situation that the charges inside matter cannot be 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.

In fact, the overview of material properties of Table 40 makes the same point even more strongly; classical electrodynamics can describe many of the effects listed, *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 mountain ascent, that on quantum theory.

See page 423



References

378 Julian Schwinger, L.L. DeRaad, K.A. Milton & W.Y. Tsai, *Classical Electrodynamics*, Perseus, Cambridge, 1998. An excellent text on the topic by one of its greatest masters.

See also the beautiful problem book by André Butoli & Jean-Marc Lévy-Leblond, *La physique en questions - électricité et magnétisme*, Vuibert, Paris, 1999. Cited on page 385.

379 A.S. Goldhaber & W.P. Trower, *Resource letter MM-1: magnetic monopoles*, American Journal of Physics **58**, pp. 429–439, 1990. Cited on page 367.

380 R. Edwards, *Filling station fires spark cars' recall*, New Scientist, pp. 4–5, 4 March 1995. Cited on page 368.

381 S. Desmet, F. Orban & F. Grandjean, *On the Kelvin electrostatic generator*, European Journal of Physics **10**, pp. 118–122, 1989. You can find construction plans for it also in various places on the internet. Cited on page 368.

382 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 373.

383 This effect has first been measured by S.J. Barnett, Philosophical Magazine **12**, p. 349, 1931. Cited on page 373.

384 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 373.

385 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 373.

386 Pierre de Maricourt, *Tractatus de magnete*, 1269. Cited on page 374.

387 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 pages 378 and 385.

388 The http://suhep.phy.syr.edu/courses/modules/MM/Biology/biology2.html web site gives an introduction into brain physiology. Cited on page 378.

389 N. Salingaros, *Invariants of the electromagnetic field and electromagnetic waves*, American Journal of Physics **53**, pp. 361–363, 1985. Cited on page 379.

390 A.C. de la Torre, *$V \leqslant c$ in 1820?*, European Journal of Physics **20**, pp. L23–L24, March 1999. Cited on page 379.

391 Mark D. Semon & John R. Taylor, *Thoughts on the magnetic vector potential*, American Journal of Physics **64**, pp. 1361–1369, 1996. Cited on pages 385 and 386.

392 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 387.

393 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 384. Challenge 875

394 H. van Dam & E.P. Wigner, ... Cited on page 385.

395 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. A good book on asymmetries in nature is H. Brunner, *Rechts oder links*, Wiley-Vch, 1999. Cited on page 389.

396 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 390.

397 For a captivating account on the history of the idea on light, see David Park, *The Fire within the Eye: a Historical Essay on the Nature and Meaning of Light*, Princeton University Press, 1997. Cited on page .

398 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 393.

399 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 pages 393 and 653.

400 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 394.

401 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 394.

402 On the birefringence of the eye, see L. BOUR, *Een eigenaardige speling der natuur*, nederlands tijdschrift voor natuurkunde **67**, pp. 362–364, december 2001. In particular, a photograph of the eye using linear polarized illumination and taken through an analyzer shows a black *cross* inside the pupil. Cited on page 394.

403 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 398.

404 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 398.

405 JAMES E. FALLER & E. JOSEPH WAMPLER, *The lunar laser reflector*, Scientific American pp. 38–49, March 1970. Cited on page 398.

406 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 398.

407 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 399.

408 The value is 4 μPa for black surfaces, and the double for mirrors. Cited on page 400.

409 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 usenet news group `sci.physics`; it is available at the http://www.desy.de/user/projects/Physics/light-mill.html web site. Cited on page 400.

410 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 400.

411 P. GALAJDA & P. ORMOS, Applied Physics Letters **78**, p. 249, 2001. Cited on page 400.

412 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, and in S.P. SMITH, S.R. BHALOTRA, A.L. BRODY, B.L. BROWN, E.K. BOYDA & M. PRENTISS, *Inexpensive optical tweezers for undergraduate laboratories*, American Journal of Physics **67**, pp. 26–35, 1999. Cited on page 401.

413 R. BETH, Physical Review **50**, p. 115, 1936. For modern measurements, see N.B. SIMPSON, K. DHOLAKIA, L. ALLEN & M.J. PADGETT, Optics Letters **22**, p. 52, 1997, and M.E.J. FRIEZE, T.A. NIEMINEN, N.R. HECKENBERG & H. RUBINSZTEIN-DUNLOP, Optics Letters **23**, p. 1, 1998. Cited on page 401.

414 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 402.

415 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 403.

416 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 pages 404 and 405.

417 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 405.

418 Y.P. TERLETSKII, *Paradoxes in the theory of relativity*, Plenum Press, New York, 1968, Cited on page 405.

419 See the excellent explanation by KIRK T. MCDONALD, *Negative group velocity*, American Journal of Physics **69**, pp. 607–614, 2001. Cited on page 405.

420 The explanation with different refraction directions was published by P. VALANJU & al., Physical Review Letters **88**, p. 187401, 2002. Two examples of materials with negative diffraction indices were 'presented' (probably by conviction, not by fraud) by David Smith and his team; even more strangely, the theoretician John Pendry had predicted that such materials would allow to make perfect lenses, without any limit on resultion, and even allowing to use evanescent waves for imaging. Today we can say that these supposed consequences just underline that negative refraction indices are impossible. Cited on page 406.

421 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 406.

422 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 408.

423 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 reference 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 408.

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.

424 A.D. Erlykin & A.W. Wolfendale, *The origin of cosmic rays*, European Journal of Physics **20**, pp. 409–418, 1999, Cited on page 381.

425 D. Singleton, *Electromagnetic angular momentum and quantum mechanics*, American Journal of Physics **66**, pp. 697–701, 1998, Cited on page 409.

426 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 409.

427 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 409.

428 This problem was suggested by Vladimir Surdin. Cited on page 409.

429 Philip Cohen, *Open wide, this won't hurt a bit*, New Scientist p. 5, 3 February 1996. Cited on page 410.

430 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 410.

431 J.E. Avron, E. Berg, D. Goldsmith & A. Gordon, *Is the number of photons a classical invariant?*, European Journal of Physics **20**, pp. 153–159, 1999. Cited on page 411.

432 H. Montgomery, *Unipolar induction: a neglected topic in the teaching of electromagnetism*, European Journal of Physics **20**, pp. 271–280, 1999. Cited on page 381.

433 D. Mugnai, A. Ranfagni & R. Ruggieri, Physical Review Letters **84**, p. 4830, 2000. Cited on page 410.

434 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. 13 151–13 167, 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 411.

435 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 412.

436 This happened to Giovanni Bellini (ca. 1430–1516) the great venetian renaissance painter, who even put this experience into writing, thus producing one of the greatest 'gaffes' ever. Cited on page 414.

437 S.R. Wilk, *How retroreflectors really work*, Optics & Photonics News, pp. 6–7, December 1993. Cited on page 414.

438 The eye sensitivity myth is debunked in detail by B.H. Soffer & D.K. Lynch, *Some paradoxes, errors, and resolutions concerning the spectral optimization of human vision*, American Journal of Physics **67**, pp. 946–953, 1999. Cited on page 416.

439 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. Their last publication is A. Roorda, A. Metha, P. Lennie & D.R. Williams, *Packing arrangement of the three cone classes in the primate retina*, Vision Research **41**, pp. 1291–1306, 2001. Cited on page 416.

440 For a recent summary, see S. Parrott, http://arxiv.org:gr.qc/9711027. See also T.A. Abbott & D.J. Griffiths, *Acceleration without radiation*, American Journal of Physics **53**, pp. 1203–1211, 1985. See alsoA. Kovetz & G.E. Tauber, American Journal of Physics **37**, p. 382, 1969. Cited on page 417.

441 C.G. Tsagas, *Magnetic tension and the geometry of the universe*, Physical Review Letters **86**, pp. 5421–5424, 2001. An overview of the topic is C.G. Tsagas, *Geometrical aspects of cosmic magnetic fields*, http.//www.arxiv.org/gr-qc/0112077. Cited on page 419.

442 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 pages 382 and 418.

443 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 419.

444 A good and short introduction is the paper F. Rohrlich, *The self-force and radiation reaction*, American Journal of Physics **68**, pp. 1109–1112, 2000. Cited on page 419.

445 An excellent review is E.H. Brandt, *Levitation in Physics*, Science **243**, pp. 349–355, 1989. Cited on pages 420 and 421.

446 See the article by ... ... in Physik in unserer Zeit February 2001. Liquid drops up to 1 g have been levitated in this way. Cited on page 420.

447 F.C. Moon & P.Z. Chang, *Superconducting Levitation – Applications to bearings and magnetic transportation*, Wiley & Sons, 1994. Cited on pages 420 and 421.

448 W.T. Scott, *Who was Earnshaw?*, ... pp. 418–419, 1959. Cited on page 420.

449 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. An example are 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 420.

450 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 421.

451 The first photographs of a single *ion* was in W. Neuhauser, M. Hohenstatt, P.E. Toschek & H. Dehmelt, *Localized visible Ba$^+$ mono-ion oscillator*, Physical Review A **22**, pp. 1137–1140, 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, 1996, D. Haubrich, H. Schadwinkel, F. Strauch, B. Ueberholz, R. Wynands & D. Meschede, Europhysics Letters **34**, p. 663, 1996. Cited on page 421.

452 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 were 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 421.

453 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 422.

454 The drill trick and the building of a Levitron are described in the beautiful lecture script by Josef Zweck, *Physik im Alltag*, Skript zur Vorlesung im WS 1999/2000 der Universität Regensburg. Cited on page 422.

455 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 422.

456 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 usenet news group `alt.magic.secrets`, in particular Tilman Hausherr, 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 422.

457 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 423.

458 HENK SWAGTEN & REINDER COEHOORN, *Magnetische tunneljuncties*, nederlands tijdschrift voor natuurkunde **64**, pp. 279–283, november 1998. Cited on page 423.

459 H. OHNO & al., Nature 21-28 December 2000. Cited on page 424.

460 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 425.

461 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 425.

462 A. AJDARI, *Electro-osmosis on inhomogeneously charged surfaces*, Physical Review Letters **75**, pp. 755–758, 1995. Cited on page 426.

463 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 427.

464 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. A good introduction is R. GRIESSEN, *Schaltbare Spiegel aus Metallhydriden*, Physikalische Blätter **53**, pp. 1207–1209, 1997. Cited on page 427.

465 See any book on thermostatics, such as LINDA REICHL, *A Modern Course in Statistical Physics*, Wiley, 2nd edition, 1998. Cited on page 429.

466 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 429.

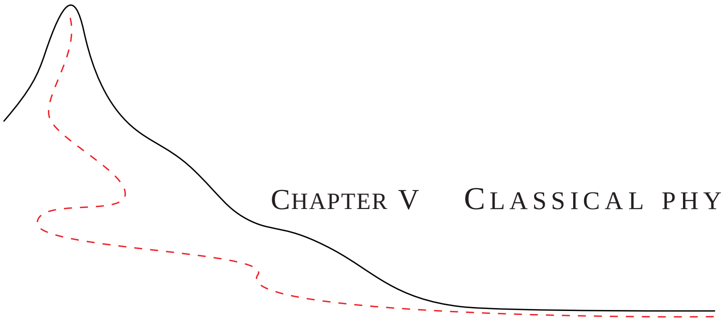
467 See its http://www.cie.co.at/cie web site. Cited on page 429.

468 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 429.

469 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 430.

470 Wolfgang Rindler, *Essential Relativity – Special, General, and Cosmological*, revised second edition, Springer Verlag, 1977, page 247. There is also the beautiful paper by M. Le Bellac & J.-M. Levy-Leblond, *Galilean electrodynamics*, Nuovo Cimento B **14**, p. 217, 1973, which explains the possibilities but also the problems appearing when trying to define the theory non-relativistically. Cited on page 431.

Typeset in
January 2003



CHAPTER V CLASSICAL PHYSICS IN A NUTSHELL

The description of general relativity and classical electrodynamics concludes our walk through classical physics.* In order to see its limits, we summarize what we have found out. In nature, we learned to distinguish and to characterize objects, radiation and space-time. All of the three can move. In all motion we distinguish the fixed, intrinsic properties from the varying state. All motion happens in a way that change is minimized.

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 thus the only localized intrinsic properties of classical, everyday objects. Both mass and electric charge are defined by the accelerations they produce around them. Both quantities are conserved and thus can be added. They are thus described by real numbers. Mass, in contrast to charge, is always positive. Mass describes the interaction of objects with their environment, charge the interaction with radiation.

All *varying* aspects of objects, i.e. their state, can be described using momentum and the position, as well as angular momentum and orientation. They can vary continuously in amount and orientation; observing how these aspects are described by different observers, we find that they are completely characterized by three-dimensional (polar or axial) vectors. The set of all possible states is continuous and is called the phase space. The state of extended objects is given by the states of all its constituent particles. These particles make up all objects and somehow interact electromagnetically.

The state of a particle depends on the observer. The state is useful to calculate the change that occurs in motion. For a given particle, the change is independent of the observer, but the states are not. The states found 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 relations*, and for different gauges they are called *gauge transformations*.

We also observe motion of a massless entity: *radiation*. Everyday types of radiation, such as light, radio waves and their related forms, are travelling electromagnetic waves. They are described by same equations that describe the interaction of charged or magnetic objects.

---

* Others prefer to include in classical physics only *special* relativity; this is a matter of personal preference.

The speed of massless entities is the maximum possible speed in nature and is the same for all observers. The *intrinsic properties* of radiation are its dispersion relation and its energy-angular momentum relation. The *state* of radiation is described by its electromagnetic field strength, its phase, its polarization, and its coupling to matter. The motion of radiation describes the motion of images.

The space-time *environment* is described by space and time coordinates. Space-time is also able to move, by changing its curvature. The intrinsic properties of space-time are the number of dimensions, its signature, and its topology. The state is given by the metric, which describes distances and thus the local warpedness. The warpedness can oscillate and propagate, so that it is fair to say that empty space can move like a wave.

Our environment is finite in age. It has a long history, and on large scales, all matter in the universe moves away from all other matter. The large scale topology of our environment is unclear, as is unclear what happens at its spatial and temporal limits.

Motion follows a simple rule. Change is always minimal. Mass is equivalent to energy. All energy moves in the way space-time tells it, and space moves the way energy tells it. This relation describes the motion of the stars, of thrown stones, of light beams and of the tides. Rest and free fall are the same, and gravity is curved space-time. Mass breaks conformal symmetry and thus distinguishes space from time. We also saw that objects speeds are bound from above by a universal constant c, and that length to mass ratios are bound from below by a universal constant $4G/c^2$.

No two objects can be at the same spot at the same time. This is the first statement about electrodynamic motion humans encounter. More detailed investigation shows that electric charge accelerates other charges, that charge is necessary to define length and time intervals, and that charges are the source of electromagnetic fields. Also light is such a field. Light travels at the maximum possible velocity. In contrast to objects, it can interpenetrate. In summary, we learned that of the two naive types of object motion, namely motion due to gravity – or space-time curvature – and motion due to the electromagnetic field, only the latter is genuine.

Above all, classical physics showed us that motion, be it linear or rotational, be it that of matter, radiation or space-time, is conserved. Motion is similar to a continuous substance: it is never destroyed, never created, but always redistributed. Due to conservation, all motion, that of objects, images and empty space, is predictable and reversible. Due to conservation of motion, time and space can be defined. In addition, we found that classical motion is right-left symmetric. In summary, despite everyday experience, classical phaysics showed us that motion is predictable: there are *no* surprises in nature.

The future of planet earth

Maybe nature shows no surprises, but it still provides many adventures. On the 8th of march 2002, a 100 m sized body almost hit the earth. It passed at a distance of only 450 000 km. On impact, it would have destroyed a region of the size of London. A few months earlier, a 300 m sized body missed the earth by 800 000 km; the record so far was in 1994, when the distance was only 100 000 km.* Several other disasters can be predicted by classical physics, as shown in Table 41. Most are problems facing humanity only in a distant future. Nevertheless, all are research topics.

Table 41 Some examples of disastrous motion of possible future importance

Critical situation	time scale in years from now
▪ end of new physics	ca. 30 (ca. year 2030)
▪ ozone shield reduction	ca. 100
▪ rising ocean levels due to greenhouse warming	ca. 100-1000
▪ explosion of volcano in Greenland, leading to darkening of sky	unknown
▪ 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 gas cloud detaches from the solar systems, changing the size of the heliosphere, and thus auroras and solar magnetic fields	ca. 3000
▪ subsequent reversal of earth's magnetic field, with increased cosmic radiation levels and thus more skin cancers and miscarriages	unknown
▪ atmospheric oxygen depletion due to forest reduction and exaggerated fuel consumption	> 1000
▪ upcoming ice age	ca. 50 000
▪ possible collision with interstellar gas cloud assumed to be crossed by the earth every 60 million years, maybe causing mass extinctions	ca. 50 000
▪ gamma ray burst from within our own galaxy, causing radiation damage to many living beings	between 0 and $5 \cdot 10^6$
▪ asteroid hitting the earth, generating tsunamis, storms, darkening sunlight, etc.	between 0 and $50 \cdot 10^6$
▪ neighbouring star approaching, starting comet shower through destabilization of Oort cloud and thus risk for life on earth	$> 10^6$
▪ 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 runs out of fuel, becomes red giant, engulfs earth	$5.0 \cdot 10^9$
▪ sun stops burning, becomes white dwarf	$5.2 \cdot 10^9$

---

* The web pages around http://cfa-www.harvard.edu/iau/lists/Closest.html provide more information on such events.

Situation	time scale in years from now
▪ 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 centre destabilizes rest of galaxy	unknown
▪ universe recollapses – if ever (see page 205 ff.)	$> 20 \cdot 10^9$
▪ matter decays into radiation – if ever (see Appendix C)	$> 10^{33}$
▪ problems with naked singularities	unknown, controversial
▪ the vacuum becomes unstable	unknown, controversial

Despite the fascination of these predictions, we leave aside these literally tremendous issues and continue in our adventure.

The essence of classical physics: infinity and the lack of surprises

We can summarize classical physics with a simple statement: nature lacks suprises because *classical physics is the description of motion using the concept of infinity*. All concepts used so far, be they for motion, space, time or observables, assume that the infinitely small and the infinitely large exist. Special relativity, despite the speed limit, still allows infinite proper velocity; general relativity, despite its black hole limit, still allows to approach it as much as possible. In the description of electrodynamics and gravitation, both integrals and derivatives are abbreviations of an infinite number of intermediate steps.

In all cases, the classical description of nature introduces infinity and then discovers that there are no surprises. The detailed study of this question we have performed leads to a simple conclusion: infinity implies determinism.[*] Surprises contradict infinity.

However, basing physics on infinity does not completely convince. The atomic structure of matter make us question the existence of the infinitely small, and the horizon of the universe makes us question the existence of the infinitely large.

Why is our mountain ascent not finished yet?

> No problem is so formidable that
> you can't walk away from it.
> George Schultz

We might think that we know nature now, like Albert Michelson and Oliver Lodge – two well-known, mainly experimental physicists working on electrodynamics – did at the end of the 19th century. They claimed that electrodynamics and Galilean physics implied that the major laws of physics were well known. Their statements are often quoted as example

---

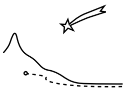
[*] No suprises also imply no miracles. Classical physics is thus in opposition to many religions; many argue that infinity is precisely the necessary ingredient to perform miracles. Classical physics shows that this is not the case.

of flawed predictions, since they reflect an incredible mental closure to the world around them.

At the end of the 19th century, the progress in technology due to the use of electricity, chemistry and vacuum technology allowed to build better and better machines and apparatuses. All were built with classical physics in mind. In the years between 1890 and 1920, these classical machines completely destroyed the foundations of classical physics. The apparatuses showed that electrical charge comes in smallest amounts, that there is no infinitely small change in nature, that nature shows surprises, and that nature behaves randomly. Like the British empire, the reign of classical physics collapsed.

Even without machines, our victorian physicists could have predicted the situation. (In fact, several more progressive minds did so.) They had overlooked a contradiction between electrodynamics and nature for which they had no excuse. In our walk so far we found that clocks and meter bars are necessarily based on matter and electromagnetism. But as we just saw, classical electrodynamics does not explain the stability of matter. Matter is made of small particles, but the relation between these particles, the smallest charges and electricity is not clear. If we do not understand matter, we do not yet understand space and time, since we defined both with measurement devices made of matter. But worse, our victorian friends overlooked a simple fact: the classical description of nature does not allow to understand *life*. The abilities of living beings – growing, feeling, seeing, hearing, moving, thinking, being health or sick, reproducing and dying – are all unexplained by classical physics. In fact, we will discover that all these abilities contradict classical physics. Understanding matter and its interactions, including life itself, is the aim of the second part of our ascent of Motion Mountain. The path will be fascinating and almost purely a sequence of surprises.

Subsequently, we need to rethink electromagnetism, as well as other interactions we will discover, in the presence of space-time curvature. We will alos explore what the new properties of matter and measuring apparatuses imply for space and time. This challenge forms the third and final part of our mountain ascent. There the adventure becomes truly mind boggling. In particular, we will need to resolve the issue we mentioned at the end of Galilean physics: we still are defining space-time with help of objects, and objects with help of space-time. Undoing this circle will be the high point of our ascent. To be well prepared, we first take a break.



INTERMEZZO: THE BRAIN, LANGUAGE AND THE HUMAN CONDITION

> 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

Typeset in January 2003

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 mechanics, general relativity and electrodynamics, we used several concepts without defining them. Examples are 'information', 'memory', 'measurement', 'set', 'number', 'infinity', 'existence', 'universe', or 'explanation'. Each of them is a common and important term. 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. Take an example: can you explain to your parents what a concept is? Challenge 876 e

The reason for studying concept definitions is simple. We need the 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 shore one result: 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.

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 we ask: 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 877 e

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

---

* 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.

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.

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. Ref. 471 A summarizing table on the history of the universe is given in the chapter on general relativity. See page 310 The almost incredible story which has lead to one's own existence starts 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 finally leads to oneself. Ref. 472

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 enrichingexperience. Challenge 878 e

Biological evolution[*] in particular tells us a few important things. Without biological evolution, we would 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 we follows our own memories backwards in time as far as we can, Ref. 473 we reach a certain stage, situated before birth, which forms the starting point of human experience. In that magic moment, we sensed somehow that apart from ourselves, there is something else. The first observation we make about the world is thus the recognition that we can distinguish two parts in it: ourselves and the rest. This distinction is an example – perhaps the first – of a large number of 'laws of nature' we stumble upon in our lifetime. By discovering more and more distinctions we bring structure in the chaos of experience. We quickly find out that the world is made of related parts, such as mama, papa, milk, earth, toys, etc.

---

[*] 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.

Later, when we learn to speak, we becomes fond of more difficult words, and we call the surroundings the *environment*. Depending on the context, we call 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 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, we are 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 mountain ascent.)

Challenge 879 n

In particular, the observation of difference between oneself and the environment goes hand in hand with the recognition that we are not independent of the environment, but that we are firmly tied to it in various inescapable ways; we 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. (Do you agree?)

See page 488

Challenge 880 e

Interactions are not arbitrary; just take touch, smell or sight as examples. They differ in reach, in strength and 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 or 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 ironical quotation marks. 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*.

---

[*] 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 the polar star. 'Nature' is also from Latin, and means 'what is born'. 'Cosmos' is from Greek χόσμος and originally means 'order'.

During childhood we learn 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 we stay 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'.

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 Ref. 475 *relations* between them. Humans call this activity *classification*. Colours, shapes, objects, mother, places, people and ideas are some of the entities humans discover first.

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.[**]

---

[*] A child not able to make this distinction among perceptions – and thus unable to lie – almost surely develops Ref. 474 or already suffers from *autism*, as recent psychological research showed.

[**] 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 our organism with the world.

Some of his opinions on the importance of language in the development are now being revised, notably Ref. 476 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.

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 is a human ready for physics. Ref. 477

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*.

Why a brain?

> Denken is bereits Plastik.[*]
> Joseph Beuys (1920–1986), sculptor. Ref. 478

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. 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' of the brain. Ref. 479 Ref. 480

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, such as 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 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 most 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.[***] Ref. 481

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

---

[*] Thinking is already sculpture.

[**] A good introduction into the study of classifiers is ...

[***] The upright posture in turn allowed humans to take breath independently of their steps, a feat which many animals cannot perform. This is turn allowed humans to develop speech. Speech in turn developed the brain.

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 channelled 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 we arrive 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.

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 modelled from the brain structure. Neurology is still making regular progress. 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. How would you argue that thought is just a physical process?

See page 659

Challenge 881 n

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

We started by saying that studying physics means to talk about motion. To talk means to transmit information. Can information be measured? Can we 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 we start 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. What is the most efficient one you can thinkof? A variation of the combination method is used in computers. For example, the story of this walk required about a thousand million bits. But since the amount of information in a normal letter depends on the set of questions we start with, it is not possible to define a precise measure for information in this way.

Challenge 882

---

* 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.

The only way to measure information precisely is to take the largest possible set of questions that can be asked about a system. In that case, the amount of unknown information is called entropy, a concept we encountered already. With this approach you should able to deduce yourself whether it is really possible to measure the advance of physics. See page 177 Challenge 883

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; however, this is a 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 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 an argument against this endeavour? Challenge 884 n

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. Records can be stored 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 inside the DNA – and especially, no fun, as proven by the sad life of those who loose their memory. Ref. 479

Obviously every record is an object. But in what case 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 we estimate the probability for a record to appear or disappear by chance? Yes, we 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 when the size, i.e. the number of components of the record is increased. In fact, the probability p_{mis} for a misreading or miswriting of a record changes as Challenge 885

$$p_{\mathrm{mis}} \sim 1/N \quad . \tag{389}$$

This relation appears because 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 the number of entities is increased. (Are you able to confirm this?) Challenge 886

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, the 'observation' is usually called a *(generalized) measurement*. Are you able to specify the bath in the case of a person watching a landscape? Challenge 887

From the preceding discussion we can deduce the following surprising statement: since we have such a good 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 observation that we have memory. (Of course, another argument proving that matter is made of smallest parts is the ubiquity of noise itself.) See page 167

A second conclusion was popularized in the late 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 888

$$S_{\text{per erased bit}} = k \ln 2 \quad . \tag{390}$$

As is well known, energy is needed to reduce the entropy of a system. Thus any system that erases memory requires energy. It is known this *dreaming* is connected with the erasing and reorganization of information. Could that be the reason that when we are very tired, without any energy left, we do not dream as much as usual? Challenge 889 n

Entropy is thus necessarily created when we forget. This is evident when we remind ourselves 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 Ref. 482

Ref. 483 * 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 magnetic tape, we have 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 magnetic tapes usually reduces their entropy.

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 is well protected for this reason. In addition, the brain literally grows connections, called *synapses*, between its various *neurons*, which are the cells doing the signal processing. The neuron is the basic processing element of the brain, performing the basic classification. It can do only two things: to fire and not to fire. (It 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 comes via the synapses from hundreds of other neurons. 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. 485, 486

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. Let us estimated this memory capacity. 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, we arrive at a storage capacity for the brain of about

$$M_{\text{rewritable}} \approx 10^{14}\,\text{bit} \approx 10^{4}\,\text{GByte} \quad . \tag{391}$$

(One *byte* is the usual name for eight bits of information.) Note that evolution has managed to put as many neurons in the brain as there are stars in the galaxy, and that if we add all the synapse lengths, we get 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,

---

such a way that fast molecules accumulate on one side, and slow molecules accumulate on the other. This is just another way to rephrase the old result of Leo Szilard, who showed that the measurements by the demon create more entropy than they can save. And every measurement apparatus contains a memory. Ref. 484

Want to play demon? Click on the http://www.wolfenet.com/~zeppelin/maxwell.htm web site.

[*] In practice, the capacity seems almost without limit, since the brain frees memory every time it needs some new space, 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

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; it cannot be of genetic origin, even allowing for smart decompression of stored information.

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 practising it.* Ref. 487

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 and its successors which mimic a psychoanalyst, is still a disappointing experience. To understand the reasons, we ask:

---

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 human life. Assuming that for each of the N cells with n connections there are fn 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 890

$$M_{\text{writeonce}} \approx 10^{16}\,\text{bit} \approx 10^{6}\,\text{GByte} \quad . \tag{392}$$

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 connections, and thus the capacity, is lower than 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 it is possible 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.

* 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.

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 wonderful 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 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, we get the requirement that the statements describing them must be translatable into all languages. But

---

[*] 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 thinker 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.

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. 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 891 e

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 42 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]

---

[*] A comprehensive list with 6 700 languages (and with 39 000 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 15 000 ± 5 000 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. Since English is the language with the largest number of words, learning it well is a greater challenge than for most of the other languages. Ref. 488

[**] Studies explore topics such as the observation 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.

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. 489

See page 477

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 892

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,we are 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.

> Every word was once a poem.
> Ralph Waldo Emerson[*]

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 mountain ascent. Since the answer to the first question is negative, we need to be prepared for a longer mountain ascent, in order

---

[*] Ralph Waldo Emerson (1803–1882), US-American essayist and philosopher.

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.

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 893 n

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 appear to be those which have no parts and no external relations, but only internal ones. Can you think of one? Challenge 894 n Only the last part of this walk will uncover the final word on the topic.

---

[*] 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*.

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 mathematics 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 from the physical one: every concept requires memory from 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

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 labelled 'definition' and others labelled '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.

---

[*] We see 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 464.

[**] 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 shown to be unfounded both by the study of child psychology (see the footnote on page 449) and by physics itself, so that these qualifiers are therefore 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 43 The defining properties of a set

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 of the sets.

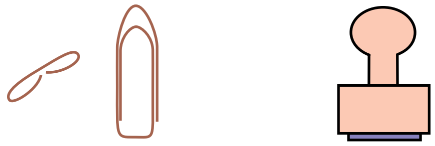


Figure 186 Devices for the definition of sets (left) and of relations (right)

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 many admitting this. In fact, mathematicians have refined the definition of the concept ‘set’ several times, because the naive definition does not work well for infinite sets. A famous example is the story about sets which do not contain themselves. Obviously, 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 895

To avoid problems with the concept of ‘set’, mathematics needs a precise definition. 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 we 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 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 we can construct all concepts used in physics.

Ref. 491

Challenge 897

The naive set definition is far from boring. To make two people happy when dividing a cake, we follow the rule: I cut, you choose. What rule is needed for three people?And for four?

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.

Challenge 898 n

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 easily be defined with help of 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.

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 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'.

---

[*] 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. 490

See page 463

Challenge 896 n

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. To find 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 mountain ascent we will meet physical concepts which are not described by sets nor by classes, containing no set at all. That is were the real fun starts.

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, we should never forget this fact. A few additional 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 a 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. We just need 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. (Can you show this?) However, *any* type of infinite set contains at least one subset which is countably infinite. Challenge 899 n

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 867

The cardinals defined in this way, including $\aleph_n$, $\aleph_\omega$, $\aleph_{\aleph_\aleph}$ are called accessible, because in the mean time, 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. Like the ordinals and the cardinals, they form examples of what are called *transfinite* numbers.

$*$ Therefore, most gods, being concepts and thus sets, are either finite, or, in case they are infinite, they are divisible. Indeed, polytheistic world views are not disturbed by this result.

$**$ 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 real numbers have the cardinality of the power set of the integers, namely $2^{\aleph_0}$. Can you show this? The result leads to 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 usual 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 – you can make the continuum hypothesis come out either right or wrong, as you prefers. Challenge 900 n

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 do continue afterwards: 'The strength of my brother is the power set of that of yours!' Rucker reports that mathematicians think there is no possible nor any conceivable end to these discussions. Ref. 492

For physicists, a simple 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 clarify your own opinion on the issue.It will be settled during the rest of our adventure. Challenge 901 e

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. We write

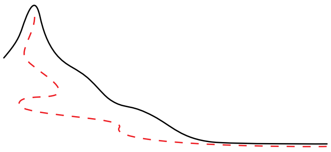
$$f : X \to Y \quad \text{and} \quad f : x \mapsto y \quad \text{or} \quad y = f(x) \tag{393}$$

where the type of arrow shows whether we are 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.*

---

* 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 ...



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

To think and talk more clearly about nature, we need 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.

A *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 $y \in X$:

$$w : X \times X \to X \quad \text{and} \quad w : (x,x) \mapsto y \quad . \tag{394}$$

Is division of numbers an operation in the sense just defined? Challenge 903 n

Now we are ready to define the first of three basic concepts of mathematics. 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 order structure and a topological structure. Ref. 493

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
> last digit is taken away and transferred to the front? Challenge 904 n

Numbers are the oldest mathematical concept and are found in all cultures. The notion of number, in Greek αριθμος, has been changed several times. Each time the aim was 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 calculating with numbers has delayed the progress of science by several centuries. (By the way, the same can be said for the affordable mass reproduction of written texts.)

Note that every category contains a set; since it is unclear whether nature contains sets, as we will discuss on page 493, 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.

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 find another?) is by identifying a natural number with the set of its predecessors. With the relation 'successor of', abbreviated S, this definition can be written as

Challenge 905 n

$$0 := \emptyset \quad , \quad 1 := S\,0 = \{0\} = \{\emptyset\} \quad ,$$
$$2 := S\,1 = \{0,1\} = \{\emptyset,\{\emptyset\}\} \quad \text{and} \quad n+1 := S\,n = \{0,\ldots,n\} \qquad (395)$$

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 44 Some large numbers

Number	examples in nature
Around us	
1	number of angels which can be in one place at the same time, following Thomas Aquinas Ref. 494
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. 495
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
10^{17}	ants in the world

---

* 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. $a+b \in R,\ a+b = b+a,\ a+0 = a,\ a+(-a) = a-a = 0$ 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 all elements a obey $1a = a$,
- 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
ca. 10^{20}	number of snowflakes falling on the earth per year
ca. 10^{24}	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 ($3 \cdot 10^9\,\mathrm{s} \cdot (1/15\,\mathrm{ms}) \cdot 2/3$ (awake) $\cdot 10^6$ (nerves to the brain) Ref. 496
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 50 000 words)
$2^{10} \cdot 3^7 \cdot 8! \cdot 12! = 4.3 \cdot 10^{19}$	possible positions of the 3 × 3 × 3 Rubik's cube Ref. 497
$5.8 \cdot 10^{78}$	possible positions of the 4 × 4 × 4 Rubik-like cube
$5.6 \cdot 10^{117}$	possible positions of the 5 × 5 × 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	
150 000 ± 50 000	hair on a healthy head
900 000	neurons in the brain of a grasshopper
$126 \cdot 10^6$	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

Number	examples in nature
$3\cdot 10^9$	letters (base pairs) in haploid human DNA
$6.1\cdot 10^9$	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 919.) 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. 498

- A number is a sequence of bits. They are usually called ups and downs, and examples are shown in Figure 187.
- 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*, that is all those 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 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, another notation is usually used, deduced from Figure 187. A surreal s is defined as the earliest number of all those between two series of earlier surreals, the left and the right series:

$$\alpha = \{a,b,c,...|A,B,C,...\} \quad \text{with} \quad a,b,c,<\alpha<A,B,C \tag{396}$$

For example, we have

$$\{0|\} = 1 \quad , \quad \{0,1|\} = 2 \quad , \quad \{|0\} = -1 \quad , \quad \{|-1,0\} = -2 \quad , \quad \{0|1\} = 1/2 \quad ,$$
$$\{0|1/2,1/4\} = 1 \quad , \quad \{0,1,3/2,25/16|41/16,13/8,7/4,2\} = 1+37/64$$

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

$$\alpha+\beta = \{...,a+\beta,...,\alpha+b,...|...,A+\beta,...,\alpha+B,...\} \quad . \tag{397}$$

---

* The surreal numbers do *not* form a set since 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.

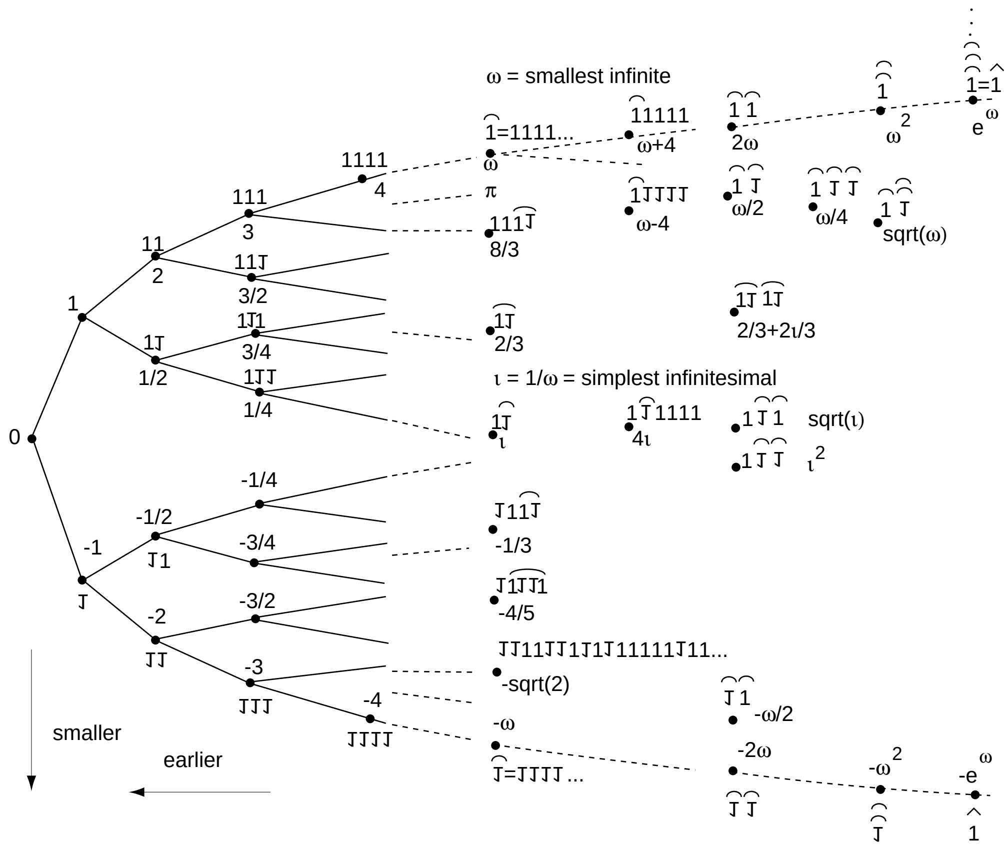


Figure 187 The surreal numbers in conventional and in bit notation

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 906 n
Multiplication is also defined recursively, namely by the expression

$$\alpha\beta = \{\ldots, a\beta + \alpha b - ab, \ldots, A\beta + \alpha B - AB, \ldots | \\ \ldots, a\beta + \alpha B - aB, \ldots, A\beta + \alpha b - Ab, \ldots\} \quad . \qquad (398)$$

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 187. However, the surreal numbers are not commonly used. More common is one of their subsets. Ref. 498

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 $0.999999\overline{9}$ from 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 907 n

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. See Appendix D

Other numbers defined with the help of reals, e.g. the complex numbers C and the quaternions H, are presented in Appendix D. A few more elaborate number systems are also presented there.

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.

Ref. 499

> 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?

Challenge 908 n

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. 500

Ref. 501

All these men forgot that numbers, as well as a large part of mathematics, 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 consequence of mathematics being the science of symbolic necessities.

Perhaps we are being too dismissive. Perhaps the cited 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 adventure.

> 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, such as wishes, ideas or feelings; just try to express the fun of swimming using mathematics. 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 us to talk about nature nor about its basic property: 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 the definition of objects with help of space and time. Physics is thus *not* modelled after mathematics, even if many physicists and mathematicians, including 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.

See page 117

If physics were an axiomatic system, it would not contain contradictions but would cease to be a language and would cease to describe nature. We will return to this issue later on.

See page 774

In short, mathematics is *not* a language, the main reason being that we 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.

---

[*] 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 both in the birth of general relativity and of quantum theory. His textbooks are still in print. His famous motto was: 'Wir müssen wissen, wir werden wissen.' (We have to know, we will know.) His famous Paris lecture is published e.g. in *Die Hilbertschen Probleme*, Akademische Verlagsgesellschaft 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.

Curiosities and fun challenges

Challenge 909 n ▪ What is the largest number which can be written with four digits of 2, and no other sign? And with four 4s?

▪ Pythagorean triplets are integers which obey $a^2 + b^2 = c^2$. Give at least ten examples. Then show the following three properties:at least one number in a triplet is a multiple of 3; at least one number in a triplet is a multiple of 4; at least one number in a triplet is a multiple of 5. Challenge 910 e

▪ The number $1/n$, when written in decimal notation, has a periodic sequence of digits. The period is at most $n-1$ digits long, as for $1/7 = 0.142857\,142857\,1428...$. Which numbers $1/n$ have periods $n-1$? Challenge 911

Physical concepts and patterns of nature

> *Die Grenzen meiner Sprache* bedeuten die Grenzen meiner Welt.[*]
> 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.

---

[*] *The limits of my language* are the limits of my world.

[**] 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*.

General statements about many examples of motion are called *rules* or *patterns*. In the past, it was often said that 'laws govern nature', using an old and inappropriate ideology. A physical 'law' is only a way to say as much as possible with as few words as possible. When saying 'laws govern nature' we actually mean to say 'being lazy, we describe observations with patterns.' Laws are the epitome of laziness. Formulating 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, we need 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', 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 we 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 mean time it even appears in Hollywood movies. We give 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), we know 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 immediately clear. (By the way, there is no known physical concept which can be called 'created' instead

---

* A logical picture of facts is a thought.

of discovered.) In fact, the requirement that people with different standpoints observing the same event from 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 strong 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.

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 mountain ascent 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. Both Occam's razor, the requirement not to introduce unnecessary concepts, and the drive for unification automatically reduce the number and the type of concepts used in physics. In other words, the progress of physical science was and is based on a program that reduces the possible choice of concepts as drastically as possible.

See page 803

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,

---

[*] Anna Wierzbicka concludes that her research clearly indicates that semantic primitives are *discovered*, in particular that they are deduced from the fundamentals of human experience, and not invented.

Ref. 489

[**] Where belief starts, science ends.

> 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 reflects a young child's attitude towards 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	analyse 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

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 fund raisers; 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, there are the people focussed on steps 1 to 3, who get others to work on steps 4 to 6; they are called *geniuses*.

---

[*] Grey, dear friend, is all theory, and green the golden tree of life.

[**] 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 Aristotle' works τά φυσικά meaning 'natural things'. Both expressions are derived from φύσις, meaning 'nature'.

Obviously an important point is hidden in step 6: how do all these people know whether their checks fail? How do they recognize truth?

> All professions are conspiracies against laymen.
> George Bernard Shaw

What is a lie?

> Get your facts straight, and then you can distort them at your 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 we 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 mountain ascent 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 kin. Children start just before their third birthday, by hiding experiences. Adults cheat on taxes. And some intellectuals even claim that truth does not exist. Ref. 474

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. If you lie in court, you better do 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

---

∗ 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*.

– remember Cassandra.* For a good lie we thus needs more than sharedness, or 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 ideas, not about observations. We have to avoid however, to fall in the opposite extreme, namely to make statements which are meaningless; the most destructive comment that can be made about a statement is the one used by the great Austrian physicist Wolfgang Pauli: 'not even wrong'. See page **??**

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 you want to lie well even with empirical statements, you need to pay attention. 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'. 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. 502

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 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 so easily spotted.

There should be no insistence on lies in physics. Unfortunately, classical physics is full of lies. We try to get rid of them during the rest of our walk.

---

[*] 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 you can deduce your own propensity to rebel, on a scale from 0 to 100%. Darwin scores 96% on that scale.

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 we 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, we need 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.

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 to keep up with them. 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 we encounter sentences starting with ‘science says ...’ or ‘science and religion do ...’, replacing ‘science’ by ‘knowledge’ or ‘experience’ is an efficient way

---

* 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. 503 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.

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. But researchers know that if one doesn't stick the neck out, it can't be a lie, nor a observation, nor a theory. (A *theory* is another name for one or several connected, not yet falsified universal statements.)[**] Lying does make 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*, we directly feel the stark contrast between the numerous boring and solid facts he collected and the vulnerable and arrogant 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 we pay 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 allows to gauge the degree that the author sticks his neck out. The more precision is demanded, the more fragile a statement become, and the more likely the fault is found, if there is one. This is the main reason that we chose the increase in precision as guide for our mountain ascent. 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 enough details

---

[*] 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, maybe all people with a passion for physics are spiritual.

[**] In other words, 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.

are collected, 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.
> Democritus

> To teach superstitions as truth is a most terrible thing.
> Hypatia of Alexandria (ca. 355–415)

> [Absolute truth:] It is what scientists say it is
> when they come to the end of their labors.
> Charles Peirce (1839–1914) Ref. 504

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 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.* Similarly undecidable statements can be constructed with terms such as ‘calculable’, ‘provable’ and ‘deducible’. Ref. 505

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’, we quickly see 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. 480 It is often used as an example for the language processing properties of the brain. But 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.

---

* 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.

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.

We 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 913 n

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.[***] Ref. 507 We found 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 that can be achieved. 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.

---

[*] 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 notoriously humour-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, and states that this 'testimony' is true. Ref. 506 But wait; there is a final twist to this story. The statement 'All cretans lie' 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 confirm this? Challenge 912 n The only *genuine* paradox is 'I am lying', to which it is indeed impossible to ascribe a truth value.

[**] 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 sizeable 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 kilometre 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. Ref. 509, 510 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 series Ref. 511 and the physics journals (Appendix E gives a general overview of information sources).

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 experimental 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 and 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 our walk.

---

[*] 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. 508

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)

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, this aspect of the house is classified 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 is a constant, the position a variable, the temperature 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 Mesures, has stopped adding new units.**

∗ 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. Physical units are presented in Appendix B.

As a note, the greatness of a physicist can be ranked 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 wave-function describing electron motion could be counted as half an observable (in fact it 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 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 all the patterns or rules of nature are known? 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.

---

[*] Is it possible to 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.

The discovery of these details will begin in the second part of the walk, the one centred on quantum theory. In the third part of our mountain ascent we will discover that there is a shortest possible averaging time, and that observations of that short duration show so many details that we cannot even distinguish particles from empty space. All our concepts of everyday life appear only after relatively long time averages. The search of an average-free description of nature is one of the big challenges remaining in our adventure.

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 which 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 154. 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. 507

---

[*] Only connexions that are *subject to law* are *thinkable*.

Challenge 914 n induction, we cannot classify observations at all! Evolution has given us memory and a brain because induction works. To criticize induction is not to criticize natural sciences, it is to criticize the use of thought in general. We 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 all, whenever in physics we make 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 we know 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.

> In the progress of physics, the exception
> always turned out to be the general case.

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; such terms contain a logical

---

* Philosophy aims at the logical clarification of thoughts.

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. 512

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 listening to crackpots or to populist politicians; it distinguishes them from more reliable thinkers. Ref. 513 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.

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. If nature were governed by 'laws', they could be changed by parliament. 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?' you recognize 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? – No emergence

> 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.

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, as theologians regularly point out. They repeatedly stressed that such a 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. By definition, this is impossible. 'Emergent' properties can only appear if interactions are approximated, negated, See page 191 or otherwise not seen as important. The idea of 'emergent' properties is a product of minds with a restricted horizon, unable to see or admit the richness of consequences that general Ref. 514 principles can produce. Whenever you meet somebody defending the idea of emergence, you know that he is belittling the importance of interactions, and thus working, in a seemingly innocuous but in fact sneaky way, against the use of reason in the study of nature. 'Emergence' is a belief.

The simple definition of interaction just given, so boring it sounds, leads to surprising conclusions. Take the atomic idea of Democritus 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 915 n

See page 745 We will resolve both paradoxes in the third part of the mountain ascent.

What is existence?

You know what I like most?
Rhetorical questions.

Ref. 515 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, we first of all need 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, we must be able to give a list of relations with the outside world. An object exists if we can interact with it. (Is observation sufficient to determine existence?) Challenge 916 n

For an abstract concept, such as 'time' or 'superstring', the definition of existence has to be refined only marginally: *(physical) existence is the effectiveness to accurately 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*.

For mathematical concepts, existence has a somewhat different meaning: a mathematical concept is said to exist if it has no built-in contradictions. This is a much weaker requirement than physical existence. It is thus incorrect to deduce physical existence from mathematical existence. This error is frequent; from Pythagoras' times onwards it was often stated that since mathematical concepts exist, they must therefore also exist in nature. Historically, this happened when it was stated that planet orbits 'must' be circles, that planet shapes 'must' be spheres or that physical space 'must' be euclidean. Today this is still happening with the statements that space and time 'must' be continuous and that nature 'must' be described by sets. In all these cases, the reasoning is wrong. In fact, the continuous attempts to deduce physical existence from mathematical existence hides that the opposite is correct: a short reflection shows that mathematical existence is a special case of physical existence. Challenge 917 n

We note that there is also a different type of existence, namely *psychological existence*. A concept can be said to exist psychologically if it describes human internal experience. Thus a concept can exist psychologically even if it does not exist physically. It is easy to find examples from the religions or from systems that describe inner experiences. Also myths, legends, and comic strips define concepts that do only exist psychologically, not physically. In our walk, whenever we talk about existence, we mean physical existence only. Challenge 918 n

Do things exist?

Wer Wissenschaft und Kunst besitzt,
Hat auch Religion;
Wer jene beiden nicht besitzt,
Der habe Religion.*
Johann Wolfgang 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 we observe 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 we observe 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 centre 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 tendency, in 1710 he published

---

* He who possesses science and art, also has religion; he who does not possess the two, better have religion.

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 other. (You may be curious to try the feat.) Challenge 919 e

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 from experience on the questions about human existence. Even though he is generally despised, 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. By religion, Berkeley does not mean either morality or spirituality; every scientists is a friend of both. By religion, Berkeley meant that the standard set of beliefs he stood for is above the deductions of reason. This widely cited statement, itself a belief, is still held dearly by many up to this day. However, when searching for the origin of motion, all beliefs 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 a physicist should be prepared to notice 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 Aristotle maintained? In the past, people used to be killed if they gave an 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 408.) 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 Democritus, the most daring thinkers of antiquity. Their speculations in turn elicited the reactionary response of Aristotle, 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, but that then it tends to move back to normal shape. Then, in the late twentieth century, quantum field theory again argued against the existence of a true vacuum and in favour of a space full of virtual particle-antiparticle pairs, culminating in the discussions around the cosmological constant.

Challenge 920 n

See page 724

The title question is settled conclusively only in the third part of this walk, in a rather surprising way.

See page 747

Is nature infinite?

> It is certain and evident to our senses, that in the world some things are in motion. Now whatever is moved is moved by another...If that by which it is moved be itself moved, then this also needs to be to be moved by another, and that by another again. But this cannot go on to infinity, because then there would be no first mover and consequently, no other mover, seeing that subsequent movers move only inasmuch as they are moved by the first mover, as the staff moves only because it is moved by the hand. Therefore it is necessary to arrive at a forst mover, moved by no other; and this everyone understands to be god.
>
> Thomas Aquinas (ca. 1225–1274) *Summa Theologiae*, I, q. 2.

Most of the modern discussions about set theory centre 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 the question from various viewpoints, we will quickly discover that it is equally simple and imprecise.

Do we need infinite quantities to describe nature? In classical and quantum physics we do indeed, e.g. in the case of space-time. Is this necessary? We can say a few things already.

Any set can be finite in one aspect and infinite in another. For example, it is possible to walk a finite distance in an infinite amount of time. It is also possible to travel along any distance whatsoever in a given amount of time, making (almost) infinite speed an option, even in relativity, as was explained there. These connections make discussions on whether humanity is near the 'end of 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 or unreachable. In practice, scientists have thus the power to

See page 214

Ref. 516

---

* Evangelista Torricelli (1608, Faenza–1647), Italian physicist, pupil and successor of Galileo. The (non-SI) pressure unit 'torr' is named after him.

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.

But is it possible at all to say of nature or of one of its aspects that it is *infinite*? Can such Challenge 921 n a statement be compatible with observations? It is evident that every statement claiming that in nature something is infinite is a belief, and not taken from observations. We will have to eliminate this belief step by step.

In short, the universe cannot be said to be infinite. On the other hand, the imprecision of calling it infinite is often small. But can nature be *finite*? At first sight, this would be the only possibility left. (It is not.) But even though many have tried to describe the universe as finite in all its aspects, they were not successful. In order to see the problems of this approach, we continue with the other question mentioned above:

Is the universe a set?

A simple observation leads us to question whether the universe is a set. For 2500 years it has Ref. 517 been 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 electrons. Ref. 518

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, we 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 a particle 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 as the totality of all space-time points and 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 and a set. 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 separable or a set 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

---

* 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.

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, often called 'laws,' expressing that the different aspects of nature form a whole, usually called the universe.

In other words, the possibility to describe observations with help of 'laws' follows from our experience of 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 us to distinguish reference frames, and thus to define all symmetries at the basis of physical descriptions. And in the same way as separability is thus necessary for *covariant* descriptions, the unity of nature is necessary for *invariant* descriptions. In other words, the so-called 'laws' of nature are based on the experience 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? We have three suspicions left: the issue of the number of particles in the universe, the circular definition of space, time and matter, and the issue whether describing nature as made of particles and void is an overdescription, an underdescription, or neither.

In short, are objects countable at all energies? We will discover in the third part of our mountain ascent, that this is not the case in nature. The consequences will be extensive and fascinating. As an example, try to answer the following: if the universe is not a set, what does that mean for space and time? See page 772 Challenge 922 n

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. There are two common definitions of the concept of 'universe'. The first is the totality of all matter, energy and space-time. But this usage 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 restricted view, namely that the universe is only the totality of all matter and energy. But also in this case it is impossible to interact with the universe. Are you able to give a few arguments? Challenge 923 n

In short, we arrive at the conclusion that the universe does not exist. We will indeed confirm this result in more detail later on in our walk. In particular, since the universe does not exist, it does not make sense to even try to answer *why* it exists. The best answer might be: because of furiously sleeping, colourless green ideas. (See page 774; Ref. 480)

What is creation?

> (Gigni) De nihilo nihilum,
> in nihilum nil posse reverti.[*]
> Persius, *Satira*, III, v. 83-84.

Ref. 519

> Anaxagoras, discovering the ancient theory that nothing comes from nothing, decided to abolish the concept of creation and introduced in its place that of discrimination; he did not hesitate to state, in effect, that all things are mixed to the others and that discrimination produces their growth.
> Anonymous fragment, middle ages.

Ref. 520

The term ‘creation’ is often heard when talking about nature. It is used in various contexts with different meanings:

One speaks 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 speaks 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 self-organization; again, they are special cases of motion.

Physicists one often say that a lamp ‘creates’ light or that a stone falling into a pond ‘creates’ water ripples. Similarly, they talk 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 550; See page 620)

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. Quantum cosmology provides more reasons that 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 322; See page 321)

---

[*] Nothing (can appear) from nothing, nothing can disappear into nothing.

In summary, we conclude 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. Creation requires something existing already, in order to take place. In addition, precise exploration shows that no physical process and no example of motion has a beginning. Our walk will show us that nature does not allow us to pinpoint beginnings. This property alone is sufficient to show that 'creation' is not a concept applicable to what happens in nature. Worse, creation is applied only to physical systems; we will discover that nature is not a system and that systems do not exist.

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, but only motion of atoms. Every transformation of matter is motion of atoms. 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. See page 165

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, we indeed experience '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.

Voltaire (1694–1778) popularized an argument against creation often used in the past: we do not know whether creation 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 mountain ascent, 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 itself; 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. It is impossible to escalate Motion Mountain without getting rid of 'creation'. That is not easy. We will encounter the next temptation to bring back creation during the study of quantum theory. See page 703

> 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 infer the creation of an object from its existence 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*.

This chain of thought contains several mistakes. First of all, beauty is not necessarily a consequence of complexity. Usually it is the opposite; indeed, the study of chaos and of self-organization demonstrated how many beautifully complex shapes and patterns can be generated with extremely simple descriptions. True, for most human artefacts, complex descriptions indeed imply complex building processes. A personal computer is a good example. But in nature, this connection does not apply. We have seen above that even 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 self-organization, chaos, turbulence, and fractal shapes. In nature, complex structures derive from *simple* processes. Beware of anybody saying that nature has 'infinite complexity': first of all, complexity is not a measurable entity, despite many attempts; in addition, all known complex system can be described by (relatively) few parameters and simple equations. Ref. 521

The second mistake is to confuse complex description with 'instruction', and maybe even to imagine 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 hidden intelligence and no instruction. 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. The genes in the tree do contain information; but no molecule is given any instructions. What seem to be instructions to us are natural movements of molecules and energy, described by the same patterns as outside living beings. The whole idea of instruction or 'law' of nature is an ideology, born from an analogy with monarchy or even tyranny, and a typical anthropomorphism.

The third important mistake is the suggestion that a complex description for a system implies an underlying design. That is not correct. A complex description only implies that the system has a long 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 (moderate) energy over time. This deduction applies to flowers, mountains, stars, life, people, watches, books, personal computers and works of art; in fact it applies to all objects in the universe. The complexity of our environment thus points out the considerable age of our environment and the shortness of our own life.

In summary, the lack of basic complexity and the lack of instructions confirms a simple result: there is not a single observation in nature which implies or requires design or creation. On the other hand, the variety and intensity of nature's phenomena inspires us with deep awe. The wild beauty of nature shows us how *small* a part of nature we actually are, both in space and in time.* Later, we will explore this experience in detail. We will find that remaining open to nature's phenomena in all their overwhelming intensity is central to the rest of our adventure.

See page 703

> There is a separation between state and church, but not yet between state and science.
> Paul Feyerabend (1924–1994)

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 489

A description is like the 'you are here' sign on a road map in a foreign city. Out of a set of possible positions, a description designates the particular situation to be specified. 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; indeed such a description 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 an observation with another *special* case; a description relates an observation with a *general* case, such as a physical theory.

> Felix qui potuit rerum cognoscere causas,
> atque metus omnis et inexorabile fatum
> subjecit pedibus strepitumque acherontis avari.
> Vergilius **

---

* The search for a 'sense' in life or in nature is a complicated way to try to face the smallness of human existence.

** 'Happy he who can know the causes of things and who, free of all fears, can lay the inexorable fate and the noise of Acheron to his feet.' (Georg. 2, 490 ss.) Publius Vergilius Maro (70–19 BCE), the great roman poet, author of the Aeneis. Acheron was the river crossed by those who just died on their way to the Hades.

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 in 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 the description of its relation with the rest of nature.[**] Ref. 522

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 and not for 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 thirty-six?' can be answered, 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 more fundamental questions about motion.

---

[*] 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, we can equally say that the future is actually a *reason* for the present and the past, a fact often forgotten.

The most general quest for an explanation derives from asking: why is the universe the way it is? The topic is covered in our mountain ascent using the two usual approaches, namely:

Unification and demarcation

Studying the properties of motion, paying incessant attention to increase the accuracy of description, we find 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 the situation were different, we would have a conclusion in contrast with observations' 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, 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 usual 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; others often call it 'anthropocentrism', or simply 'arrogance'.

When we discover that light travels in such a way to take the shortest possible time to its target, or when we describe motion by a principle of least action, or when we discover that trees are branched in such a way that they achieve the largest effect with the smallest effort, we are 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. We 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*.

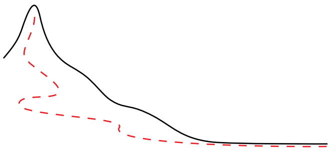
However, neither demarcation nor unification can explain the universe. Can you see why? Challenge 926 n In fact, apart from unification and demarcation, there is a third possibility which merges

---

Challenge 924 n * 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.



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

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 927 n

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 we follow the consequences of different rules of nature until we find them in contradiction with the most striking observation: our own human existence. 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 appearance 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 numbers 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 most hydrogen, helium or lithium atoms, are formed in stars through fusion. While studying the mechanisms of fusion in 1953, the well-known 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. And indeed, the excited state was found a few months later by Willy Fowler.[****] Ref. 523

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. In the popular literature however, Ref. 524 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

---

[*] The most important instrument of a scientist is the waste paper 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– 2001), important British astronomer and astrophysicist. He was the first and maybe only physicist who ever made a concrete prediction – namely the existence of an exited state of carbon nucleus – from the simple fact that humans exist. A permanent maverick, he coined the term ‘big bang’ even though he did not accept the evidence for it, and proposed another model, the ‘steady state’. His most important and well-known research was on the formation of atoms inside stars. He also propagated the belief that life was brought to earth from extraterrestrial microbes.

[****] William A. Fowler (1911–1995) shared the Nobel prize in physics for this and related discoveries.

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 we distinguish between the serious and the perverted form? We get exactly the same rules and patterns of nature if we would use as starting point the existence of pigs or of monkeys. In other words, if we would get *different* conclusions by using the *porcine principle* or the *simian principle*, we are using the perverted form of the anthropic principle, otherwise we are using using the serious form. (The carbon-12 story is thus an example of the serious form.) This test is 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.[*]

Er wunderte sich, daß den Katzen genau an den Stellen Löcher
in den Pelz geschnitten wären, wo sie Augen hätten.
Georg Christoph Lichtenberg[**]

Does one need cause and effect in explanations?

In nature there are neither rewards nor punishments
– there are consequences.
Ivan Illich (1926, Vienna –2002)

The world owes you nothing.
It was there first.
Mark Twain (1835–1910)

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, Ref. 525 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; this avoids pseudo-explanations such as the famous example by Molière when the doctor explains to his patient in elaborate terms that sleeping pills work because they contain a dormitive virtue.

---

[*] 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.

[**] ‘He was amazed that cats had holes cut into their fur precisely in those places where they had eyes.’ Georg Christoph Lichtenberg (1742–1799), German physicist and intellectual, professor in Göttingen, still famous today for his extremely numerous and witty aphorisms and satires. Among others, already in his time, Lichtenberg was making fun of all those who maintained that the universe was made exactly to the measure of man, a frequently encountered idea in the foggy world of the anthropic principle.

But in physics, the two concepts of cause and effect are not used at all. That miracles do not appear is expressed every time we use symmetries and conservation theorems. The observation 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. 526 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 in particular one's inner mechanisms and motivations. Consciousness is mainly the ability of introspection. For this reason, consciousness is *not* a prerequisite for studying motion. Indeed, animals, plants or machines are also able to observe motion. For the same reason, consciousness is not necessary to observe quantum mechanical motion. 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.

---

* Change pleases.

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. 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, Ref. 527 it does so voluntarily: rats get *addicted* to the feelings connected 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 is found 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 guiding Ref. 528 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. Epicurus (Epikuros) (341–271 BCE) maintained that this feeling, φα υμάζειν, 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 these feelings, 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 Ref. 529 seeking.)

Perhaps the most beautiful moments in the study of physics are those appearing after new observations have shaken our previously held thinking habits, have forced to give up a previously held conviction, and have engendered the feeling of being lost. When, in this moment of crisis, we finally discover the more adequate, more precise description of the observations providing a better insight into the world, we are 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

---

[*] 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.

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 well-known Austrian-born physicist Victor Weisskopf (1908–2002) liked 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 Epicurus, 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, Epicurus 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.

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–)

---

[*] 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 moment: saturation, incubation, illumination, and verification. Ref. 530

[**] The unveiled secret takes avenge.

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 we need more. Like any enterprise, also curiosity requires courage, and complete curiosity, as aimed for in our quest, requires complete courage. In fact, it is easy to get discouraged from this trip. It is often dismissed by others as useless, uninteresting, childish, confusing, damaging or most often, evil. Indeed, 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. Occasionally, this is still so, as the newspapers show. Ref. 536

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 statements with observations. As mentioned above, this implies living with superstitions or beliefs. Through the refusal inherent in them, superstitions and beliefs produce fear. Fear is the basis of all unjust authorities. As a consequence, curiosity and science are 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 has 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 curiosity and injustice has hindered or even suppressed completely the development of physics and the other sciences, with extremely negative economic, social and cultural consequences.

When embarking on this ascent, we need to be conscious of what we are doing. In fact, external obstacles can be avoided or at least largely reduced by keeping the project secret. Other difficulties still remain, this time of personal nature. Many tried to embark in this adventure with some hidden or explicit intention, usually of ideological nature, and then got tangled up by it before reaching the end. Some were not prepared to the humility required for such an endeavour. Others were not prepared for the openness required, which can shatter

---

* '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.

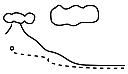
deeply held beliefs. Still others were not ready to continually turn towards the unclear, the dark and the unknown, confronting it at every occasion.

On the other hand, the dangers are worth it. By taking curiosity as a maxim, facing disinformation and fear with all courage, one achieves freedom from all beliefs. In exchange, one gets to savour the fullest pleasures and the deepest satisfaction that life has to offer.

After this look to the basics, we continue our hike. The trail towards the top of Motion Mountain leads us towards the next 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

471 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 447.

472 A simple description is MALCOLM ROSS MACDONALD, *The Origin of Johnny*, Jonathan Cape, 1976. Cited on page 447.

473 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 447.

474 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 pages 449 and 477.

475 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 456). 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 interactions, 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 449.

476 For a view of the mental abilities different from that of Piaget (described on page 449), 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 RENÉ VAN DER VEER & JAAN VALSINER, *Understanding Vigotsky: 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 449.

477 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 450.

478 Quoted in V. HARLEN, R. RAPPMANN & P. SCHATA, *Soziale Plastik – Materialien zu Joseph Beuys*, Achberger Verlag, 1984, on p. 61. Cited on page 450.

479 The problems appearing 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 anthropologist 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 http://ari.ss.uci.edu/cogsci/personnel/hoffman.html web site. Cited on pages 450 and 452.

480 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 pages 450, 481 and 495.

481 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 450.

482 More about the connection between entropy and computers can be found in the classic paper by R. LANDAUER, *Irreversibility and heat generation in the computing process*, IBM Journal of Research and Development **5**, pp. 183–191, 1961, and in C.H. BENNETT & R. LANDAUER, *The fundamental physical limits of computation*, Scientific American **253**, pp. 48–56, 1985. Cited on page 453.

483 W.H. ZUREK, *Thermodynamic cost of computation, algorithmic complexity and the information metric*, Nature **341**, pp. 119–124, 14 August 1989. Cited on page 453.

484 L. SZILARD, *Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen*, Zeitschrift für Physik **53**, p. 840, 1929. This classic paper can also be found in English translation in the collected works by Leo Szilard. Cited on page 454.

485 J.J. Hopfield/ Nature **376**, pp. 33–36, 1995. This paper by one of the fathers of the field presents one possibility by which the *timing* of nerve signals, instead of the usually assumed firing frequency, could also carry information. Cited on page 454.

486 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 to simulate a neuron's electrical behaviour using a silicon circuit. Cited on page 454.

487 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, and **78**, pp. 533–542, 1991. Cited on page 455.

488 However, the language with the largest available dictionary is Dutch, with the 40 volumes of the *Wordenboek der Nederlandsche Taal*, which appeared between 1864 and 1998. It has almost 400 000 entries. Cited on page 457.

489 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 pages 458 and 475.

490 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 462.

491 See the beautiful article by IAN STEWART, *Fair shares for all*, New Scientist pp. 42–46, 17 June 1995. Cited on page 462.

492 RUDY RUCKER, *Infinity and the mind – the science and philosophy of the infinite*, Bantam, Toronto, 1983. Cited on page 464.

493 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 466.

494 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 PETER THE LOMBARD, *Liber Sententiarum*. Cited on page 467.

See page 573

495 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 467.

496 H. DAVSON, *The Eye*, Academic Press, 1962. Cited on page 468.

497 See the http://akbar.marlboro.edu/~mahoney/cube/NxN.txt web site. Cited on page 468.

498 An introduction to the surreal numbers 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 available on the http://www-cs-faculty.stanford.edu/~knuth/sn.html web site. The usually quoted references on the topic 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 pages 469 and 470.

499 This beautiful 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. In four dimensions, the answer is known to lie somewhere between 50 000 and 100 000, whereas the five-dimensional answer is conjectured to be 'never'. Cited on page 471.

500 A. Pais, *Niels Bohr's times: in physics, philosophy, and polity*, Oxford University Press, 1991, page 176. Cited on page 471.

501 Eugene Wigner, *Symmetries and reflections*, Indiana University Press, 1962. Cited on page 471.

502 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 478.

503 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 *Grand dictionnaire universel du XIXe siècle*, by Pierre Larousse, also contains a recipe; it has been again shown to the public in the 1980s by Henri Broch. A wonderful and classic text is Harry Houdini, *Miracle Mongers and their Methods*, Prometheus Books, Buffalo, 1981. The original, written in 1920, by the world famous magician called 'The Great Houdini', is also available on the http://etext.lib.virginia.edu/toc/modeng/public/HouMirM.html web site. The milk drinking Indian statues were common around the world in 1994 and 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 479.

504 John Horgan, *The End of Science – Facing the Limits of Knowledge in the Twilight of the Scientific Age*, Broadway Books, 1997, pp. 31–32, and chapter 2, note 2. Cited on page 481.

505 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 481.

506 See the book by J. Barwise & J. Etchemendy, *The liar*, Oxford University Press, New York, 1987. Cited on page 482.

507 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 pages 482 and 486.

508 Cited by Max Dresden, *The Klopsteg memorial lecture*, American Journal of Physics **66**, pp. 468–482, 1998. Cited on page 483.

509 Well-known books are e.g. Friedrich Kohlrausch, *Praktische Physik*, Teubner, 24. Auflage, 1996. Cited on page 483.

510 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 483.

511 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 483.

512 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 488.

513 See the nice collection of cranks on the http://www.crank.net web site. Cited on page 488.

514 As an example, the opposite view on the emergence of properties is strongly defended by ... in Physics Today 1998. Cited on page 489.

515 See page 133 of the bibliography by John Bowlby, *Charles Darwin*, Pimlico, 1991. Cited on page 489.

516 John Horgan, *The end of science – Facing the limits of knowledge in the twilight of scientific age*, Addison Wesley, 1996. Cited on page 492.

517 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 Democritus and the other presocratics. Cited on page 493.

518 The famous quote is found at the beginning of chapter XI, 'The Physical Universe', in Arthur Eddington, *The Philosophy of Physical Science*, Cambridge, 1939. Cited on page 493.

519 Giuseppe Fumagalli, *Chi l'ha detto?*, Hoepli, Milano, 1983. Cited on page 495.

520 See Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 653, 1991. Cited on page 495.

521 For a beautiful text on fractals, see the footnote on page 47. Cited on page 497.

522 As has been pointed out by René Descartes. Cited on page 499.

523 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 502.

524 An extensive overview of the topic is given in the thick book 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 502.

525 Voltaire, *Candide ou l'optimisme*, 1759. See the footnote on page 136. 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 503.

526 The number of books on consciousness is large and the contents not always interesting, such as Karl R. Popper & John Eccles, *The self and its brain - an argument for interactionism*, Rutledge, 1993. Cited on page 504.

527 See e.g. the *Encyclopedia Britannica*, Macropaedia, in the entry on animal behaviour. Cited on page 505.

528 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 505.

529 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 505.

530 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 506.

531 For an exploration of ‘inner’ motions, see the beautiful text by RICHARD SCHWARTZ, *Internal Family Systems Therapy*, The Guilford Press, 1995. Cited on page .

532 A clear overview of philosophy of science, often called *epistemology*, without unnecessary detail, is given by ROBERT BLANCHÉ, *L’Epistémologie*, Presses Universitaires de France, 1972. Cited on page .

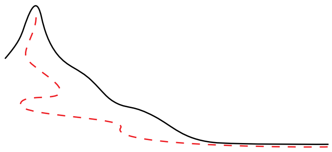
533 An introduction to neurology ... Cited on page .

534 An introduction to computer science ... Cited on page .

535 ALFRED TARSKI, *Introduction to modern logic*, Dover, 19▷▷. See also the famous children book by the mathematician and phototgrapher LEWIS CARROLL, *Alice in wonderland*. Cited on page .

536 For example, one nees the courage to face envy. About this topic see the classic text by HELMUT SCHOECK, *Der Neid*, 1966,, published in English as *Envy: A Theory of Social Behavior*, 1969. It is the standard work in the field. Cited on page 507.

Second Part

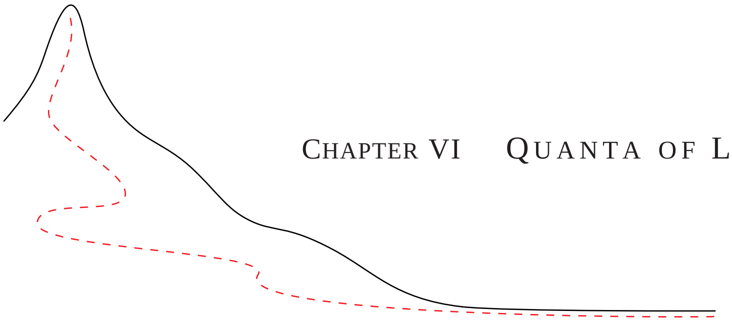


QUANTUM THEORY

WHAT IS MATTER?
WHAT ARE INTERACTIONS?

Where the existence of a minimal change is deduced,
implying that motion is fuzzy,
that matter is not permanent,
that boxes are never tight,
that matter is composed of elementary units
and that light and interactions are streams of particles,
thus explaining why antimatter exists,
why the floor does not fall but keeps on carrying us,
why particles are unlike condoms,
why empty space pulls mirrors together
and why the stars shine.

Typeset in
January 2003



Chapter VI Quanta of Light and Matter

16. An appetizer – quantum theory for poets and lawyers

> Natura [in operationibus suis] non facit saltus.[*]
> 15th century

Ref. 537

Escalating Motion Mountain up to this point, we have completed three main legs. We first encountered Galileo's mechanics, the description of motion for kids, then Einstein's relativity, the description of motion for science fiction enthusiasts, and finally Maxwell's electrodynamics, the description of motion valuable to craftsmen and businessmen.

Figure of orchidea to be inserted

Figure 188 An example of a quantum system

These three classical descriptions of motion are impressive, beautiful, and useful. However, they also have a small problem: they are wrong. The reason is simple: none of them describes *life*. When we observe a flower, we enjoy its bright colours, its wild smell, its soft and delicate shape or the fine details of its symmetry. None of the three classical descriptions can explain any of these properties nor the working of our senses. Classical physics can partly describe them, but it cannot explain their origins. For an explanation, we need *quantum theory*. In fact we will discover that *every type of pleasure* in life is an example of quantum motion. Just try; take any example of a pleasant situation, such as a beautiful evening sky, a waterfall, a caress, or a happy child. Classical physics is not able to explain it. Challenge 928

In the beginning of physics this limitation was not seen as a shortcoming: in those times neither senses nor material properties were imagined to be related to motion. And of course, in older times the study of pleasure was not deemed a serious topic of investigation for a respectable researcher. However, in the meantime we learned that the senses of touch, smell See page 185

---

[*] Nature [in its workings] makes no jumps.

and sight are first of all *detectors of motion*. Without motion, no senses! In addition, all detectors are built of matter. In the chapter on electromagnetism we started to understand that all properties of matter are due to motion of charged constituents. Density, stiffness, colour and all other material properties result from the electromagnetic behaviour of the Lego bricks of matter, namely the molecules, the atoms, and the electrons. Thus, also matter properties are consequences of motion. In addition, we saw that these tiny constituents are *not* correctly described by classical electrodynamics. We even found that light itself behaves unclassically. Therefore the inability of classical physics to describe matter and the senses is indeed due to its intrinsic limitations. See page 429

In fact, every failure of classical physics can be traced back to a single, fundamental discovery made in 1899 by Max Planck:* Ref. 538

▷ *In nature, actions smaller than the value* $\hbar/2 = 0.53 \cdot 10^{-34}\,\mathrm{Js}$ *are not observed.*

All experiments trying to do so invariably fail. In other words, in nature there is always some action – like in a good movie. This existence of a minimal action, the *quantum principle*, is in *full* contrast with classical physics. (Why?)However, it has passed the largest imaginable number of experimental confirmations, many of which we will encounter in this second part of our mountain ascent. Planck discovered the principle when studying the properties of incandescent light, i.e. the light emanating from hot bodies. But the quantum principle also applies to motion of matter, and even, as we will see later, to motion of space-time. By the way, the factor $1/2$ results from the historical accidents in the definition of the constant $\hbar$, which is read as 'eitch-bar'. Despite the missing factor, the constant $\hbar$ is called the *quantum of action* or also, after its discoverer, *(reduced) Planck's constant*. Challenge 929 See page 429

The quantum principle states that no experiment whatsoever can measure an action value smaller than $\hbar/2$. For a long time, even Einstein tried to devise experiments to overcome the limit. But he failed: nature does not allow it.

Interestingly, since action in physics, like action in the movie industry, is a way to measure the *change* occurring in a system, a minimum action implies that *there is a minimum change in nature.* The quantum of action thus would be better named the *quantum of change*. Comparing two observations, there always is change. Before we cite all the experiments confirming this statement, we give an introduction to some of its more surprising consequences.

Since action measures change, a minimum observable action means that two subsequent observations of the same system always differ by at least $\hbar/2$. In every system, there is always *something* happening. As a consequence, in nature *there is no rest*. Everything moves, all the time, at least a little bit. Natura facit saltus. True, it is only a tiny bit, as the value of $\hbar/2$ is so small. For example, the quantum of action implies that in a mountain, a system at rest if there is any, all atoms and all electrons are continuously buzzing around. Rest can be observed only macroscopically, and only as a long time or many particle average. See page 515

---

[*] This somewhat unconventional, but useful didactic approach is almost never found in the literature and used in a teaching text for the first time here. Ref. 539

About Max Planck and his accomplishments, see the footnote on page 429. In fact, the cited quantum principle is a simplification; the constant originally introduced by Planck was the (unreduced) constant $h = 2\pi\hbar$. The factors 2π and 1/2 leading to the final quantum principle were found somewhat later, by other researchers.

Since there is a minimum action for all observers, and since there is no rest, in nature *there is no perfectly straight and no perfectly uniform motion*. Forget all you have learned so far. Every object moves in straight and uniform motion only approximately, and only when observed over long distances or long times. We will see later that the more massive the object is, the better the approximation is. Can you confirm this?As a consequence, macroscopic observers can still talk about space-time symmetries. *Special* relativity can thus be reconciled with quantum theory. Challenge 930

Obviously, also free fall, i.e. motion along geodesics, exists only as a long time average. In this sense, *general* relativity, being based on the existence of freely falling observers, cannot be correct when actions of the order of $\hbar$ are involved. Indeed, the reconciliation of the quantum principle with *general* relativity and thus with curved space is a big challenge. The issues are so mind-shattering that the topic forms a separate, third part of this mountain ascent.

Have you ever wondered why leaves are green? Probably you know that they are green because they absorb blue light, of small wavelength, and red light, of large wavelength, and let green, medium wavelength light undisturbed. How can a system filter out the small and the large, and let the middle go through? To do so, leaves must somehow *measure* the wavelength. But we have seen that classical physics does not allow to measure length or time intervals, as any measurement requires a measurement unit, and classical physics does not allow to define units for them. On the other hand, it takes only a few lines to confirm that with help of the quantum of action $\hbar$ (and the Boltzmann constant k, which Planck discovered at the same time), fundamental measurement units of *all* measurable quantities can be defined, including length and thus wavelength. Can you find a combination of c, G and $\hbar$ giving a length?It only will take a few minutes. When Planck found the combination, he was happy like a child; he knew straight away that he had made a fundamental discovery, even though in 1899 quantum theory did not exist yet. He even told his seven year old son Erwin abut it, while walking with him through the forests around Berlin. Planck knew that he had found the key to understand most of the effects which were unexplained so far. In particular, without the quantum of action, colours would not exist. Every colour is a quantum effect.* Challenge 931

Planck realized that the quantum of action allows to understand the *size* of all things. With the quantum of action, it was finally possible to answer the question on the maximum size of mountains, of trees, and of humans. Planck knew that the quantum of action confirmed the answer Galileo had deduced already long before him: size are due to fundamental, minimal scales in nature. The way the quantum of action allows to understand the sizes of physical systems will be uncovered step by step in the following pages. See page 166

The size of objects is related to the size of atoms; however, the *size of atoms* is a *direct consequence* of the quantum of action. Can you deduce an approximation for the size of atoms, knowing that it is given by the motion of electrons of mass m_e and charge e, constrained by the quantum of action?This connection, a simple formula, was discovered in Challenge 933

---

* It is also possible to define all units using c, G, and e, the electron charge. Why is this notsatisfactory? Challenge 932

1910 by Arthur Erich Haas, 15 years before quantum theory was formulated; at the time, everybody made fun of him. Nowadays, the expression is found in all textbooks.*

Thus the quantum of action has the important consequence that Gulliver's travels are impossible. There are no tiny people and no giant ones. Classically, nothing speaks against the idea; but the quantum of action does. Can you provide the detailed argument? Challenge 934 n

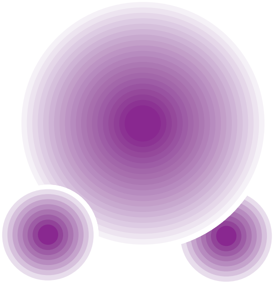


Figure 189 An (average) water particle

But if rest does not exist, how can *shapes* exist? Any shape, also that of a flower, is the result of body parts remaining *at rest* with respect to each other. Now, all shapes result from the interactions of matter constituents, as shown most clearly in the shape of molecules. But how can a molecule, such as the water molecule H_2O, have a shape? In fact, it does not have a *fixed* shape, but its shape fluctuates, as expected from the quantum of action. Despite the fluctuations it does have an *average* shape, because different angles and distances correspond to different energies. And again, these average length and angle values only result because the quantum of action leads to fundamental length scales in nature. Without the quantum of action, there would be *no* shapes in nature.

As we will discover shortly, quantum effects surround us from all sides. However, since the minimum action is so small, its effects *on motion* appear mostly, but not exclusively, in *microscopic* systems. The study of such systems has been called *quantum mechanics* by Max Born, one of the main figures of the field.** Later on, the term *quantum theory* became more popular. In any case, quantum physics is the description of microscopic motion. But when is quantum theory necessary? Table 45 shows that all processes on atomic and molecular scale, including biological and chemical ones, involve action values near the quantum of action. So do processes of light emission and absorption. All these phenomena can be described *only* with quantum theory.

The term 'quantum' theory, by the way, does not mean that all measurement values are *multiples* of a smallest one; this is correct only in certain cases. Quantum theory means the existence of *minimum* measurable values, precisely in the way that Galileo already speculated about in the 17th century. As mentioned in detail earlier on, it was Galileo's insistence on these 'piccolissimi quanti' that got him condemned to lifelong imprisonment, and not, See page 165

---

[*] Before the discovery of $\hbar$, the only simple length scale for the electron was the combination $e^2/mc^2 \approx 3\,\mathrm{fm}$; this is ten thousand times smaller than an atom.

[**] Max Born (1882, Breslau–1970) first studied mathematics, then turned to physics. Professor in Göttingen, he made the city one of the world centres of physics. He developed quantum mechanics with his assistants Werner Heisenberg and Pascual Jordan, then applied it to scattering, to solid state physics, to optics, and to liquids. He was the physicist who first understood that the state function describes a probability amplitude. He is one of the authors of the famous Born & Wolf textbook on optics; it still remains the main book of the field. He attracted to Göttingen the most brilliant talents of the time, receiving as visitors Hund, Pauli, Nordheim, Oppenheimer, Goeppert-Mayer, Condon, Pauling, Fock, Frenkel, Tamm, Dirac, Mott, Klein, Heitler, London, von Neumann, Teller, Wigner, and dozens of others. Being Jewish, Max Born lost his job in 1933; he emigrated and became professor in Edinburgh, where he stayed for twenty years. Physics at Göttingen university never recovered from this loss. For his elucidation of the meaning of the wave function he received the 1954 Nobel prize in physics. Ref. 540

System & change	typical action	motion type
Light		
Smallest amount of light absorbed by a coloured surface	$1\hbar$	quantum
Smallest hit when light reflects from mirror	$2\hbar$	quantum
Smallest visible amount of light	ca. $5\hbar$	quantum
Smallest amount of light absorbed in flower petal	ca. $1\hbar$	quantum
Blackening of photographic film	ca. $3\hbar$	quantum
Photographic flash	ca. $10^{17}\hbar$	classical
Electricity		
Electron ejected from atom	ca. $1-2\hbar$	quantum
Electron added to molecule	ca. $1-2\hbar$	quantum
Electron extracted from metal	ca. $1-2\hbar$	quantum
Signal transport in nerves, from one molecule to the next	ca. $5\hbar$	quantum
Current flow in lighting bolt	ca. $10^{38}\hbar$	classical
Material science		
Tearing apart two neighbouring iron atoms	ca. $1-2\hbar$	quantum
Breaking a steel bar	ca. $10^{35}\hbar$	classical
Basic process in superconductivity	$1\hbar$	quantum
Basic process in transistors	$1\hbar$	quantum
Basic process in magnetic effects	$1\hbar$	quantum
Chemistry		
Atom collisions in liquids at room temperature	ca. $1\hbar$	quantum
Shape oscillation of water molecule	ca. $1-5\hbar$	quantum
Shape change of molecule, e.g. in chemical reaction	ca. $1-5\hbar$	quantum
Single chemical reaction curling a hair	ca. $2-6\hbar$	quantum
Tearing apart two mozzarella molecules	ca. $300\hbar$	quantum
Smelling one molecule	ca. $10\hbar$	quantum
Burning fuel in a cylinder in an average car engine explosion	ca. $10^{37}\hbar$	classical
Life		
Air molecule hitting ear drum	ca. $2\hbar$	quantum
Smallest sound signal detectable by the ear	challenge	classical
DNA duplication step in cell division	ca. $100\hbar$	quantum
Ovule fecundation	ca. $10^{14}\hbar$	classical
Smallest step in molecular motor	ca. $5\hbar$	quantum
Sperm motion by one cell length	ca. $10^{15}\hbar$	classical
Cell division	ca. $10^{19}\hbar$	classical
Fruit fly's wing beat	ca. $10^{24}\hbar$	classical
Person walking one body length	ca. $2\cdot10^{36}\hbar$	classical
Nuclei and stars		
Nuclear fusion reaction in star	ca. $1-5\hbar$	quantum
Particle collision in accelerator	ca. $1\hbar$	quantum
Explosion of gamma ray burster	ca. $10^{80}\hbar$	classical

Table 45 Some small systems in motion and the observed action values for the changes they undergo

as is usually told, his ideas on the motion of the earth. Of course, we will discover that only the idea of a smallest change leads to a precise and accurate description of nature.

Table 45 also shows that the term ‘microscopic’ has a different meaning for a physicist and for a biologist. For a biologist, a system is microscopic if it requires a *microscope* for its observation. For a physicist however, a system is microscopic if its characteristic action is of the order of the quantum of action. In short, for a physicist, a system is microscopic if it is *not* visible in a (light) microscope. To increase the confusion, some quantum physicists nowadays call their own class of microscopic systems ‘mesoscopic,’ whereas many classical, macroscopic systems are now called ‘nanoscopic’. Both names mainly help to attract funding.

There is another way to characterize the difference between a microscopic or quantum system on one side and a macroscopic or classical system on the other. A minimum action implies that the difference of action S between two successive observations of the same system, spaced by a time Δt, is limited. Therefore the system follows

$$S(t+\Delta t)-S(t)=(E+\Delta E)(t+\Delta t)-Et=E\Delta t+t\Delta E+\Delta E\Delta t\geqslant\frac{\hbar}{2} \quad . \tag{399}$$

Since the value of the energy E and of the time t – but not that of ΔE or of Δt – can be set to zero if one chooses a suitable observer, we follow that the existence of a quantum of action implies that in any system the evolution is constrained by

$$\Delta E\Delta t\geqslant\frac{\hbar}{2} \quad . \tag{400}$$

By a similar reasoningwe find that for any system the position and momentum values are constrained by Challenge 935

$$\Delta x\Delta p\geqslant\frac{\hbar}{2} \quad . \tag{401}$$

These two famous relations were called *indeterminacy relations* by their discoverer, Werner Heisenberg.[*] The name was translated incorrectly into English, where they often are called ‘uncertainty relations’. However, this latter name is wrong: the quantities are not uncertain, but *undetermined*. Due to the quantum of action, system observables have *no* definite value. There is *no* way to ascribe a precise value to momentum, position and other observables of a quantum system. Any system whose indeterminacies are of the order of $\hbar$ is a quantum system; if the uncertainty product is much larger, the system is classical, and classical physics

---

[*] One often hears the myth that the indeterminacy relation for energy and time has another weight than the one for momentum and position. That is wrong; it is a myth propagated by the older generation of physicists. This myth survived through many textbooks for over 70 years; just forget it, as it is incorrect. It is essential to remember that all four quantities appearing in the inequalities are quantities describing the *internal* properties of the system. In particular, it means that t is some time variable deduced from changes observed *inside* the system and *not* the external time coordinate measured by an outside clock, in the same way that the position x is *not* the external space coordinate, but the position characterizing the system. Ref. 541

Werner Heisenberg (1901–1976) was an important German theoretical physicist and an excellent table tennis and tennis player. In 1925, as a young man, he developed, with some help by Max Born and Pascual Jordan, the first version of quantum theory; from it he deduced the indeterminacy relations. For these achievements he received the Nobel prize for physics in 1932. He also worked on nuclear physics and on turbulence. During the second world war, he worked in the German nuclear fission program. After the war, he published several successful books on philosophical questions in physics and he unsuccessfully tried, with some half-hearted help by Wolfgang Pauli, to find a unified description of nature based on quantum theory, the ‘world formula’.

is sufficient for its description. In other words, even though classical physics assumes that there are *no* measurement uncertainties in nature, a system is classical only if its uncertainties are *large* compared to the minimum possible ones. As a result, quantum theory is also necessary in all those cases in which one tries to measure some quantity as precisely as possible.

The indeterminacy relations again show that *motion cannot be observed to infinite precision.* In other words, the microscopic world is *fuzzy*. This strange result has many important and many curious consequences. For example, if motion cannot be observed with infinite precision, the very concept of motion needs to be used with great care, as it cannot be applied in certain situations. In a sense, the rest of our quest is an exploration of the implications of this result. In fact, as long as space-time is *flat*, it turns out that we *can* keep motion as a concept describing observations, provided we remain aware of the limitations of the quantum principle.

In particular, the quantum of action implies short-time deviations from energy, momentum, and angular momentum conservation in microscopic systems. Now, in the first part of our mountain ascent we realized that any type of nonconservation implies the existence of surprises in nature. Well, here are some more.

See page 114

Since uniform motion does not exist in the precise meaning of the term, a system moving in one dimension only, such as the hand of a clock, always has a possibility to move a bit in the opposite direction, thus leading to incorrect readings. Indeed, quantum theory predicts that *clocks have limits*, and that perfect clocks do not exist. In fact, quantum theory implies that strictly speaking, one-dimensional motion does not exist.

Challenge 936

Obviously, the limitations apply also to meter bars.Thus the quantum of action is responsible on one hand for the possibility to perform measurements at all, and on the other hand for their limitations.

In addition, it follows from the quantum of action that any observer must be *large* to be inertial or freely falling, as only large systems approximate inertial motion. *An observer cannot be microscopic*. If humans were not macroscopic, they could neither observe nor study motion.

Due to the finite accuracy with which microscopic motion can be observed, faster than light motion should be possible in the microscopic domain. Quantum theory thus predicts *tachyons*, at least over short time intervals. For the same reason, also *motion backwards in time* should be possible over microscopic times and distances. In short, a quantum of action implies the existence of microscopic time travel.

But there is more: the quantum of action implies that *there is no permanence in nature.* Imagine a moving car suddenly disappearing for ever. In such a situation neither momentum nor energy would be conserved. The action change for such a disappearance is large compared to $\hbar$, so that its observation would contradict even classical physics, as you might want to check.However, the quantum of action allows that a *microscopic* particle, such as an electron, disappears for a *short* time, provided it reappears afterwards.

Challenge 937

The quantum of action also implies that *the vacuum is not empty*. If one looks at empty space twice in a row, the two observations being spaced by a tiny time interval, some energy will be observed the second time. If the time interval is short enough, due to the quantum of action, matter particles will be observed. Indeed, particles can appear anywhere from nowhere, and disappear just afterwards, as the action limit requires it. In other words, classical physics' idea of an *empty* vacuum is correct only when observed over *long* time scales. In summary, nature shows short time appearance and disappearance of matter.

The quantum of action also implies that compass needles cannot work. If one looks twice in a row at a compass needle or even at a house, we usually observe that they stay oriented in the same direction. But since physical action has the same unit as angular momentum,a minimum value for action also means a minimum value for angular momentum. Therefore, every macroscopic object has a minimum value for its rotation. In other words, quantum theory predicts that *in everyday life, everything rotates*. Lack of rotation exists only approximately, when observations are spaced by long time intervals. Challenge 939

For microscopic systems, the situation is more involved. If their rotation angle *can* be observed, such as for molecules, they behave like macroscopic objects: their position and their orientation are fuzzy. But for those systems whose rotation angle *cannot* be observed, the quantum of action turns out to have somewhat different consequences. Their angular momentum is limited to values which are multiples of $\hbar/2$. As a result, all microscopic bound systems, such as molecules, atoms, or nuclei, contain rotational motion and rotating components.

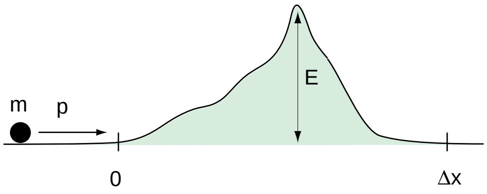


Figure 190 Hills are never high enough

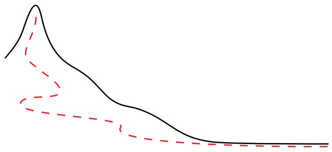
But there is more to come. A minimum action implies that cages in zoos are dangerous and banks are not safe. A cage is a feature requiring a lot of energy to be overcome. Mathematically, the wall of a cage is an energy hill, similar to the one shown in Figure 190. If a particle on one side of the hill has momentum p, it is simple to show that the particle can be observed on the other side of the hill, at position Δx, even if its kinetic energy $p^2/2m$ is *smaller* than the height E of the hill. In everyday life this is impossible. But imagine that the missing momentum $\Delta p = \sqrt{2mE - p^2}$ to overcome the hill satisfies $\Delta x \Delta p \geqslant \hbar/2$. The quantum of action thus implies that a hill of width

$$\Delta x \leqslant \frac{\hbar/2}{\sqrt{2mE - p^2}} \tag{402}$$

is *not* an obstacle to the particle. But this is not all. Since the value of the particle momentum p is itself undetermined, a particle can overcome the hill even if the hill is *wider* than value (402), though the broader it is the smaller the probability is. As a result, any particle can overcome *any* obstacle. This effect, for obvious reasons, is called the *tunnelling effect*. In short, the minimum action principle implies that there are no safe boxes in nature. Due to tunnelling, *matter is not impenetrable*, in contrast to everyday, classical observation. Can you explain why lion cages work *despite* the quantum of action? Challenge 940

By the way, the quantum of action also implies that a particle with a kinetic energy larger than the energy height of a hill can get reflected by the hill. Classically this is impossible. Can you explain the observation? Challenge 941

The minimum action principle also implies that book shelves are dangerous. Shelves are obstacles to motion. A book in a shelf is in the same situation as the mass in Figure 191;



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

the mass is surrounded by energy hills hindering its escape to the outer, lower energy world. Now, due to the tunnelling effect, escape is always possible. The same picture applies to a branch of a tree, a nail in a wall, or to anything attached to anything else. Fixing things to each other is never for ever. We will find out that every example of light emission, and even radioactivity, results from this effect. The quantum of action thus implies that *decay is part of nature*. In short, there are no stable excited systems in nature. For the same reason by the way, *no* memory can be perfect. (Can you confirm the deduction?)Note that decay Challenge 942 often appears in everyday life, where it just has a different name: *breaking*. In fact, all cases in which something breaks require the quantum of action for their description. Obviously, the cause of breaking is often classical, but the *mechanism* of breaking is always quantum. Only objects following quantum theory can break.

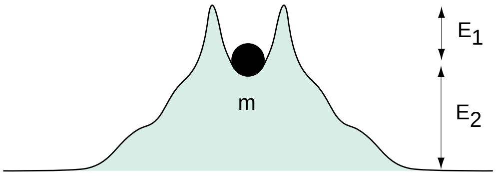


Figure 191 Leaving enclosures

Taking a more general view, also *aging* and *death* result from the quantum of action. Death, like aging, is a sum of breaking processes; breaking is a form of decay, which is due to tunnelling. Death is thus a quantum process. Classically, death does not exist. Might this be the reason that so many believe inimmortality or eternal youth? Challenge 943

A minimum action also implies that matter cannot be continuous, but must be composed of smallest entities. Indeed, the flow of a truly continuous material would contradict the quantum principle. Can you give the precise argument?Of course, at this point of our adventure, the non-continuity of matter is no Challenge 944 news any more. In addition, the quantum of action implies that even *radiation* cannot be continuous. As Albert Einstein stated clearly for the first time, light is made of quantum particles. More generally, the quantum of action implies that in nature *all flows and all waves are made of microscopic particles*. The term ‘microscopic’ or ‘quantum’ is essential, as such particles do *not* behave like little stones. We have already encountered several differences, and will encounter more of them shortly. For these reasons, microscopic particles should bear a special name; but all proposals, of which *quanton* is the most popular, have not caught on yet.

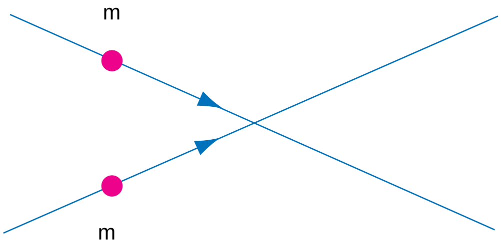


Figure 192 Identical objects with crossing paths

The quantum of action has several strange consequences for microscopic particles. Take two of them with the same mass and the same composition. Imagine that their paths cross, and that at the crossing they approach each other to small distances, as shown in Figure 192. A minimum action implies that in such a situation, if the distance becomes small enough, the two particles can switch role without anybody being able to avoid or to ever notice it. For example, in a gas it is *impossible*, due to the quantum of action, to follow particles moving around and to say which particle is which. Can you confirm this deduction and specify the conditions using the indeterminacy relations?In summary, in nature Challenge 945 *it is impossible to distinguish identical particles*. Can you guess what happens in the case of light? Challenge 946

But matter deserves still more attention. Imagine two particles, even two different ones, approaching each other to small distances, as shown in Figure 193. We know that if the approach distance gets small, things get fuzzy. Now, if something happens in that small domain in such a way that the resulting outgoing products have the same total momentum and energy as the incoming ones, the minimum action principle makes such processes possible. Indeed, ruling out such processes would imply that arbitrary small actions could be observed, thus eliminating nature's fuzziness, as you might want to check Challenge 947 by yourself.In short, *a minimum action allows transformation of matter*. One also says that the quantum of action allows particle *reactions*. In fact, we will discover that *all* kinds of reactions in nature, including chemical and nuclear ones, are only due to the existence of the quantum of action.

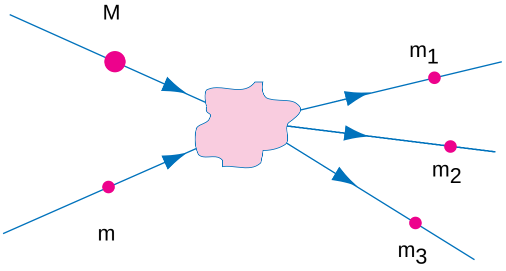


Figure 193 Transformation through reaction

But there is more. Due to the indeterminacy relations, it is impossible to give a definite value to both the momentum and the position of a particle. Obviously, this is also impossible for all the components of a measurement set-up or an observer. This implies that initial conditions – both for a system *and* for the measurement set-up – cannot be exactly duplicated. A minimum action thus implies that whenever an experiment on a microscopic system is performed twice, the outcome will be *different*. The result would be the same only if both the system and the observer would be in exactly the same condition in both situations. This turns out to be impossible, both due to the second principle of thermodynamics and due to the quantum principle. Therefore, *microscopic systems behave randomly*. Obviously, there will be some *average* outcome; nevertheless, microscopic observations are probabilistic. Albert Einstein found this conclusion of quantum theory the most difficult to swallow, as this randomness implies that the behaviour of quantum systems is strikingly different from that of classical systems. But the conclusion is unavoidable: nature behaves randomly.

A good example is given by trains. Einstein used trains to develop and explain relativity. But trains are also important for quantum physics. Everybody knows that one can use a train window to look either at the *outside* landscape or, by concentrating on the reflected image, to observe some interesting person *inside* the carriage. In other words, glass reflects some of the light particles and lets some others pass through. A random selection of light particles, yet with constant average, is reflected by the glass. Partial reflection is thus similar to the tunnel effect. Indeed, the partial reflection of glass for photons is a result of the quantum of action. Again, the average situation *can be described* by classical physics, but the mechanism of random photon reflection *cannot be explained* without quantum theory. Neither can the precise amount of partial reflection be deduced with classical physics. Without the quantum of action, train trips would be much more boring.

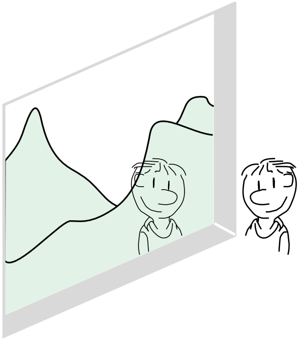


Figure 194 How do train windows manage to show two superimposed images?

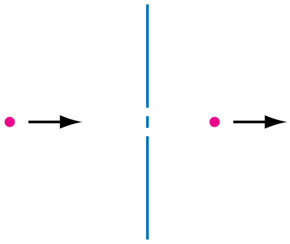


Figure 195 A particle and a screen with two nearby slits

Finally, the quantum of action implies a famous result about the *path* of particles. If a particle travels from a point to another, there is no way to say which path it has taken in between. Indeed, in order to distinguish among the possible paths, actions smaller than $\hbar/2$ would have to be measured. In particular, if a particle is sent through a screen with two nearby slits, it is usually impossible to say through which slit it passed to the other side. The impossibility is fundamental. As we will find out soon, this impossibility leads to particle interference.

We will also discover that the quantum of action is the origin for the importance of the action observable in classical physics. In fact, the existence of a *minimal* action is the reason for the *least* action principle of classical physics.

We will then discover that the quantum of action implies that macroscopic physical systems cannot be copied or 'cloned', as quantum theorists like to say. Nature does not allow to copy objects. Copying machines do not exist. The quantum of action makes it impossible to gather and use all information in a way to produce a perfect copy. As a result, we will deduce that the precise order of measurements plays a role in experiments; in short, physical observables do not commute.

We will also find out that the quantum of action implies that systems are not always independent, but can be 'entangled'. This term, introduced by Erwin Schrödinger, describes the most absurd consequences of quantum theory. Entanglement makes everything in nature connected to everything else. Entanglement produces effects which look (but are not) faster than light. Entanglement produces a (fake) from of non-locality. Entanglement also implies that trustworthy communication does not exist. Ref. 542

Don't all these deductions look wrong or at least crazy? In fact, if you or your lawyer made any of these statements in court, maybe even under oath, you would be likely to end up in prison! However, all above statements are correct, as they are all confirmed by experiment. And the surprises are by far not finished. You might have noticed that so far, no situation related to electricity, to the nuclear interactions, or to gravity was included. In these domains the surprises are even more astonishing; the observation of antimatter, of electric current flow without resistance, of the motion inside muscles, of vacuum energy, of nuclear reactions in stars, and maybe soon of boiling empty space, will fascinate you as much as they have fascinated and still fascinate thousands of researchers.

In particular, the consequences of the quantum of action on the early universe are simply Challenge 948 mind-boggling. Just try to explore for yourself its consequences for the big bang.Together, all these topics will lead us towards the top of Motion Mountain. The topics are so strange, so incredible, and at the same time so numerous that quantum physics can be rightly called the description of motion for *crazy* scientists. In a sense, this is the generalization of the previous definition, when we called quantum physics the description of motion related to pleasure.

In order to continue towards the top of Motion Mountain, our next task will be the study of our classical standard of motion: the motion of light.

> Nie und nirgends hat es Materie ohne Bewegung gegeben, oder kann es sie geben.
> Friedrich Engels[*]

17. Light – the strange consequences of the quantum of action

> ... alle Wesen leben vom Lichte,
> jedes glückliche Geschöpfe ...
> Friedrich Schiller[**]

What is colour?

If all the colours of materials are quantum effects, as just argued, it becomes even more Ref. 543 interesting to study the properties of light in the light of the quantum of action. If in nature there is a minimum change, there should also be a minimum illumination. Such a result has been indirectly predicted by Epicurus (341–271 BCE) already in ancient Greece, when he stated that light is a stream of little particles.

But our eye does not detect light particles. We need devices to help us. A simple way is to start with a screen behind a prism illuminated with white light. The light is split into colours. When the screen is put further and further away, the illumination intensity cannot become infinitely small, as that would contradict the quantum of action. To check this prediction, we only need some black and white photographic film. Everybody knows that film is blackened by daylight of any colour; at medium light intensities it becomes dark grey and at lower intensities light grey. Looking at an extremely light grey film under the microscope, we discover that even under uniform illumination the grey shade actually is a more or less dense collections of black spots. Exposed film does *not* show a homogeneous colour; on the contrary, it reacts as if light is made of small particles. In fact, this is a general observation:

Ref. 558 [*] 'Never and nowhere has matter existed, or can it exist without motion.' Friedrich Engels (1820–1895) was one of the theoreticians of communism.

[**] '... all beings live of light, every happy creature ...' Friedrich Schiller (1759, Marbach–1805, Weimar), important German poet, playwright, and historian.

whenever sensitive light detectors are constructed with the aim to ‘see’ as accurately as possible, as e.g. in dark environments, one always finds that light manifests itself as a stream of *particles*. They are called *photons*, a term that appeared in 1926, superseding the term ‘light quanta’ in usage before. A low or high light intensity is simply a small or high number of photons.

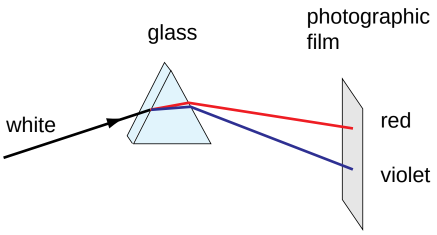


Figure 196 Illumination by pure-colour light

Figure 197 Observation of photons

These experiments thus show that the continuum description of light is *not correct* for small intensities. More precise measurements confirm the role of the quantum of action: every photon leads to the *same* amount of change. This amount of change is the *minimal* amount of change that light can produce. Indeed, if a minimum action would not exist, light could be packaged into arbitrary small amounts. On the contrary, the classical description of light by a *continuous* state function $A(t,x)$ or $F(t,x)$, See page 387 whose evolution is described by a principle of least action, is *wrong*, as it does *not* describe the observed particle effects. Another, modified description is required. The modification has to be important only at low light intensities, since at high intensities the classical Lagrangian accurately describes all experimental observations.[*]

At which intensities does light cease to behave as a continuous wave? Our eye can help us to find a limit. Human eyesight does not allow to consciously distinguish single photons, even though experiments show that the hardware of the eye is able to do this. Ref. 544 The faintest stars which can be seen at night produce a light intensity of about 0.6 $\mathrm{nW/m^2}$. Since the pupil of the eye is quite small, and as we are not able to see individual photons, photons must have small energy indeed.

In today’s laboratory experiments, recording and counting individual photons is standard practice. Photon counters are part of many spectroscopy set-ups, such as those used to measure smallest concentration of materials. For example, they help to detect drugs in human hair. All these experiments thus prove directly that light is a stream of particles, as Epicurus had advanced in ancient Greece.

---

[*] This transition from the classical case to the quantum case used to be called *quantization*. The concept and the ideas behind it are only of historical interest today.

This and many other experiments show that a beam of light of frequency f, which determines its colour, is accurately described as a stream of photons, each with the *same* energy E given by

$$E = \hbar 2\pi f = \hbar\omega \quad . \tag{403}$$

This shows that for light, the smallest measurable action is given by the quantum of action $\hbar$. This is *twice* the smallest action observable in nature; the reasons and implications will unfold during the rest of our walk. In summary, *colour* is a property of photons. A coloured light beam is a hailstorm of the corresponding photons.

The value of *Planck's constant* can be determined from measurements of black bodies or other light sources. The result See page 429

$$\hbar = 1.1 \cdot 10^{-34}\,\mathrm{Js} \tag{404}$$

is so small that we understand why photons go unnoticed by humans. Indeed, in normal light conditions the photon numbers are so high that the continuum approximation for the electromagnetic field is of high accuracy. In the dark, the insensivity of the signal processing of the human eye, in particular the slowness of the light receptors, makes photon counting impossible. The eye is not far from maximum possible sensitivity though; from the numbers given above about dim stars we can deduce that humans are able to see *consciously* flashes of about half a dozen detected photons. Ref. 544 Challenge 949

What other properties do photons have? Quite a collection, as we will see. We will deduce them systematically in the following, using the data collected in classical physics, while taking the quantum of action firmly into account. For example, photons have no mass[*] and no electric charge. Can you confirm this? Challenge 950

We know that light can *hit* objects. Since the energy and the speed of photons is known, we guess that the photon momentum obeys Challenge 951

$$p = \hbar\frac{2\pi}{\lambda} \quad \text{or} \quad \mathbf{p} = \hbar\mathbf{k} \quad . \tag{405}$$

In other words, if light is made of particles, we should be able to play billiard with them. This is indeed possible, as Arthur Compton showed in a famous experiment in 1923. He directed X-rays, which are high energy photons, onto graphite, a material in which electrons move almost freely. He found that whenever the electrons in the material get hit by the X-ray photons, the deflected X-rays change colour. As expected, the strength of the hit depends on the deflection angle of the photon. From the colour change and the reflection angle, Compton confirmed that the photon momentum obeys the above expression. All other experiments agree that photons have momentum. For example, when an atom *emits* light, the atom feels a *recoil*; the momentum again turns out to be given by the same value (405). In short, every photon has momentum. Ref. 546

The value of photon momentum also respects the indeterminacy principle; in the same way that it is impossible to measure exactly both the wavelength of a wave and the position of its crest, it becomes impossible to measure both the momentum and the position of a photon. Can you confirm this?In other words, the value of the photon momentum can also Challenge 952

---

[*] The present upper limit for the mass of a photon is 10^{-51} kg. Ref. 545

be seen as a direct consequence of the quantum of action.

From our study of classical physics we know that light has more properties than its colour: light can be *polarized*. That is only a complicated way to say that light can *turn* objects it shines on. Or again, light has an angular momentum oriented along the axis of propagation. What about photons? Measurements consistently find that each light particle carries an *angular momentum* of $\hbar$, also called its *spin*. Photons somehow 'turn'. The direction of the spin is either parallel or antiparallel to the direction of motion. Again, the magnitude of the photon spin is not a surprise; it confirmsthe classical relation $L = E/\omega$ between energy and angular momentum that we found in the section on classical electrodynamics. Note that in contrast to intuition, the angular momentum of a photon is fixed, and thus *independent* of its energy. Even the photons with the highest energy have $L = \hbar$. Of course, the value of the spin also respects the limit given by the quantum of action. The spin value $\hbar$ – instead of $\hbar/2$ – has important consequences; they will become clear shortly.

See page 401

Challenge 953

What is light? – Again

> La lumière est un mouvement luminaire de corps lumineux.[*]

In the 17th century, Blaise Pascal[**] used this sentence to make fun about certain physicists. He ridiculed (rightly so) the blatant use of a circular definition. Of course, he was right; in his time, the definition was indeed circular, as no meaning could be given to any of the terms. But as usual, whenever an observation is studied with care by physicists, they give philosophers a beating. All those originally undefined terms now have a definite meaning: light is indeed a type of motion, this motion can rightly be called luminary because in opposition to motion of material bodies, it has the unique property $v = c$, and the luminous bodies, today called photons, are characterized and differentiated from all other particles by the dispersion relation $E = \hbar k$, by their spin $L = \hbar$, by the vanishing value of all other quantum numbers, and by being the quanta of the electromagnetic field.

In short, *light is a stream of photons*. The existence of photons is the first example of a general property of the world on small scales: *all waves and all flows in nature are made of particles*. In the old days of physics, books used to discuss at length a so-called *wave-particle duality*. Large numbers of microscopic particles do indeed behave as waves. We will see shortly that this is the case even for matter. The fundamental constituents of *all* waves are quantum particles. There is *no* exception. The everyday, continuum description of light is thus similar in many aspects to the description of water as a continuous fluid; photons are atoms of light, and continuity is an approximation for large particle numbers.

However, a lot is not clear yet. Where inside matter do these monochromatic photons come from? Even more interestingly, if light is made of particles, all electromagnetic fields, even static ones, must be made of photons as well. However, in static fields nothing is flowing. How is this apparent contradiction solved? And what effects does the particle aspect have on these static fields?

---

[*] Light is the luminary movement of luminous bodies.

[**] Blaise Pascal (1623, Clermont–1662, Paris) important French mathematician and physicist up to the age of twenty-six; he then turned theologian and philosopher.

An even more important question remains: photons clearly do *not* behave like small stones at all: stones do not form waves and do not interfere. The properties of photons thus require some more careful study. Let us go on.

Size of photons

First of all, we might ask: what are these photons made of? All experiments so far, performed down to the present limit of about 10^{-20} m, give the same answer: 'we can't find anything'. That is consistent both with a vanishing mass and a vanishing size of photons; indeed, one intuitively expects any body with a finite size to have a finite mass. Thus, even though experiments give only an upper limit, it is consistent to claim that a photon has *no size*.

A particle with no size cannot have any constituents. A photon thus cannot be divided into smaller entities. For this reason people refer to photons as *elementary* particles. We will give some strong additional arguments for this deduction soon. (Can you find one?)This is a strange result. How can a photon have vanishing size, have no constituents, and still be *something*? The answer will appear later on. At the moment we simply have to accept the situation as it is. We therefore turn to an easier question.

Challenge 954

Are photons countable? – Squeezed light

Above we showed that in order to count photons, the simplest way is to absorb them on a screen. Everybody knows that a light beam crossing a dark room cannot be seen – except if it hits dust or some object, which means absorption in the detection device, such as the eye. How can one count photons without destroying them?

It is possible to reflect photons on a mirror, and to measure the recoil of the mirror. This seems almost unbelievable, but nowadays the effect is becoming measurable even for small number of photons. For example, it is becoming of importance in the mirrors used in gravitational wave detectors, where the position of laser mirrors has to be measured to high precision.

Another way of counting photons without destroying them uses special high quality laser cavities. Using smartly placed atoms inside such a cavity, it is possible to count the number of photons by the effect they have on these atoms.

However, once we know how to count photons and start doing so, the next difficulty appears straight away. Measurements show that even the best light beams, from the most sophisticated lasers, *fluctuate* in intensity. This does not come as a surprise: if a steady beam would *not* fluctuate, observing it twice in a row would yield a vanishing value for the action. However, there is a minimum action in nature, namely $\hbar/2$. Thus any beam and any flow in nature fluctuates. But that is not all.

A light beam is described by its intensity and its phase. The change – or action – occurring while a beam moves is given by the variation in the product of intensity and phase. Experiments confirm the obvious deduction: intensity and phase of beams behave like momentum and position of particles: they obey an indeterminacy relation. You can deduce it yourself, in the same way we deduced Heisenberg's relations. Using as characteristic intensity $I = E/\omega$

the energy per circular frequency, and calling the phase φ, we get[*]

$$\Delta I \, \Delta\varphi \geqslant \frac{\hbar}{2} \quad . \tag{407}$$

For light from lamps, the product of the left side is much larger than the quantum of action. On the other hand, laser beams can (almost) reach the limit. Among these, light beams in which the two uncertainties strongly differ from each other are called *nonclassical light* or *squeezed light*; they are used in many modern research applications. Such light beams have to be treated carefully, as the smallest disturbances transform them back into usual laser beams, where the two uncertainties have the same value. An example of non-classical light are those beams with a given, fixed photon number, thus with an extremely large phase uncertainty.

The observation of nonclassical light points to a strange consequence valid even for classical light: the number of photons in a light beam is *not* a defined quantity. In general it is *undetermined*, and it *fluctuates*. The number of photons at the beginning of a beam is not necessarily the same as at the end of the beam. Photons, in contrast to stones, cannot be counted precisely – as long as they move. Only within the limit set by indeterminacy an approximate number can be measured.

A limit example are those beams with an (almost) fixed phase. In them, the photon number fluctuates from zero to infinity. In other words, to produce a *coherent* laser beam, approximating a pure sine wave as perfectly as possible, one must build a source in which the photon number is as undetermined as possible.

The other extreme is a beam with a fixed number of photons; in such a beam, the phase fluctuates erratically. Most daily life situations, such as the light from incandescent lamps, lie somewhere in the middle: both phase and intensity uncertainties are of similar magnitude.

As an aside, it turns out that in deep, dark intergalactic space, far from every star, there still are about 400 photons per cubic centimetre. But also this number, like the number of photons in a light beam, has its measurement indeterminacy. Can you estimate it? Challenge 955

In summary, unlike little stones, *photons are not countable*. And that it not the last difference between photons and stones.

Position of photons

Where is a photon when it moves in a beam of light? Quantum theory gives a simple answer: nowhere in particular. The proof is given most spectacularly by experiments which show that even a beam made of a *single* photon can be split, be led along two different paths, and

---

[*] A large photon number is assumed in the expression; this is obvious, as Δφ cannot grow beyond all bounds. The exact relations are

$$\begin{aligned} \Delta I \, \Delta\cos\varphi &\geqslant \frac{1}{2} |<\sin\varphi>| \\ \Delta I \, \Delta\sin\varphi &\geqslant \frac{1}{2} |<\cos\varphi>| \end{aligned} \tag{406}$$

where $< x >$ is the expectation value of the observable x.

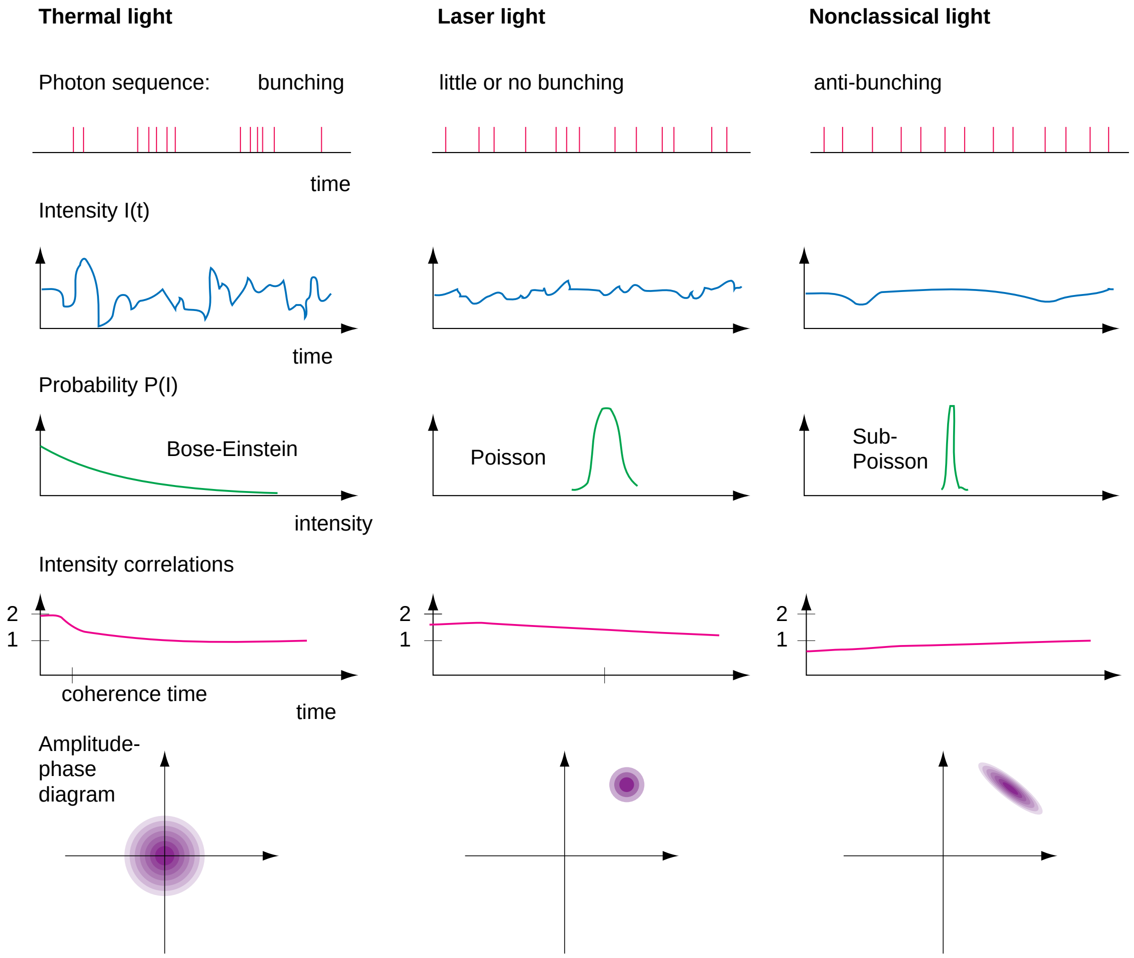


Figure 198 Various types of light

then be recombined. The resulting interference shows that the single photon cannot be said to have taken either of the two paths. If one of the two paths is blocked, the pattern on the screen changes. In other words, the photons somehow must have taken both paths at the same time. Photons cannot be localized. They have no position.*

This impossibility of localization can be clarified somewhat in more detail. It is impossible to localize photons in the direction *transverse* to the motion. It is less difficult to localize photons *along* the motion direction. In the latter case, the quantum of action implies that the longitudinal position is uncertain within a value given by the wavelength of the corresponding colour. Can you confirm this? Challenge 956

In particular, this means that photons *cannot* be simply visualized as short wave trains. Photons are truly *unlocalizable* entities specific to the quantum world.

---

* This conclusion cannot be avoided by saying that photons are split at the beam splitter: if one puts a detector into each arm, one finds that they never detect a photon at the same time. Photons cannot be divided.

Now, if photons can be almost localized along their motion, we can ask the following question: How are photons lined up in a light beam? Of course, we just saw that it does not make sense to speak of their *precise* position. But are photons in a perfect beam arriving in almost regular intervals or not?

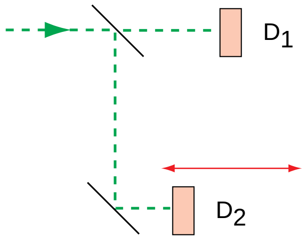


Figure 199 How to measure photon statistics

To the shame of physicists, the study of this question was started by two astronomers, Robert Hanbury Brown and Richard Twiss, in 1956. They used a simple method to measure the probability that a second photon in a light beam arrives at a given time after a first one. They simply split the beam, put one detector in the first branch and varied the position of a second detector in the other branch. Ref. 547

Hanbury Brown and Twiss found that for coherent light the clicks in the two counters, and thus the photons, are *correlated*. First of all, this result is in complete contrast with classical electrodynamics. Photons are indeed necessary to describe light. In more detail, he experiment showed that whenever a photon hits, the probability that a second one hits just afterwards is highest. Photons in beams are thus *bunched*. Every light beam shows an upper limit time to bunching, called the *coherence time*. For times larger than the coherence time, the probability for bunching is low and independent of the time interval, as shown in Figure 199. The coherence time characterizes every light beam, or better, every light source. In fact, it is more obvious to use the concept of *coherence length*, as it gives a clearer image for a light beam. For thermal lamps, the coherence length is only a few micrometers, a small multiple of the wavelength. The largest coherence lengths, up to over 100 000 km, are realized in research lasers. Interestingly, coherent light is even found in nature; several special stars have been found to emit coherent light. Ref. 548

Even though the intensity of a good laser light is almost constant, laser beam photons still do not arrive in regular intervals. Even the best laser light shows bunching, though with different statistics and to a lower degree than lamp light. Light for which photons arrive regularly, thus showing so-called photon *anti-bunching*, is obviously nonclassical in the sense defined above; such light can be produced only by special experimental arrangements. The most extreme example is pursued at present by several research groups; they aim to construct light sources which emit one photon at a time, at regular time intervals, as often as possible.

In summary, experiments force one to conclude that light is made of photons, but that photons *cannot* be localized in light beams. It makes no sense to talk about the position of a photon in general; the idea makes only sense in some special situations, and then only approximately and as a statistical average.

Are photons necessary?

Also gibt es sie doch.
Max Planck[*]

In light of the results uncovered so far, the answer to the title question is obvious. On the other hand, the *photoelectric effect* is usually cited in school books as the first and most obvious experimental proof of the existence of photons. For certain metals, such as lithium or caesium, incident light leads to the emission of electrons. It is observed that the energy of the ejected electrons is *not* dependent on the intensity of the light, but only dependent on the difference between $\hbar$ times its frequency and a material dependent threshold energy. In 1905, Albert Einstein predicted this result from the assumption that light is made of photons of energy $E = \hbar\omega$. He imagined that this energy is used partly to extract the electron over the threshold, and partly to give it kinetic energy. More photons only lead to more electrons, not to faster ones.

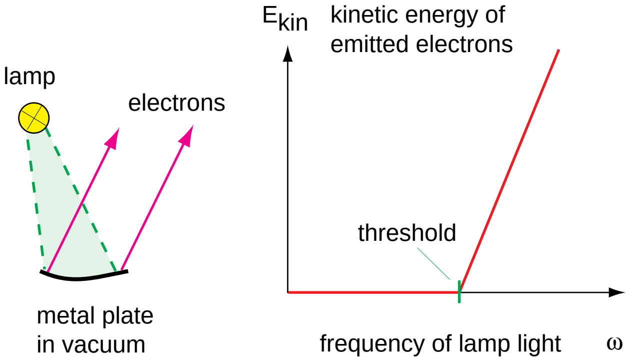


Figure 200 The kinetic energy of electrons emitted in the photoelectric effect

Einstein received the Nobel price for this explanation. But Einstein was a *genius*; that means he deduced the correct result by a somewhat incorrect reasoning. The (small) mistake was the prejudice that a classical, continuous light beam would produce a different effect. It does not take a lot to imagine that a classical, continuous electromagnetic field interacting with discrete matter, made of discrete atoms containing discrete electrons, leads to exactly the same result, if the motion of electrons is described by quantum theory. Several researchers confirmed this point already early in the twentieth century. The photoelectric effect by itself does *not* require photons for its explanation. Ref. 549

Indeed, many were not convinced. Historically, the most important argument for the *necessity* of light quanta was given by Henri Poincaré. In 1911 and 1912, at age 57 and only a few months before his death, he published two influential papers proving that the radiation law of black bodies, the one in which the quantum of action had been discovered by Max Planck, *requires* the introduction of photons. He also showed that the amount of radiation emitted by a hot body is *finite* only due to the *quantum* nature of the processes leading to light emission. A description of the processes by classical electrodynamics would lead Ref. 550

---

[*] 'Thus they do exist after all.' Max Planck, in later years, said this after standing silently, for a long time, in front of an apparatus which counted single photons by producing a click for each one it detected. It is not a secret that for a large part of his life, Planck was not a friend of the photon concept, even though his own results were the starting point for its introduction.

to *infinite* amounts of radiated energy. These two influential papers convinced most of the sceptic physics researchers at the time that it was worthwhile to study quantum phenomena in more detail.

Poincaré did not know about $S \geqslant \hbar/2$; yet his argument is based on the observation that light of a given frequency always has a minimum intensity, namely one photon. Splitting such a one photon beam into two beams, e.g. using a half-silvered mirror, does produce two beams. However, there is no way to find more than one photon in those two beams together.

Another interesting experiment requiring photons is the observation of 'molecules of photons'. In 1995, Jacobson et al. predicted that the de Broglie wavelength of a *packet* of photons could be observed. Following quantum theory it is given by the wavelength of a single photon divided by the number of photons in the packet. The team argued that the packet wavelength could be observable if one would be able to split and recombine such packets without destroying the cohesion within the packet. In 1999, this effect was indeed observed by de Pádua and his brazilian research group. They used a nonlinear crystal to create what they call a *biphoton*, and observed its interference properties, finding a reduction of the effective wavelength by the predicted factor of two. Ref. 552

Another argument for the necessity of photons is the mentioned recoil felt by atoms emitting light. The recoil measured in these cases is best explained by the emission of a photon in a particular direction. Classical electrodynamics predicts the emission of a spherical wave, with no preferred direction.

Obviously, the observation of *nonclassical light*, also called *squeezed light*, also argues for the existence of photons, as squeezed light proves that photons indeed are an *intrinsic* aspect of light, necessary even when no interactions with matter play a role. The same is true for the Hanbury Brown-Twiss effect. Ref. 551

In summary, the concept of photon is indeed *necessary* for a precise description of light, but the details are often subtle, as its properties are unusual and require a change in thinking habits. To avoid this, all high school books stop discussing photons after the photoelectric effect. That is a pity; things get interesting only after that. To savour the fascination, ponder the following issue. Obviously, all electromagnetic fields are made of photons. Photons can be counted for gamma rays, X-rays, ultraviolet light, visible light and infrared light. However, for lower frequencies, photons have not been detected yet. Can you imagine whatwould be necessary to count the photons emitted from a radio station? Challenge 957

The issue directly leads to the most important question of all:

How can a wave be made of particles?

> Fünfzig Jahre intensiven Nachdenkens haben mich der Antwort
> auf die Frage 'Was sind Lichtquanten?' nicht näher gebracht.
> Natürlich bildet sich heute jeder Wicht ein, er wisse die Antwort.
> Doch da täuscht er sich.
> Albert Einstein, 1951 *

If a light wave is made of particles, one must be able to explain each and every wave properties with help of photons. The experiments mentioned above already hinted that this is possible only because photons are *quantum* particles. Let us take a more detailed look at this connection.

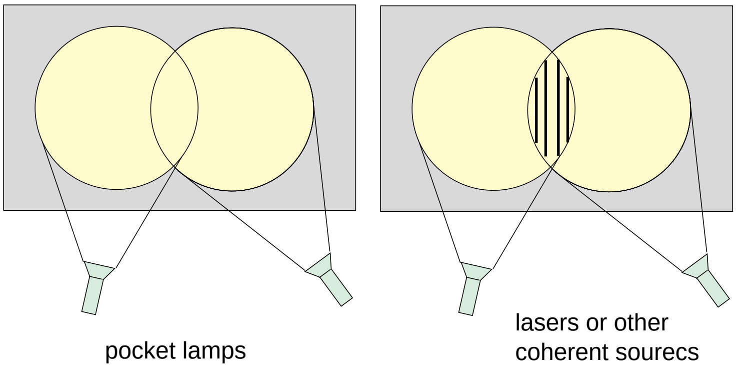


Figure 201 Light crossing light

Light can *cross* other light undisturbed. This observation is not hard to explain with photons; since photons do not interact with each other, and since they are point-like, they 'never' hit each other. In fact, there is an extremely small probability for their interaction, as will be found below, but this effect is not observable in everyday life.

But the problems are not finished yet. If two light beams of identical frequencies and fixed phase relation cross, we observe alternating bright and dark regions, so-called *interference fringes*. How do these interference fringes appear? Obviously, photons are not detected in the dark regions. How can this be? There is only one possible way to answer: the brightness gives the *probability* for a photon to arrive at that place. Some additional thinking leads to the following conclusion:

(1) The probability of a photon arriving somewhere is given by the *square of an arrow*;
(2) the final arrow is the *sum* of all arrows getting there, taking all possible paths;
(3) the arrow's direction stays fixed in space when photons move;
(4) the length of an arrow shrinks with the square of the travelled distance;
(5) photons emitted by one-coloured sources are emitted with arrows of constant length pointing in direction ωt; in other words, such monochromatic sources spit out photons with a *rotating* mouth.

---

* 'Fifty years of intense reflection have not brought me nearer to the answer of the question 'What are light quanta?' Of course nowadays every little mind thinks he knows the answer. But he is wrong.'

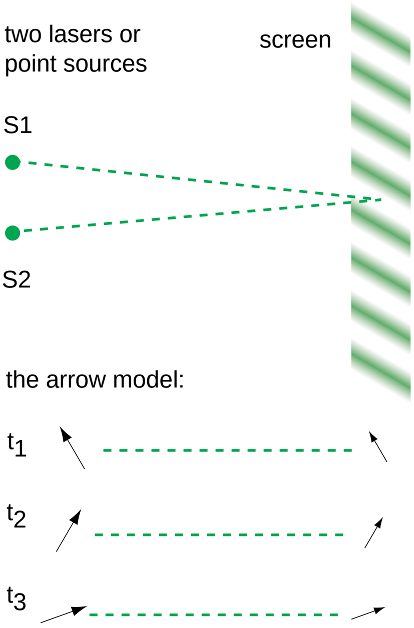


Figure 202 Interference and the description of light with arrows (at one particular instant of time)

(6) photons emitted by thermal sources, such as pocket lamps, are emitted with arrows of constant length pointing in *random* directions.

With this model[*] we can explain the stripes seen in laser experiments, such as those of Figure 201 and Figure 202. You can check that in some regions, the two arrows travelling through the two slits add up to zero *for all times*. No photons are detected there. In other regions, the arrows always add up to the maximal value. These regions are always bright. In between regions give in between shades. Obviously, for the case of pocket lamps the brightness also behaves as expected: the averages then simply add up, as in the common region in the left case of Figure 201.

You might want to calculate the distance of the lines when the source distance, the colour, and the distance to the screen is given. Challenge 958

Obviously, the photon model implies that interference patterns are built up as the sum of a large number of one-photon hits. Using low intensity beams, we should therefore be able to see how these little spots slowly build up an interference pattern, by accumulating at the bright spots, and never hitting the dark regions. That is indeed the case. Many experiments have confirmed this description.

It is important to stress that interference of two light beams is *not* the result of two different photons cancelling out or adding each other up. The cancelling would be against energy and momentum conservation. Interference is an effect valid for each photon separately, because each photon is spread out over the whole set-up; each photon takes all possible paths and interferes. As Paul Dirac said, *each photon interferes only with itself*. Interference only works because photons are quantons, and not at all classical particles. Ref. 553

Dirac's widely cited statement leads to a famous paradox: if a photon can interfere only with itself, how can two laser beams from two different lasers show interference? The answer of quantum physics is simple but strange: in the region where the beams interfere, there is no way to say from which source a photon is arriving. Photons are quantons; the photons in the crossing region *cannot* be said to come from a specific source. Photons in the interference region are quantons on their own right, which indeed interfere only with themselves.[**] In that region, one cannot honestly say that light is a flow of photons. That is the strange result of the quantum of action.

Waves also show *diffraction*. To understand this phenomenon with photons, let us start with a simple mirror and study *reflection* first. Photons (like any quantum particle) move

---

[*] The model gives a correct description of light with the exception that it neglects polarization.

[**] Despite regular claims of the contrary, Dirac's statement is correct.

from source to detector in *all* ways possible. As the discoverer of this explanation, Richard Feynman,* likes to stress, the term ‘all’ has to be taken literally. This was not a big deal in the explanation of interference. But in order to understand a mirror we have to include all possibilities, as crazy as they seem, as shown in Figure 203.

For a mirror, we have to add up the arrows arriving at the same time at the location of the image. They are shown, for each path, below the corresponding segment of the mirror. The arrow sum shows that light indeed does arrive at the image. It also shows that most of the contributions is coming from those paths near the middle one. If we were to perform the same calculation for another direction, (almost) no light would get there. In summary, the rule that reflection occurs with incoming angle equal to the outgoing angle is an approximation only; the rule follows from the arrow model of light.

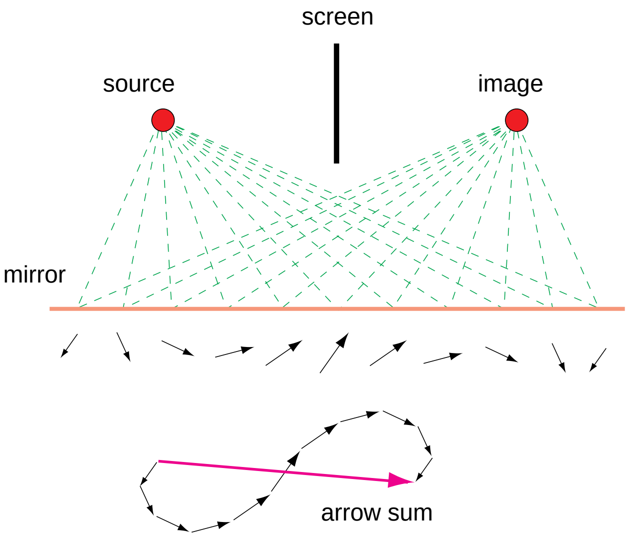


Figure 203 Light reflected by a mirror and the corresponding arrows (at one particular instant of time)

In fact, a detailed calculation, with more arrows, shows that the approximation is quite precise; the errors are much smaller than the wavelength of the light used.

The proof that light does indeed take all these strange paths is given by a more specialized mirror. As show in Figure 204, one can repeat the experiment with a mirror which reflects only along certain stripes. In this case, the stripes were carefully chosen such that the arrows reflected there all show a bias to one direction, namely to the left. The same calculation now shows that such a specialized mirror, usually called a *grating*, allows light to be reflected in unusual directions. And indeed, this behaviour is standard for waves, and called *diffraction*. In short, the arrow model for photons does allow to describe this wave property of light, provided that photons follow the mentioned crazy probability scheme. Do not get upset; as said before, quantum theory *is* the theory of crazy people.

---

* Richard (‘Dick’) Phillips Feynman (1918, New York City–1988) US American physicist. One of the founders of quantum electrodynamics, he discovered the ‘sum-over-histories’ reformulation of quantum theory, made important contributions to the theory of the weak interaction and of quantum gravity, and coauthored a famous physics textbook, the *Feynman lectures of physics*. He was famously arrogant and disrespectful of authorities, deeply dedicated to physics and to enlarging knowledge in his domain, a well known collector of surprising explanations, and an author of several popularizing texts on his work and his life. He shared the 1965 Nobel prize in physics for his work on quantum electrodynamics.

If you are interested, you might want to check that the arrow model, with the approximations it generates by summing over all possible paths, automatically ensures that the quantum of action is indeed the smallest action that can be observed. Challenge 959

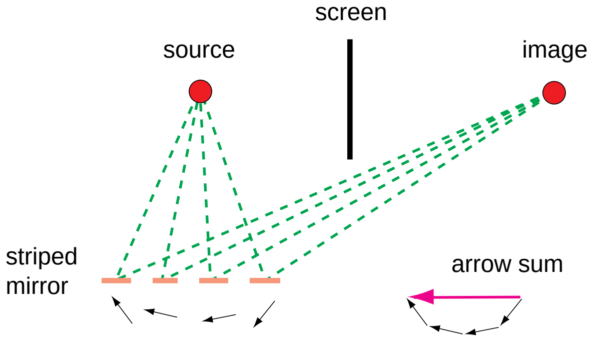


Figure 204 Light reflected by a badly placed mirror and by a grating

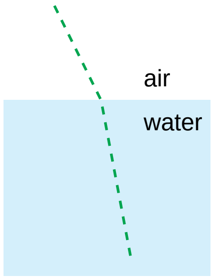


Figure 205 If light were made of little stones, they would move faster inside water

All waves have a *signal velocity*. As a consequence, waves show *refraction* when they move from one medium into another with different signal velocity. Interestingly, the naive particle picture of photons as little stones would imply that light is faster in materials with high indices of refraction, the so-called *dense* materials. Just try it.However, experiments show that light in dense materials moves *slowly*. The wave picture has no difficulties explaining this observation. (Can you confirm it?)Historically, this was one of the arguments *against* the particle theory of light. However, the arrow model of light presented above is able to explain refraction properly. Itis not difficult doing so. Challenge 960 Challenge 961 Challenge 962

Waves also *reflect partially* from materials such as glass. This is one of the toughest properties of waves to be explained with photons. The issue is important, as it is one of the few effects that is *not* explained by a classical wave theory of light. However, it *is* explained by the arrow model, as we will find out shortly. Partial reflection confirms the description of the rules (1) and (2) of the arrow model. Partial reflection shows that photons indeed behave *randomly:* some are reflected and other are not, without any selection criterion. The distinction is purely statistical. More about this issue shortly.

In waves, the fields *oscillate in time and space*. One way to show how waves can be made of particles is to show once for all how to build up a sine wave using a large number of photons. A sine wave is a coherent state of light. The way to build them up was explained by Glauber. In fact, to build a pure sine wave, one needs a superposition of a beam with one photon, a beam with two photons, a beam with three photons, continuing up to a beam with an infinite number of them. Together, they give a perfect sine wave. As expected, its photon number fluctuates to the highest degree possible. Ref. 554

If we repeat the calculation for non-ideal beams, we find that the indeterminacy relation for energy and time is respected; every emitted wave will possess a certain spectral width.

Purely monochromatic light does not exist. Similarly, no system which emits a wave *at random* can produce a monochromatic wave. All experiments confirm these results.

Waves can be *polarized*. So far, we disregarded this property. In the photon picture, polarization is the result of carefully superposing beams of photons spinning clockwise and anticlockwise. Indeed, we know that linear polarization can be seen as a result of superposing circularly polarized light of both signs, using the proper phase. What seemed a curiosity in classical optics turns out to be the fundamental explanation of quantum theory.

Photons are *indistinguishable*. When two photons of the same colour cross, there is no way to say, after the crossing, which of the two is which. The quantum of action makes this impossible. The indistinguishability of photons has an interesting consequence. It is impossible to say which emitted photon corresponds to which arriving photon. In other words, there is *no* way to follow the path of a photon in the way we are used to follow the path of a billiard ball.

Particles which behave in this way are called bosons. We will discover more details about the indistinguishability of photons in the next chapter.

In summary, we find that light *can* indeed be built of particles. However, this is only possible under the condition that photons are not precisely countable, that they are not localizable, that they have no size, no charge, and no mass, that they carry an (approximate) phase, that they carry spin, that they are indistinguishable bosons, that they can take any path whatsoever, that one cannot pinpoint their origin, and that their probability to arrive somewhere is determined by the square of the sum of amplitudes for all possible paths. In other words, light can be made of particles only under the condition that these particles have extremely special, *quantum* properties. Only these quantum properties allow them to behave like waves, in the case that they are present in large numbers.

Quantons are thus quite different from usual particles. In fact, one can argue that the only particle aspects of photons are their quantized energy, momentum, and spin. In all other aspects photons are *not* like little stones. It is more honest to say that *photons are calculating devices to precisely describe observations about light*. Ref. 555

This strange conclusion is the reason that earlier attempts to describe light as a stream of particles, such as the one by Newton, failed miserably, under the rightly deserved ridicule of all other scientists. Indeed, Newton upheld his idea against all experimental evidence, especially that on light's wave properties, something a physicist should never do. Only when people accepted that light is a wave, and then discovered and understood that quantum particles are different from everyday particles was the approach successful.

To separate between wave and particle descriptions, we can use the following criterion. Whenever matter and light interact, it is more appropriate to describe electromagnetic radiation as a wave if

$$\lambda \gg \frac{\hbar c}{kT} \quad , \tag{408}$$

where $k = 1.4 \cdot 10^{-23}\,\mathrm{J/K}$ is Boltzmann's constant. If the wavelength is much *smaller* than the right hand side, the particle description is most appropriate. If the two sides are of the same order of magnitude, both effects play a role.

Can light move faster than light? – Virtual photons

Light can move faster than c in vacuum, as well as slower than c. The quantum principle even explains the details. As long as the quantum principle is obeyed, the speed of a short light flash can differ a bit from the official value, though only a *tiny* bit. Can you estimate the allowed difference in arrival time for a light flash from the dawn of times? Challenge 963

The little arrow explanation also gives the same result. If one takes into account the crazy possibility that photons can move with any speed, one finds that all speeds very different from c cancel out. The only variation that remains, translated in distances, is the uncertainty of about one wavelength in the longitudinal directionwhich we mentioned already above. Challenge 964

However, the most absurd results of the quantum of action appear when one studies *static* electric fields, such as the field around a charged metal sphere. Obviously, such a field must also be made of photons. How do they move? It turns out that static electric fields are built of *virtual* photons. In the case of static electric fields, virtual photons are *longitudinally* polarized, do *not* carry energy away, and cannot be observed as free particles.

In fact, the vector potential A allows *four* polarizations, corresponding to the four coordinates (t,x,y,z). For the photons one usually talks about, the free or *real* photons, the polarizations in t and z direction cancel out, so that one observes only the x and y polarizations. For bound or *virtual* photons, the situation is different.

– CS – more to be written – CS –

In short, static electric and magnetic fields are continuous flows of virtual photons. Virtual photons can have mass, can have spin directions not pointing along the motion path, and can have momentum opposite to their direction of motion. All these properties are different from real photons. In this way, exchange of virtual photons leads to the *attraction* of bodies of different charge. In fact, virtual photons necessarily appear in any description of electromagnetic interactions; more about their effects, such as the famous attraction of neutral bodies, will be discussed later on.

In summary, light can indeed move faster than light, though only in amounts allowed by the quantum of action. For everyday situations, i.e. for cases with a high value of the action, all quantum effects average out, including light velocities different from c.

A different topic also belongs into this section. Not only the position, but also the energy of a single photon can be undefined. For example, certain materials split one photon of energy $\hbar\omega$ into two photons, whose two energies sum up to the original one. Quantum mechanics makes the strange prediction that the precise way the energy is split is known only when the energy of one of the two photons is measured. Only at that very instant the energy of the second photon is known. Before that, both photons have undefined energies. The process of energy fixing takes place *instantaneously*, even if the second photon is far away. We will explain below the background of this and similar strange effects, which seem to be faster than light but which are not. Indeed, such effects do not transmit energy or information faster than light. Challenge 965

Ref. 556

See page 588

Electric fields

We saw that the quantum of action implies an uncertainty for light intensity. That implies a similar limit for electric and magnetic fields. This conclusion was first drawn by Bohr and Ref. 557 Rosenfeld, in 1933. They started from the definition of the fields using a test particle of mass m and charge q, namely

$$m\mathbf{a} = q(\mathbf{E}+\mathbf{v}\times\mathbf{B}) \quad . \tag{409}$$

Since it is impossible to measure momentum and position of a particle, an indeterminacy Challenge 966 for the electrical field

$$\Delta E = \frac{\hbar}{q\,\Delta x\,T} \tag{410}$$

follows, where T is the measurement time. The value of electric fields, and similarly that of magnetic fields, is thus affected with an indeterminacy. The state of the electromagnetic fields behaves like the state of matter in this aspect.

18. Motion of matter – the end of classical physics

Wine glasses and pencils

All great things begin as blasphemies.
George Bernard Shaw

A simple consequence of the quantum of action is the impossibility of completely filling a glass of wine. If we call 'full' a glass at maximum capacity (including surface tension effects, to make the argument precise), we immediately see that the situation requires com- Challenge 967 n plete rest of the liquid's surface; however, the quantum of action forbids this. Indeed, a completely quiet surface would allow two subsequent observations which differ by less than $\hbar/2$. In other words, the quantum of action proves the old truth that a glass of wine is always partially empty *and* partially full.

The quantum of action has many similar consequences for everyday life. For example, a pencil on its tip *cannot* stay vertical, even if it is isolated from all disturbances, such as vibrations, air molecules, and ther- Challenge 968 mal motion. Are you able to confirm this?In fact, it is even possible to calculate the time after which a pencil must have fallen over.*

But the quantum of action has more important effects. As Table 45 shows, the fundamental processes of material science, of chemistry and

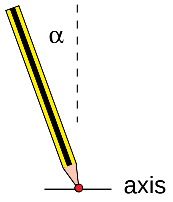


Figure 206 A falling pencil

---

* That is not easy, but neither too difficult. For an initial orientation close to the vertical, the fall time T turns out to be

$$T = \frac{1}{2\pi}T_\mathrm{o}\ln\frac{8}{\alpha} \tag{411}$$

Ref. 559 where α is the starting angle, and a fall by π is assumed. Here T_o is the oscillation time of the pencil for small Challenge 969 angles. (Can you determine it?)

Now the indeterminacy relation for the tip of the pencil gives a minimum angle, because the momentum Challenge 970 uncertainty cannot be made as large as wanted. You should be able to provide an upper limit.Once the angle is known, you can calculate the maximum time.

of life have action values around $\hbar$. This is also the case for all light emission processes.

All the unexplained effects of classical physics take place in domains where the action is near the minimum observable one. Thus we need quantum theory to understand these situations. We begin the exploration with the study of the motion of a *single* particle. Later on we will briefly expand this to a few situations with a higher number.

Matter particles

> Die Bewegung ist die Daseinsform der Materie.
> Friedrich Engels, *Anti-Dühring*.[*]

Experiments show that perfect rest is never observed. Whenever the position of a system is determined to high precision, we need a high energy probe. Only a high energy probe has a wavelength small enough to allow high precision. As a result, the system is disturbed, so that it will not be found in the same position at the next measurement. In addition, the disturbance itself is also found to be unprecisely measurable. There is thus no way to determine the original position even by taking the disturbance itself into account. In short, perfect rest cannot be observed. Experiments confirm this result for single electrons, for atoms, for molecules and for metal bars with a mass of a ton. All systems who have ever been observed with high precision confirm that perfect rest does not exist.

Quantum theory provides the general and simple rule. The smallest change in nature is the quantum of action. Thus two observations of the same system always differ in position and momentum. The product of the differences is at least the quantum of action. Rest is impossible in nature. As a result, even at lowest temperatures, particles inside matter are in motion. The fundamental lack of rest is often called *zero-point fluctuations*. Their magnitude is regularly confirmed. A good example are the recent measurements of Bose-Einstein condensates, systems with a small number of atoms (between ten and a few million) at lowest temperatures (around 1 nK) which in addition can be observed with high precision. Using tricky experimental techniques, Bose-Einstein condensates can be put into states for which $\Delta p \Delta x$ is almost exactly equal to $\hbar/2$, though never lower than this value.

But not only rest is made impossible be the quantum of action; the same applies to any situation which does not change in time. The most important example is flow.

All flows are made of particles. This statement was deduced from the quantum of action. Two flows ask for direct confirmation: flows of matter, like that of a liquid, and flows of electricity. That matter is made of particles is not new. We mentioned in the first part that a consequence of liquids being made of molecules is that even in the smoothest of pipes, even oil or any other smoothest liquid still produces noise when it flows through the pipe. We mentioned that the noise we hear in our ears in situations of absolute silence, such as in a snowy landscape in the mountains, is due to the granularity of matter.

---

[*] 'Motion is matter's way of being.'

The quantum of action also implies that electrical current cannot be a continuous flow; otherwise it would be possible to observe actions as small as desired. The simplest of these experiments was discovered only in the 1990s: Take two metal wires on the table, crossing each other. If current is made to flow from one wire to the other, via the crossover, a curve like the one shown in Figure 207 is found. The current increases or decreases in steps.

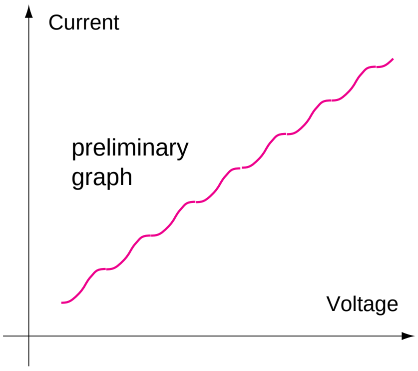


Figure 207 Steps in the flow of electrons

Many other experiments confirm the result: there is a *smallest* charge in nature. This smallest charge has the same value as the charge of an electron. Indeed, electrons turn out to be part of every atom, in a complex way to be explained shortly. In addition, a number of electrons can move freely in metals; that is the reason that metal conduct electricity so well.

Also flow of matter shows smallest units. They are called atoms or molecules, depending on the material. Electrons, atoms and molecules are *quantum* particles or *quantons*; they show some of the aspects of everyday particles, but show many other aspects which are different from what is expected from little stones. Let us have a rapid tour. Everyday matter has size, structure, mass, shape, colour, position and momentum. What about matter quantons?

First of all, matter quantons do have mass. Secondly they move in such a way that the quantum of action is respected. Matter quantons, like stones and in contrast to photons, *can* be localized. However, there is no way to ascribe them a specific momentum and a specific position at the same time. The limits are easily experienced in experiments.

Figure to be included

Figure 208 Trying to measure position and momentum

Some attempts are shown in Figure 208.

In each case, there is no way to measure indeterminacy products of position and momentum which are smaller than the quantum of action. In short, position and momentum are not exactly defined for microscopic systems. The more accurately one quantity is known, the less accurately the other is.

– CS – the rest of quantum theory will appear in the next version – CS –

No north pole

> Tristo quel discepolo che non avanza il suo maestro.
> Leonardo da Vinci[*]

The quantum of action has also important consequences for rotational motion. We saw above that due to the quantum of action, in the same way that for every object the momentum is fuzzy, also its angular momentum is. But there is more.

Classically speaking, the poles of the earth are spots which do not move, when seen form a non-rotating observer. Therefore at those spots matter would have a defined position and a defined momentum. However, the quantum of action forbids this. In short, there is no north pole on earth. More exactly, the idea of a *rotation axis* is an approximation not valid in general.

The quantum of action has other consequences for rotation. Even more interesting are the effects on microscopic particles, such as atoms, molecules or nuclei. To begin with, we note that action and angular momentum have the same units. The precision with which angular momentum can be measured depends on the precision of the rotation angle. But if a *microscopic* particle rotates by an angle, this rotation might be unobservable, a situation in fundamental contrast with the case of *macroscopic* objects. Experiments indeed confirm that many microscopic particles have unobservable rotation angles. For example, in many, but not all cases, an atomic nucleus rotated by half a turn cannot be distinguished from the unrotated nucleus.

If a microscopic particle has a *smallest* unobservable rotation angle, the quantum of action implies that the angular momentum of that particle *cannot* be zero. It must always be rotating. Therefore we need to check for each particle what its smallest unobservable angle of rotation is. Experiments provide the following unobservable values when all particles in nature are checked: 0, 4π, 2π, $4\pi/3$, π, $4\pi/5$, $2\pi/3$, etc.

Let us take an example. Let us take the mentioned nucleus for which the smallest unobservable rotation angle is *half* a turn. That would be the case for a nucleus that looks like a rugby ball turning around the short axis. Both the largest observable rotation and the uncertainty are thus a *quarter* turn. Since the change or action produced by a rotation is the

---

[*] 'Sad is that disciple who does not surpass his master.' The statement is painted in large letters in the aula magna of the University of Rome.

number of turns times the angular momentum, we find that the angular momentum of this nucleus is $2 \cdot \hbar$.

As a general result we deduce that the angular momentum of a microscopic particle can be 0, $\hbar/2$, $\hbar$, $3\hbar/2$, $2\hbar$, $5\hbar/2$, $3\hbar$, etc. In other words, the intrinsic angular momentum of particles, usually called their *spin*, is an integer multiple of $\hbar/2$. Spin describes how a particle behaves under rotations.

How can a electron rotate? At this point we do not know yet how to *picture* the rotation. But we can *feel* it. This is done in the same way we showed that light is made of rotating entities: all matter, including electrons, can be *polarized*. This was shown most clearly by the famous Stern-Gerlach experiment. Stern and Gerlach found that a beam of silver atoms can be spilt into two beams by using an inhomogeneous magnetic field. Ref. 560

– CS – the rest of quantum theory will appear in the next version – CS –

19. Colours and other interactions between light and matter

After the description of the motion of matter and radiation, the next step is the description of their interactions. In other words, how do charged particles react to electromagnetic fields and vice versa? There are a number of surprising effects, most of which appear when the problem is treated taking special relativity into account.

What are stars made of?

Figure of the spectrum of sunlight to be added

Figure 209 The spectrum of daylight: a section of the rainbow

In the beginning of the eighteenth century the english physicist William Wollaston and again the bavarian instrument maker Joseph Fraunhofer[*] noted that the rainbow lacks certain colours. These colours appear as black lines in Figure 209 and are called *Fraunhofer lines* today. In 1860, Gustav Kirchhoff and Robert Bunsen showed that the missing colours were exactly those colours that certain elements emitted when heated. With a little of experimenting they managed to show that sodium, calcium, barium, nickel, magnesium, zinc, copper and iron existed on the sun. They were unable to attribute 13 of the 476 lines they knew. In 1868, Jules Janssen and Joseph Lockyer independently predicted that the lines were from a new element; it was eventually found also on earth, by William Ramsay in 1895. Obviously it was called ‘helium’, from the greek word ‘helios’ – sun. Today we know that it is the second ingredient of the sun, in order of frequency, and of the universe, after hydrogen.

Understanding the colour lines produced by each element had started to become of interest already before this discovery; the interest rose even more afterwards, due to the applications of colours in chemistry, physics, technology, crystallography, biology and lasers.

---

[*] Born as Joseph Fraunhofer (1787, Straubing–1826). Bavarian, orphan at 11, he learns lens polishing at that age; autodidact, he studies optics from books. He enters an optical company at age 19, ensuring the success of the business, by producing the best available lenses, telescopes, micrometers, optical gratings and optical systems of his time. He invents the spectroscope and the heliometer. He discovers and counts 476 lines in the spectrum of the sun, today named after him. They are used as standards for various measurements. Physicists across the world buy their equipment from him, visit him, and ask for copies of his publications. Even after his death, his instruments remain unsurpassed. With his telescopes, in 1837 Bessel is able to measure the first parallax of a star and in 1846 Johann Gottfried Galle discovers Neptune. Fraunhofer became professor in 1819; he died young, from the consequences of the years spent working with lead and glass powder.

Classical electrodynamics cannot explain the sharp lines. Only quantum theory achieved this.

What determines the colour of atoms?

The speed of electrons is

$$v_\mathrm{n} = \frac{e^2}{2n\varepsilon_\mathrm{o} h} \approx \frac{2.2\,\mathrm{Mm/s}}{n} \approx \frac{0.007\,c}{n} \tag{412}$$

This can be checked! Putting a muon into an atom, instead of an electron, gives a time dilation of its lifetime due to its speed. The numbers coincide. Ref. 581

The orbital frequency of electrons is

$$f_\mathrm{n} = \frac{e^4 m}{4n^3 \varepsilon_\mathrm{o}^2 h^3} \approx \frac{\ldots\ldots}{n^3} \tag{413}$$

and orbital radius is

$$r_\mathrm{n} = \frac{h^2 \varepsilon_\mathrm{o} n^2}{\pi m e^2} \approx 53\,\mathrm{pm}n^2 \quad . \tag{414}$$

Quantum theory thus implies that a hydrogen atom in level $n = 500$ is about $12\,\mu$m in size, larger than many bacteria! This has indeed been achieved, even though such blown-up atoms, usually called Rydberg atoms, are extremely sensitive to perturbations. Ref. 570

The energy levels for the orbiting electrons are

$$r_\mathrm{n} = \frac{-me^4}{8n^2 \varepsilon_\mathrm{o}^2 h^2} \approx \frac{\ldots\,\mathrm{aJ}}{n^2} \approx \frac{13.6\,\mathrm{eV}}{n^2} \quad . \tag{415}$$

Some details of this approach are interesting. Electrons can go in ellipses; the highest eccentricity corresponds to the minimum value of the azimuthal quantum number, whereas the case $l = n$ correspond to circles. Ref. 571

Relativistic wave equations

> The equation was more intelligent than I was.
> Paul Dirac about his equation, repeating
> a statement made by Heinrich Hertz.

A few years after Max Planck had discovered the quantum of action, Albert Einstein published the theory of special relativity. The first question Planck asked himself was whether the quantum of action would be independent of the observer. For this reason, he invited Einstein to Berlin, and then made the then unknown patent office clerk famous in physicist circles.

A moving object implies a certain amount of change. For a comoving observer, the object would be at rest and not change at all. How does a minimum change, as described by the quantum of action, change this picture?

– CS – This section is still missing – CS –

In summary, as far as is known today, the relativistic description of the motion of charged matter and electromagnetic fields is *perfect:* no differences between theory and experiment have ever been found, despite intensive searches and despite a high reward for anybody who would find one. All known predictions completely correspond with the measurement results. In the most spectacular cases, this is true with a precision of fourteen digits. But the precision of QED is less interesting than those of its features which are missing in classical electrodynamics. Let's have a quick tour.

Antimatter

Antimatter is now a household term. Interestingly, the term was formed *before* any experimental evidence for it was known. The antimatter companion of the electron was predicted in 1926 by Paul Dirac from his equation. Without knowing this prediction, C.D. Anderson discovered it in 1932 and called it *positron*, even though 'positon', without the 'r', would have been the correct name. Anderson was studying cosmic rays and noticed that some 'electrons' were turning the wrong way in the magnetic field he had applied to his apparatus. He checked everything in his machine and finally deduced that he found a particle with the same mass as the electron, but with positive electric charge.

The existence of positrons has many strange implications. Already in 1928, before their discovery, the swedish theorist Oskar Klein had pointed out that Dirac's equation for electrons makes a strange prediction: when an electron hits a sufficiently steep potential wall, the reflection coefficient is larger than unity. Such a wall will reflect *more* than what is thrown at it. In 1935, after the discovery of the positron, Werner Heisenberg and Hans Euler explained the paradox. They found that the Dirac equation predicts a surprising effect: if an electric field exceeds the critical value of Ref. 575

$$E_c = \frac{m_e c^2}{e\lambda_e} = \frac{m_e^2 c^3}{e\hbar} = 1.3\,\mathrm{EV/m} \quad , \tag{416}$$

the vacuum will spontaneously generate electron-positron pairs, which then are separated by the field. As a result, the original field is reduced. This so-called *vacuum polarization* is also the reason for the reflection coefficient greater than unity found by Klein, since steep potentials correspond to high electric fields.

Truly gigantic examples of vacuum polarization, namely around charged black holes, will be described later on. See page 640

We note that such effects show that the *number* of particles is not a constant in the microscopic domain, in contrast to everyday life. Only the *difference* between particle number and antiparticle number turns out to be conserved. This topic will be expanded in the chapter on the nucleus.

Of course, the generation of electron-positron pairs is not a *creation* out of nothing, but a *transformation* of energy into matter. Such processes are part of every relativistic description of nature. Unfortunately, physicists have the habit to call this transformation 'creation', and thus confuse this issue somewhat. Vacuum polarization is a process transforming, as we will see, *virtual* photons into matter. That is not all: the same can also be done with *real* photons.

Transforming light into matter

Everybody who consumes science fiction nowadays knows that matter and antimatter annihilate and transform into pure light. In more detail, a matter particle and an antimatter particle annihilate into two or more photons. More interestingly, quantum theory predicts that the opposite process is also possible: photons hitting photons can produce matter!

Ref. 576 In 1997, this was finally confirmed experimentally. At the Stanford particle accelerator, photons from a high energy laser pulse were bounced off very fast electrons. In this way, the reflected photons had a very large energy when seen in the inertial frame of the experimenter. The original pulse, of 527 nm or 2.4 eV green light, had a peak power density of 10^{22} W/m^2, about the highest achievable so far. To give an idea, that is a photon density of 10^{34} /m^3 and an electric field of 10^{12} V/m, both of which are record values.

When this laser pulse was reflected off a 46.6 GeV electron beam, the returning photons had an energy of 29.2 GeV and thus had become intense gamma rays.They then collided with the other incoming green photons and produced electron-positron pairs by the reaction Challenge 971

$$\gamma_{29.2} + n\,\gamma_{\text{green}} \rightarrow \mathrm{e}^+ + \mathrm{e}^- \qquad (417)$$

for which both final particles were detected by special apparatuses. The experiment thus showed that light hitting light is possible in nature, and above all, that doing so can produce matter. This is the nearest one can get to the science fiction idea of light swords or of laser swords banging onto each other.

We will describe a few more successes of quantum theory shortly. Before we do that, we settle one important question.

Compositeness

When is an object composite? Quantum theory gives several practical answers. The first one is somewhat strange: an object is composite when its gyromagnetic ratio g is different than the one predicted by QED. The *gyromagnetic ratio* is defined as

$$g = \frac{2m}{q}\,\frac{\mu}{L} \qquad (418)$$

where μ is the magnetic moment, L the angular momentum, and m and q denote mass and electric charge of the object.

If this ratio is *different* than the value predicted by QED, about 2.0, the object is *composite*. For example, a $^4\mathrm{He}^+$ helium ion has a spin $1/2$ and a g value of $14.7 \cdot 10^3$ .Indeed, the radius of the helium ion is $3 \cdot 10^{-11}$ m, finite, and the ion is a composite entity. For the proton, one finds a g value of about 5. Indeed, one measures a proton radius of about 0.9 fm. Challenge 972

Also the neutron, which has a magnetic moment despite being neutral, must therefore be composite. Indeed, its radius is approximately that of the proton. Similarly, molecules, mountains, stars, and people must be composite. Following this first criterion, the only elementary particles are *leptons* – i.e. electrons, muons, tauons, and neutrinos –, *quarks*, and *intermediate bosons* – i.e. photons, W-bosons, Z-bosons, and gluons. More details on these particles will be uncovered in the chapter on the nucleus. See page 657

Another simple criterion for compositeness has just been mentioned: *any object with a measurable size is composite*. This criterion is related to the previous one. Indeed, the simplest models for composite structures make the prediction that Ref. 578

$$g-2=\frac{R}{\lambda_C} \qquad (419)$$

where λ_C is the Compton wavelength of the system. The expression is surprisingly precise for helium 4 ions, helium 3, tritium ions, and protons, as you might want to check.This criterion produces the same list of elementary particles as the first. Challenge 973

A third criterion for compositeness is more general: *any object larger than its Compton length is composite*. The background idea is simple. An object is composite if one can detect *internal* motion, i.e. motion of some components. Now the action of any part with mass m_{part} moving inside a composed system of size r follows

$$S_{\text{part}} < 2\pi r\, m_{\text{part}}\, c < \pi r m c \qquad (420)$$

where m is the mass of the *composite* object. On the other hand, following the principle of quantum theory, this action, to be observable, must be larger than $\hbar/2$. Inserting this condition, we find that for any composite object*

$$r > \frac{\hbar}{2\pi m c} \quad . \qquad (421)$$

The right hand side differs only by a factor $4\pi^2$ from the so-called *Compton (wave)length*

$$\lambda=\frac{h}{mc} \quad . \qquad (422)$$

of an object. Any object *larger* than its own Compton wavelength is thus composite. Any object *smaller* than the right hand side of expression (421) is thus elementary. Again, only leptons, including neutrinos, quarks, and intermediate bosons pass the test. All other objects are composite, as the tables in Appendix C make clear. This third criterion produces the same list as the previous ones. Can you explain the reason? Challenge 975

Interestingly, the topic is not over yet. Even stranger statements about compositeness will appear when gravity is taken into account. Just be patient; it is worth it.

Curiosities and challenges in quantum theory

Quantum theory is so full of strange results that all of it could be titled ‘curiosities’. A few of the prettier cases are given here.

Challenge 974 * Can you find the missing factor of 2?And is the assumption valid that the components must always be lighter than the composite?

- Can atoms rotate? Can an atom that falls on the floor roll under the table? Can atoms be put into high speed rotation? The answer is no to all questions, because angular momen- Ref. 579 tum is quantized and because atoms are not solid objects. The macroscopic case of an object turning slower and slower until it stops does not exist in the microscopic world. Can you Challenge 976 explain how this follows from the quantum of action?

To be included

Figure 210 The result of moving atoms one by one on a metal surface

- If atoms contain orbiting electrons, the rotation of the earth, via the Coriolis acceleration, should have an effect on their motion. This beautiful prediction is due to Mark Silverman; the effect is so small however, that is has not been mea- Ref. 581 sured yet.
- Do hydrogen atoms exist? Most types of atoms have been imaged with microscopes, photographed under illumination, levitated one by one, and even moved with needles, one by one, as the picture shows. Others have moved single atoms using laser beams to push Ref. 583 them. However, no such experiments exist of hydrogen atoms. Is that a reason to doubt the Challenge 977 n existence of hydrogen atoms? Taking seriously this not-so-serious discussion can be a lot of fun.
- Light is diffracted by material gratings. Can matter be diffracted by light gratings? Surprisingly, it actually can, as predicted by Dirac and Kapitza in 1937. It was proven for atoms Ref. 580 in 1986. For free electrons the feat is more difficult. The clearest confirmation came in 2001, when the technology advances for lasers were used to perform a beautiful measurement of the typical diffraction maxima for electrons diffracted by a light grating.
- Light is refracted when entering dense matter. Do matter waves behave similarly? Yes, they do. In 1995, David Pritchard showed this for sodium waves entering helium and xenon Ref. 582 gas.
- Light is totally reflected when it is directed to a dense material under an angle so large that it cannot enter it any more. Several russian physicists have shown that in the case that Ref. 581 the material is excited, the totally reflected beam can be amplified.
- 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 978 momentum p_y. Can you give an explanation?
- Space and time translations commute. Why then do the momentum operator and the Challenge 979 Hamiltonian not commute in general?
- Small changes in the strength of electromagnetic attraction between electrons and protons would have numerous important consequences. Can you describe what would happen to the size of people, to the colour of objects, to the colour of the sun or to the workings of computers if the strength would double? And if it would drop to half the usual value over Challenge 980 time?
- Ref. 584 With two mirrors and a few photons, it is possible to capture an atom and keep it floating between the two mirrors. This feat, one of the several ways to isolate single atoms, is now Challenge 981 standard practice in laboratories. Can you imagine how to achieve this?
- For a bound system in a nonrelativistic state with no angular momentum, one has the Ref. 585 relation

$$< rr >< T > \geqslant \frac{9\hbar^2}{8m} \quad , \tag{423}$$

where m is the reduced mass and T the kinetic energy of the components, and r the size of the system. Can you deduce the result and check it for hydrogen? Challenge 982 n

- The quantum of action means that there are no fractals in nature. Can you confirm this result? Challenge 983
- Electrons don't like high magnetic fields. When a magnetic field is too high, electrons are squeezed into a small space, in the transversal direction. If this spacing becomes smaller than the Compton wavelength, something happens. Physicists express this also by saying that the Landau levels spacing then becomes larger than the electron rest energy. The corresponding limit field is called the quantum *critical magnetic field*. Its value is about 4.4 GT. Challenge 984 n However, in magnetars, fields over 20 times as high have been measured. How is this possible?

The strength of electromagnetism

The great Wolfgang Pauli used to say that after his death, the first question he would ask would be an explanation of Sommerfeld's fine structure constant. (People used to comment that after the devil will have explained it to him, he will think a little, and then snap 'Wrong!') The name *fine structure constant* was given by Arnold Sommerfeld to the constant See page 405

$$\alpha = \frac{e^2}{4\pi\varepsilon_0\hbar c} \approx \frac{1}{137.035\,999\,76(50)} \approx 0.007\,297\,352\,533(27) \tag{424}$$

because it appears in explanations for the fine structure of certain atomic colour spectra. Sommerfeld was the first to understand its importance. The number is central to quantum electrodynamics for several reasons. First of all, it describes the *strength* of electromagnetism. Since all charges are multiples of the electron charge, a higher value would mean a stronger attraction or repulsion between charged bodies. The value of α thus determines the size of atoms, and thus the size of all things, as well as all colours.

Secondly, only because this number is quite a bit smaller than unity are we able to talk about particles at all. The argument is somewhat involved; it will be detailed later on. In any case, only the small value of the fine structure constant makes it possible to distinguish particles from each other. If the number were near or larger than one, particles would interact so strongly that it would *not* be possible to observe or to talk about particles at all.

This leads to the third reason that the constant is important. Since the fine structure constant is a dimensionless number, it implies some yet unknown mechanism fixing its value. Uncovering this mechanism is one challenge left over. As long as the mechanism remains unknown, we do not understand the colour and size of a single thing.

Explaining the number is the most famous and the toughest challenge of modern physics since the issue appeared in the 1920s. It is the reason for Pauli's statement cited above. In 1946, during his Nobel Prize lecture, he repeated the statement that a theory that does not determine this number cannot be complete. The challenge is so tough that for the first 50 Ref. 586 years there were only two classes of physicists: those who did not even dare to take on the challenge, and those who had no clue. This fascinating story still awaits us.

To continue with the highest efficiency on our path through quantum theory, we first look at two important topics: the issue of indistinguishability and the issue of interpretation of its probabilities.

References

537 Giuseppe Fumagalli, *Chi l'ha detto?*, Hoepli, Milano, 1983. Cited on page 515.

538 The quantum of action is introduced in Max Planck, *Über irreversible Strahlungsvorgänge*, Sitzungsberichte der Preußischen Akademie der Wissenschaften, Berlin, pp. 440–480, 1899. In this paper, Planck used the letter *b* for what nowadays is called *h*. Cited on page 516.

539 For some the rare publications emphasizing the quantum of action see M.B. Mensky, *The action uncertainty principle in continuous quantum measurements*, Physics Letters A **155**, pp. 229–235, 1991, M.B. Mensky, Physics Letters A **162**, p. 219, 1992, and M.B. Mensky, *The action uncertainty principle in continuous quantum measurements*, Physics Letters A **155**, pp. 229–235, 1991.

There is a large number of general textbooks on quantum theory. The oldest textbooks are obviously German or English. The choice should be taken by the reader, considering his own preferences.

A well-known *conceptual* introduction is Jean-Marie Levy-Leblond & Francoise Balibar, *Quantique – Rudiments*, Masson, 1997, translated into English as *Quantics*, North-Holland, 1990.

One of the most beautiful books is Julian Schwinger, *Quantum Mechanics - Symbolism of Atomic Measurements*, edited by Berthold-Georg Englert, Springer Verlag, 2001.

A modern approach with a beautiful introduction is M. Schubert & G. Weber, *Quantentheorie*, Spektrum Akademischer Verlag, 1993.

A standard beginner's text is C. Cohen-Tannoudji, B. Diu & F. Laloë, *Mécanique quantique I et II*, Hermann, Paris, 1977. It is also available in several translations. Cited on page 516.

540 Max Born, *Zur Quantenmechanik der Stoßvorgänge (vorläufige Mitteilung)*, Zeitschrift für Physik **37**, pp. 863–867, 1926, Max Born, *Quantenmechanik der Stoßvorgänge*, Zeitschrift für Physik **38**, pp. 803–827, 1926. Cited on page 518.

541 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 520.

542 R. Clifton, J. Bub & H. Halvorson, *Characterizing Quantum Theory in terms of Information-Theoretic Constraints*, http://www.arxiv.org/quant-ph/0211089. Cited on page 526.

543 Rodney Loudon, *The Quantum Theory of Light*, Oxford University Press, 2000. Cited on page 527.

544 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 in the retina detect, in colour, light intensities in the uppermost 7 or 8 decades, whereas the rods detect, in black and white, the lower light intensities. Cited on pages 528 and 529.

545 E. Fischbach, H. Kloor, R.A. Langel, A.T.Y. Lui & M. Peredo, *New geomagnetic limit on the photon mass and on long-range forces coexisting with electromagnetism*, Physical Review Letters **73**, pp. 514–517, 1994. Cited on page 529.

546 A.H. Compton, *The scattering of X-rays as particles*, American Journal of Physics **29**, pp. 817–820, 1961. Cited on page 529.

547 The famous paper is R. Hanbury Brown & R.Q. Twiss, Nature **178**, p. 1046, 1956. They got the idea to measure light in this way from their earlier work, which used the same method with radiowaves: R. Hanbury Brown & R.Q. Twiss, Nature **177**, p. 27, 1956, Cited on page 534.

548 Science 8 September 1995 Cited on page 534.

549 See the summary by P.W. Milonni, *Answer to question 45: What (if anything) does the photoelectric effect teach us?*, American Journal of Physics **65**, pp. 11–12, 1997. Cited on page 535.

550 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 535.

551 For an introduction to spueezed light, see L. Mandel, *Non-classical states of the electromagnetic field*, Physica Scripta T **12**, pp. 34–42, 1986. Cited on page 536.

552 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. Monken & 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 536.

553 The famous quote on single photon interference is found on page 9 of P.A.M. Dirac, *The Principles of quantum mechanics*, Clarendon Press, Oxford, 1930. It is also discussed, somewhat confusely, in the otherwise informative article by H. Paul, *Interference between independent photons*, Reviews of Modern Physics **58**, pp. 209–231, 1986. Cited on page 538.

554 The original papers on coherent states are three R.J. Glauber, *The quantum theory of optical coherence*, Physical Review **130**, pp. 2529–2539, 1963, J.R. Klauder, *Continuous-representation theory*, I and II, Journal of Mathematical Physics **4**, pp. 1055–1058, 1963, and E.C.G. Sudarshan, *Equivalence of semiclassical and quantum mechanical descriptions of statistical light beams*, Physical Review Letters **10**, p. 227, 1963. Cited on page 540.

555 R. Kidd, J. Aedini & A. Anton, *Evolution of the modern photon*, American Journal of Physics **57**, pp. 27–35, 1989, Cited on page 541.

556 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 542.

557 N. Bohr & L. Rosenfeld, *Zur Frage der Meßbarkeit der elektromagnetischen Feldgrößen*, Mat.-fys. Medd. Danske Vid. Selsk. **12**, p. 8, 1933. The results were later published in English language as N. Bohr & L. Rosenfeld, *Field and charge measurements in quantum electrodynamics*, Physical Review **78**, pp. 794–798, 1950. Cited on page 543.

558 The quotes on motion are found in chapter VI of F. ENGELS, *Herrn Eugen Dührings Umwälzung der Wissenschaft*, Verlag für fremdsprachliche Literatur, Moskau, 1946. The book is commonly called *Anti-Dühring*. Cited on page 527.

559 E.I. BUTIKOV, *The rigid pendulum – an antique but evergreen physical model*, European Journal of Physics **20**, pp. 429–441, 1999. Cited on page 543.

560 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 547.

561 LOUIS DE BROGLIE, ... Cited on page .

562 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 .

563 P. CARRUTHERS & M.M. NIETO, *Phase and angle variables in quantum mechanics*, Review of Modern Physics **40**, pp. 411–440, 1968. Cited on page .

564 The uncertainty principle for rotational motion is well explained by ... Cited on page .

565 The whole bunch of atoms behaves as one single molecule; one speaks of a Bose-Einstein condensate. The first observations, worth a Nobel prize, were by M.H. ANDERSON *&* al., *Observation of Bose-Einstein condensation in a dilute atomic vapour*, Science **269**, pp. 198–201, 1995, C.C. BRADLEY *&* al., Physical Review Letters **75**, p. 1687, 1995, K.B. DAVIS *&* al., Physical Review Letters **75**, p. 3969, 1995. For a simple introduction, see W. KETTERLE, *Experimental studies of Bose-Einstein condensation*, Physics Today pp. 30–35, December 1999/ Cited on page .

566 R.H. DICKE & J.P. WITTKE, *Introduction to quantum theory*, Addison-Wesley, Reading, Massachusetts, 1960. Cited on page .

567 G. PAPINI, *Shadows of a maximal acceleration*, http://arxiv.org/gr-qc/0211011. Cited on page .

568 Y. AHARONOV & D. BOHM, *Significance of electromagnetic potentials in the quantum theory*, Physical Review **115**, pp. 485–491, 1959. Cited on page .

569 K. VON MEYENN & E. SCHUCKING, *Wolfgang Pauli*, Physics Today pp. 43–48, February 2001. Cited on page .

570 J. NEUKAMMER *&* al., *Spectroscopy of Rydberg atoms at $n \sim 500$*, Physical Review Letters **59**, pp. 2947–2950, 1987. Cited on page 549.

571 This is explained by J.D. HEY, *Mystery error in Gamow's Tompkins reappears*, Physics Today pp. 88–89, May 2001. Cited on page 549.

572 L.L. FOLDY & S.A. WOUTHUYSEN, *On the Dirac theory of spin 1/2 particles and its nonrelativistic limit*, Physical Review **78**, pp. 29–36, 1950. Cited on page .

573 L.L. FOLDY, *The electromagnetic properties of Dirac particles*, Physical Review **83**, pp. 688–693, 1951. L.L. FOLDY, *The electron-neutron interaction*, Physical Review **83**, pp. 693–696, 1951. L.L. FOLDY, *Electron-neutron interaction*, Review of Modern Physics **30**, pp. 471–481, 1952. Cited on page .

574 JOHN P. COSTELLA & BRUCE H.J. MC KELLAR, *The Foldy-Wouthuysen transformation*, American Journal of Physics **63**, pp. 1119–1121, 1995. Cited on page .

575 H. EULER & B. KOCKEL, Naturwissenschaften **23**, p. 246, 1935, H. EULER, Annalen der Physik **26**, p. 398, 1936, W. HEISENBERG & H. EULER, *Folgerung aus der Diracschen*

Theorie des Electrons, Zeitschrift für Physik **98**, pp. 714–722, 1936. Cited on pages 550 and 628.

576 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 551.

577 About tachyons, see ... Cited on page .

578 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 552.

579 J.P. Woerdman, G. Nienhuis & I. Kuscer, *Is it possible to rotate an atom?*, Optics Communications **93**, pp. 135–144, 1992. We are talking about atoms rotating around their centre of mass; atoms can of course rotate around other bodies, as discussed by M.P. Silverman, *Circular birefringence of an atom in uniform rottion: the classical perspective*, American Journal of Physics **58**, pp. 310–317, 1990. Cited on page 553.

580 For the atomic case, see P.L. Gould, G.A. Ruff, D.E. Pritchard, *Diffraction of atoms by light: the near resonant Kapitza-Dirac effect*, Physical Review Letters **56**, pp. 827–830, 1986. Many ealry experimental attempts, in particular those performed in the 1980s, for the diffraction of electrons by light were controversial; most showed only the deflection of electrons, as explained by H. Batelaan, Contemporary Physics **41**, p. 369, 2000. Later on, he and his group performed the newest and most spectacular experiment, demonstrating real diffraction, including interference effects; it is described in D.L. Freimund, K. Aflatooni & H. Batelaan, *Observation of the Kapitza-Dirac effect*, Nature **413**, pp. 142–143, 2001. Cited on page 553.

581 Mark P. Silverman, *And yet it moves: strange systems and subtle questions in physics*, Cambridge University Press 1993. A beautiful book from an expert of motion. Cited on pages 549 and 553.

582 J. Schmiedmayer & al., Physical Review Letters July-August 1995. Cited on page 553.

583 The nearest to an image of a hydrogen atom is found in A. Yazdani, *Watching an atom tunnel*, Nature **409**, pp. 471–472, 2001. The experiments on Bose-Einstein condensation are also candidates for images of hydrogen atoms. The company Hitachi made a fool of itself in 1992 by claiming in a press release that its newest electron microscope could image hydrogen atoms. Cited on page 553.

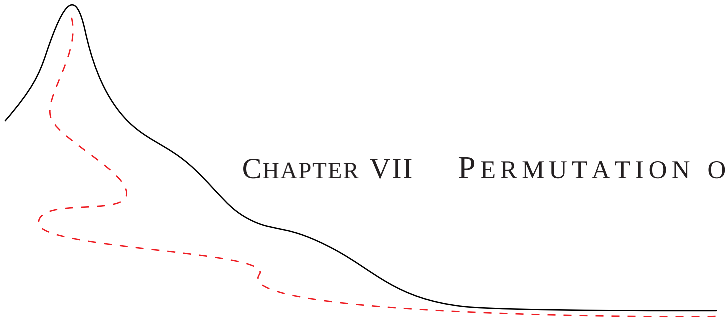
584 An introduction is given by P. Pinkse, G. Rempe, *Wie fängt man ein Atom mit einem Photon?*,Physikalische Blätter **56**, pp. 49–51, 2000. Cited on page 553.

The nearest to an image of a hydrogen atom is found in A. Yazdani, *Watching an atom tunnel*, Nature **409**, pp. 471–472, 2001. The experiments on Bose-Einstein condensation are also candidates for images of hydrogen atoms. The company Hitachi made a fool of itself the world over by claiming in a 1992 press release that its newest electron microscope could image hydrogen atoms.

585 A.M. Wolsky, *Kinetic energy, size, and the uncertainty principle*, American Journal of Physics **42**, pp. 760–763, 1974. Cited on page 553.

586 Wolfgang Pauli, *Exclusion principle and quantum mechanics*, Nobel lecture, 13 december 1946, in *Nobel Lectures, Physics*, Volume 3, 1942-1962, Elsevier, Amsterdam, 1964. Cited on page 554.

Typeset in
January 2003



CHAPTER VII PERMUTATION OF PARTICLES

Why are we able to distinguish twins from each other? Why can we distinguish what looks alike, such as a copy from an original? In everyday life, copies always differ somewhat from originals. But think about any method that allows to distinguish objects: you will find that it runs into trouble for point-like particles. Challenge 985 n Therefore in the quantum domain something must change about our ability to distinguish objects. Let us explore the issue.

20. Are particles like condoms?

Some usually forgotten properties of objects are highlighted by studying a pretty combinatorial puzzle: the *condom problem*. Ref. 587 It asks:

> *How many condoms are necessary if w women and m men want to encounter each other in a hygienical way, so that nobody gets in contact with the body fluids of anybody else?*

The same problem also appears in other settings. For example, it also applies to surgical gloves, doctors and patients, or to computers, interfaces and computer viruses. Nevertheless, the phrasing above is the most common found in the literature. And obviously, the optimal number of condoms or gloves is *not* the product wm. In fact, the problem has three subcases.

- Challenge 986 e The simple case $m = w = 2$ already provides the most important ideas needed.Are you able to find the optimal solution and procedure?
- In the case $w = 1$ and m odd or the case $m = 1$ and w odd, the solution is $(m+1)/2$ condoms.This is the optimal solution, as you can easily check yourself. Challenge 987 e
- Ref. 588 A solution with a simple procedure 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. (However, this formula does not always give the optimal solution; better values exist in certain subcases.) Challenge 988 e

Two basic properties of condoms determine the solution to the puzzle. Firstly, condoms have two sides, an interior and an exterior one. Secondly, condoms can be distinguished from each other. Do these two properties also apply to particles? We will discuss the issue of double-sidedness in the third part of the mountain ascent. See page 713 In fact, the question whether particles can be turned inside out will be of great importance for their description and their motion. In the present chapter we concentrate on the second issue, namely whether objects and particles can always be distinguished. We will find that *elementary* particles do not behave like condoms but in an even more surprising manner.

In everyday life, distinction of objects can be achieved in two ways. We are able to distinguish objects – or people – from each other because they differ in their *intrinsic properties*, such as their mass, colour, size or shape. In addition, we are also able to distinguish objects if they have the *same* intrinsic properties. Any game of billiard suggests that by following the path of each ball, we can distinguish it from the others. In short, objects with identical properties can also be distinguished using their *state*.

The state of a billiard ball is given by its position and momentum. In the case of billiard balls, the state allows distinction 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 the case. First of all, atoms or other microscopic particles of the same type have the same intrinsic properties. To distinguish them in collisions, we would need to keep track of their motion. But we have no chance to achieve this. Already in the nineteenth century it was shown experimentally that even nature itself is not able to do this. This result was discovered studying systems which incorporate a large number of colliding atoms of the same type: *gases*.

The calculation of the entropy of a simple gas, made of N simple particles of mass m moving in a volume V, gives

See page 178

$$S = k \ln \left[V \left(\frac{mkT}{2\pi\hbar^2} \right)^{3/2} \right]^N + \frac{3}{2} kN + k \ln \alpha \qquad (425)$$

where k is the Boltzmann constant and T the temperature. In this formula, the quantity α is equal to 1 if the particles are distinguishable, and equal to $1/N!$ if they are not. Measuring the entropy thus allows us to determine α and therefore whether particles are distinguishable. It turns out that only the second case describes nature. This can be checked with a simple test:only in the second case does the entropy of two volumes of identical gas *add* up.[*] The result, often called *Gibbs' paradox*,[**] thus proves that the microscopic components of matter are *indistinguishable*: in a system of microscopic particles, there is no way to say which particle is which. Indistinguishability is an experimental property of nature.[***]

Challenge 989 e

Ref. 589

The properties of matter would be completely different without indistinguishability. For example, we will discover that without it, knifes and swords would not cut. In addition, the soil would not carry us; we would fall right through it. To illuminate the issue in more detail, we explore the next question.

Challenge 990

[*] Indeed, the entropy values observed by experiment are given by the so-calledSakur-Tetrode formula

$$S = kN \ln \left[\frac{V}{N} \left(\frac{mkT}{2\pi\hbar^2} \right)^{3/2} \right] + \frac{5}{2} kN \qquad (426)$$

which follows when $\alpha = 1/N!$ is inserted above.

[**] Josiah Willard Gibbs (1839–1903), US-American physicist who was, with Maxwell and Planck, 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 of atoms, as decay seems to single out certain atoms compared to others. But quantum theory then showed that this is not the case and that atoms remain indistinguishable.

Why does indistinguishability appear in nature?

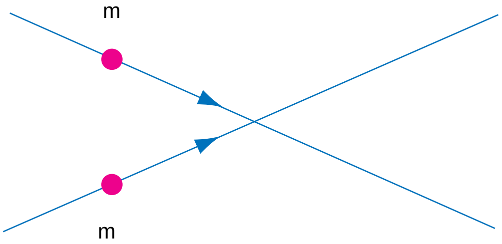


Figure 211 Identical objects with crossing paths

Take two microscopic particles with the same mass, the same composition, and the same shape, such as two atoms. Imagine that their paths cross, and that they approach each other to small distances at the crossing, as shown in Figure 211. Both the collision of atoms in a gas or a near miss are examples. Experiments show that at small distances the two particles can switch roles, without anybody being able to avoid it. This is the basic process that makes it *impossible* in a gas to follow particles moving around and to determine which particle is which.

For a path approaching particles to small distances a role switch requires only a small amount of change, i.e. only a small (physical) action. However, we know that there is a smallest observable action in nature. Keeping track of each particles at small distances would require action values *smaller* than the minimal action observed in nature. The existence of a smallest action makes it thus impossible to keep track of microscopic particles when they come too near to each other.

In short, indistinguishability is a consequence of the existence of a minimal action in nature. This result leads straightaway to the next question:

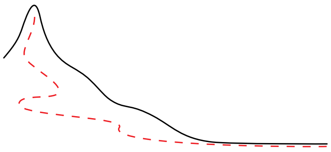
Can particles be counted?

In everyday life, objects can be counted because they can be distinguished. Since quantum particles cannot be distinguished, we need some care in determining how to count them. The first step is the definition of what is meant by a situation without any particle at all. This seems an easy thing to do, but later on we will encounter situations where already this step runs into difficulties. In any case, the first step is thus the *specification of the vacuum*. Any counting method requires that situations with particles be clearly separated from situations without particles.

The second step is the specification 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.* Counting is then performed by measuring the total charge and dividing by the unit charge.

This method has several advantages. First of all, it is not important whether particles are distinguishable or not; it works in either case. Secondly, virtual particles are not counted. This is a welcome state of affairs, as we will see, because for virtual particles, i.e. for particles for which $E^2 \neq p^2c^2 + m^2c^4$, there is *no way* to define a particle number anyway.

∗ In everyday life, the weight or mass is commonly used as observable. Obviously it cannot be used in the quantum domain, except for simple cases. Can you give at least two reasons, one from special relativity, and one from general relativity? Challenge 992 n



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

The other side is that antiparticles count negatively! Also this consequence is a result of the quantum of action. We saw above that the quantum of action implies that even in vacuum, particle-antiparticle pairs are observed at sufficiently high energies. As a result, an antiparticle must count as minus one particle. In other words, any way of counting particles can produce an error due to this effect. In everyday life this limitation plays no role, as there is no antimatter around us. The issue does play a role at higher energies, however. It turns out that there is no general way to count the exact number of particles and antiparticles at the same time. In short, quantum theory shows that particle counting is never perfect.

In summary, nature does provide a way to count particles even if they cannot be distinguished, though only for everyday, low energy conditions; due to the quantum of action, antiparticles count negatively, and provide a limit to the counting of particles at high energies.

What is permutation symmetry?

Since particles are indistinguishable but countable, there exists a symmetry of nature for systems composed of several identical particles. *Permutation symmetry*, also called *exchange symmetry*, is the property of nature that observations are unchanged under exchange of identical particles. Like space-time symmetry, gauge symmetry, and the not yet encountered renormalization symmetry, permutation symmetry forms one of the four pillars of quantum theory. Permutation symmetry is a property of *composed* systems, i.e. of systems made of many (identical) subsystems. Only for such systems indistinguishability plays a role.

In other words, 'indistinguishable' is not the same as 'identical'. Two particles are not the *same*; they are more like *copies* of each other. On the other hand, everyday life experience shows us that two copies can always be distinguished under close inspection, so that the term is not fully appropriate either. In the microscopic domain, particles are countable and completely indistinguishable.* Particles are perfect copies of each other.

We will discover shortly that permutation is partial rotation. Permutation symmetry thus is a symmetry under partial rotations. Can you find out why? Challenge 993 e

Indistinguishability and symmetry

The indistinguishability of particles leads to important conclusions about the description of their state of motion. This happens because it is impossible to formulate a description of motion that includes indistinguishability right from the start. Are you able to confirm this? Challenge 994 n As a consequence we describe a n-particle state with a state $\Psi_{1...i...j...n}$ which assumes that distinction is possible, as expressed by the ordered indices in the notation, and we introduce the indistinguishability afterwards. Indistinguishability means that the exchange of any two particles describes the same physical system.** Now, two quantum states have the same

---

* The word 'indistinguishable' is so long that many physicists sloppily speak of 'identical' particles nevertheless. Take care.

** 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.

physical properties if they differ at most by a phase factor; indistinguishability thus requires

$$\Psi_{1\ldots i\ldots j\ldots n} = e^{i\alpha}\, \Psi_{1\ldots j\ldots i\ldots n} \qquad (427)$$

for some unknown angle α. Applying this expression twice, by exchanging the same couple of indices again, allows us to conclude that $e^{2i\alpha} = 1$. This implies that

$$\Psi_{1\ldots i\ldots j\ldots n} = \pm \Psi_{1\ldots j\ldots i\ldots n} \quad , \qquad (428)$$

in other words, a wavefunction is either *symmetric* or *antisymmetric* under exchange of indices. Quantum theory thus 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*.**

Experiments show that the behaviour depends on the *type* of particle. Photons are bosons. On the other hand, electrons, protons and neutrons are found to be fermions. Also about half of the atoms are found to behave as bosons (at moderate energies). In fact, a composite of an *even* number of fermions (at moderate energies) – or of any number of bosons (at any energy) – turns out to be a boson; a composite of an *odd* number of fermions is (always) a fermion. For example, almost all of the known molecules are bosons (electronically speaking). Fermionic molecules are rather special and even have a special name in chemistry; they are called *radicals* and are known for their eagerness to react and to form normal bosonic molecules. Inside the human body, too many radicals can have adverse effects on health; it is well known that vitamin C is important because it is effective in reducing the number of radicals.

To which class of particles do mountains, trees, people, and all other macroscopic objects belong? Challenge 995

The behaviour of photons

A simple experiment allows to determine the behaviour of photons. Take a source that emits two photons of identical frequency and polarization at the same time, as shown in figure 212. In the laboratory, such a source can be realized with a down-converter, a material that converts a photon of frequency 2ω into two photons of frequency ω. Both photons, after having travelled exactly the same distance, are made to enter the two sides of a beam splitter. At the two exits of the beam splitter are two detectors. Experiments show that both photons are always detected together on the *same* side, and never separately on opposite sides. This result shows that photons are bosons. Fermions behave in exactly the opposite Ref. 590

---

[*] This conclusion applies to three-dimensional space only. In two dimensions there are more possibilities.

[**] The term 'fermion' 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 mainly worked on nuclear and elementary particle physics, on spin and on statistics. For his experimental work he was called 'quantum engineer'. He is also famous for his lectures, which are still published in his own hand-writing, and his beautiful approach to physical problems. 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 incorrect.

'Bosons' are named after the Indian physicist Satyenra Nath Bose (1894, Calcutta–1974, Calcutta) who first described the statistical properties of photons, later expanding his work in collaboration with Albert Einstein.

Figure will be added in the future

Figure 212 Two photon interference

way; two fermions are always detected separately on opposite sides, never together on the same side.

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, we deduce that they do not form two separate entities, but 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 impossibility has been checked experimentally with all elementary particles, with nuclei, with atoms, and with numerous molecules.

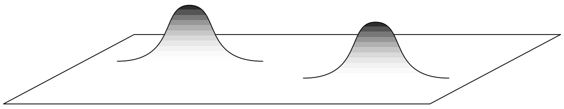


Figure 213 Particles as localized excitations

How does this fit with everyday life, i.e. with classical physics? Photons do not worry us much here. Let us focus the discussion on matter particles. We know to be able to distinguish electrons by pointing to the wire in which they flow, and we can distinguish our fridge from that of our neighbour. While the quantum of action makes distinction impossible, everyday life allows it. The simplest explanation is to imagine a microscopic particle, especially an elementary one, as a bulge, i.e. as a localized excitation of the vacuum. Figure 213 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; we cannot say any more which is which.

The bulge image shows that either for large distances or for high potential walls separating them, distinction of identical particles does become possible. In such situations, measurements allowing to track them independently do exist. In other words, we can specify a limit energy at which permutation symmetry of objects or particles separated by a distance
Challenge 996 d becomes important. It is given by

$$E = \frac{c\,\hbar}{d} \quad . \tag{429}$$

Challenge 997 Are you able to confirm the expression?For example, at everyday temperatures we *can* distinguish atoms inside a solid from each other, since the energy so calculated is much higher than the thermal energy of atoms. To have fun, you might want to determine at what energy two truly identical human twins become indistinguishable. Estimating at what energies the statistical character of trees or fridges will become apparent is then straightforward. Challenge 998

The bulge image of particles thus purveys the idea that distinguishability exists for objects in everyday life but not for particles in the microscopic domain. To sum up, in daily life we are able to distinguish objects and thus people for two reasons: because they are made of *many* parts, and because we live in a *low energy* environment.

The energy issue immediately adds a new aspect to the discussion. How can we describe fermions and bosons in the presence of virtual particles and of antiparticles?

Indistinguishability in quantum field theory

Quantum field theory, as we will see shortly, simply puts the bulge idea of Figure 213 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 each particle is an excitation of the same basic substrate and each excitation has the same properties. A situation with one particle is then described by a vacuum state acted upon by a *creation operator*. Adding a second particle is described by adding a second 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 must behave to describe observations.* It arrives at the following conclusions:

- Fields with half-integer spin are fermions and imply (local) anticommutation.
- Fields with integer spin are bosons and imply (local) commutation.
- For all fields at spacelike separations, the commutator – respectively anticommutator – vanishes.
- 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. They describe how particles are identical. But why are they? 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 mountain ascent.

---

* Whenever the relation

$$[b,b^\dagger] = bb^\dagger - b^\dagger b = 1 \tag{430}$$

holds between the creation operator $b^\dagger$ and the annihilation operator b, the operators describe a *boson*. If the operators for particle creation and annihilation anticommute

$$\{d,d^\dagger\} = dd^\dagger + d^\dagger d = 1 \tag{431}$$

they describe a *fermion*. The so defined bracket is called the *anticommutator bracket*.

How accurately is permutation symmetry verified?

A simple but effective experiment testing the fermion behaviour of electrons was carried out by Ramberg and Snow. Ref. 591 They sent an electric current of 30 A through a copper wire for one month and looked for X-ray emission. They did not find any. They concluded that electrons are always in an antisymmetric state, with a symmetric component of less than

$$2 \cdot 10^{-26} \tag{432}$$

of the total state. Electrons are thus fermions.

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 energy level of a copper atom, leading to X-ray emission. The lack of such X-rays implies that electrons are fermions to a very high accuracy. In particular, two electrons cannot be in the same state; this is the so-called Pauli exclusion principle, our next topic.

Copies, clones, and condoms

Can classical systems be indistinguishable? They can: large molecules are examples – provided they are made of exactly the same isotopes. Can *large* classical systems, made of a mole or more particles be indistinguishable? This simple question effectively asks whether a perfect copy, or (physical) *clone* of a system is possible.

It could be argued that any factory for mass-produced goods, such as one producing shirt buttons or paper clips, shows that copies are possible. But the appearance is deceiving. In 1982, the Dutch physicist Dennis Dieks and the US-American physicists Wootters and Zurek published simple proofs that quantum systems cannot be copied. Ref. 593 A copying machine is a machine which takes an original, leaves it unchanged, and produces a copy. However, we know that if we extract information from an original, we have to interact with it. As a result, the system will change at least by the quantum of action. Copies and originals can never be identical.

Simply stated, if a copying machine would be able to copy originals either in state $|A>$ or in state $|B>$, it could not decide whether the state $|A+B>$ should be copied into $|A+B>$ or into $|A>+|B>$, as both results must apply for such a machine, which is necessarily linear.* Another way to put the result: if we could clone systems, we would be able to measure a variable on one system, and a second variable on the copy at the same time. If we took two conjugate variables, we would be thus able to beat the indeterminacy relation. This is impossible. Copies are always imperfect.

Other researchers then explored how perfect a copy can be, especially in the case of classical systems. Ref. 594 To make a long story short, these investigations show that also the copying or cloning of macroscopic systems is impossible. In simple words, copy machines do not exist. Copies can always be distinguished from originals if observations are made with

---

* This no-cloning theorem puts severe limitations on quantum computers, as computations often need copies of intermediate results. It also shows that faster-than-light communication is impossible in EPR experiments. In compensation, quantum cryptography becomes possible – at least in the laboratory –, as it is impossible to copy a result without being noticed.

sufficient care. In particular, this is the case for biological clones; biological clones are identical twins born following separate pregnancies. They differ in their finger prints, iris scans, physical and emotional memories, brain structures and in many other aspects. (Can you specify a few more?) Biological clones, like identical twins, are not copies of each other. Challenge 999 n

The lack of quantum mechanical copying machines is disappointing. Such machines, or teleportation devices, could be fed with two different inputs, such as a lion and a goat, and produced a superposition: a chimaera. Quantum theory shows that all these imaginary beings cannot be realized.

In summary, everyday life objects such as photocopies, billiard balls or twins are always distinguishable. There are two reasons: firstly, quantum effects play no role in everyday life, so that there is no danger of unobservable exchange; secondly, perfect clones of classical systems do not exist anyway, so that there always are tiny differences between any two objects, even if they look identical at first sight. Condoms can always be distinguished.

21. Rotations and statistics – visualizing spin

We saw above that spin is the observation that matter rays, like light rays, can be polarized. Spin thus describes how particles behave under rotations, and it proves that particles are not simple spheres shrunk to points. We also saw that spin describes a fundamental difference between quantum systems and condoms: spin specifies the indistinguishability of quantum systems. Let us explore this connection in more detail.

The general background for the appearance of spin was elucidated by Eugene Wigner in 1939.[*] He started by recapitulating that any quantum mechanical particle, if elementary, must behave like an irreducible representation of the set of all viewpoint changes. This set forms the symmetry group of flat space-time, the so-called *inhomogeneous Lorentz group*. We have seen in the chapter on classical mechanics how this connection between elementarity and irreducibility arises. To be of physical relevance for quantum theory, representations have to be *unitary*. The full list of irreducible unitary representations of viewpoint changes thus provides the range of possibilities for any particle that wants to be *elementary*. Ref. 595 See page 147

Cataloguing the possibilities, one finds first of all that every elementary particle is described by four-momentum – no news so far – and by an internal angular momentum, the *spin*. Four-momentum results from the translation symmetry of nature, and spin from its rotation symmetry. The momentum value describes how a particle behaves under translation, i.e. under position and time shift of viewpoints. The spin value describes how an object behaves under rotations in three dimensions, i.e. under orientation change of viewpoints.[**] As is well known, the magnitude of four-momentum is an invariant property, given by the mass, whereas its orientation in space-time is free. Similarly, the magnitude of spin is an in-

---

[*] Eugene Wigner (1902, Budapest–1995), Hungarian-american theoretical physicist, Nobel prize for physics in 1993. He wrote over 500 papers, many about symmetry in physics. He was also famous for being the most polite physicist in the world.

[**] The group of physical rotations is also called SO(3), since mathematically it is described by the group of **S**pecial **O**rthogonal **3** by 3 matrices.

variant property, and its orientation has various possibilities with respect to the direction of motion. In particular, the spin of massive particles behaves differently from that of massless particles.

For *massive* particles, the inhomogeneous Lorentz group implies that the invariant magnitude of spin is $\sqrt{J(J+1)}\hbar$, often simply written J. Since the value specifies the magnitude of the angular momentum, it gives the representation under rotations of a given particle type. The spin magnitude 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. Experiments show that electrons, protons and neutrons have spin $1/2$, the W and Z particles spin 1, and helium atoms spin 0. In addition, the representation of spin J is $2J+1$ dimensional, meaning that the spatial orientation of the spin has $2J+1$ possible values. For electrons there are thus two possibilities; they are usually 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. It can take the same values as in the massive case. For example, photons and gluons have spin 1. For massless particles, the 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. Massless particles can have positive or negative helicity, often also called right-handed and left-handed. There is no other freedom for the orientation of spin in the massless case.

The symmetry investigations lead to the classification of particles by their mass, their momentum, and their spin. To complete the list, the remaining symmetries must be included. These are motion inversion parity, spatial parity and charge inversion parity. Since these symmetries are parities, each elementary particle has to be described by three additional numbers, called T, C, an P, each of which can take values of either $+1$ or -1. Being parities, they must be *multiplied* to yield the value for a composed system.

See page 889

A list of the values observed for all elementary particles in nature is given in Appendix C. Spin and parities together are called *quantum numbers*. As we will discover later on, additional interaction symmetries will lead to additional quantum numbers. But let us return to spin.

The main result is that spin 1/2 is a possibility in nature, even though it does not appear in everyday life. Spin 1/2 means that only a rotation of 720 degrees is equivalent to one of 0 degrees, while one of 360 degrees is not, as explained in Table 473. The mathematician Hermann Weyl used a simple image explaining this connection.

See page 42

Take two cones, touching each other at their tips as well as 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.*

To sum up, the 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,

---

* If one imagines the cone angle to vary continuously, this image also shows that a 720 degree rotation can be continuously deformed into a 0 degree one, whereas a 360 degree rotation cannot.

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. 597

Table 46 Irreducible representations of the rotation group

and all bosons have integer spin? 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.

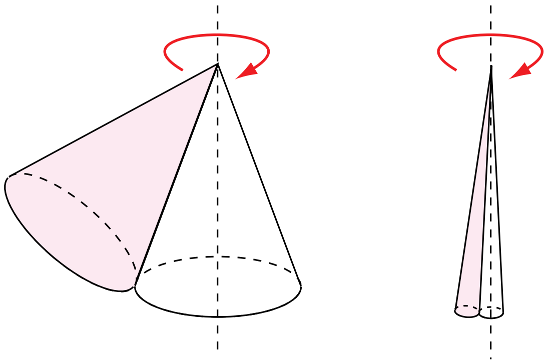


Figure 214 An argument showing why rotations by 4π are equivalent to no rotation at all

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 were hidden. It took quite some years to convince everybody that a simple observation about belts was the central part of the proof. Ref. 596

The belt trick

The well-known *belt trick* was often used by Dirac to explain the features of spin 1/2. Taking Figure 213, which models particles as indistinguishable excitations, it is not difficult to imagine a sort of sheet Ref. 598

connecting them, similar to a belt connecting the two parts of the buckle, as shown in Figure 215. 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 1000 e

In addition, 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 we take 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.

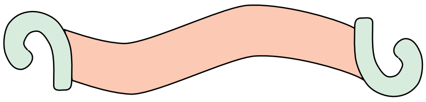


Figure 215 A belt visualizing two spin 1/2 particles

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.

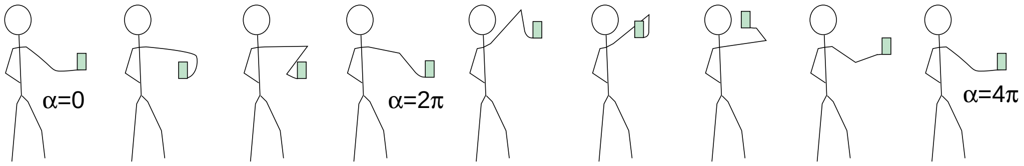


Figure 216 The human arm as spin 1/2 model

The trick is even more impressive when many tails are used. In fact, there are two ways to do this. One way is connect two buckles with *many* bands or threads, like in Figure 217. Both a rotation by 2π of one end or an exchange of the two ends produces quite a tangle; nevertheless, in both cases a second operation leads back to the original situation.

There is a second, 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 of a human being, 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, as shown in Figure 218, all twists and tangles can be made to disappear, without moving or turning the object. You really have to experience this in order to believe it. And the trick really works with *any* number of bands glued to the object.

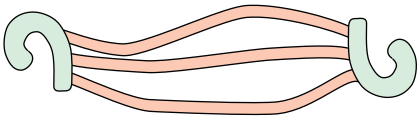


Figure 217 Another model for two spin 1/2 particles

Even more astonishing is the other half of the experiment. Take *two* particles as shown in the left of Figure 218. If you exchange the positions of two such spin 1/2 particles, always keeping the ends at infinity fixed, a tangled mess is created. But incredibly, if you exchange the objects a second time, everything untangles neatly, independently of the number of attached strings. You might want to test yourself that the behaviour is still the same with sets of three or more particles. Challenge 1001 e

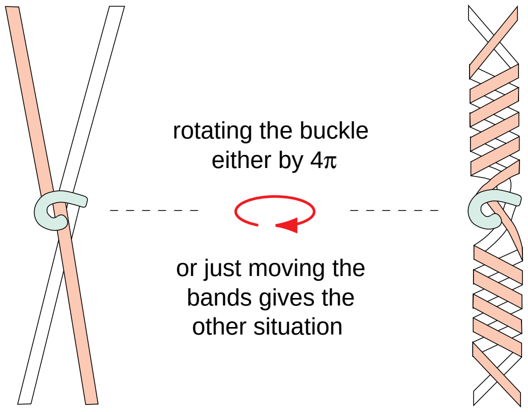


Figure 218 The extended belt trick, modelling a spin 1/2 particle: the two situations can be transformed into each other either by rotating the central object by 4π or by keeping the central object fixed and moving the bands around it

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 behaviour of the configuration space of particles. These investigations result in the following statements: Ref. 599

- ▷ *Objects of spin 1/2 are fermions.*[*]
- ▷ *Exchange and rotation of spin 1/2 particles are similar processes.*

Note that all these arguments require three dimensions, because there are no tangles (or knots) in fewer dimensions.[**] And indeed, spin exists only in three or more spatial dimensions.

Pauli's exclusion principle and the hardness of matter

Why are we able to knock on a door? Why can stones not fly through tree trunks? How does the mountain we are walking on carry us? In classical physics, we avoided this issue, by taking solidity as a defining property of matter. But doing so, we cheated: we have seen that matter consists mainly of empty space, so that we have to study the issue without any sneaky way out. The answer is not clear: penetration is made impossible by Pauli's exclusion principle between the electrons inside atoms.

---

[*] 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*. We will introduce it later on. Challenge 1002 e

[**] Of course, knots and tangles do exist in higher dimensions. Instead of considering knotted one-dimensional lines, one can consider knotted planes or knotted higher-dimensional hyperplanes. For example, deformable planes can be knotted in four dimensions and deformable 3-spaces in five dimensions.

Why do electrons and other fermions obey the Pauli exclusion principle? The answer can be given with a beautifully simple argument. We know that exchanging two fermions produces a minus sign. Imagine these two fermions being, as a classical physicist would say, located at the same spot, or as a quantum physicist would say, in the same state. If that could be possible, an exchange would change nothing in the system. But an exchange of fermions must produce a minus sign for the total state. Both possibilities – no change at all as well as a minus sign – cannot be realized at the same time. There is only one way out: two fermions must avoid to ever be in the same state. Ref. 600

The exclusion principle is the reason that two pieces of matter in everyday life cannot penetrate each other, but have to repel each other. For example, bells only work because of the exclusion principle. Bells would not work if the colliding pieces that produce the sound would interpenetrate. But in any example of two interpenetrating pieces the electrons in the atoms would have to be in similar states. This is forbidden. For the same reason we do not fall through the floor, even though gravity pulls us down, but remain on the surface. In other words, the exclusion principle implies that matter cannot be compressed indefinitely, as at a certain stage an effective Pauli pressure appears, so that a compression limit ensues. For this reason for example, planets or neutron stars do not collapse under their own gravity.

The exclusion principle also answers the question about how many angels can dance on the top of a pin. (Note that angels must be made of fermions, as you might want to deduce from the information known about them.) Challenge 1003 n Both theory and experiment confirm the answer already given by Thomas Aquinas in the middle ages: only one. Ref. 592 The fermion exclusion principle could also be called 'angel exclusion principle'. To stay in the topic, the principle also shows that *ghosts* cannot be objects, as ghosts are supposed to be able to traverse walls.

Whatever the interpretation, the exclusion principle keeps things in shape; without it, there would be no three-dimensional objects. Only the exclusion principle keeps the cloudy atoms of nature from merging, holding them apart. Shapes are a direct consequence of the exclusion principle.

As a result, when we knock on a table or on a door, we show that both objects are made of fermions.

Since permutation properties and spin properties of fermions are so well described by the belt model, we could be led to the conclusion that these properties might really be 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. In the third part of this walk we will discover whether this idea is correct.

So far, we have only considered spin $1/2$ particles. We will not talk much about systems with odd spin of higher value, such as $3/2$ or $5/2$. Such systems can be seen as being composed of spin $1/2$ entities. Can you confirm this? Challenge 1004

We did not talk about lower spins than $1/2$ either. A famous theorem states that a positive spin value below $1/2$ is impossible, because the largest angle that can be measured in three dimensions is 4π. There is no way to measure a larger angle;[*] The quantum of action makes this impossible. Thus there cannot be any spin value between 0 and $1/2$.

---

[*] This is possible in two dimensions though.

Integer spin

Under rotations, integer spin particles behave differently from half-integer particles. Integer spin particles 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 219. Exploring their properties in the same way as above, we arrive 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.*

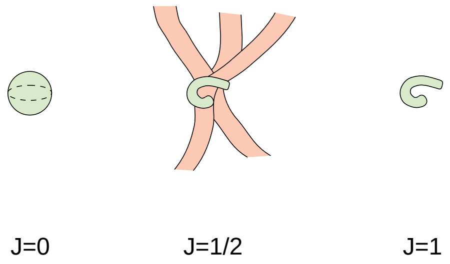


Figure 219 Some visualizations of spin representations

You might prove by yourself that this suffices to show that Challenge 1005

> ▷ *Composites of bosons, as well as composites of an even number of fermions (at low energy), 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.

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 as part of angular momentum, is described like angular momentum, and has a name synonymous with angular momentum. Despite all this, for many decades a strange myth was spread in physics courses and textbooks around the world, namely that spin $1/2$ is *not* a rotation about an axis. The myth maintains that any rotating object must have integer spin. Since half integer spin is not possible in classical physics, it is argued that such spin is not due to rotation. It is time to finish with this example of muddled thinking.

Electrons do have spin $1/2$, and are charged. Electrons and all other charged particles with spin $1/2$ do have magnetic momentum.[**] Magnetic momentum is expected for any rotating charge. In other words, spin $1/2$ does behave like rotation. However, assuming that a particle consists of a continuous charge distribution in rotational motion gives the wrong value for the magnetic momentum. In the early days of the twentieth century, when physicists were still thinking in classical terms, they concluded that spin $1/2$ particles thus cannot be rotating. This myth has survived through many textbooks. The correct deduction

---

Challenge 1006 [*] This sentence implies that spin 1 and higher can also be achieved *with* tails; can you find such a representation?

Note that composite fermions can be bosons only up to that energy at which the composition breaks down. Otherwise, by packing fermions into bosons, we could have fermions in the same state.

Challenge 1007 [**] This can easily be measured in a an experiment; however, not one of the Stern-Gerlach type. Why?

though is that the assumption of continuous charge distribution is wrong. Indeed, charge is quantized; nobody today expects that elementary charge is distributed in space, as that would contradict its quantization.

Let us remember what rotation is. Both the belt trick for spin 1/2 as well as the integer spin case remind us: 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 body continuously exchanging the positions of its parts. Rotation and exchange are the same.

Above we found that spin is exchange behaviour. Since rotation is exchange and spin is exchange, it follows that spin *is* rotation. Since we deduced, like Wigner, spin from rotation invariance, this consequence is not a surprise.

The belt model of a spin 1/2 particle tells us that such a particle can rotate continuously without any hindrance.In short, we are allowed to maintain that spin is rotation about an axis, without any contradiction to observations, even for spin 1/2. The belt model helps to keep two things in mind: we must assume that in the belt model only the buckles can be observed and do interact, not the belts, and we must assume that elementary charge is pointlike and cannot be distributed.*

Challenge 1008

Ref. 601

Why is fencing with laser beams impossible?

When a sword is approaching dangerously, we 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. Fermions make matter impenetrable. On the other hand, photons are bosons. Bosons can be in the same state; they allow penetration. Matter is impenetrable because at the fundamental level it is composed of fermions. Radiation is composed of bosons. The distinction between fermions and bosons thus explains why objects can be touched while images cannot. In the first part of our mountain ascent we started by noting this difference; now we know its origin.

See page 60

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 requires antimatter. Taking these three statements together, the conclusion of the title is not surprising any more. Interestingly, there is a simple argument making the same point directly, without any help of quantum theory, when the belt model is extended from space alone to full *space-time*.

See page 226

---

* Obviously, the detailed structure of the electron still remains unclear at this point. Any angular momentum S is given classically by $S = \Theta\omega$; however, neither the moment of inertia Θ, connected to the rotation radius and electron mass, nor the angular velocity ω are known at this point. We have to wait quite a while, until the third part of our adventure, to find out more.

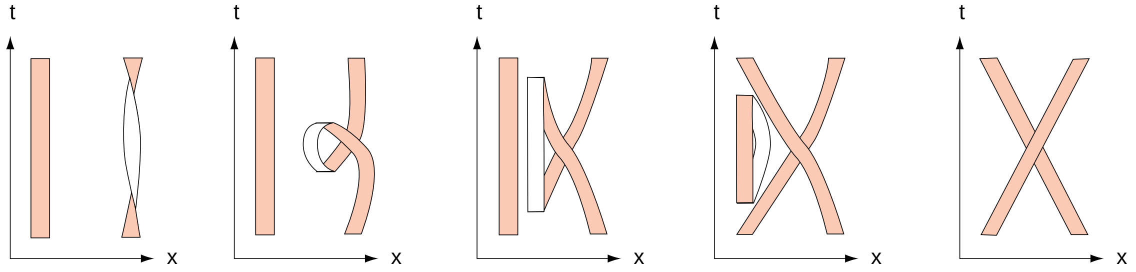


Figure 221 Belts in space-time: rotation and antiparticles

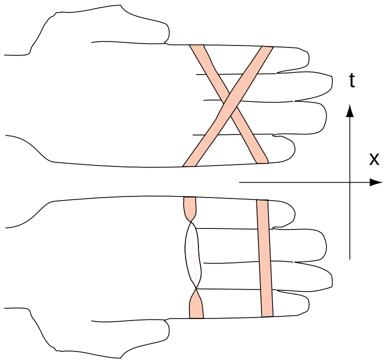


Figure 220 Equivalence of exchange and rotation in space-time

To learn how to think in space-time, let us take a particle spin 1, i.e. a particle looking like a detached belt buckle in three dimensions. When moving in a 2+1 dimensional space-time, it is described by a ribbon.Playing around with ribbons in space-time, instead of belts in space, provides many interesting conclusions. For example, Figure 220 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.* Both sides of the hand transform the same initial condition, at one border of the hand, to the same final condition at the other border. We have thus successfully extended a known result from space to space-time. Interestingly, we can also find a smooth sequence of steps realizing this equivalence. Challenge 1009

When particles in space-time are described as ribbons, Figure 221 shows the intermediate steps allowing to identify a rotation with an exchange. The sequence requires the use of a particle-antiparticle pair. Without antiparticles, the equivalence of rotation and exchange would not hold in space-time. Rotation in space-time indeed requires antiparticles.

Limits and open questions

The topic of statistics is an important research field in theoretical and experimental physics. In particular, researchers have searched and still are searching for generalizations of the possible exchange behaviours of particles.

In two spatial dimensions, the result of an exchange of the wavefunction is not described by a sign, but by a continuous phase. Tow-dimensional objects behaving in this way, 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

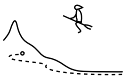
---

∗ Obviously, the next step would be to check the full spin 1/2 model of Figure 218 in four-dimensional space-time. But thisis not an easy task; there is no generally accepted solution yet. Challenge 1010 fractional quantum Hall effect, perhaps the most interesting discovery of modern experimental physics, has pushed anyons onto the stage of modern research. See page 618

Other theorists generalized the concept of fermions in other ways, introducing parafermions, parabosons, plektons and other hypothetical concepts. O.W. Greenberg has spent most of his professional life on this issue. His conclusion is that in $3+1$ space-time dimensions, only fermions and bosons exist. (Can you show that this implies that the ghosts appearing in scottish tales do not exist?) Ref. 602 Challenge 1011 n

From a different viewpoint, the above belt model invites 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, which states that particles, especially at high energies, are not point-like, but extended entities.

Still another generalization of statistical behaviour at high energies is the concept of quantum group, which we will encounter later on. In all of these cases, the quest is to understand what happens to permutation symmetry in a unified theory of nature. A glimpse of the difficulties appears already above: how can Figures 213, 218 and 221 be reconciled? We will settle this issue in the third part of our mountain ascent.



References

587 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 559.

588 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 559.

589 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 560.

590 C.K. HONG, Z.Y. OU & L. MANDEL, *Measurement of subpicosecond time intervals between two photons by interference*, Physical Review Letters **59**, pp. 2044–2046, 1987. Cited on page 564.

591 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 567.

592 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 573.

593 The original no-cloning theorem is by D. Dieks, *Communication by EPR devices*, Physics Letters A **92**, pp. 271–272, 1982, and by W.K. Wootters & W.H. Zurek, *A single quantum cannot be cloned*, Nature **299**, pp. 802–803, 1982. For a discussion of photon and multiparticle cloning, see N. Gisin & S. Massar, *Optimal quantum cloning machines*, Physics Review Letters **79**, pp. 2153–2156, 1997. The whole topic has been presented in detail by V. Buzek & M. Hillery, *Quantum cloning*, Physics World **14**, pp. 25–29, November 2001. Cited on page 567.

594 The most recent experimental and theoretical results on physical cloning are described in A. Lamas-Linares, C. Simon, J.C. Howell & D. Bouwmeester, *Experimental quantum cloning of single photons*, Science **296**, pp. 712 – 714, 2002, D. Collins & S. Popescu, *A classical analogue of entanglement*, preprint arXiv:quant-ph/0107082, 2001, and A. Daffertshofer, A.R. Plastino & A. Plastino, *Classical no-cloning theorem*, Physical Review Letters **88**, p. 210601, 2002. Cited on page 567.

595 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 him the Nobel Prize for physics. Cited on page 568.

596 Wolfgang Pauli, ..., ... Cited on page 570.

597 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 570.

598 This famous explanation is given *e.g.* in the book by C.W. Misner, K.S. Thorne & J.A. Wheeler, *Gravitation*, Freeman, 1973. It is called the *scissor trick* on page 43 of volume 1 of R. Penrose & W. Rindler, *Spinors and Spacetime*, 1984. It is also cited and discussed by R. Gould, *Answer to question #7*, American Journal of Physics **63**, p. 109, 1995. Cited on page 570.

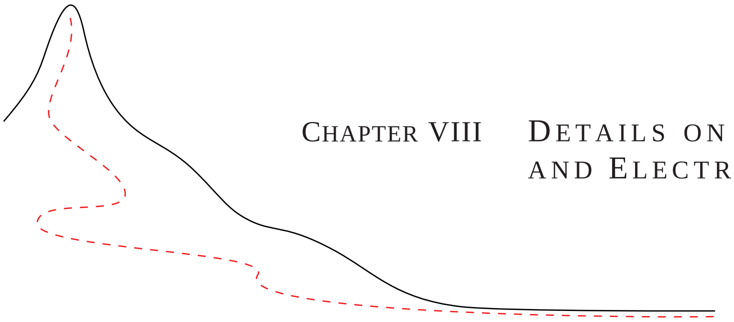
599 M.V. Berry & J.M. Robbins, *Indistinguishability for quantum particles: spin, statistics, and the geometric phase*, Proceedings of the Royal Society in London A **453**, pp. 1771–1790, 1997. See also the comments to this result by J. Twamley, *Statistics given a spin*, Nature **389**, pp. 127–128, 11 September 1997. Their newer results are M.V. Berry & J.M. Robbins, *Quantum indistinguishability: alternative constructions of the transported basis*, Journal of Physics A (Letters) **33**, pp. *L*207–*L*214, 2000, and M.V. Berry & J.M. Robbins, in *Spin-statistics*, eds. R. Hilborn & G. Tino, (American Institute of Physics), 2000, pp. 3–15. See also Michael Berry's home page on the http://www.phy.bris.ac.uk/staff/berry_mv.html web site. Cited on page 572.

600 R.W. Hartung, *Pauli principle in Euclidean geometry*, American Journal of Physics **47**, pp. 900–910, 1979. Cited on page 573.

601 The point that spin can be seen as a rotation was already 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 Sowjetunion **12**, p. 466, 1937. Cited on page 575.

602 Generalizations of bosons and fermions are reviewed in the (serious!) paper by O.W. Greenberg, D.M. Greenberger & T.V. Greenbergest, *(Para)bosons, (para)fermions, quons and other beasts in the menagerie of particle statistics*, preprint http://arxiv.org/abs/hep-th/9306225, 1993. A newer summary is O.W. Greenberg, *Theories of violation of statistics*, preprint http://arxiv.org/abs/hep-th/0007054, 2000. Cited on page 577.

Typeset in
January 2003



CHAPTER VIII DETAILS ON QUANTUM THEORY AND ELECTROMAGNETISM

22. Superpositions and probabilities – quantum theory without ideology

> 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 physical issue 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 who follow 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 since quantum theory appeared.

Why are people either dead or alive?

The evolution equation of quantum mechanics is linear in the wavefunction; thus we 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

---

* Niels Bohr (1885, Copenhagen–1962) made Copenhagen university into one of the centres 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 flee 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.

would it evolve in time? Similarly, we 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.

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 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 physical action 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.

no figure yet

Figure 222 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 about 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

$$\begin{aligned} \rho_{\text{pure}} = \psi \otimes \psi^\dagger &= |a|^2 \psi_a \otimes \psi_a^\dagger + |b|^2 \psi_b \otimes \psi_b^\dagger + a\, b^* \psi_a \otimes \psi_b^\dagger + a^*\, b\, \psi_b \otimes \psi_a^\dagger \\ &= (\psi_a, \psi_b) \begin{pmatrix} |a|^2 & a\, b^* \\ a^*\, b & |b|^2 \end{pmatrix} \begin{pmatrix} \psi_a^\dagger \\ \psi_b^\dagger \end{pmatrix} \quad . \end{aligned} \tag{433}$$

We 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.

---

[*] Most what can be said about this topic has been said by two people: John von Neumann, who in the nineteen thirties stressed the differences between evolution and decoherence, and by Hans Dieter Zeh, who in the nineteen seventies stressed the importance of baths in this process. Ref. 603 Ref. 604

[**] Using the density matrix, we can rewrite the evolution equation of a quantum system:

$$\dot{\psi} = -iH\psi \quad \text{becomes} \quad \frac{d\rho}{dt} = -\frac{i}{\hbar}[H, \rho] \quad . \tag{434}$$

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 the two state example

$$\begin{aligned}\rho = {} & |a|^2\psi_a \otimes \psi_a^\dagger + |b|^2\psi_b \otimes \psi_b^\dagger \\ = {} & (\psi_a,\psi_b)\begin{pmatrix} |a|^2 & 0 \\ 0 & |b|^2 \end{pmatrix}\begin{pmatrix} \psi_a^\dagger \\ \psi_b^\dagger \end{pmatrix}\end{aligned} \qquad (435)$$

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 its state is *not known*, or equivalently, to be in a *(phase) incoherent superposition*, because interference effects cannot be observed 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 definition of its entropy and of all its other thermodynamic quantities. These studies show that Ref. 605 Challenge 1012

$$S = -k\,\mathrm{tr}(\rho \ln \rho) \qquad (436)$$

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 (436) of entropy that the loss of off-diagonal elements 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. Challenge 1013

Where is the bath interacting with the system? It obviously must be 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 47.

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.

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 we have 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 under investigation.

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 set-ups. That would imply that the time for decoherence can be made arbitrary large. Let us 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 colliding particles, or by modelling it as a continuous field.

Table 47 Some common and less common baths with their main properties

Bath type	temperature T	wavelength λ_{eff}	particle flux φ	hit time $t_{\text{hit}} = 1/\sigma\varphi$ 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	10^{32} K	10^{-35} m		$>10^{..}$ 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 calculationshows that the decoherence time t_d is in Challenge 1014 any case smaller than this time interval, so that

$$t_d \leqslant t_{\rm hit} = \frac{1}{\varphi\sigma} \quad , \tag{437}$$

where φ is the flux of particles and σ is the cross section for the hit.* Typical values are given in Table 47. We easily note 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 $E_{\rm hit}$ until the relaxation process is completed.

The objects of interest in this discussion, like the mentioned cat, 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.

---

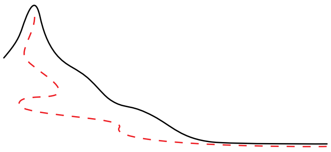
* The decoherence time is derived by studying the evolution of the density matrix $\rho(x,x')$ of objects localized at two points x and x'. One finds that the off-diagonal elements follow $\rho(x,x',t) = \rho(x,x',0)e^{-\Lambda t(x-x')^2}$, where the localization rate Λ is given by

$$\Lambda = k^2\varphi\sigma_{\rm eff} \tag{438}$$

where k is the wave number, φ the flux, and $\sigma_{\rm eff}$ the cross section of the collisions, i.e. usually the size of the macroscopic object. Ref. 606

One also finds the surprising result that a system hit by a particle of energy $E_{\rm hit}$ collapses the density matrix roughly down to the de Broglie (or thermal de Broglie) wavelength of the hitting particle. Both results together give the formula above. Ref. 607

** 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'.



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

The initial coherence of the superposition, so disturbingly in contrast with our everyday experience, disappears exponentially within a *decoherence time* t_d given by[*] Ref. 608

$$t_d = t_r \frac{E_{\text{hit}}}{E_r} \frac{e^{E_{\text{hit}}/kT} - 1}{e^{E_{\text{hit}}/kT} + 1} \tag{441}$$

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 any more 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, we 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. We further assume 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 we get Ref. 610

$$t_d = t_r \frac{E_{\text{hit}}^2}{2E_r kT} = t_r \frac{\hbar^2}{2mkTl^2} = t_r \frac{\lambda_T^2}{l^2} \tag{442}$$

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 $m = 1$ g, $l = 1$ mm and $T = 300$ K we get $t_d/t_r = 1.3 \cdot 10^{-39}$. Even if the interaction between the system and the environment would

---

Ref. 609 [*] This result is derived as in the above case. A system interacting with a bath always has an evolution given by the general form

$$\frac{d\rho}{dt} = -\frac{i}{\hbar}[H,\rho] - \frac{1}{2t_o}\sum_j [V_j\rho, V_j^\dagger] + [V_j, \rho V_j^\dagger] \quad . \tag{439}$$

Challenge 1016 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

$$t_o = t_r \frac{E_{\text{hit}}}{E_r} = t_{\text{hit}} \tag{440}$$

or some variations of it, as in the example above.

be so weak that the system would have as relaxation time the age of the universe, which is about $4 \cdot 10^{17}$ s, the time t_d would still be shorter than $5 \cdot 10^{-22}$ s, which is over a million times faster than the oscillation time of a beam of light (about 2 fs for green light). For Schrödinger's cat, the decoherence time would be even shorter. These times are so short that we 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 we have $m = 9.1 \cdot 10^{-31}$ kg, and typically $l = 1$ nm and $T = 4$ K; we then get $t_d \approx t_r$ 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 will mention a few more below. Ref. 611

In 1996 the first actual measurement of decoherence times was published by the Paris team around Serge Haroche. It confirmed the relation between the decoherence time and the relaxation time, thus showing that the two processes have to be distinguished at microscopic scale. In the meantime, other experiments also confirmed the presentation given above, both for small and large values of t_d/t_r, including the evolution equations for the decoherence process itself. Ref. 612 Ref. 613

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 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 *microscopic* or *quantum mechanical* 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. 614

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, some systems 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, 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 little distance away from his body ...
> James Joyce, *A Painful Case*

We asked about non-locality also in general relativity. Let us study the situation in quantum mechanics. We 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 223. A movie of the same process can be seen in the lower right corners on the pages of the present, second part of our mountain ascent. 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 time. Does this happen faster than light? Yes, it does. But is it a problem? See page 336

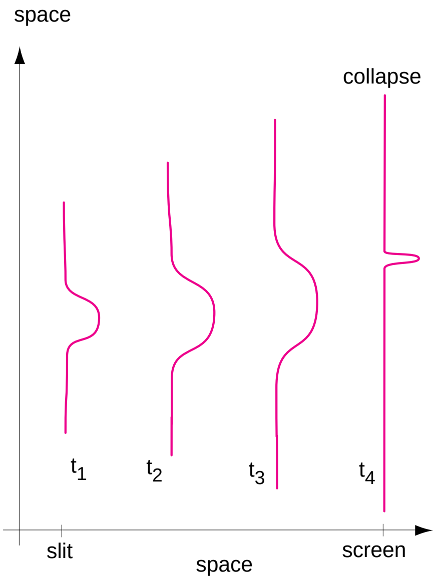


Figure 223 Quantum mechanical motion: an electron wave function (actually its module squared) from the moment it passes a slit until it hits a screen

A situation is called *acausal* or *nonlocal* if energy is transported faster than light. Using Figure 223 youcan determine the energy velocity involved, using the results on signal propagation. The result is a value smaller than c. A wavefunction maximum moving faster than light does *not* imply energy motion faster than light.[*] Challenge 1017 See page 403

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. 615, 616

We note again that no energy is transported faster than light. No non-locality 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 406

Such experiments have actually been performed. The first and most famous was the one performed in 1982, with photons instead of electrons by Alain Aspect. Like all latter ones, it has fully confirmed quantum mechanics. Ref. 617

In fact, such experiments just confirm that it is not possible 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 of apparent non-locality 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 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

---

[*] In classical electrodynamics, the same happens with the scalar and the vector potential, if the Coulomb gauge is used.

[**] David Joseph Bohm (1917–1992) American-British physicist, codiscovered the Aharonov-Bohm effect; he spent a large part of his life investigating the connections between quantum physics and philosophy.

second atom can be excited only after a time $t = l/c$ has been elapsed, i.e. after the photon has travelled to the second atom.

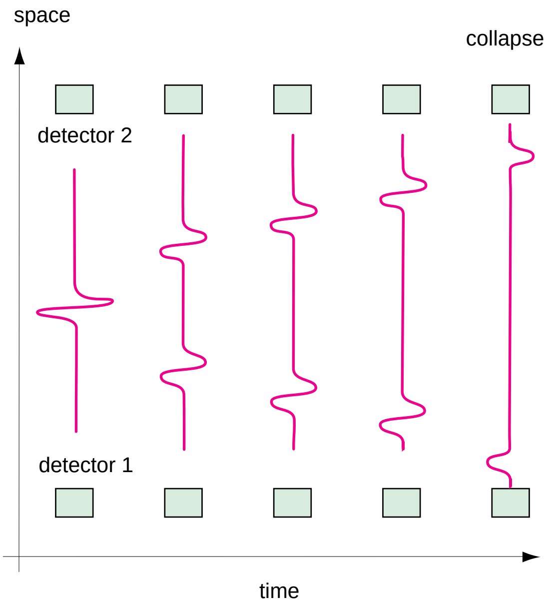


Figure 224 Bohm's Gedankenexperiment

Surprisingly, this conclusion is wrong. The atom in its ground state has a non-zero 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 Ref. 618 confirmed experimentally.

More careful studies show that 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. This pretty conclusion again avoids non-locality.

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 Ref. 610 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. 625
- Obviously, superpositions of magnetization in up and down direction for several materials have also be observed. Ref. 619
- 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. Ref. 621 Experiments with single atoms in superpositions of states are among the most popular ones. Ref. 622
- In 1997, coherent atom waves were extracted from a cloud of sodium atoms. Ref. 623
- 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 manifest themselves differently, since matter can stay at rest but light cannot, and because light is made of bosons, but matter 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. A beam is thus incoherent if its particles arrive randomly in time and in frequency. In everyday life, the rarity of observation of coherent matter superpositions has the same origin as the rarity of observation of coherent light. See page 575

▪ We will discuss the relation between the environment and the *decay* of unstable systems later on. The phenomenon is completely described by the concepts given here. See page ??

▪ Another conclusion deserves to be mentioned: *teleportation contradicts correlations*. Can you confirm it? Challenge 1018

▪ Some people say that quantum theory could be used for quantum computing, by using superpositions of wavefunctions. Can you give a reason that makes this aim very difficult, even without knowing how such a quantum computer might work? Ref. 627 Challenge 1019 n

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, we speak 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 453 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 also its particle aspect appear. Experiment shows 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 us to use probabilities for the description of microsystems. We find 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

$$P_n = |\varphi_n^{\dagger}\psi|^2 \quad , \tag{443}$$

where φ_n is the eigenfunction of the operator A corresponding to the eigenvalue a_n. See page 916

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 generalized to the more general cases with degenerate and continuous eigenvalues. Ref. 624

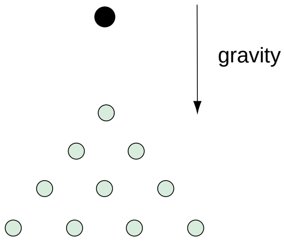


Figure 225 A system showing probabilistic behaviour

At first sight, the sort of probabilities encountered in quantum theory are different from the probabilities we encounter 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 uncontrollable variations of the starting conditions every time the experiment is repeated.[*]

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 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 See page 452

---

[*] 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 … …,

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 many components any recording device is made of. But 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, 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 we 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. Ref. 620 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.

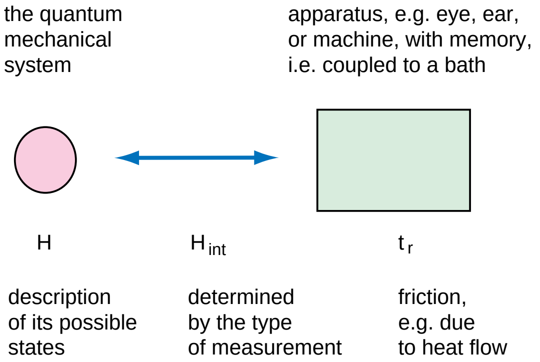


Figure 226 The concepts used in the description of measurements

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 a 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

---

[*] Since baths imply friction, we can also say: memory needs friction.

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 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 our 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 localized 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. In the time before the interaction between the particle and the detector, the combined system was in the initial state ψ_i given simply by

$$\psi_i = \psi_p \chi_{\text{start}} = \sum_n c_n \varphi_n \psi_{\text{other}} \chi_{\text{start}} \quad . \tag{446}$$

---

* 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

$$\rho_{\text{f}} = T \rho_{\text{i}} T^\dagger \tag{444}$$

Challenge 1020 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

$$[H + H_{int}, A] = 0 \quad . \tag{445}$$

After the interaction, using the just mentioned characteristics of the apparatus, the combined state ψ_a is

$$\psi_a = \sum_n c_n \varphi_n \psi_{\mathrm{other},n} \chi_n \quad . \tag{447}$$

This evolution from ψ_i to ψ_a follows from the evolution equation applied to the particle detector combination. Now the state ψ_a is a superposition of macroscopically distinct states, as it is a superposition of distinct macroscopic states of the detector. 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. Let us see why. The density matrix ρ_a of this situation, given by

$$\rho_a = \psi_a \otimes \psi_a^\dagger = \sum_{n,m} c_n c_m^* (\varphi_n \psi_{\mathrm{other},n} \chi_n) \otimes (\varphi_m \psi_{\mathrm{other},m} \chi_m)^\dagger \quad , \tag{448}$$

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

$$\rho_f = \sum_n |c_n|^2 (\varphi_n \psi_{\mathrm{other},n} \chi_n) \otimes (\varphi_n \psi_{\mathrm{other},n} \chi_n)^\dagger \tag{449}$$

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 $\varphi_n \psi_{\mathrm{other,n}}$ as well as the detector in the state χ_n. But this is precisely what the two properties of quantum measurements state.

We therefore find that describing a measurement as an evolution of a quantum system interacting with a macroscopic detector, itself containing a bath, we 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:

$$t_{\mathrm{collapse}} = t_{\mathrm{d}} < t_r \tag{450}$$

We thus have a formula for the time the wavefunction takes to collapse. The first experimental measurements of the time of collapse are appearing, and confirm these results. Ref. 626

---

* 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.

Hidden variables

Obviously a large number of people are not satisfied with the arguments 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.

▪ Clearly, hidden variables controlling the evolution of microscopic system would contradict the result that action values below $\hbar/2$ cannot be detected. This minimum observable action is the reason for the random behaviour of microscopic systems.

▪ Historically, the first argument against hidden variables was given by John von Neumann.[*]

– CS – to be written – CS –

▪ 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. 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. Ref. 628

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.

The starting idea is to distinguish quantum theory and locally realistic theories using hidden variables by measuring the polarizations of two correlated photons. Quantum theory says that the polarization of the photons is fixed only at the time it is measured, whereas local realistic theories say that it is fixed already in advance. The correct description can be found by experiment.

---

[*] John von Neumann (1903, Budapest–1957, Washington DC) mathematician, one of the fathers of the modern computer.

[**] Which leads to the definition: one zillion is 10^{23}.

[***] John Stewart Bell (1928–1990), theoretical physicist who worked mainly on the foundations of quantum theory.

Imagine the polarization is measured at two distant points A and B, each observer can measure 1 or -1 in each of his favourite direction. Let each observer choose two directions, 1 and 2, and call their results a_1, a_2, b_1, and b_2. Since the measurement results all are either 1 or -1, the value of the specific expression $(a_1+a_2)b_1+(a_2-a_1)b_2$ has always the value ± 2.

Ref. 629 Imagine you repeat the experiment many times, assuming that the hidden variables appear statistically. You then can deduce (a special case of) Bell's inequality for two hidden variables

Challenge 1021 e

$$|(a_1b_1)+(a_2b_1)+(a_2b_2)-(a_1b_2)| \leqslant 2 \qquad (451)$$

where the expressions in brackets are the averages of the measurement products over a large number of samples. This result holds independently of the directions of the involved polarizers.

On the other hand, if the polarizers 1 and 2 at position A and the corresponding ones at position B are chosen with angles of $\pi/4$, quantum theory predicts that the result is

$$|(a_1b_1)+(a_2b_1)+(a_2b_2)-(a_1b_2)| = 2\sqrt{2} > 2 \qquad (452)$$

which is in complete contradiction with the hidden variable result.

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.

Another measurable contradiction between quantum theory and locally realistic theories has been predicted by Greenberger, Horn and Zeilinger. Experiments trying to check the result are being planned. No deviation from quantum theory is expected.

Conclusions on probabilities and determinism

> Geometric demonstramus quia facimus;
> si physics demonstrare possemus, faceremus.
> Giambattista Vico[*]

From the arguments presented here we draw a number of conclusions which we need for the rest of our mountain ascent. Note that these conclusions are not shared by all physicists! The whole topic is still 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

---

[*] 'We are able to demonstrate geometrical matters because we make them; if we could prove physical matters we would be able to make them.' Giovanni Battista Vico (1668, Napoli– 1744, Napoli) important Italian philosopher and thinker. In this famous statement he points out a fundamental distinction between mathematics and physics.

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 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, we 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 modelled by the change of the state of the system from ψ_{other} to $\psi_{\text{other,n}}$. Without such a change of state, without a disturbance of the quantum system, a measurement is impossible.

Challenge 1022 * Actually, there are more reasons; can youname 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 1023 n

- Since the complete measurement is described by quantum mechanics, unitarity is and remains the basic property of evolution. 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 499
- 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 for such disinformation is the following quote.

> Nobody understands quantum mechanics.
> Richard Feynman

What is the difference between space and time?

More specifically, why are objects localized 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 centred 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 'localization' 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 interactions do *not* commute with

---

[*] This implies that the so-called 'many worlds' interpretation is wishful thinking. One also reaches this conclusion when studying the details of this religious approach. Ref. 630

[**] 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.

time. In fact, this is 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 1024

In other words, in contrast to general relativity, quantum theory produces a distinction between space and time. In fact, we can *define* position as what commutes with interaction Hamiltonians. Note that 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.

Are we good observers?

Are humans 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. Our senses are classical machines, in the sense described above. Our brain is also a classical apparatus, but the fact is secondary; our sensors are the key. Ref. 631

In addition, we have stressed several times that any observing entity needs memory, which means it needs to incorporate a bath. That means that observers have to be made of matter; an observer cannot be made of 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 basic anthropomorphism? We will uncover more details as we continue.

What is the connection between information theory, cryptology and quantum theory?

Physics means talking about observations of nature. Like any observation, also measurements produce information. It is thus possible to translate much (but not all) of quantum theory into the language of information theory. In particular, the existence of a minimal change in nature implies that the information about a physical system can never be complete, that information transport has its limits and that information can never be fully trusted. The details of these studies form a fascinating way to look at the microscopic world. The studies become even more interesting when the statements are translated into the language of cryptology. Cryptology is the science of transmitting hidden messages that only the intended receiver can decrypt. Ref. 632 In our modern times of constant surveillance, cryptology is an important tool to protect personal freedom.[*] Due to the quantum of action, nature provides limits on the possibility of sending encrypted messages. The statement of these limits is (almost) equivalent to the statement that change in nature is limited by the quantum of action.

The connection between information theory and quantum theory has also a dark side, often found in the media. The media do not encourage clear thinking. Stating that the universe is information or that the universe is a computer is as devoid of good sense as saying that See page 773 the universe is an observation, that it is a measurement apparatus, or that it is a clockwork. Any expert of motion should beware of these and similarly fishy statements; people who use them either deceive themselves or try to deceive others.

---

[*] For a good textbook on cryptology, see e.g. the text by ...

Does the ‘wavefunction of the universe’ exist?

This expression is frequently heard in discussions about quantum mechanics. Numerous conclusions are drawn from it, e.g. about the irreversibility of time, the importance of initial conditions, the decoherence of the universe, about changes required to quantum theory, changes necessary to thermodynamics or the importance of the mind. Are these arguments correct?

See page 307

The first thing to clarify is the meaning of ‘universe’. As already explained the term can have two meanings: either the collection of all matter and radiation, or this collection *plus* all of space-time. Secondly, we have to recall the meaning of ‘wavefunction’: it describes the *state* of a system. The state distinguishes two otherwise identical systems; for example, position and velocity distinguish two otherwise identical ivory balls on a billiard table. Alternatively and equivalently, the state describes changes in time.

See page 33

Does the universe have a state? If we take the wider meaning of universe, obviously it does not. Talking about the state of the universe is a contradiction: by definition, the concept of state, defined as the non-permanent aspects of an object, is applicable only to *parts* of the universe.

We can take the narrower sense of ‘universe’, as sum of all matter and radiation only, without space and time, and ask the question again. To determine its state, we need a possibility to measure it: we need an environment. But the environment of the smaller universe is space-time only; initial conditions cannot be determined since we need measurements to do this, and thus an apparatus, i.e. a material system with a bath attached to it.

In short, standard quantum theory does not allow for measurements of the universe; therefore it has no state. Summing up, beware of anybody who claims to know something about the wavefunction of the universe. Just ask him: If you know the wavefunction of the universe, why aren’t you rich?

Several famous physicists have proposed evolution equations for the wavefunction of the universe! It seems a silly point, but the predictions of these equations cannot be compared to experiments; the arguments just given even make this impossible in principle. The pursuits in this directions, so interesting they are, must therefore be avoided if we want to reach the top of Motion Mountain.

There are many more twists to this story. One possibility is that space-time itself, even without matter, is a bath. This speculation will be shown to be correct later on and seems to allow speaking of the wavefunction of all matter. But then again, it turns out that time is *undefined* at the scales where space-time would be an effective bath; this means that the concept of state is not applicable there.

We can retain as result, valid even in the light of the latest results of physics: there is *no* wavefunction of the universe, independently of what is meant by ‘universe’. Before we go on studying the more complicated consequences of quantum theory for the whole universe, we first continue a bit with the consequences of quantum theory for our everyday observations.

23. Applications of quantum mechanics – life, pleasure and their means

Now that we can look at quantum effects without ideological baggage, let us have some serious fun in the world of quantum theory. The quantum of action has important consequences for biology, chemistry, technology and science fiction. We will only explore a cross section of these topics, but it will be worth it.

Biology

Typeset in January 2003

A special form of electromagnetic motion is of importance to humans: life. We mentioned at the start of quantum theory that life cannot be described by classical physics. Life is a quantum effect. Let us see why.

Living beings can be described as objects showing metabolism, information processing, information exchange, reproduction and motion. Obviously, all these properties follow from a single one, to which the others are enabling means:

▷ *Living beings are objects able to reproduce.*[*]

This definition implies several consequences. In order to reproduce, living beings must be able to move in self-directed ways. An object able to perform self-directed motion is called a *machine*. All self-reproducing beings are machines.

Since reproduction is simpler the smaller the system is, most living beings are extremely small machines for the tasks they perform, especially when compared to human made machines. This is the case even though the design of human machines has considerably fewer requirements: human-built machines do not need to be able to reproduce; as a result, they do not need to be made of a single piece of matter, as all living beings have to. But despite all the restrictions nature has to live with, living beings hold many miniaturization world records:

- The brain has the highest processing power per volume of any calculating device so far. Just look at the size of chess champion Gary Kasparov and the size of the computer against which he played.
- The brain has the densest and fastest memory of any device so far. The set of compact disks (CDs) or digital versatile disks (DVDs) that compare with the brain is many thousand times larger.

See page 454

- Motors in living beings are many orders of magnitude smaller than human-built ones. Just think about the muscles in the legs of an ant.
- The motion of living beings beats the acceleration of any human-built machine by orders of magnitude. No machine moves like a grasshopper.
- Living being's sensor performance, such as that of the eye or the ear, has been surpassed by human machines only recently. For the nose this feat is still far away. Nevertheless, the sensor sizes developed by evolution – think also about the ears or eyes of a common fly – is still unbeaten.

---

[*] However, there are examples of objects which reproduce and which nobody would call living. Can you find some examples, together with a sharper definition?

Challenge 1025 n

- Flying, swimming or crawling living beings – such as a fruit fly, plankton or amoebas – are still thousands of times smaller than anything built by humans.
- Challenge 1026 n Can you spot more examples?

The superior miniaturization of living beings is due to their continuous strife for efficient construction. In the structure of living beings, everything is connected to everything: each part influences many others. Indeed, the four basic processes in life, namely metabolic, mechanical, hormonal and electrical, are intertwined in space and time. For example, breathing helps digestion; head movements pump liquid through the spine; a single hormone influences many chemical processes. Furthermore, all parts in living systems have more than one function. For example, bones provide structure and produce blood; fingernails are tools and shed chemical waste.

The miniaturization, the reproduction and the functioning of living beings all rely on the quantum of action. Let us see how.

Reproduction

> Life is a sexually transmitted disease.
> Anonymous

All the astonishing complexity of life is geared towards reproduction. *Reproduction* is the ability of an object to build other objects similar to itself. Quantum theory old us that only a *similar* object is possible, as an exact copy would contradict the quantum of action, as we found out above. See page 567

Since reproduction requires mass increase, reproducing objects show both metabolism and growth. In order that growth leads to an object similar to the original, a construction plan is necessary. This plan must be similar to the plan used by the previous generation. Organizing growth with a construction plan is only possible if nature is made of smallest entities which can be assembled following that plan.

We can thus deduce that reproduction implies that matter is made of smallest entities. If matter were not made of smallest entities, there would be no way to realize reproduction. Reproduction thus requires quantum theory. Indeed, without the quantum of action there would be no DNA molecules and there would be no way to inherit our own properties – our own construction plan – to children.

Passing on a plan requires that living beings have ways to store information. Living beings must have some built-in memory. We know already that a system with memory must be made of many particles. There is no other way to store information. The large number of particles is mainly necessary to protect the information from the influences of the outside world.

Our own construction plan, made of what biologists call genes, is stored in DNA molecules. Reproduction is thus first of all a transfer of parent’s genes to the next generation. We will come back to the details below. We first have a look on how our body moves itself and its genes around.

Quantum machines

Living beings are machines. How do these machines work? From a physical point of view, we need only a few sections of our walk so far to describe them: universal gravity and QED. Simply stated, life is an electromagnetic process taking place in weak gravity.[*] But the details of this statement are tricky and interesting. Table 504 gives an overview of motion processes in living beings. Interestingly, all motion in living beings can be summarized in a few classes by asking for the motor driving it.

Nature only needs few small but powerful devices to realize all motion types used by living beings. Given the long time that living systems have been around, these devices are extremely efficient. In fact, ion pumps, chemical pumps, rotational and linear molecular motors are all specialized molecular motors. Ion and chemical pumps are found in membranes and transport matter. Rotational and linear motor move structures against membranes. In short, all motion in living beings is due to molecular motors. Even though there is still a lot to be learned about them, what is known already is spectacular enough.

How do we move? – Molecular motors

How do our muscles work? What is the underlying motor? One of the beautiful results of modern biology is the elucidation of this issue. It turns out that muscles work because they contain molecules which change shape when supplied with energy. This shape change is repeatable. A clever combination and repetition of these molecular shape changes is then used to generate macroscopic motion. There are three basic classes of molecular motors: linear motors, rotational motors, and pumps.

Linear motors are at the basis of muscle motion; other linear motors separate genes during cell division. They also move organelles inside cells and displace cells through the body during embryo growth, when wounds heal, or in other examples of cell motility. A typical molecular motor consumes around 100 to 1000 ATP molecules per second, thus about 10 to 100 aW. The numbers are small; however, we have to take into account that the power white noise of the surrounding water is 10 nW. In other words, in every molecular motor, the power of the environmental noise is eight to nine orders of magnitude higher than the power consumed by the motor. The ratio shows what a fantastic piece of machinery such a motor is. Challenge 1027 n

We encountered rotational motors already above. Nature uses them to rotate the cilia of many bacteria as well as sperm tails. Researchers have also discovered that evolution produced molecular motors which turn around DNA helices like a motorized bolt would

See page 59

Ref. 640

---

[*] In fact, also the nuclear interactions play some role for life: cosmic radiation is one source for random mutations, which are so important in evolution. Plant growers often use radioactive sources to increase mutation rates. But obviously, radioactivity can also terminate life. Ref. 636

The nuclear interactions are also implicitly involved in several other ways. They were necessary to form the materials – carbon, oxygen, etc. – required for life. Nuclear interactions are behind the main mechanism for the burning of the sun, which provides the energy for plants, for humans and for all other living beings (except a few bacteria in inaccessible places).

Summing up, the nuclear interactions play a role in the appearance and in the in destruction of life; but they play no (known) role for the actions of particular living beings.

Motion type	examples	main involved devices
Growth	collective molecular processes in cell growth	ion pumps
	gene turn-on and turn-off	linear molecular motors
	aging	linear molecular motors
Construction	material types and properties (polysaccharides, lipids, proteins, nucleic acids, others)	material transport through muscles
	forces and interactions between biomolecules	cell membrane pumps
Functioning		
	details of metabolism (respiration, digestion)	muscles, ion pumps
	energy flow in biomolecules	
	thermodynamics of whole living system and of its parts	muscles
	muscle working	linear molecular motors
	nerve signalling	ion motion, ion pumps
	brain working	ion pumps
	illnesses	cell motility, chemical pumps
	viral infection of a cell	rotational molecular motors for RNA transport
Defence		
	the immune system	cell motility, linear molecular motors
Reproduction		
	information storage and retrieval	linear molecular motors inside cells, sometimes rotational motors, as in viruses
	cell division	linear molecular motors inside cells
	sperm motion	rotational molecular motors
	courting	muscles, brain, linear molecular motors
	evolution	muscles, linear molecular motors

Table 48 Motion and motors in living beings

turn around a screw. Such motors are attached at the end of some viruses and insert the DNA into virus bodies when they are being built by infected cells, or extract the DNA from the virus after it has infected a cell. Another rotational motor, the smallest known so far – 10 nm across and 8 nm high – is ATP synthase, a protein that synthesizes most ATP in cells. Ref. 637

Ref. 638 The ways molecules produce movement in linear motors was uncovered during the 1990s. The results then started a wave of research on all other molecular motors found in nature. All molecular motors share a number of characteristic properties. There are no temperature gradients involved, as in car engines, no electrical currents, as in electrical motors, and

Figure missing

Figure 227 Myosin and actin: the building bricks of a simple linear molecular motor

no concentration gradients, as found in chemically induced motion. The central part of linear molecular motors is a combination of two protein molecules, namely myosin and actin. Myosin changes between two shapes and literally *walks* along actin. It moves in regular small steps. The motion step size has been measured with beautiful experiments to always be an integer multiple of 5.5 nm. A step, usually forward, but sometimes backwards, Ref. 638 results whenever an ATP (adenosine triphosphate) molecule, the standard biological fuel, hydrolyses to ADP (adenosine diphosphate), thus releasing its energy. The force generated is about 3 to 4 pN; the steps can be repeated several times a second. Muscle motion is the result of thousand of millions of such elementary steps taking place in concert.

How do molecular motors work? These motors are are so small that the noise due to the molecules of the liquid around them is not negligible. But nature is smart: with two tricks it takes advantage of Brownian motion and transforms it into macroscopic molecular motion. Molecular motors are therefore also called *Brownian motors*. The transformation of disordered molecular motion into ordered macroscopic motion is one of the great wonders of nature. The first trick of nature is the use of an asymmetric, but periodic potential, a so-called *ratchet*.* The second trick of nature is a temporal variation of the potential, together with an energy input to make it happen. The most important realizations are shown in Figure 228. Ref. 639

The periodic potential variation allows that for a short time the Brownian motion of the moving molecule – typically 1 μm/s – affects its position. Then the molecule is fixed again. In most of these short times of free motion, the position will not change. But if the position does change, the intrinsic asymmetry of the ratchet shape ensures that in most cases the molecule advances in the preferred direction. Then the molecule is fixed again, waiting for the next potential change. On average, the myosin molecule will thus move in one direction. Nowadays the motion of single molecules can be followed in special experimental set-ups. These experiments confirm that muscles use such a ratchet mechanism. The ATP molecule

---

* It was named by Walt Disney after by Ratchet Gearloose, the famous inventor from Duckburg.

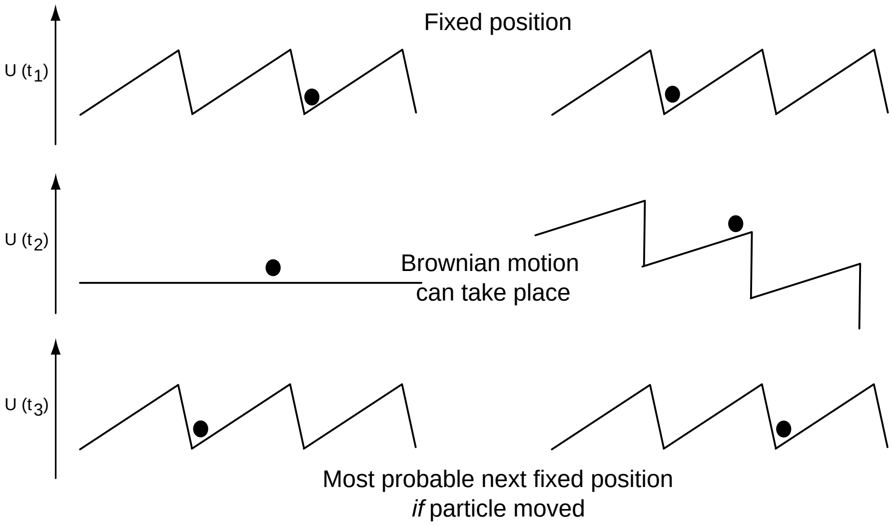


Figure 228 Two types of Brownian motors: switching potential (left) and tilting potential (right)

adds energy to the system and triggers the potential variation through the shape change it induces in the myosin molecule. That is how our muscles work.

Another well-studied linear molecular motor is the kinesin-microtubule system which carries organelles from one place to the other within a cell. As in the previous example, also in this case chemical energy is converted into unidirectional motion. Researchers were able to attach small silica beads to single molecules and to follow their motion. Using laser beams, they could even apply forces to these single molecules. Kinesin was found to move with around 800 nm/s, in steps lengths which are multiples of 8 nm, using one ATP molecule at a time, and exerting a force of about 6 pN.

Quantum ratchets also exist as human built systems, such as electrical ratchets for electron motion or optical ratchets that drive small particles. Extensive experimental research is going on in the field.

Curiosities and challenges of biology

The physics of life is still not fully explored.

Challenge 1028 n ▪ How would you determine which of two identical twins is the father of a baby?

▪ Can you give at least five arguments to show that a human clone, if there will ever be one, is a completely different person that the original? Challenge 1029

▪ Many molecules found in living beings, such as sugar, have mirror molecules. However, in all living beings only one of the two sorts is found. Life is intrinsically asymmetric. How can this be? Challenge 1030

▪ How can it be that the genetic difference between man and chimpanzee is regularly given as about 1%, whereas the difference between man and woman is one chromosome in 46, in other words, about 2.2%? Challenge 1031 n

Challenge 1032 ▪ Could life come to earth from outer space?

Challenge1033 h ▪ How did life start?

- Life is not a clearly defined concept. The definition used above, the ability to reproduce, has its limits when applied to old animals, to a hand cut off by mistake, to sperm or to ovules. It also gives problems when trying to apply it to single cells. Is the definition of life as 'self-determined motion in the service of reproduction' more appropriate? Is the definition of living beings as 'what is made of cells' more appropriate? Challenge 1034
- Mammals have a narrow operating temperature. In contrast to machines, humans function only if the internal temperature is within a narrow range. Why? Does this requirement apply to extraterrestrials? Challenge 1035 d

The physics of pleasure

> What is mind but motion in the intellectual sphere?
> Oscar Wilde (1854–1900) *The Critic as Artist.*

Pleasure is a quantum effect. The reason is simple. Pleasure comes from the senses. All senses measure. And all measures rely on quantum theory. Indeed, the human body, like an expensive car, is full of sensors. Evolution has build these sensors in such a way that they trigger pleasure sensations whenever we do with our body what we are made for.

Among the most astonishing aspects of our body sensors is their sensitivity. The *ear* is so sensitive and at the same time so robust against large signals that the experts are still studying how it works. No known sensor can cover an energy range of 10^{13}; the detected intensity ranges from 1 $\mathrm{pW/m^2}$ to 10 $\mathrm{W/m^2}$, the corresponding air pressure variations from 20 μPa to 60 Pa.

Audible sound wavelengths span from 17 m (20 Hz) to 17 mm (20 kHz). In this range, the ear is able to distinguish at least 1500 pitches with its 16 000 to 20 000 hair cells. But the ear is also able to distinguish 400 from 401 Hz using a special pitch sharpening mechanism.

The *eye* is a position dependent photon detector. Each eye contains around 126 million separate detectors on the retina. Their spatial density is the highest possible that makes sense, given the diameter of the lens of the eye. There are about 6 million less sensitive colour detectors, the cones, and 120 million highly sensitive general light intensity detectors, the rods. This sensitivity difference is the reason that at night all cats are grey. Until recently, human built light sensors with the same sensitivity as rods had to be helium cooled, because technology was not able to build sensors at room temperature as sensitive as the human eye.

The *touch sensors* are distributed over the skin, with a surface density which varies from one region to the other. It is lowest on the back, and highest in the face and on the tongue. There are separate sensors for pressure, for deformation, for vibration and for tickling. Some react proportionally to the stimulus intensity, some differentially.

The taste mechanisms of *tongue* are only partially known. The tongue produces five taste signals* – sweet, salty, bitter, sour, proteic – and the mechanisms are just being unravelled. No sensors with a distinguishing ability of the same degree have been built by humans so far.

---

* Taste is *not* distributed on the tongue in distinct regions; this is an incorrect idea that has been copied from book to book for over a hundred years. You can perform a falsification by yourself, using sugar or salt grains. Challenge 1036 n

The *nose* can distinguish numerous smells. Together with the five signals that the sense of taste can produce (sweet, salty, bitter, sour, proteic) the nose produces a vast range of sensations. It protects against chemical poisons, such as smoke, or biological poisons, such as faecal matter. In contrast, artificial gas sensors exist only for a small range of gases. Artificial taste and smell sensors would allow to check wine or cheese during their production, thus making its inventor extremely rich. Challenge1037 h

The body also contains orientation sensors in the ear, extension sensors in each muscle, pain sensors almost all over the body, heat sensors and coldness sensors on the skin. Other animals feature additional types. Sharks can feel electrical fields, snakes have sensors for infrared, and pigeons can feel magnetic fields. Many birds can see UV light. Bats are able to hear ultrasound up to 100 kHz and more. Whales can detect and localize infrasound signals.

We can conclude that the sensors nature provides us with are state of the art; their sensitivity and ease of use is the highest possible. Since all sensors trigger pleasure or help to avoid pain, nature obviously wants us to enjoy life with the most intense pleasure possible. Studying physics is one way to do this. Ref. 641

> There are two things that make life worth living:
> Mozart and quantum mechanics.
> Victor Weisskopf[*]

The nerves and the brain

> There is no such thing as perpetual tranquillity of mind while we live here;
> because life itself is but motion, and can never be without desire,
> nor without fear, no more than without sense.
> Thomas Hobbes (1588–1679) *Leviathan*.

The main unit processing all these signals, the brain, is another of the great wonders of nature. The human brain has the highest complexity of all brains known;[**] its processing power and speed is orders of magnitude larger than any device build by man.

See page **??** We saw already how electrical signals from the sensors are transported into the brain. This is the main process at the basis of perception.

In the brain, the arriving signals are classified and stored, sometimes for a short time, sometimes for a long time. The details of the various storage mechanisms, essentially taking place in the structure and the connection strength between brain cells, are roughly known. See page **??** The remaining issue is the process of classification. For certain low level classifications,

---

[*] Victor Weisskopf (Vienna, 1908 – 2002), acclaimed theoretical physicist who worked with Einstein, Born, Bohr, Schrödinger and Pauli. He worked on the Manhattan project but later in life intensely campained against the use of nuclear weapons. He was professor at MIT and for many years director of CERN, in Geneva. He wrote several successful physics textbooks. The author heard him making the above statement in Geneva, during one of his lectures.

[**] This is not in contrast with the fact that one or two whale species have brains with a slightly larger mass. The larger mass is due to the protection these brains require against the high pressures which appear when whales dive (some dive to depths of 1 km). The number of neurons in whale brains is considerably smaller than in human brains.

such as colours or geometrical shapes for the eye or sound harmonies for the ear, the mechanisms are known. But for high-level classifications, such as the ones used in conceptual thinking, the aim is not yet achieved. It is not well known how to describe the processes of reading, understanding and talking in terms of signal motions. Research is still in full swing and will remain so for the largest part of the 21st century.

In the following we have a look at a few abilities of our brain, of our body and of other bodies which are important for the study of motion.

Clocks in quantum mechanics

> L'horologe fait de la réclame pour le temps.*
> Georges Perros

Most clocks used in everyday life are electromagnetic. (Do you know an exception?) Any clock on the wall, be it mechanical, quartz controlled, radio or solar controlled, or of any other type, is based on electromagnetic effects. There are even clocks of which we do not even know how they work. Just look at singing. We know from everyday experience that humans are able to keep the beat to within a few per cent for a long time. Also when we sing a musical note we reproduce the original frequency with high accuracy. In many movements humans are able to keep time to high accuracy, e.g. when doing sport or when dancing. (For shorter or longer times, the internal clocks are not so precise.) The way the brain achieves these feats is still subject of research; it is even possible that electrical signals are running in circles inside the grey matter, around the basal ganglia at the centre of each brain half, and that this cycle times provide the basis for the timing feats that are so obvious in our daily life.

Challenge 1038 n

Ref. 642

Do clocks exist?

> Die Zukunft war früher auch besser.**
> Karl Valentin, German writer.

In general relativity, we found that clocks do not exist, because there is no unit of time that can be formed using the constants c and G. Clocks, like any measurement standard, need matter and non-gravitational interactions to work.. This is the domain of quantum theory. Let us see what the situation is in this case.

See page 335

First of all, the time operator, or any operator proportional to it, is not an observable. Indeed, the time operator is not hermitean, as any observable must be. In other words, there is no physical observable whose value is proportional to time.

On the other hand, clocks are quite common; for example, the sun or Big Ben work to everybody's satisfaction. Nature thus encourages us to look for an operator describing the position of the hands of a clock. However, if we look for such an operator we find a strange result. Any quantum system having a Hamiltonian bounded from below – having a lowest

---

* Clocks are ads for time.

** Also the future used to be better in the past.

energy – lacks a hermitean operator whose expectation value increases monotonically with time. This result can be proven rigorously. In other words, time cannot be measured. Challenge 1039 d

That time cannot be measured is not a surprise at all. The meaning of this statement is that every clock needs to be wound up after a while. Take a mechanical pendulum clock. Only if the weight driving it can fall forever, without reaching a bottom position, can the clock go on working. However, in all clocks the weight has to stop, or the battery is empty. In other words, in all real clocks the Hamiltonian is bounded from below.

In short, quantum theory says that any clock can only be *approximate*. Quantum theory shows that exact *clocks do not exist in nature*. Obviously, this result can be of importance only for high precision clocks. What happens if we try to increase the precision of a clock as much as possible?

High precision implies high sensitivity to fluctuations. Now, all clocks have a motor inside that makes them work. A high precision clock thus needs a high precision motor. In all clocks, the position of the motor is read out and shown on the dial. The quantum of action implies that a precise clock motor has a position indeterminacy. The clock precision is thus limited. Worse, like any quantum system, the motor has a small, but finite probability to stop or to run backwards for a while.

You can check this prediction yourself. Just have a look at a clock when its battery is almost empty, or when the weight driving the pendulum has almost reached the bottom position. It will start doing funny things, like going backwards a bit or jumping back and forward. When the clock works normally, this behaviour is only strongly reduced in amount; however, it is still present. Even for the sun. Challenge 1040

In other words, clocks necessarily have to be macroscopic in order to work properly. A clock must be large as possible, in order to average out its fluctuations. Astronomical systems are examples. A good clock must also be well-isolated from the environment, such as a freely flying object whose coordinate is used as time variable, as is done in certain optical clocks.

How big is the problem we have thus discovered? What is the actual error we make when using clocks? Given the various properties of quantum theory, what is the ultimate precision of a clock?

To start with, the uncertainty relation provides the limit that the mass M of a clock must be larger than Challenge 1041

$$M > \frac{\hbar}{c^2\tau} \qquad (453)$$

which is obviously always fulfilled in everyday life.But we can do better. Like for a pendulum, we can relate the accuracy τ of the clock to its maximum reading time T. The idea was first published by Salecker and Wigner. They argued that Challenge 1042 e Ref. 643

$$M > \frac{\hbar}{c^2\tau}\frac{T}{\tau} \qquad (454)$$

where T is the time to be measured. You might check that this directly requires that any clock must be *macroscopic*. Challenge 1043 e

Let us play with this formula by Salecker and Wigner. One way to rephrase it is the following. They showed that for a clock which can measure a time t, the size l is connected to the mass m by

$$l > \sqrt{\frac{\hbar t}{m}} \quad . \tag{455}$$

How close can this limit be achieved? It turns out that he smallest clocks known, as well as the clocks with most closely approach this limit are bacteria. The smallest bacteria, the *mycoplasmas*, have a mass of about $8 \cdot 10^{-17}$ kg, and reproduce every 100 min, with a precision of about 1 min. The size predicted from expression (455) is between 0.09 μm and 0.009 μm. The observed size of the smallest mycoplasmas is 0.3 μm. The fact that bacteria can come so close to the clock limit shows us again what a good engineer evolution has been. Ref. 645

Note that the requirement by Salecker and Wigner is not in contrast with the possibility to make the *oscillator* of the clock very small; people have built oscillators made of a single atom. In fact, such oscillations promise to be the most precise human built clocks. Ref. 644

In the real world, the expression can be stated even more strictly. The whole mass M cannot be used in the above limit. For clocks made of atoms, only the binding energy between atoms can be used. This leads to the so-called *standard quantum limit for clocks*

$$\frac{\delta\nu}{\nu} = \sqrt{\frac{\Delta E}{E_{\text{tot}}}} \tag{456}$$

where $\Delta E = \hbar/T$ is the energy uncertainty stemming from the finite measuring time T and $E_{\text{tot}} = NE_{\text{ryd}}$ is the total binding energy of the atoms in the meter bar. However, the quantum limit has not been achieved for clocks, even though experiments are getting near to it.

In summary, clocks exist only in the limit of $\hbar$ being negligible. In practice, the errors made by using clocks and meter bars can be made as small as required; it suffices to make the clocks large enough. We can thus continue our investigation into the details of matter without much worry. Only in the third part of our mountain ascent, where the precision requirements will be higher and general relativity will limit the size of physical systems, things will get much more interesting: the impossibility to build clocks will then become a central issue.

Metre bars

For length measurements, the situations is similar to that for time measurements. The limit by Salecker and Wigner can also be rewritten for length measurement devices. Are you able to do it? Challenge 1044

In general relativity we found that we need matter for any length measurement. Quantum theory, our description of matter, again shows that meter bars are only approximately possible, but with errors which are negligible if the device is macroscopic.

– CS – Several sections on time, Zeno-effect and more will be added here – CS –

Consciousness: a result of the quantum of action

See page 504 Consciousness is our ability to observe what is going on in our mind. This activity, like any type of change, can itself be observed and studied. Obviously, consciousness takes place in the brain. If it were not, there would be no way to keep it connected with a given person. We know that each brain moves with over one million kilometres per hour through the cosmic background radiation; we also observe that consciousness moves along with it.

The brain is a quantum system; it is based on molecules and electrical currents. The changes in consciousness that appear when matter is taken away from the brain – in operations or accidents – or when currents are injected into the brain – in accidents, experiments or misguided treatments – have been described in great detail by the medical profession. Also the observed influence of chemicals on the brain – from alcohol to hard drugs – makes the same point. The brain is a quantum system.

Magnetic resonance imaging can detect which parts of the brain work when sensing, remembering or thinking. Not only is sight, noise and thought processed in the brain; we can follow the processing on computer screens. The other, more questionable experimental method, positron tomography, works by letting people swallow radioactive sugar. It confirms the findings on the location of thought and on its dependence on chemical fuel. In addition, we already know that memory depends on the particle nature of matter. All these observations depend on the quantum of action.

Not only the consciousness of others, also your own consciousness is a quantum process. Can you give some arguments? Challenge 1045

In short, we know that thought and consciousness are examples of motion. We are thus in the same situation as material scientists were before quantum theory: they knew that electromagnetic fields influence matter, but they could not say how electromagnetism was involved in the build-up of matter. We know that consciousness is made from the signal propagation and signal processing in the brain; we know that consciousness is an electrochemical process. But we do not know yet the details of how the signals make up consciousness. Unravelling the workings of this fascinating quantum system is the aim of neurological science. This is one of the great challenges of twenty-first century science.

It is sometimes claimed that consciousness is not a physical process. Every expert of motion should be able to convincingly show the opposite, even though the details are not clear yet. Can you add arguments to the ones given here? Challenge 1046

Why can we observe motion?

Studying nature can be one of the most intense pleasures of life. All pleasure is based on the ability to observe motion. Our human condition is central to this ability. In our adventure so far we found that we experience motion 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 a mixture of liquids and solids, 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 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; we would have no fun studying physics.

In addition, we saw that we have these abilities only because our forefathers lived on earth, only because life evolved here, only because we live in a relatively quiet region of our galaxy, and only because the human species evolved long after than the big bang.

If any of these conditions were not fulfilled, or if we were not humans (or animals), motion would not exist. In many ways motion is thus an illusion, as Zeno of Elea had claimed. To say the least, the observation of motion is due to the limitations of the human condition. A complete description of motion and nature must take this connection into account. Before we do that, we explore a few details of this connection.

Curiosities and challenges of quantum experiences

- Are ghost images in TV sets, often due to spurious reflections, examples of interference? Challenge 1047 n
- What happens when two monochromatic electrons overlap? Challenge 1048 h

Material science

> Did you know that one cannot use a boiled egg as a toothpick?
> Karl Valentin

It was mentioned several times that the quantum of action explains all properties of matter. Many researchers from physics, chemistry, metallurgy, engineering, mathematics and biology have cooperated in the proof of this statement. In our mountain ascent we have little time to explore this vast topic. Let us walk across a selection.

Why does the floor not fall?

We do not fall through the mountain we are walking on. Some interaction keeps us from falling through. In turn, the continents keep the mountains from falling through them. Also the liquid magma in the earth's interior keeps the continents from sinking. All these statements can be summarized. Atoms do not penetrate each other. Despite being mostly empty clouds, atoms keep a distance. All this is due to the Pauli principle between electrons. the fermion character of electrons avoids that atoms interpenetrate. At least on earth. See page 573

Not all floors keep up due to the fermion character of electrons. Atoms are not impenetrable at all pressures. They can collapse, and form new types of floors. Some floors are so exciting to study that people have spent their whole life to understand why they do not fall, or when they do, how it happens: the surfaces of stars.

In most stars, the radiation pressure of the light plays only a minor role. Light pressure does play a role in determining the size of red giants, such as Betelgeuse; but for average stars, light pressure is negligible.

In most stars, such as in the sun, the gas pressure takes the role which the incompressibility of solids and liquids has for planets. The pressure is due to the heat produced by the nuclear reactions.

The next star type appears whenever light pressure, gas pressure, and the electronic Pauli pressure cannot keep atoms from interpenetrating. In that case, atoms are compressed until all electrons are pushed into the protons. Protons then become neutrons, and the whole star has the same mass density of atomic nuclei, namely about $\ldots \cdot 10^{\cdot\cdot}\,\mathrm{kg/m^3}$. A spoonful weighs about ...tons. In these so-called *neutron stars*, the floor – or better, the size – is also determined by Pauli pressure; however, it is the Pauli pressure between neutrons, triggered by the nuclear interactions. These neutron stars are all around 10 km in radius.

See page 672

If the pressure increases still further the star becomes a black hole, and never stops collapsing. Black holes have no floor at all; they still have a constant size though, determined by the horizon curvature.

See page 338

The question whether other star types exist in nature, with other floor forming mechanisms – such as quark stars – is still a topic of research.

How can one look through matter?

Quantum theory showed us that all obstacles have only finite potential heights. That leads to a question: Is it possible to look through matter? For example, can we see what is hidden inside a mountain? To be able to do this, we need a signal which fulfils two conditions: it must be able to *penetrate* the mountain, and it must be scattered in a *material-dependent* way. Table 49 gives an overview of the possibilities.

Table 49 Signals penetrating mountains and other matter

Signal	penetration depth in stone	achieved resolution	material dependence	use
matter				
diffusion of water or liquid chemicals	ca. 5 km	ca. 100 m	medium	mapping hydrosystems
diffusion of gases	ca. 5 km	ca. 100 m	medium	studying vacuum systems
electromagnetism				
sound, explosions	0.1 – 10 m	ca. $l/100$	high	oil and ore search
ultrasound		1 mm	high	medical imaging, acoustic microscopy
infrasound and earthquakes	100 000 km	100 km	high	mapping of earth crust and mantle
static magnetic fields			medium	cable search, cable fault localisation
static electric fields				
electrical currents				soil investigations, search for tooth decay
X rays	a few metre	5 μm	high	medicine, material analysis, airports
visible light	ca. 1 cm	0.1 μm	medium	imaging of many sorts
IR	ca. 1 cm	0.1 μm	medium	mammography of the future
mm and THz waves	below 1 mm	1 mm		see through clothes Ref. 649

Signal	penetration depth in stone	achieved resolution	material dependence	use
radio waves	10 m	1 m to 1 mm	small	soil radar, magnetic imaging, research into solar interior
weak interactions				
neutrino beams	light years	zero	very weak	studies of sun
strong interactions				
cosmic radiation	1 m to 1 km			
radioactivity	1 mm to 1 m			airports
gravitation				
change of gravitational acceleration		50 m	low	oil & ore search

We see that many signals are able to penetrate a mountain. However, only sound or radio waves provide the possibility to distinguish different materials, or to distinguish solids from liquids and from air. In addition, any useful method requires a large number of signal sources and of signal receptors, and thus a large amount of cash. Will there ever be a simple method allowing to look into mountains as precisely as X-rays allow to study human bodies? For example, will it ever be possible to map the interior of the pyramids? A motion expert like the reader should be able to give a definite answer. Challenge 1049 n

One of the high points of twentieth century physics was the development of the best method so far to look into matter with dimensions of about a meter or less: magnetic resonance imaging. We will discuss it later on. See page 659

The other modern imaging technique, ultrasound imaging, is getting more and more criticized. It is much used for prenatal diagnostics of embryos. However, studies have found that ultrasound produces extremely high levels of audible sound to the baby, especially when the ultrasound is switched on or off, and that babies react negatively to this loud noise.

What is necessary to make matter invisible?

You might have already imagined what adventures would be possible if you could be invisible for a while. Some years ago, a team of Dutch scientists found a material than can be switched from mirror mode to transparent mode using an electrical signal. This seems a first step to realize the dream to become invisible at will. Ref. 648

Nature shows us how to be invisible. An object is invisible if it has no surface, no absorption and small size. In short, invisible objects are either small clouds or composed of them. Most atoms and molecules are examples. Homogeneous non-absorbing gases also realize these conditions. That is the reason that air is (usually) invisible. When air is not homogeneous, it can be visible, e.g. above hot surfaces.

In contrast to gases, solids or liquids do have surfaces. Surfaces are usually visible, even if the body is transparent, because the refractive index changes there. For example, quartz can be made so transparent that one can look through 1 000 km of it; pure quartz is thus more transparent than usual air. Still, objects made of pure quartz are visible to the eye due to the index change at the surface. Quartz can be invisible only when submerged in liquids with the same refractive index.

In other words, to become invisible, we must transform ourselves into a diffuse cloud of non-absorbing atoms. On the way to become invisible, we would loose all memory and all genes, in short, we would loose all our individuality. But an individuum cannot be made of gas. An individuum is defined through its boundary. There is no way that we can be invisible; a reversible way to perform the feat is also impossible. In summary, quantum theory shows that only the dead can be invisible.

Curiosities and challenges of material science

Material science is not a central part of this walk. A few curiosities can give a taste of it.

▪ What is the maximum height of a mountain? This question is of course of interest to Ref. 650 all climbers. Many effects limit the height. The most important is the fact that under heavy Challenge 1050 pressure, solids become liquid. For example, on earth this happens at about 27 km.This is quite a bit more than the highest mountain known, which is the volcano Mauna Kea in Hawaii, whose top is about 9.45 km above the base. On Mars gravity is weaker, so that mountains can be higher. Indeed the highest mountain on Mars, Olympus mons, is 80 km Challenge 1051 n high. Can you find a few other effects limiting mountain height?

▪ Do you want to become rich? Just invent something that can be produced in the fac- Challenge1052 h tory, is cheap, and can substitute duck feathers in bed covers and sleeping bags. Another industrial challenge is to find an artificial substitute for latex.

Challenge 1053 ▪ What is the difference between solids, liquids, and gases?

▪ At low temperatures of about 2 mK, helium-3 becomes a *superfluid*. It is even able, after an initial kick, to flow over obstacles, such as glass walls.

▪ What is the difference between the makers of bronze age knifes and the builders of the Eiffel tower? Only their control of dislocation distributions.

Challenge 1054 n ▪ Quantum theory shows that tight walls do not exist. Every material is penetrable. Why?

▪ Quantum theory shows that even if tight walls would exist, the lid of such a box can Challenge 1055 n never be tightly shut. Can you provide the argument?

▪ Quantum theory predicts that heat transport at a given temperature is quantized. Can Challenge 1056 you guess the unit of thermal conductance?

▪ Robert Full has shown that van der Waals forces are responsible for the way that geckos Ref. 651 walk on walls and ceilings. The gekko, a small reptile with a mass of about 100 g, uses an elaborate structure on its feet to perform the trick. Each foot has 500 000 hairs each split in up to 1000 small spatulae, and each spatula uses the van-der-Waals force to stick to the surface. As a result, the gekko can walk on vertical glass walls or even on glass ceilings; the sticking force can be as high as 100 N per foot.

▪ Millimetre waves or terahertz waves are emitted by all bodies at room temperature. Modern cameras allow to image them. In this way, it is possible to see through clothes. This might be used in future to detect hidden weapons in airports. But the development of a practical and affordable detector which can be handled as easily as a binocular is still underway.

Quantum technology

> I were better to be eaten to death with a rust
> than to be scoured to nothing with perpetual motion.
> William Shakespeare (1564–1616) *King Henry IV*.

Quantum effects do not appear only in microscopic systems. Several quantum effects are important in modern life; transistors, lasers, superconductivity and a few other effects and systems are worth knowing.

Quantized conductivity

In 1996, the Spanish physicist J.L. Costa-Krämer and his colleagues performed a simple experiment. They put two metal wires on top of each other on a kitchen table and attached a battery, a 10 kΩ resistor and a storage oscilloscope to them. They then measured the electrical current while knocking on the table. In the last millisecond before the wires detach, the conductivity and thus the electrical current diminished in regular steps of a few 7 μA, as can easily be seen on the oscilloscope. This simple experiment could have beaten, if it had been performed a few years earlier, a number of enormously expensive experiments which discovered this quantization at costs of several million euros, using complex set-ups and low temperatures. Ref. 654

In fact, quantization of conductivity appears in any electrical contact with a small cross section. In such situations the quantum of action implies that conductivity can only be a multiple of $2e^2/\hbar \approx 1/12\,906\,1/\Omega$. Can you confirm this result? Challenge 1057

Note that electrical conductivity can be as small as required; only *quantized* electrical conductivity has the minimum value of $2e^2/\hbar$.

The fractional quantum Hall effect

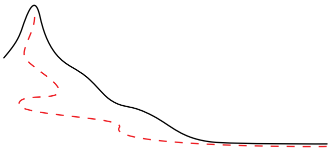
The fractional quantum hall effect is one of the most intriguing discoveries of material science. In 1982, Robert Laughlin predicted that in this system one should be able to observe objects with electrical charge $e/3$. This strange prediction was indeed verified in 1997. Ref. 655

The story begins with the discovery of Klaus von Klitzing of the quantum Hall effect. In 1980 he found that in two-dimensional systems at low temperatures – about 1 K – the electrical conductance is quantized in multiples of the quantum of conductance

$$S = n\,\frac{e^2}{\hbar} \quad . \qquad (457)$$

The explanation is straightforward and is the quantum analogue of the classical Hall effect, which describes how conductance varies with applied magnetic field. Von Klitzing received the Nobel prize in physics for the discovery, as it was completely unexpected, allows a highly precise measurement of the fine structure constant and allows to detect the smallest voltage variations measurable.

Two years later, it was found that in extremely strong magnetic fields the conductance could vary in steps one third that size. Other, even stranger fractions were also found. Robert Laughlin explained all these results by assuming that the electron gas could form collective



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

state showing quasiparticle excitations with a charge $e/3$. This was confirmed 15 years later, and earned him a Nobel price as well. We have seen in several occasions that quantization is best discovered through noise measurements; also in this case, the clearest confirmation came from electrical current noise measurements.

How can we imagine these excitations?

– CS – explanation to be inserted – CS –

What do we learn from this result? Systems in two dimensions have states which follow different rules than systems in three dimensions. Obviously, the first question we ask is whether we can infer something about quarks from this result. Quarks are the constituents of protons and neutrons, and have charges $e/3$ and $2e/3$. At this point we need to stand the suspense; we come back to this issue later on.

Lasers and other spin one vector boson launchers

In 19..., Albert Einstein showed that there are two types of light sources, both of which actually 'create' light. He showed that every lamp whose brightness is turned up high enough will show special behaviour when a certain intensity threshold is passed. The main mechanism of light emission then changes from spontaneous emission to *stimulated emission*. Nowadays such a special lamp is called a *laser*. (The letters 'se' in laser are an abbreviation of 'stimulated emission'.) In summary, Einstein and others showed that lasers are lamps which are sufficiently turned up. Lasers consist of some light producing and amplifying material together with a mechanism to pump energy into it. The material can be a gas, a liquid or a solid; the pumping process can use electrical current or light. Usually, the material is put between two mirrors, in order to improve the efficiency of the light production. Common lasers are semiconductor lasers (essentially highly pumped LEDs or light emitting diodes), He-Ne lasers (highly pumped neon lamps), liquid lasers (essentially highly pumped fire flies) and ruby lasers (highly pumped luminescent crystals).

Lasers produce radiation in the range from microwaves and extreme ultraviolet. They have the special property of emitting *coherent* light, usually in a collimated beam. Therefore lasers achieve much higher light intensities than lamps, allowing their use as tools. In modern lasers, the coherence length, i.e. the length over which interference can be observed, can be thousands of kilometres. Such high quality light is used e.g. in gravitational wave detectors.

People have become pretty good at building lasers. Lasers are used to cut metal sheets up to 10 cm thickness, others are used instead of knives in surgery, others increase surface hardness of metals or clean stones from car exhaust pollution. Other lasers drill holes in teeth, measure distances, image biological tissue or grab living cells.

Some materials amplify light so much that end mirrors are not necessary. This is the case for nitrogen lasers, in which nitrogen, or simply air, is used to produce a UV beam. Even a laser made of a single atom (and two mirrors) has been built; in this example, only eleven photons on average were moving between the two mirrors. Quite a small lamp. Also lasers emitting in two dimensions have been built. They produce a light plane instead of a light beam. Ref. 670

Can two photons interfere?

In 1930, Dirac made the famous statement already mentioned above:*

> Each photon interferes only with itself. Interference between two different photons never occurs.

Often this statement is misinterpreted as implying that two separate photon *sources* cannot interfere. It is almost unbelievable how this false interpretation has spread through the Ref. 667 literature. Everybody can check that this statement is incorrect with a radio: two distant radio stations transmitting on the same frequency lead to beats in amplitude, i.e. to *wave interference*. (This should not to be confused with the more common *radio interference*, with usually is simply a superposition of intensities.) Radio transmitters are coherent sources of photons, and any radio receiver shows that two such sources can indeed interfere.

In 1949, interference of two different sources has been demonstrated with microwave beams. Numerous experiments with two lasers and even with two thermal light sources Ref. 667 have shown light interference from the fifties onwards. Most cited is the 1963 experiment by Magyar and Mandel; they used two ruby lasers emitting light pulses and a rapid shutter camera to produce spatial interference fringes.

However, all these experimental results do not contradict the statement by Dirac. Indeed, two photons cannot interfere for several reasons.

- Interference is a result of space-time propagation of waves; photons appear only when the energy-momentum picture is used, mainly when interaction with matter takes place. The description of space-time propagation and the particle picture are mutually exclusive – this is one aspect of the complementary principle. Why does Dirac seem to mix the two in his statement? Dirac employs the term ‘photon’ in a very general sense, as quantized state of the electromagnetic field. When two coherent beams are superposed, the quantized entities, the photons, cannot be ascribed to either of the sources. Interference results from superposition of two coherent states, not of two particles.
- Interference is only possible if one cannot know where the detected photon comes from. The quantum mechanical description of the field in a situation of interference never allows to ascribe photons of the superposed field to one of the sources. In other words, if you can say from which source a detected photon comes from, you *cannot* observe interference.
- Interference between two beams requires a fixed phase between them, i.e. an uncertain particle number; in other words, interference is only possible if the photon number for each of the two beams is unknown.

A better choice of words is to say that interference is always between two (indistinguishable) states, or if one prefers, between two possible (indistinguishable) histories, but never between two particles. In summary, two different electromagnetic beams can interfere, but not two different photons.

Figure yet to be included

Figure 229 An electron hologram

Can two electron beams interfere?

Do coherent electron sources exist? Yes, as it is possible to make holograms with electron beams. However, electron coherence is only transversal, not longitudinal. Transversal coherence is given by the possible size of wavefronts with fixed phase. The limit of this size is given by the interactions such a state has with its environment; if the interactions are weak, matter wave packets of several metres of size can be produced, e.g. in particle colliders, where energies are high and interaction with matter is low.

Actually, the term transversal coherence is a fake. The ability to interfere with oneself is not the definition of coherence. Transversal coherence only expresses that the source size is small. Both small lamps (and lasers) can show interference when the beam is split and recombined; this is not a proof of coherence. Similarly, monochromaticity is not a proof for coherence either.

A state is called *coherent* if it possesses a well-defined phase throughout a given domain of space or time. The size of that region or of that time interval defines the degree of coherence. This definition yields coherence lengths of the order of the source size for small 'incoherent' sources. Nevertheless, the size of an interference pattern, or the distance d between its maxima, can be much larger than the coherence length l or the source size s.

In summary, even though an electron can interfere with itself, it cannot interfere with a second one. Uncertain electron numbers are needed to see a macroscopic interference pattern. That is impossible, as electrons (at usual energies) carry a conserved charge.

– CS – sections on transistors and superconductivity to be added – CS –

Challenges and dreams of quantum technology

Many challenges in applied quantum physics remain, as quantum effects seem to promise to realize many age-old technological dreams.

- Is it possible to make A4-size, thin, and flexible colour displays for an affordable price? Challenge 1059 d
- Will there ever be desktop laser engravers for 2000 Euro? Challenge1060 h
- Will there ever be room-temperature superconductivity? Challenge1061 h

---

[∗] See the famous, beautiful but difficult textbook P.A.M. Dirac, *The Principles of Quantum Mechanics*, Clarendon Press, Oxford, 1930, page 9.

- Will there ever be teleportation of everyday objects? Challenge 1062 n
- Will there ever be applied quantum cryptology? Challenge 1063 d
- Will there ever be printable polymer electronic circuits, instead of lithographically patterned silicon electronics as is common now? Challenge 1064 d
- Will there be radio-controlled flying toys in the size of insects? Challenge1065 h

24. Quantum electrodynamics – the origin of virtual reality

The central concept the quantum theory introduces in the description of nature is the idea of *virtual particles*. Virtual particles are short-lived particles; they owe their existence exclusively to the quantum of action. Because of the quantum of action, they do not need to follow the energy-mass relation that special relativity requires of normal, real particles. Virtual particles can move faster than light and can move backward in time. Despite these strange properties, they have many observable effects.

Typeset in January 2003

Ships, mirrors and the Casimir effect

When two parallel ships roll in a big swell, *without* even the slightest wind blowing, they will attract each other. This effect was well known up to the 19th century, when many places still lacked harbours. Shipping manuals advised captains to let the ships be pulled apart using a well-manned rowing boat. Ref. 658

Figure will be added in the future

Figure 230 Ships in a swell

Wave induce oscillations of ships because a ship absorbs energy from the waves. When oscillating, the ship also emits waves. This happens mainly towards the two sides of the ship. As a result, for a single ship, the wave emission has no net effect on its position. Now imagine that two parallel ships oscillate in a long swell, with a wavelength much larger than the distance between the ships. Due to the long wavelength, the two ships will oscillate in phase. The ships will thus not be able to absorb energy from each other. As a result, the energy they radiate towards the outside will push them towards each other.

The effect is not difficult to calculate. The energy of a rolling ship is

$$E = mgh\,\alpha^2/2 \tag{458}$$

where α is the roll angle amplitude, m the mass of the ship, and $g = 9{,}8\,\mathrm{m/s^2}$ the acceleration due to gravity. The *metacentric height* h is the main parameter characterizing a ship, especially a sailing ship; it tells with what torque the ship returns to the vertical when inclined by an angle α. Typically, one has $h = 1.5$ m.

When a ship is inclined, it will return to the vertical by a damped oscillation. A damped oscillation is characterized by a period T and a quality factor Q. The quality factor is the number of oscillations the system takes to reduce its amplitude by a factor $e = 2.718$. If the quality factor Q of an oscillating ship and its oscillation period T are given, the radiated power W is

$$W = 2\pi \frac{E}{QT} \quad . \tag{459}$$

We saw above that radiation pressure is W/c, where c is the wave propagation velocity. For water waves, we have the famous relation

$$c = \frac{gT}{2\pi} \quad . \tag{460}$$

Assuming that for two nearby ships each one completely absorbs the power emitted from the other, we find that the two ships are attracted towards each other following

$$ma = m 2\pi^2 \frac{h\alpha^2}{QT^2} \quad . \tag{461}$$

Inserting typical values such as $Q = 2.5$, $T = 10\,\mathrm{s}$, $\alpha = 0.14\,\mathrm{rad}$, and a ship mass of 700 tons, we get about 1.9 kN. Long swells thus make ships attract each other. The intensity of the attraction is comparatively small and can indeed be overcome with a rowing boat. On the other hand, even the slightest wind will damp the oscillation amplitude and have other effects that will avoid the observation of the attraction.

Sound waves or noise in air can have the same effect. It is sufficient to suspend two metal plates in air and surround them by loudspeakers. The sound will induce attraction (or repulsion) of the plates, depending on whether the sound wavelength cannot (or can) be taken up by the other plate. Ref. 659

In 1948, the Dutch physicist Hendrik Casimir made one of the most spectacular predictions of quantum theory: he predicted a similar effect for metal plates in vacuum. Casimir, who worked at the Dutch Electronics company Philips, wanted to understand why it was so difficult to build television tubes. Television screens are made by deposing small neutral particles on glass, but Casimir observed that the particles somehow attracted each other. Casimir got interested in understanding how neutral particles interact. During these theoretical studies he discovered that two neutral mirrors (or metal plates) would attract each other even in complete vacuum. This is the famous *Casimir effect*. Casimir also determined the attraction strength between a sphere and a plate, and between two spheres. In fact, all conducting bodies attract each other in vacuum, with a force depending on their geometry.

In all these situations, the role of the sea is taken by the zero-point fluctuations of the electromagnetic field, the role of the ships by the mirrors. Casimir understood that the space between two parallel mirrors, due to the geometrical constraints, had different zero-point fluctuations that the free vacuum. Like two ships, the result would be the attraction of the mirrors.

Casimir predicted that the attraction for two mirrors of mass m and surface A is given by

$$\frac{ma}{A} = \frac{\pi^3}{120} \frac{\hbar c}{d^4} \quad . \tag{462}$$

The effect is a pure quantum effect; in classical electrodynamics, two neutral bodies do not attract. The effect is small; it takes some dexterity to detect it. The first experimental check

Ref. 660 was by Spaarnay, Casimir's colleague at Philips, in 1958. Two beautiful high-precision measurements of the Casimir effect were performed in 1997 by Lamoreaux and in 1998 by Mohideen and Roy; they confirmed Casimir's prediction with a precision of 5% and 1% respectively. Ref. 661

In a cavity, spontaneous emission is suppressed, if it is smaller than the wavelength of the emitted light! This effect has also been observed. If confirms the old saying that spontaneous emission is emission stimulated by the zero point fluctuations.

The Casimir effect thus confirms the existence of the zero-point fluctuations of the electromagnetic field. It confirms that quantum theory is valid also for electromagnetism.

The Casimir effect between two spheres is proportional to $1/r^7$ and thus is much weaker than between two parallel plates. Despite this strange dependence, the fascination of the Casimir effect led many amateur scientists to speculate that a mechanism similar to the Casimir effect might explain gravitational attraction. Can you give at least three arguments why this is impossible, even if the effect had the correct distance dependence? Challenge 1066 n

Like the case of sound, the Casimir effect can also produce repulsion instead of attraction. It is sufficient that one of the two materials be perfectly permeable, the other a perfect conductor. Such combinations repel each other, as Timothy Boyer discovered in 1974. Ref. 662

The Casimir effect bears another surprise: between two metal plates, the speed of light changes and can be larger than c. Can you imagine what exactly is meant by 'speed of light' in this context? Ref. 663 Challenge 1067

The QED Lagrangian

– CS – section on the QED Lagrangian to be added – CS –

Interactions and virtual particles

The electromagnetic interaction is exchange of virtual photons. So how can the interaction be attractive? At first sight, any exchange of virtual photons should drive the electrons from each other. However, this is not correct. The momentum of virtual photons does not have to be in the direction of its energy flow; it can also be in opposite direction.* Obviously, this is only possible within the limits provided by the indeterminacy principle.

Moving mirrors

Mirrors also work when in motion; in contrast, walls that produce echoes do not work at all speeds. Walls do not produce echoes if one moves faster than sound. However, mirrors always produce an image. This observation shows that the speed of light is the same for any observer. Can you detail the argument? Challenge 1068 n

---

[*] One of the most beautiful booklets on quantum electrodynamics which makes this point remains the text by RICHARD FEYNMAN, *QED: the strange theory of light and matter*, Penguin books, 1990.

Mirrors also differ from tennis rackets. We saw that mirrors cannot be used to change the speed of the light they hit, in contrast to what tennis rackets can do with balls. This observation shows that the speed of light is also a limit velocity. In short, the simple existence of mirrors is sufficient to derive special relativity.

But there are more interesting things to be learned from mirrors. We only have to ask whether mirrors work when their motion is accelerated. It turns out that in general they don't.

In the 1970s, quite a number of researchers found that there is no vacuum for accelerated observers. This effect is variously called *dynamical Casimir effect* or *Davies-Unruh effect*.

As a consequence, a mirror in accelerated motion reflects the fluctuations it encounters and reflects them. In short, an accelerated mirror emits light. Unfortunately, the intensity is so weak that it has not been measured up to now. Can you explain why accelerated mirrors emit light, but not matter?

Challenge 1069 n

Photon-photon scattering

When virtual particles are taken into account, light beams can 'bang' onto each other. This result is in contrast to classical electrodynamics. Indeed, QED shows that the virtual electron-positron pairs allow photons to hit each other. And such pairs are found in any light beam.

However, the cross section is small. When two beams cross, most photons will pass undisturbed. The cross section A is approximately

$$A \approx \frac{973}{10\,125\pi}\alpha^4 \left(\frac{\hbar}{m_e c}\right)^2 \left(\frac{\hbar\omega}{m_e c^2}\right)^6 \qquad (463)$$

for the case that the energy $\hbar\omega$ of the photon is much smaller than the rest energy $m_e c^2$ of the electron. This value is about 18 orders of magnitude smaller than what was measurable in 1999; future will show whether the effect can be observed for visible light. However, for high energy photons these effects are observed daily in particle accelerators. In these cases one observes not only interaction through virtual electron-antielectron pairs, but also through virtual muon-antimuon pairs, virtual quark-antiquark pairs, and much more.

Is the vacuum a bath?

If the vacuum is a sea of virtual photons and particle-antiparticle pairs, vacuum could be suspected to act as a bath. In general, the answer is negative. Quantum field theory works because the vacuum is *not* a bath for single particles. However, there is always an exception. For dissipative systems made of many particles, such as electrical conductors, the vacuum *can* act as a viscous fluid. Irregularly shaped, neutral, but conducting bodies can emit photons when accelerated, thus damping such type of motion. This is due to the Davies-Unruh effect, also called dynamical Casimir effect, as described above. The damping depends on the shape and thus also on the direction of the body's motion.

Ref. 664

See page 176

In 1998, Gour and Sriramkumar even predicted that Brownian motion should also appear for an imperfect, i.e. partly absorbing mirror placed in vacuum. The fluctuations of the vacuum should produce a mean square displacement

Ref. 665

$$< d^2 > = \hbar/mt \qquad (464)$$

increasing with time; however, the extremely small displacements produced this way seem out of experimental reach so far. But the result is not a surprise. Are you able to give another, less complicated explanation for it? Challenge 1070

Renormalization – Why is an electron so light?

– CS – section on renormalization to be added – CS –

Some curiosities and challenges of quantum electrodynamics

Motion is an interesting topic, and when a curious person asks a question about 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. Let us have a look at some of them.

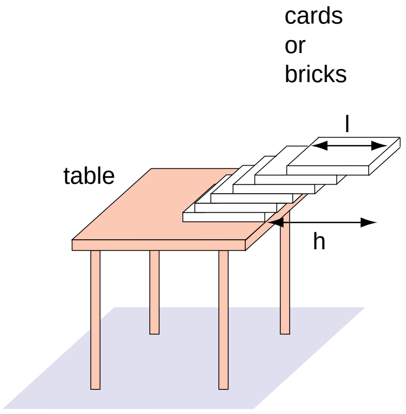


Figure 231 What is the maximum possible value of h/l?

- There is a famous riddle asking how far the last card (or the last brick) of a stack can hang over the edge of a table. Of course, only gravity, no glue or any other means is allowed to keep the cards on the table. After you solved the riddle, can you give the solution in case that the quantum of action is taken into account? Challenge 1071 n
- 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 as charged black holes, as we will see shortly. Ref. 575 Challenge 1072 n See page 640
- Taking 91 of the 92 electrons off an uranium atom allows researchers to check whether the innermost electron still is described by QED. The electric field near the uranium nucleus, 1 EV/m is near the threshold for spontaneous pair production. The field is the highest constant field producible in the laboratory, and an ideal testing ground for precision QED experiments. The effect of virtual photons is to produce a Lamb shift; even in these extremely high fields, the value fits with the predictions. Ref. 668
- Is there a critical magnetic field in nature, like there is a critical electric field? Challenge 1073
- In classical physics, the field energy of a point-like charged particle, and hence its mass, was predicted to be infinite. QED effectively *smears out* the charge of the electron over its Compton wavelength, so that in the end the field energy contributes only a small correction to its total mass. Can you confirm this? See page 419 Challenge 1074

- Microscopic evolution can be pretty slow. Light, especially when emitted by single atoms, is always emitted by some metastable state. Usually, the decay times, being induced by the vacuum fluctuations, are much shorter than a microsecond. However, 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 $3\hbar$; this is an extremely unlikely process. Ref. 669
- 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 1075 n
- 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. In short, both quantum mechanics and special relativity show that rigid bodies do not exist, albeit for different reasons.
- Have you ever admired a quartz crystal or some other crystalline material? The beautiful shape and atomic arrangement has 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 complex and interesting.

For example, are regular crystal lattices *energetically* optimal? This simple question leads to a wealth of problems. We 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 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. Challenge 1076 n Ref. 671

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 temperatures*, regular sphere arrangements indeed show the largest possible entropy. At low temperatures, spheres in a crystal can oscillate around their average position and be thus more disordered than if they were in a liquid; in the liquid state the spheres would block each other's motion and would not allow to show disorder at all.

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. For the same reason, one gets the surprising result that even slightly repulsing spheres (or atoms) can form crystals, and melt at higher temperatures. These are beautiful examples of how classical thinking can explain certain material properties, using from quantum theory only the particle model of matter. Ref. 672

But the energetic side of crystal formation provides other interesting questions. 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. 673

For an exact study of crystal energy, the interactions between all atoms have to be included. The simplest 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; the answer is still open.

Another question is the mechanism of face formation in crystals. Can you confirm that crystal faces are those planes with the *slowest* growth speed, because all fast growing planes are eliminated? The finer details of the process form a complete research field in itself. Challenge 1077 n

Ref. 674

Finally, there remains the question of symmetry: why are crystals often symmetric, such as snowflakes, instead of asymmetric? This issue is a topic of self-organization, as mentioned already in the section of classical physics. It turns out that the symmetry is an automatic result of the way molecular systems grow under the combined influence of diffusion and nonlinear processes. The issue is still a topic of research. See page 186

- A similar breadth of physical and mathematical problems are encountered in the study of liquids and polymers. The ordering of polymer chains, the bubbling of hot water, the motion of heated liquids and the whirls in liquid jets show complex behaviour that can be explained with simple models. Turbulence and self-organization will be a fascinating research field for many years to come. Ref. 675
- 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 in physics 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 can behave like a Foucault pendulum. With a clever arrangement, it was possible to measure the rotation of the helium in the ring using phonon signals, and to show the rotation of the earth. Ref. 676
- If an electrical wire is sufficiently narrow, its electrical conductance is quantized in steps of $2e^2/\hbar$. The wider the wire, the more such steps are added to its conductance. Can you explain the effect?By the way, quantized conductance has also been observed for light and for phonons. Challenge 1078 Ref. 677
- 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. They appear when a completely ionized atom, i.e. one without any electrons, is brought in contact with a metal. The acquired electrons then orbit on the outside, leaving the inner shells empty, in stark contrast with usual atoms. Such hollow atoms can also be formed by intense laser irradiation. Ref. 678
- In the past, the description of motion with formulas was taken rather seriously. Before computers appeared, only those examples of motion 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.
- Can you explain why mud is not clear? Challenge 1079 n

▪ Photons not travelling parallel to each other attract each other through gravitation, and thus deflect each other. Could two such photons form a bound state, a sort of atom of light, in which they would circle each other, provided there were enough empty space for this to happen? Challenge 1080 n

▪ Can the universe ever have been smaller than its own Compton wavelength? Challenge 1081

In fact, quantum electrodynamics, or QED, provides a vast number of curiosities, and every year there is at least one interesting new discovery. We now conclude the theme with a more general approach.

How can one move on perfect ice? – The ultimate physics test

In our quest, we have encountered motion of many sorts. Therefore, the following test – not to be taken too seriously – is the ultimate physics test, allowing 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? Any method, so tiny its effect may be, is allowed.

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 making use of the location of the surface on earth? Challenge 1082 n 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 domains? Challenge 1083 n

General relativity showed that turning one arm will emit gravitational radiation unsymmetrically, leading to motion as well. Can you find at least two better methods? Challenge 1084 n

Quantum theory offers a wealth of methods. Of course, quantum mechanics shows that we actually are 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 able to find at least four methods of moving on perfect ice due to quantum effects? Challenge 1085 n

Material science, geophysics, atmospheric physics, and astrophysics also provide ways to move, such as cosmic rays or solar neutrinos. Can you find four additional methods? Challenge 1086 n

Self-organization, chaos theory, and biophysics also provide ways to move, when the inner workings of the human body are taken into account. Can you find at least two methods? Challenge 1087 n

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? Challenge 1088 n Can you find at least two other methods to move yourself using quantum gravity effects? Can you find one from string theory?

If you want points for the test, the marking 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 the following: *matter is made of charged particles which interact through photon exchange in the way as described by Figure 232.*

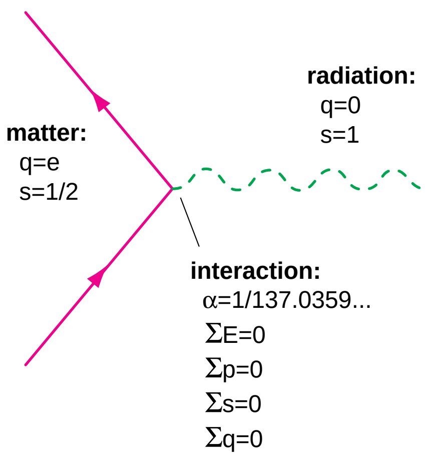


Figure 232 QED as perturbation theory in space-time

No additional information is necessary. 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. *Interactions* between charged particles are described as the exchange of virtual photons, and *decay* is described as the interaction with the virtual photons of the vacuum.

All physical results of QED can be calculated by using the single diagram of Figure 232. As QED is a perturbative theory, the diagram directly describes the first order effects and its composites describe effects of higher order. QED is a perturbative theory.

QED describes all everyday properties of matter and radiation. It describes the *divisibility* down to the smallest constituents, the *isolability* from the environment and the *impenetrability* of matter. It also describes the *penetrability* of radiation. All these properties are due to electromagnetic interactions of constituents and follow from Figure 232. 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, in particular because matter constituents are fermions. Radiation is divisible into photons, and is penetrable because photons are bosons and first order photon-photon interactions do not exist.

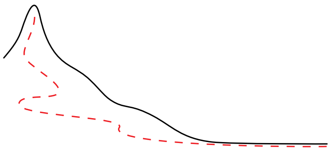
Both matter and radiation are made of elementary constituents. These elementary constituents, whether bosons or fermions, are indivisible, isolable, indistinguishable, and point-like.

To describe observations, it is necessary to use quantum electrodynamics in all those situations for which the characteristic dimensions d are of the order of the Compton wavelength

$$d \approx \lambda_C = \frac{h}{m\,c} \quad . \tag{465}$$

---

* The author keeps track of all answers on the http://www.motionmountain.org web site.



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

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:

$$d \approx \lambda_{\mathrm{dB}} = \frac{h}{m\,v} \quad . \qquad (466)$$

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 and all images are described by it, including their properties, their shape, their transformations and their other changes. This includes self-organization and chemical or biological. In other words, QED gives us 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. In fact, all examples of electrification, such as the charging of amber through rubbing, the experiment which gave electricity its name, are still poorly understood. Ref. 680 Ref. 681

Material science in all its breadth, including the study of solids, fluids, and plasmas, as well as biology and medicine, still provides many topics of research. In particular, the 21st century will undoubtedly be the century of the life sciences.

The study of 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. 683

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 mechanisms. Ref. 682

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 higher order interaction diagrams built using the fundamental diagram of Figure 232 contain relations to the Ref. 684

---

* 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. 679

theory of knots. This research topic will provide even more interesting results in the near future.

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 and where particle numbers are not conserved, promise a wealth of new insights. We will return to the topic later on.

High energies provide many more questions. So far, the description of motion was based on the idea that measurable quantities can be multiplied and added. This always happens at one space-time point. In mathematical jargon, observables form a *local* algebra. Thus the structure of an algebra contains, implies, and follows from the idea that local properties *lead* to local properties. We will discover later on that this basic assumption is wrong at high energies.

Challenge 1090 d We defined special relativity using $v \leqslant c$, general relativity using $L/M \geqslant 4G/c^2$ and quantum theory using $S \geqslant \hbar/2$. How can we define electromagnetism in one statement? Nobody knows yet.

Many other open issues of more practical nature have not been mentioned. Indeed, by far the largest numbers of physicists get paid for some form of applied QED. However, our quest is the description of the *fundamentals* of motion. So far, we have not achieved it. For example, we still need to understand motion in the realm of atomic nuclei. But before we do that, we take a first glimpse of the strange issues appearing when gravity and quantum theory meet.

25. Quantum mechanics with gravitation – first approaches

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 a list of important issues.

A collapsing house

Figure to be inserted

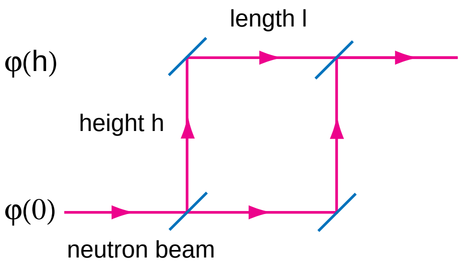


Figure 233 The weakness of gravitation

Challenge 1091 In the chapter on general relativity we already mentioned that light frequency changes with height. But gravity also changes the *phase* of matter wavefunctions. Can you imagine why?The effect was first confirmed in 1975 with help of neutron interferometers, where neutron beams are brought to interference after having climbed some height h at two different locations. The experiment is shown schematically in Figure 233; it exactly confirmed the predicted phase difference

$$\delta\varphi = \frac{mghl}{\hbar v} \qquad (467)$$

where l is the distance of the two climbs, and v and m are the speed and mass of the neutrons. These beautifully simple experiments have confirmed the formula within experimental errors.[*] Ref. 685

In the 1990s, similar experiments have even been performed with complete atoms. These set-ups allow to build interferometers so sensitive that local gravity g can be measured with a precision of more than eight significant digits. Ref. 687

Corrections to the Schrödinger equation

In 2002, the first observation of actual quantum states due to gravitational energy was performed. Any particle above the floor should feel the effect of gravity. Ref. 688

In a few words, one can say that because the experimenters managed to slow down neutrons to the incredibly small value of 8 m/s, using grazing incidence on a flat plate they could observe how neutrons climbed and fell back due to gravity with speeds below a few cm/s.

Obviously, the quantum description is a bit more involved. The lowest energy level for neutrons due to gravity is $2.3 \cdot 10^{-31}$ J, or 1.4 peV. To get an impression of it smallness, we can compare it to the value of $2.2 \cdot 10^{-18}$ J or 13.6 eV for the lowest state in the hydrogen atom.

A rephrased large number hypothesis

Despite its weakness, gravitation provides many puzzles. Most famous are a number of curious coincidences that can be found when quantum mechanics and gravitation are combined. They are usually called 'large number hypotheses' because they usually involve large dimensionless numbers. A pretty, but less well known version connects the Planck length, the cosmic horizon, and the number of baryons: Ref. 706

$$(N_\mathrm{b})^3 \approx \left(\frac{R_\mathrm{o}}{l_\mathrm{Pl}}\right)^4 = \left(\frac{t_\mathrm{o}}{t_\mathrm{Pl}}\right)^4 \approx 10^{244} \qquad (468)$$

in which $N_\mathrm{b} = 10^{81}$ and $t_\mathrm{o} = 1.2 \cdot 10^{10}$ a were used. There is no known reason why the number of baryons and the horizon size should be related in this way. This coincidence is equivalent to the one originally stated by Dirac,[**] namely

$$m_\mathrm{p}^3 \approx \frac{\hbar^2}{Gct_\mathrm{o}} \quad . \qquad (470)$$

---

[*] Due to the influence of gravity on phases of wavefunctions, some people who do not believe in bath induced decoherence have even studied the influence of gravity on the decoherence process of usual quantum systems in flat space-time. Predictably, the results have not convinced. Ref. 686

[**] The equivalence can be deduced using $Gn_\mathrm{b}m_\mathrm{p} = 1/t_\mathrm{o}^2$, which, as Weinberg explains, is required by several cosmological models. Indeed, this can be rewritten simply as Ref. 707 See page 769

$$m_\mathrm{o}^2/R_\mathrm{o}^2 \approx m_\mathrm{Pl}^2/R_\mathrm{Pl}^2 = c^4/G^2 \quad . \qquad (469)$$

where m_p is the proton mass. This approximate equality seems 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 speculations, especially since the time dependence of the two sides differs. Some people even speculate whether relations (468) or (470) express some long-sought relation between local and global topological properties Ref. 708 of nature. Up to this day, the only correct statement seems to be that they are *coincidences* connected to the time at which we happen to live. But gravity also leads to other quantum surprises.

Limits to disorder

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

We have already encountered the famous statement by Clausius, the father of the term See page 329 'entropy'. Strangely, for over hundred years nobody asked whether there actually exists a theoretical maximum for entropy. This changed in 1973, when Jakob Bekenstein found Ref. 689 the answer while investigating the consequences gravity has for quantum physics. He found that the entropy of an object of energy E and size L is bound by

$$S \leqslant EL\frac{k\pi}{\hbar c} \qquad (471)$$

for all physical systems. In particular, he deduced that (nonrotating) black holes saturate the Challenge 1092 bound, with an entropy given by

$$S = \frac{kc^3}{G\hbar}\frac{A}{4} = \frac{kG}{\hbar c}4\pi M^2 \qquad (472)$$

where A is now the area of the *horizon* of the black hole, given by $A = 4\pi R^2 = 4\pi(2GM/c^2)^2$. In particular, the result implies that every black hole has an entropy. Black holes are thus disordered systems described by thermostatics. Black holes are the most disordered systems known.

As an interesting note, the maximum entropy also gives a memory limit for memory Challenge 1093 chips. Can you find out how?

Which 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 Ref. 690 of the horizon itself, and that they are due to the diffeomorphism invariance at this boundary. As Gerard 't Hooft explains, the entropy expression implies that the number of degrees of Challenge 1094 freedom of a black hole is about (but not exactly) one per Planck area of the horizon.

Together with the definition of the baryon density $n_b = N_b/R_0^3$ one gets Dirac's large number hypothesis, substituting protons for pions. Note that the Planck time and length are defined as $\sqrt{\hbar G/c^5}$ and $\sqrt{\hbar G/c^3}$ and are the natural units of length and time. We will study them in detail in the third part of the mountain ascent.

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

If black holes have entropy, they must have a temperature. What does this temperature 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, even though, incredibly, nobody had thought about them for over 40 years.

How to measure acceleration with a thermometer – Davies-Unruh radiation

In 1973, Paul Davies and William Unruh independently made a theoretical discovery while studying quantum theory: if an inertial observer observes that he is surrounded by vacuum, a second observer *accelerated* with respect to the first does not: he observes black body radiation, with a spectrum corresponding to the temperature Ref. 691

$$T = a\frac{\hbar}{2\pi kc} \quad . \tag{473}$$

The result means that there is no vacuum on earth, because an observer on its surface can maintain that he is accelerated with $9.8\,\mathrm{m/s^2}$, thus leading to $T = 40$ zK! We can thus measure gravity, at least in principle, using a thermometer. However, even for the largest practical accelerations the temperature values are so small that it is questionable whether the effect will ever be confirmed experimentally. But if it will, it will be a great discovery. Ref. 692

When this effect was predicted, people studied the argument from all sides. For example, it was then found that the acceleration of a *mirror* leads to radiation emission! Mirrors are thus harder to accelerate than other bodies of the same mass.

When the acceleration is high enough, also *matter* particles can be detected. If a particle counter is accelerated sufficiently strongly across the vacuum, it will start counting particles! We see that the difference between vacuum and matter becomes fuzzy at large energies.

For completeness, we mention that also an observer in rotational motion detects radiation following expression (473).

Black holes aren't black

In 1974, the English physicist Stephen Hawking, famous for the courage with which he fights a disease which forces him into the wheelchair, surprised the world of general relativity with a fundamental theoretical discovery. He found that if a virtual particle-antiparticle pair appeared in the vacuum near the horizon, there is a finite chance that one particle escapes as a real particle while the antiparticle is captured by the black hole. This is true both for fermions and for bosons. From far away this effect looks like the emission of a particle. Hawking's detailed investigation showed that black holes *radiate* as black bodies.

Black hole radiation confirms both the result on black hole entropy by Bekenstein as the effect for observers accelerated in vacuum found by Davies and Unruh. When all this became clear, a beautiful Gedankenexperiment (thought experiment) was published by William Unruh and Robert Wald, showing that the whole result could have been deduced already 50 years before! Ref. 693

Shameful as this delay of the discovery is for the community of theoretical physicists, the story itself remains beautiful. It starts in the early 1970s, when Robert Geroch studied the issue shown in Figure 234. Imagine a mirror box full of heat radiation, thus light. The mass of the box is assumed to be negligible, such as a box made of thin aluminium paper. We lower the box, with all its contained radiation, from a space station towards a black hole. On the space station, lowering the weight of the heat radiation allows to generate energy. Obviously, when the box reaches the black hole horizon, the heat radiation is red shifted to infinite wavelength. At that point, the full amount of energy originally contained in the heat radiation has been provided to the space station. We can now do the following: we can open the box on the horizon, let drop out whatever is still inside, and wind the empty and massless box back up again. As a result, we have completely converted heat radiation into mechanical energy. Nothing else has changed: the black hole has the same mass as beforehand.

But this result contradicts the second principle of thermodynamics! Geroch concluded that something must be wrong. We must have forgotten an effect which makes this process impossible.

In the 1980s, Unruh and Wald showed that black hole radiation is precisely the forgotten effect that puts everything right again. Because of black hole radiation, the box feels buoyancy, so that it cannot be lowered down to the horizon. It floats somewhat above it, so that the heat radiation inside the box has not yet *zero* energy when it falls out of the opened box. As a result, the black hole does increase in mass, and thus in entropy. In summary, when the empty box is pulled up again, the final situation is thus the following: only part of the energy of the heat radiation has been converted into mechanical energy, part of the energy went into the increase of mass and thus of entropy of the black hole. The second principle of thermodynamics is saved.

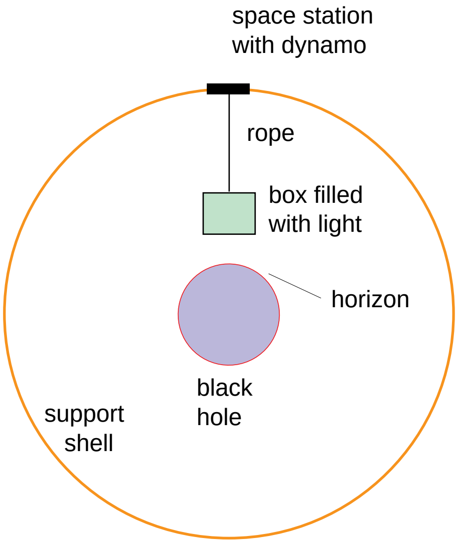


Figure 234 A *Gedankenexperiment* allowing to deduce the existence of black hole radiation

Well, it is only saved if the heat radiation has precisely the right energy density at the horizon and above. Let us have a look. The centre of the box can only be lowered up to a hovering distance d above the horizon, where the acceleration due to gravity is $g = c^2/4GM$. The energy E gained by lowering the box is

$$E = mc^2 - mg\frac{d}{2} = mc^2(1 - \frac{dc^2}{8GM}) \qquad (474)$$

The efficiency of the process is $\eta = E/mc^2$. To be consistent with the second law of thermodynamics, this efficiency must obey

$$\eta = \frac{E}{mc^2} = 1 - \frac{T_{\mathrm{BH}}}{T}e \qquad (475)$$

We thus find a black hole temperature T_{BH} given by the hovering distance d. That hovering distance d is roughly given by the size of the box. The box size in turn must be at least the wavelength of the thermal radiation; in first approximation, Wien's relation gives $d \approx \hbar c/kT$.

The more precise calculation, first performed by Hawking, gives the result

$$T_{\mathrm{BH}} = \frac{\hbar c^3}{8\pi kGM} = \frac{\hbar}{2\pi kc}\, g_{\mathrm{surf}} \quad \text{with} \quad g_{\mathrm{surf}} = \frac{c^4}{4GM} \tag{476}$$

where M is the mass of the black hole. It is either called the black-hole temperature or Bekenstein-Hawking temperature. As an 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. All known black hole candidates have masses in the range from a few to a few million solar masses. The radiation is thus extremely weak, the reason being that the emitted wavelength is of the order of the black hole radius, as you might want to check.The radiation emitted by black holes is often also called *Bekenstein-Hawking radiation*. Challenge 1095

This rather academic effect leads to a luminosity Challenge 1096

$$L \sim \frac{1}{M^3} \quad \text{or} \quad L = nA\sigma T^4 = \frac{n\pi^3 k^4}{15c^2\hbar^3} T^4 \tag{477}$$

where n is the number of particle degrees of freedom that can be radiated; if only photons are radiated, we have $n = 2$. (Actually, massless neutrinos are emitted more frequently than photons.) *Black holes thus shine*, and the more the smaller they are. This is a genuine *quantum* effect, since classically, black holes cannot emit any light. Ref. 694

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 1097

$$t = M^3 \frac{20\,480\,\pi G^2}{\hbar c^4} \approx M^3\, 3.4 \cdot 10^{-16}\,\mathrm{s/kg^3} \tag{478}$$

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. In any case, *black holes evaporate*. However, this extremely slow process for usual black holes determines their lifetime only if no other, faster process comes into play. We will present a few such processes shortly. Hawking radiation is the weakest of all known effects. It is not masked by stronger effects only if the black hole is non-rotating, electrically neutral and with no matter falling into it from the surroundings.

So far, none of these quantum gravity effects has been confirmed experimentally, as the values are much too small to be detected. However, the deduction of a Hawking temperature has been beautifully confirmed 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. A second type of analogue system, namely *optical black holes*, are also being investigated. Ref. 695 Ref. 696

In 1975, a much more dramatic radiation effect than black hole radiation was predicted for *charged* black holes by Damour and Ruffini. Charged black holes have a much shorter Ref. 697

lifetime than just presented, because during their formation a second process takes place. In a region surrounding them the electric field is larger than the so-called vacuum polarization value, so that large numbers of electron-positron pairs are produced, which then almost all annihilate. This process effectively reduces the charge of the black hole to a value for which the field is below critical everywhere, while emitting large amounts of high energy light. It turns out that the mass is reduced by up to 30% in a time of the order of seconds. That is quite shorter than 10^{68} years. This process thus produces an extremely intense *gamma ray burst*.

Such gamma ray bursts had been discovered in the late 1960s by military satellites which Ref. 698 were trying to spot nuclear explosions around the world through their gamma ray emission. The satellites found about two such bursts per day, coming from all over the sky. Another satellite, the Compton satellite, confirmed that they were extragalactic in origin, and 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* in the X-ray domain of many hours, sometimes of days. In 1997 afterglow was discovered also in the optical domain. The satellite also allowed to find the corresponding X-ray, optical, and radio sources for each burst. These measurements in turn allowed to determine the distance of the burst Ref. 699 sources; red shifts between 0.0085 and 4.5 were measured. In 1999 it also became possible to detect *optical* bursts corresponding to the gamma ray ones.*

All this data together shows that the gamma ray bursts have energies ranging from 10^{40} W to $3 \cdot 10^{47}$ W. The larger value is (almost) the same brightness as that of all stars of the Challenge 1098 n whole visible universe taken together! Put differently, it is the same amount of energy that is 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. This limit is given by

$$L < L_{\mathrm{Pl}} = \frac{c^5}{2G} = 1.8 \cdot 10^{52}\,\mathrm{W} \quad , \tag{479}$$

Challenge 1099 as you might want to check yourself.In short, the sources of gamma ray bursts are the Ref. 699 biggest bombs found in the universe. In fact, more detailed investigations of experimental data confirm that gamma ray bursts are ‘primal screams’ of black holes in formation.

With all this new data, Ruffini took up his 1975 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 afterglow discovered by the BeppoSAX satellite. Among others, Ruffini’s team speculates that

---

* For more about this fascinating topic, see the http://www.aip.de/˜jcg/grb.html web site by Jochen Greiner.

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. 700

Other processes leading to emission of radiation appear when matter falls into the black hole and heats up, when matter is ejected from rotating black holes through the Penrose process, 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. They are assumed to be black holes surrounded by matter, in the development stage following gamma ray bursters. The details of what happens in quasars, the enormous voltages (up to 10^{20} V) and magnetic fields generated, as well as their effects on the surrounding matter are still object of intense research in astrophysics. See page 345

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 object. For example, 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 is given by

$$\rho = \frac{1}{M^2}\frac{3c^6}{32\pi G^3} \tag{480}$$

and 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

$$g_{\rm surf} = \frac{1}{M}\frac{c^4}{4G} = \frac{c^2}{2R} \tag{481}$$

which is still 15 $\mathrm{km/s^2}$ for an air density black hole. Challenge 1100

Obviously, the black hole temperature is related to the entropy by its usual definition

$$\frac{1}{T} = \left.\frac{\partial S}{\partial E}\right|_\rho = \left.\frac{\partial S}{\partial (Mc^2)}\right|_\rho \tag{482}$$

All other thermal properties can be deduced by the standard relations from thermostatics.

In particular, it looks as if black holes are the matter states with the largest possible entropy. Can you confirm this statement? Challenge 1101

It also turns out that black holes have a *negative* heat capacity: when heat is added, they cool down. In other words, black holes cannot achieve equilibrium with a bath. This is not a real 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 1102 Challenge 1103 n Ref. 702

Black holes, like any object, oscillate when slightly perturbed. These vibrations have also been studied; their frequency is proportional to the mass of the black hole. Ref. 703

Nonrotating black holes have no magnetic field, as was established already in the 1960s by Russian physicists. On the other hand, black holes have something akin to a finite elec- Ref. 694

trical conductivity and a finite viscosity. Some of these properties can be understood if the horizon is described as a membrane, even though this model is not always applicable. In any case, one can study and describe macroscopic black holes like any other macroscopic material body. The topic is not closed. Ref. 701

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. Expression (478) for the lifetime, applied to a black hole of Planck mass, yields a value of over sixty thousand Planck times. A surprising large value. What happens in those last instants of evaporation?

A black hole approaching the Planck mass at some time will get 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 these final evaporation steps take 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), how the emitted radiation deviates from black body radiation. There is still enough to study. 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, or equivalently, that entropy increases.

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, we need to look at it with precision, laying all prejudice aside. Three intermediate questions can help us 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 1104 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 – or low entropy – object is thrown into the sun, the information contained is radiated away. The information 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, comparing the entropy of a room temperature book and the information contained in it, shows that these effects are extremely small and difficult to measure.

A clear exposition of the topic was given by Don Page. He calculated what information would be measured 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. The result is simple. Even if a system is large – consisting of many degrees of freedom – and in pure state, any smaller *sub*system nevertheless 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 entropy S_m given by Ref. 704 Challenge 1105

$$S_m = \frac{1-m}{2n} + \sum_{k=n+1}^{mn} \frac{1}{k} \tag{483}$$

which is approximately given by

$$S_m = \ln m - \frac{m}{2n} \quad \text{for} \quad m \gg 1 \quad . \tag{484}$$

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 (484) 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 state; it looks like a thermal system even though it is not.

A calculation shows that the second, small term on the right of equation (484) 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 not only it is lost in measurements; it is also lost in the usual, perturbative calculations for physical systems.

The question whether any radiated information could be measured can now be answered directly. As Don Page showed, 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. In summary, at a given instant, the amount of information radiated by a black hole is negligible when compared with the total black hole radiation, and is practically impossible to detect by measurements or even by usual calculations. Ref. 705

More paradoxes

A black hole is a macroscopic object, similar to a star. Like all objects, it can interact with its environment. It has the special property to swallow everything that falls into them. This immediately leads us to ask if we can use this property to cheat around the usual everyday 'laws' of nature. Some attempts have been studied in the section on general relativity and above; here are a few more. See page 346

- Apart from the questions of entropy, we can look for methods to cheat around conservation of energy, angular momentum, or charge.Every Gedankenexperiment comes to the same conclusions. No cheats are possible; in addition, 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 for quite some time. Challenge 1106
- If a twin falls into a large black hole, according to general relativity he makes no special observations. However, in his brother's view he burns on the horizon.As explained before, Challenge 1107

while burning, he – or his ashes – also spread all over the horizon. We thus find a strange situation: adding quantum theory to the study of black holes shows that in contrast to what general relativity says, there is no way to cross the horizon unharmed. Quantum theory and general relativity contradict each other. We will study more such puzzles later on.

▪ 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 1108

Quantum mechanics of gravitation

Let us take a conceptual step at this stage. So far, we looked at quantum theory *with* gravitation; now we have a glimpse at quantum theory *of* gravitation.

If we focus on the similarity between the electromagnetic field and the gravitational 'field,' we should try to find the quantum description of the latter. Despite attempts 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 gravitational Bohr radius of Challenge 1109

$$r_{\text{gr.B.}} = \frac{\hbar^2}{G m_e^2 m_p} = 1.1 \cdot 10^{29}\,\text{m} \qquad (485)$$

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 quest these questions cannot be answered. Worse, the weakness of gravity even means that with high probability, future experiments will provide little additional data helping to decide among competing answers. The only help is careful thought. Challenge 1110

Decoherence of space-time

If the gravitational field evolves like a quantum system, we encounter all issues found in other quantum systems. General relativity taught us that the gravitational field and space-time are the same. As a result, we may ask why no superpositions of different macroscopic space-times are observed.

Ref. 709 The discussion is simplified for the simplest case of all, namely the superposition, in a vacuum region of size l, of a homogeneous gravitational field with value g and one with value g'. As in the case of a superposition of macroscopic distinct wavefunctions, such a superposition *decays*. In particular, it decays when particles cross the volume. A short calculation yields a decay time givenby See page 584 Challenge 1111

---

* Modern approaches take another direction, as explained in the third part of the mountain ascent.

$$t_d = (\frac{2kT}{\pi m})^{3/2} \frac{nl^4}{(g-g')^2} \quad , \tag{486}$$

where n is the particle number density, kT their kinetic energy, and m their mass. Inserting typical numbers, we find that the variations in gravitational field strength are *extremely* small.In fact, the numbers are so small that we can deduce that the gravitational field is the *first* variable which behaves classically in the history of the universe. Quantum gravity effects for space-time will thus be extremely hard to detect. Challenge 1112

In short, matter not only tells space-time how to curve, it also tells it to behave with class. This result calls for the following question.

Do gravitons exist?

Quantum theory says that everything that moves is made of particles. What kind of particles are gravitational waves made of? If the gravitational field is to be treated quantum mechanically like the electromagnetic field, its waves should be quantized. Most properties of these quanta can be derived in a straightforward way.

The $1/r^2$ dependence of universal gravity, like that of electricity, implies that the particles have *vanishing* mass and move at light speed. The independence of gravity from electromagnetic effects implies a *vanishing* electric charge.

The observation that gravity is always attractive, never repulsive, means that the field quanta have *integer* and *even* spin. Vanishing spin is ruled out, since it implies no coupling to energy. To comply with the property that 'all energy has gravity', $S = 2$ is needed. In fact, it can be shown that *only* the exchange of a massless spin 2 particle leads, in the classical limit, to general relativity. Ref. 710

The coupling strength of gravity, corresponding to the fine structure constant of electromagnetism, is given either by

$$\alpha_{G1} = \frac{G}{\hbar c} = 2.2 \cdot 10^{-15}\,\mathrm{kg}^{-2} \quad \text{or by} \quad \alpha_{G2} = \frac{Gmm}{\hbar c} = \left(\frac{m}{m_{\mathrm{Pl}}}\right)^2 = \left(\frac{E}{E_{\mathrm{Pl}}}\right)^2 \tag{487}$$

However, the first expression is not a pure number; the second expression is, but depends on the mass one inserts. These difficulties reflect the fact that gravity is not properly speaking an interaction, as became clear in the section on general relativity. It is often argued 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 indeed weak compared to electromagnetism, for which $\alpha_{\mathrm{em}} = 1/137.04$.

If all this is correct, *virtual* field quanta would also have to exist, to explain static gravitational fields.

However, up to this day, the so-called *graviton* has not yet been detected, and there is in fact little hope that it ever will. On the experimental side, nobody knows yet how to build a graviton detector. Just try! On the theoretical side, the problems with the coupling constant probably make it impossible to construct a *renormalizable* theory of gravity; the lack of renormalization means the impossibility to define a perturbation expansion, and thus to define particles, including the graviton. In summary, it may be that the particle concept has to be changed before applying quantum theory to gravity. The issue is still open at this point. Challenge 1113 n

Space-time foam

The indeterminacy relation for momentum and position also applies to the gravitational field. As a result, it leads to an expression for the indeterminacy of the metric tensor g in a region of size L, which is given by

$$\Delta g \approx 2 \frac{{l_{\mathrm{Pl}}}^2}{L^2} \quad , \tag{488}$$

Challenge 1114 where $l_{\mathrm{Pl}} = \sqrt{\hbar G/c^3}$ is the Planck length. Can you deduce the result?Quantum theory thus shows that like the momentum or the position of a particle, also the metric tensor g is a *fuzzy* observable.

But that is not all. 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

$$g \geqslant \frac{2\hbar G}{c^3} \frac{1}{L^2} \quad . \tag{489}$$

Can you confirm this? The result has far-reaching consequences. A minimum value for the metric depending inversely on the region size implies that it is impossible to say what happens to the shape of space-time at extremely small dimensions. In other words, at extremely high energies, the concept of space-time itself becomes fuzzy. John Wheeler introduced the term *space-time foam* to describe this situation. The term makes clear that space-time is not continuous nor a manifold in those domains. But this was the basis on which we built our description of nature so far! We are forced to deduce that our description of nature is built on sand. This issue will form the start of the third part of our mountain ascent. See page 726

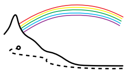
No particles

Gravity has another important consequence for quantum theory. To count and define particles, quantum theory needs a defined vacuum state. However, the vacuum state cannot be defined when the curvature radius of space-time, instead of being larger than the Compton wavelength, becomes comparable to it. In such highly curved space-times, particles cannot be defined. The reason is the impossibility to distinguish the environment from the particle in these situations: in the presence of strong curvatures, the vacuum is full of spontaneously generated matter, as black holes show. Now we just saw that at small dimensions, space-time fluctuates wildly; in other words, space-time is highly curved at small dimensions or high energies. In other words, strictly speaking particles cannot be defined; the particle concept is only a low energy approximation! We will explore this strange conclusion in more detail in the third part of our mountain ascent.

See page 561

No cheating any more

This short excursion into the theory of quantum gravity showed that a lot of trouble is waiting. The reason is that up to now, we deluded ourselves. In fact, it was more than that: we cheated. We carefully hid a simple fact: quantum theory and general relativity *contradict* each other. That was the real 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. However, it does contradict *general* relativity. The issues are so dramatic, changing everything from the basis of classical physics to the results of quantum theory, that we devote the beginning of the third part only to the exploration of the contradictions. There will be surprising consequences on the nature of space-time, particles and motion. But before we study these issues, we complete the theme of the the present, second part of the mountain ascent, namely the essence of matter and interactions.



References

603 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 581.

604 H.D. Zeh, *On the interpretation of measurement in quantum theory*, Foundations of Physics **1**, pp. 69–76, 1970. Cited on page 581.

605 Linda Reichl, *A modern course in statistical physics*, Wiley, 2nd edition, 1998. An excellent introduction into thermodynamics. Cited on page 582.

606 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 584.

607 M. Tegmark, *Apparent wave function collapse*, Foundation of Physics Letters, **6**, pp. 571–590, 1993. Cited on page 584.

608 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 Hans Dieter Zeh. Cited on page 586.

609 G. Lindblad, *On the generators of quantum dynamical subgroups*, Communications in Mathematical Physics **48**, pp. 119–130, 1976. Cited on page 586.

610 Wojciech 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 pages 586 and 590.

611 John Bardeen, his talk in Japan. Cited on page 587.

612 The first decoherence measurement was performed by M. Brune & al., *Observing progressive decoherence of the "meter" in a quantum measurement*, Physical Review Letters **77**, pp. 4887–4890, 9 December 1996. Cited on page 587.

613 Other experiments confirming the numerical predictions from decoherence were published by C. Monroe & al., Science **272**, p. 1131, 1996, W.P. Schleich & al., Nature **403**, p. 256, 2000, C.J. Myatt & al., Nature **403**, p. 269, 2000. See also the summary by W.T. Strunz, G. Alber & F. Haake, *Dekohärenz in offenen Quantensystemen*, Physik Journal **1**, pp. 47–52, November 2002. Cited on page 587.

614 K. Baumann, *Quantenmechanik und Objektivierbarkeit*, Zeitschrift für Naturforschung **25a**, pp. 1954–1956, 1970. Cited on page 588.

615 David Bohm, *Quantum Theory*, Prentice-Hall, 1951, pp. 614–622. Cited on page 589.

616 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 589.

617 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 589.

618 John Maddox, ... Cited on page 590.

619 On the superposition of magnetization in up and down directions there are numerous papers. 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 590.

620 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 593.

621 For example, superpositions were observed 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 590.

622 S. Haroche, *Entanglement, decoherence, and the quantum-classical transition*, Physics Today **51**, pp. 36–42, July 1998. An experiment putting atom at two places at once, distant about 80 nm, was published by C. Monroe & al., Science **272**, p. 1131, 1996. Cited on page 590.

623 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 590.

624 The most famous reference on the wave function collapse is chapter IV of the book by Kurt 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 592.

625 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 590.

626 Collapse times have been measured ... Cited on page 595.

627 A clear discussion can be found in S. Haroche & J.-M. Raimond, *Quantum computing: dream or nightmare?*, Physics Today **49**, pp. 51–52, 1996, as well as the comments in Physics Today **49**, pp. 107–108, 1996. Cited on page 591.

628 S. Kochen & E.P. Specker, *The problem of hidden variables in quantum mechanics*, **17**, pp. 59–87, 1967. Cited on page 596.

629 J.F. Clauser, M.A. Horne, A. Shimony & R.A. Holt, Physical Review Letters **23**, p. 880, 1969. The more general and original result is found in J.S. Bell, Physics **1**, p. 195, 1964, Cited on page 597.

630 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 599.

631 M. 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 600.

632 Connections between quantum theory and information theory can be found in X. Boy, ... Cited on page 600.

633 E. Fermi, *Quantum theory of radiation*, Review of Modern Physics **4**, p. 87, 1932. Cited on page .

634 G. Hegerfeldt, *Causality problems for Fermi's two-atom system*, Physical Review Letters **72**, pp. 596–599, 1994, Cited on page .

635 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 .

636 The use of radioactivity for breeding is described by ... Cited on page 604.

637 S.M. Block, *Real engines of creation*, Nature **386**, pp. 217–219, 1997. Cited on page 605.

638 Early results and ideas on molecular motors are summarized by Barbara Goss Levi, *Measured steps advance the understanding of molecular motors*, Physics Today pp. 17–19, April 1995. Newer results are described in R.D. Astumian, *Making molecules into motors*, Scientific American pp. 57–64, July 2001. Cited on pages 605 and 606.

639 R. Bartussek & P. Hänggi, *Brownsche Motoren*, Physikalische Blätter **51**, pp. 506–507, 1995. Cited on page 606.

640 The motorized screw used by viruses was described by … …, Nature **408**, p. 745, 2000. Cited on page 604.

641 To learn to enjoy life to the maximum there is no standard way. A good foundation can be found in those books which teach the ability to those which have lost it. Cited on page 609.

The best experts are those who help others to overcome traumas. Peter A. Levine, *Waking the tiger - ...*, ...Geoff Graham, *How to Become the Parent You Never Had - a Treatment for Extremes of Fear, Anger and Guilt*, Real Options Press, 1986.

A good complement to these texts is the systemic approach, as presented by Bert Hellinger, *Zweierlei Glück*, Carl Auer Verlag, 1997. Some of his books are also available in English. The author explains how to achieve a life of contentness, living with the highest possible responsibility of one's actions, by reducing entaglements with one's past and with other people. He presents a simple and efficient technique for realizing this disentanglement.

The next step, namely full mastery in the enjoyment of life, can be copied from any book written by somebody who has achieved mastery in any one topic. The topic itself is not important, only the passion is.

Plato, *Phaedrus*, A beautiful text ...

A. De La Garanderie, ... and his other books. The author is expert on teaching and learning, especially on the importance of the evocation, imagination and motivation.

Françoise Dolto, ..., and her other books. The author, child psychiatrist, is one of the world experts on the growth of the child; her main theme is that growth is only possible by giving the highest possible responsibility to every child during its evolution.

In the domain of art, many had the passion to achive full pleasure. A good piece of music, a beautiful painting, an expressive statue or a good movie can show it. On a smaller scale, the art to typeset beautiful books, so different from what many computer programs do by default, the best introduction is by Jan Tschichold (1902–1974), the undisputed master of the field. Among the many books he designed are the beautiful Penguin books of the late 1940s; he also was a type designer, e.g. of the *Sabon* typeface. A beautiful summary of his views is the short but condensed text Jan Tschichold, *Ausgewählte Aufsätze über Fragen der Gestalt des Buches und der Typographie*, Birkhäuser Verlag, Basel, 1993. An extensive and beautiful textbook on the topic is Hans Peter Willberg & Friedrich Forssman, *Lesetypographie*, Verlag Hermann Schmidt, Mainz, 1997.

Many scientists passionately enjoyed their occupation. Any biography of Charles Darwin will purvey his fascination for biology, of Friedrich Bessel for astronomy, of Albert Einstein for physics and of Linus Pauling for chemistry.

642 The group of John Wearden in Manchester has shown by experiments with persons that the accuracy of a few per cent is possible for any action with a duration between a tenth of a second and a minute. See J. McCrone, *When a second lasts forever*, New Scientist pp. 53–56, 1 November 1997. Cited on page 610.

643 E.J. Zimmerman, American Journal of Physics **30**, 1962, pp. 97–105. Cited on page 611.

644 The potential of single-atom clocks is explained in ... Cited on page 612.

645 Peter D. Peşić, *The smallest clock*, European Journal of Physics **14**, pp. 90–92, 1993. Cited on page 612.

646 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 .

647 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 .

648 The switchable mirror 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. A good introduction is R. Griessen, *Schaltbare Spiegel aus Metallhydriden*, Physikalische Blätter **53**, pp. 1207–1209, 1997. Cited on page 616.

649 The present attempts to build cheap THz wave sensors – which allow to see through clothes – is described by D. Clery, *Brainstorming their way to an imaging revolution*, Science **297**, pp. 761–763, 2002. Cited on page 615.

650 V.F. Weisskopf, *Of atoms, mountains, and stars*, Science **187**, p. 605, 1975. Cited on page 617.

651 R. Full, **Nature**, **405**, p. 681, ... Cited on page 617.

652 For good figures of atomic orbitals, take any modern chemistry text. See for example the text by ... Cited on page .

653 For experimentally determined pictures of the orbitals of dangling bonds, see for example F. Giessibl & al., *Subatomic features on the silicon (111)-(7x7) surface observed by atomic force microscopy*, Science **289**, pp. 422–425, 2000. Cited on page .

654 J.L. Costa-Krämer, N. Garcia, P. Garcia-Molchales, P.A. Serena, *Conductance in disordered nanowires: foward and backscattering*, Physical Review B **53**, p. 102658, 1996. Cited on page 618.

655 The original prediction, which earned him a Nobel prize for physics, is in R.B. Laughlin, *Anomalous quantum Hall effect: an incompressible quantum fluid with fractionally charged* *excitations*, Physical Review Letters **50**, pp. 1395–1398, 1982. The first experimental evidence is by V.J. Goldman & B. Su, *Resonance tunneling in the fractional quantum Hall regime: measurement of fractional charge*, Science **267**, pp. 1010–1012, 1995. The first unambiguous measurements are by R. de Picciotto & al., *Direct observation of a fractional charge*, Nature **389**, pp. 162–164, 1997, and L. Saminadayar, D.C. Glattli, Y. Jin & B. Etienne, *Observation of the e*,3 fractionally charged Laughlin quasiparticle/ Physical Review Letters **79**, pp. 2526–2529, 1997. or http://arxiv.org/cond-mat/9706307. Cited on page 618.

The fractional quantum Hall effect was discovered by D.C. Tsui, H.L. Stormer & A.C. Gossard, *Two-dimensional magnetotransport in the extreme quantum limit*, Physical Review Letters **48**, pp. 1559–1562, 1982.

The discovery of the original quantum Hall effect, which earned von Klitzing a Nobel prize, was published as K. von Klitzing & G. Dorda, M. Pepper, *New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance*, Physical Review Letters **45**, pp. 494–497, 1980.

656 H.B.G. Casimir, *On the attraction between two perfectly conducting bodies*, Proc. Kon. Ned. Acad. Wetensch. **51**, pp. 793–795, 1948. Cited on page .

657 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 .

658 S.L. Boersma, *Aantrekkende kracht tussen schepen*, nederlands tidschrift voor natuurkunde **67**, pp. 244–246, *Augustus* 2001. See also his paper S.L. Boersma,*A maritime analogy of the Casimir effect*, American Journal of Physics **64**, pp. 539–541, 1996. Cited on page 624.

659 A. Larraza & B. Denardo, *An acoustic Casimir effect*, Physics Letters A **248**, pp. 151–155, 1998. See also A. Larraza, *A demonstration apparatus for an acoustic analog of the Casimir effect*, American Journal of Physics **67**, pp. 1028–1030, 1999. Cited on page 625.

660 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 626.

661 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 626.

662 T.H. Boyer, *Van der Waals forces and zero-point energy for dielectric and permeable materials*, Physical Review A **9**, pp. 2078–2084, 1974. This was taken up again in O. Kenneth, I. Klich, A. Mann & M. Revzen, *Repulsive Casimir forces*, Physical Review Letters **89**, p. 033001, 2002. However, none of these effects has been verified yet. Cited on page 626.

663 The effect was discovered by K. Scharnhorst, *On propagation of light in the vacuum between plates*, Physics Letters B **236**, pp. 354–359, 1990. See also G. Barton, *Faster-than-c light between parallel mirrors: the Scharnhorst effect rederived*, Physics Letters B **237**, pp. 559–562, 1990, and P.W. Milonni & K. Svozil, *Impossibility of measuring faster-than-c signalling by the Scharnhorst effect*, Physics Letters B **248**, pp. 437–438, 1990. The latter corrects an error in the first paper. Cited on page 626.

664 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 627.

665 G. Gour & L. Sriramkumar, *Will small particles exhibit Brownian motion in the quantum vacuum?*, http://www.arxiv.org/quant/phys/9808032 Cited on page 627.

666 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 .

667 Misleading statements are given in the introduction and in the conclusion of the review by H. Paul, *Interference between independent photons*, Review of Modern Physics, volume 58, pages 209-231; however, in the bulk of the article the author in practice retracts the statement, e.g. on page 221. Cited on page 621.

668 Thomas Stoehlker & al., Physical Review Letters 9 October 2000. Cited on page 628.

669 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 629.

670 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 620.

671 See the web site by Tom Hales on http://www.math.lsa.umich.edu/~hales/countdown. An earlier book on the issue is the delightful text by Max Leppmaier, *Kugelpackungen von Kepler bis heute*, Vieweg Verlag, 1997. Cited on page 629.

672 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**, pp. 53–56, 1999. Cited on page 629.

673 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 629.

674 See for example the textbook by A. Pimpinelli & J. Villain, *Physics of crystal growth*, Cambridge University Press, 1998. Cited on page 630.

675 P.G. de Gennes, *Soft matter*, Reviews of Modern Physics **63**, p. 645, 1992. Cited on page 630.

676 The fluid helium based quantum gyroscope was first built by the Czech group ... and the Berkeley group, Nature 10. April 1997. Cited on page 630.

677 K. Schwab & al., Nature **404**, p. 974, 2000. For the optical analog, see reference 399. Cited on page 630.

678 J.P. Briand & al., *Production of hollow atoms by the excitation of highly charged ions in interaction with a metallic surface*, Physical Review Letters **65**, pp. 159–162, 1990. Cited on page 630.

See also C. Rhodes, *Hollow atoms*, ..., and ...Physikalische Blätter Oktober 1996.

679 The present knowledge on how humans move is summarized by ... Cited on page 634.

680 See e.g. D.A. Bryant, *Electron acceleration in the aurora*, Contemporary Physics **35**, pp. 165–179, 1994. Cited on page 634.

681 The discharges above clouds are described e.g. by ... Cited on page 634.

682 See e.g. L.O'C. Drury, *Acceleration of cosmic rays*, Contemporary Physics **35**, pp. 231–242, 1994. Cited on page 634.

683 About Coulomb explosions, see ... Cited on page 634.

684 See the text by Dirk Kreimer, ..., 2000. Cited on page 634.

685 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 636.

686 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 636.

687 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 636.

688 V.V. NESVIZHEVSKY & al., *Quantum states of neutrons in the Earth's gravitational field*, Nature **415**, p. 297, 17 January 2002. Cited on page 636.

689 JAKOB BEKENSTEIN, *Black holes and entropy*, Physical Review D p. 2333, 1973. Cited on page 637.

690 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 637.

691 PAUL DAVIES, ..., WILLIAM G. UNRUH, *Notes on black hole evaporation*, Physical Review D **14**, pp. 870–892, 1976. Cited on page 638.

692 About the possibility to measure Davies-Unruh radiation, see for example the paper by H. ROSU, *On the estimates to measure Hawking effect und Unruh effect in the laboratory*, International Journal of Modern Physics D3 p. 545, 1994. http://arxiv.org:gr-qc/9605032 or the paper P. CHEN & T. TAJIMA, *Testing Unruh radiation with ultra-intense lasers*, Physical Review Letters **83**, pp. 256–259, 1999. Cited on page 638.

693 W.G. UNRUH & R.M. WALD, *Acceleration radiation and the generalized second law of thermodynamics*, Physical Review D **25**, pp. 942–958 , 1982. Cited on page 638.

694 For a delightful popular account from a black hole expert, see IGOR NOVIKOV, *Black holes and the universe*, Cambridge University Press 1990. Cited on pages 640 and 642.

695 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 640.

696 Optical black holes ... Cited on page 640.

697 T. DAMOUR & R. RUFFINI, Physical Review Letters **35**, p. 463, 1975. Cited on page 640.

698 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 641.

699 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 641.

700 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, *The dyadosphere of black holes and gamma-ray bursts*, Astronomy and Astrophysics **338**, pp. $L87$–$L90$, 1998, R. RUFFINI, J.D. SALMONSON, J.R. WILSON & S.-S. XUE, *On the pair electromagnetic pulse of a black hole with electromagnetic structure*, Astronomy and Astrophysics **350**, pp. 334–343, 1999, R. RUFFINI, J.D. SALMONSON, J.R. WILSON & S.-S. XUE, *On the pair electromagnetic pulse from an electromagnetic black hole surrounded by a baryonic remnant*, Astronomy and Astrophysics **359**, pp. 855–864, 2000, and C.L. BIANCO, R. RUFFINI & S.-S. XUE, *The elementary spike produced by a pure e^+e^- pair-electromagnetic pulse from a black hole: the PEM pulse*, Astronomy and Astrophysics **368**, pp. 377–390, 2001. 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 642.

701 About the membrane despription of black holes, see ... Cited on page 643.

702 Negative heat capacity has also been found in atom clusters and in nuclei. See e.g. M. Schmidt & al., *Negative heat capacity for a cluster of 147 sodium atoms*, Physical Review Letters **86**, pp. 1191–1194, 2001. Cited on page 642.

703 H.-P. Nollert, *Quasinormal modes: the characteristic 'sound' of black holes and neutron stars*, Classical and Quantum Gravity **16**, pp. R159–R216, 1999. Cited on page 642.

704 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 his electronic preprint at http://www.arxiv.org/abs/hep-th/9305040. Cited on page 644.

705 See Don N. Page, ..., Cited on page 644.

706 The author speculates that this version of the coincidences could be original; he has not found it in the literature. Cited on page 636.

707 Steven Weinberg, *Gravitation and Cosmology*, (Wiley, 1972). See equation 16.4.3 on page 619 and also page 620. Cited on page 636.

708 It could be that knot theory provides a relation between local knot invariant and a global one. Cited on page 637.

709 E. Joos, *Why do we observe a classical space-time?*, Physics Letters A **116**, pp. 6–8, 1986. Cited on page 645.

710 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 646.

711 An overview on squeezed light is given in ... Cited on page .

712 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 .

713 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, *Zum Heisenbergschem Unschärfeprinzip*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse Berlin, pp. 296–303, 1930. Cited on page .

714 W. Heisenberg, *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, Zeitschrift für Physik **43**, pp. 172–198, 1927. Cited on page .

715 M.G. Raymer, *Uncertainty principle for joint measurement of noncommuting variables*, American Journal of Physics **62**, pp. 986–993, 1994. Cited on page .

716 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 .

717 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 .

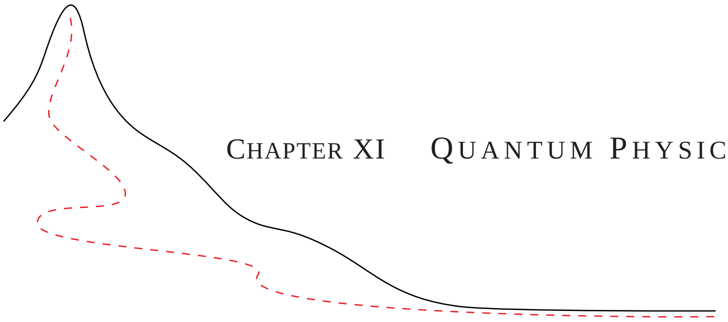
718 MARK P. SILVERMAN, *And yet it moves: strange systems and subtle questions in physics*, Cambridge University Press, 1993. It is a treasure chest for anybody interested in the details of modern physics.

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 and insights about physics. Cited on page .

719 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 .

720 The famous quantum theory book is P.A.M. DIRAC, *Quantum mechanics*, 4th revised edition, Oxford University Press, 1967. Cited on page .

Typeset in
January 2003



CHAPTER XI   QUANTUM PHYSICS IN A NUTSHELL

Achievements in precision

Compared to the classical description of motion, quantum theory is remarkably more complex. The basic idea is simple: in nature there is a minimum change, or a minimum action. The minimum action results in all the strange observations made in the microscopic domain, such as wave behaviour of particles, tunnelling, uncertainty relations, randomness in measurements, quantization of angular momentum, pair creation, decay, indistinguishability and reactions of particles. The mathematics is often disturbingly involved. 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 – that can be measured today. The limit is *not* given by the theory, it is given by the measurement accuracy. In other words, the agreement is only limited by the amount of money the experimenter is willing to spend, Table 50 shows this in more detail.

Table 50 A few comparisons between quantum theory and experiment

Observable	Classical prediction	Prediction of quantum theory[a]	Measurement	Cost[b], estimated
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$
Tunnelling 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$

Observable	Classical prediction	Prediction of quantum theory[a]	Measurement	Cost[b], estimated
Colours of objects				
Lamb shift	none	$\Delta\lambda = 1057.86(1)\,\mathrm{MHz}$	$(1 \pm 10^{-6})\ \Delta\lambda$	50 k$
Rydberg constant	none	$R_\infty = m_e c\alpha^2/2h$	$(1 \pm 10^{-9})\ R_\infty$	50 k$
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	...	...	50 M$
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	$(1 \pm 5\%)\ m_p$	$m_p = 1.67\ \mathrm{yg}$	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	0	25 J/K	$< 10^{-3}$ 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$

a. These predictions are calculated from the values of Table 51. For more precise values, see Appendix B. See page 701 See page 872

b. Sometimes the cost for the calculation of the prediction is higher than that its measurement. (Can you spot the examples?) The sum of the two is given. Challenge 1119 n

We notice that the predicted 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, these unexplained observations provide the input for the calculations just cited; we list them in detail below, in Table 51. See page 701

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 of quantum theory

Deorum offensae diis curae.
Voltaire, *Traité sur la tolérance*.

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.*

The second statement in fact results from the first. The existence of a smallest action in nature directly leads to the main lesson we learned about motion in the second part of our adventure:

▷ *If it moves, it is made of particles.*

This statement applies to everything, thus to all objects and to all images, i.e. to matter and to radiation. Moving stuff is made of *quanta*. Stones, water waves, light, sound waves, earthquakes, gelatine and everything else we can interact with is made of particles. We started the second part of our mountain ascent with the title question: what is matter and what are interactions? Now we know: they are composites of elementary particles.

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. See page 889

Moving entities are made of particles. To see how deep this result is, you 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 1120 e

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

$$\Delta x \Delta p \geqslant \hbar/2 \quad \text{with} \quad \hbar = 1.1 \cdot 10^{-34}\,\mathrm{Js} \qquad (525)$$

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. It results from the interactions with the baths in the environment of any 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 a single particle through

$$\lambda = \frac{h}{p} = \frac{2\pi\hbar}{p} \qquad (526)$$

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. All waves interfere, refract and diffract. This applies to electrons, atoms, photons, and molecules. 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 moved, 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 boundaries, since there is a finite probability to overcome any obstacle. This process is called tunnelling when seen from the spatial point of view and is called decay when seen from the temporal point of view. Tunnelling 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, specifying their behaviour under rotations. Bosons have integer spin, fermions have half 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. Solids are impenetrable because of the fermion character of its electrons in the atoms.
- 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 failed to detect 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, the requirements of space-time (Poincaré) symmetry, permutation 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

$$\lambda_C = \frac{h}{mc} = \frac{2\pi\hbar}{mc} \quad . \qquad (527)$$

At the latest at these distances we 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 small 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.

▪ Quantum theory implies, through the appearance of Planck's constant $\hbar$, that length scales exist in nature. Quantum theory introduces a fundamental jitter in every example of motion. Thus the infinitely small is eliminated. In this way, 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 in an atom small creatures live in the same way that humans live on the earth circling the sun. Quantum theory shows the impossibility of Lilliput.

▪ Clocks and meter bars have finite precision, due to the existence of a smallest action and due to their interactions with baths. On the other hand, all measurement apparatuses must 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 photon, the eight gluons, and the two weak interaction bosons.

See page 889

▪ Quantum chromodynamics, the field theory of the strong interactions, explains the masses of mesons and baryons through its descriptions as bound quark states. At fundamental scales, the strong interaction is mediated by the elementary gluons. At femtometer scales, the strong interaction effectively acts through the exchange of spin 0 pions, and is thus strongly attractive.

▪ 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 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.

See page 889

▪ 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.
- Quantum physics leaves no room for cold fusion, astrology, teleportation, telekinesis, supernatural phenomena, multiple universes, or faster than light phenomena, the EPR paradox notwithstanding. Ref. 738
- 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 we are fooled only if we forget our own limitations. Once we start to see ourselves as part of the game, we start 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, our human condition permanently fools us. Quantum theory answers the title question affirmatively; that is its main attraction.

The dangers of buying a can of beans

The ultimate product warning, which according to certain well-informed lawyers should be printed on ever package, in particular on cans of beans, gives another summary of our walk so far. It shows in detail how deeply our human condition fools us. Ref. 739

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 **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 kilometres 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.

Warning: care should be taken when **transporting** this product:

- The force needed depends on its velocity, as does its weight.
- This product 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 kilometres, over time cosmic radiation will render this product radioactive.
- This product may disintegrate in the next 10^{35} years.
- It could cool down and lift itself into the air.
- 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.
- This product can disappear from its present location and reappear at any random place in the universe, including your neighbour's garage.

Warning: care should be taken when **travelling 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.

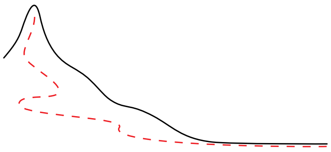
The impression of a certain paranoid side to physics is purely coincidental.

The essence of quantum theory

We can summarize quantum physics with a simple statement: *quantum physics is the description of matter and radiation* without *the concept of infinity*.

---

[*] A standard nuclear warhead has an explosive power of about 0.2 megatons (implied is the standard explosive trinitrotoluene or TNT). A megaton is defined as 1 Pcal=4.2 PJ, even though TNT delivers about 5% slightly less energy that this value. In other words, a megaton is the energy content of about 47 g of matter. That is less than a handful for most solids or liquids.



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

Matter and radiation are described by finite quantities. On the other hand, some remainders of infinities had to be retained in our description of nature, namely in the description of space or time, and in topics related to them, such as renormalization. We did not manage to eliminate all infinities. We are thus not yet at the end of our quest. Soon we shall find out that a completely finite description of all of nature is impossible because a completely finite description of nature, like one using infinities, cannot be accurate.

What is unexplained by quantum mechanics and general relativity?

The material gathered in this second part of our mountain ascent, together with the earlier summary of general relativity, allows us to give a complete answer to this question. Even though the available concepts and theories allow us to describe all observed phenomena connected to motion, there remain some *unexplained* properties of nature. Whenever we ask 'why?' and continue doing so after each answer, we arrive at one of the points in Table 51.

Table 51 *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 (or 6, if neutrinos have 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 (or 7, if neutrinos have masses)
θ_{CP}	the value of the CP parameter
θ_{st}	the value of the strong topological angle
3	the number of particle generations
$3+1$	the number of space and time dimensions
$0.5\,\mathrm{nJ/m}^3$	the value of the observed vacuum energy density or cosmological constant
Global quantities, from general relativity	
$1.2(1)\cdot 10^{26}\,\mathrm{m}$ (?)	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 or stars (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

Observed value	Property unexplained so far
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. The two theories 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. Our 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. Finding the answers will require effort.

On the other hand, the list is also *short*.** The description of nature our adventure 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 almost unnecessary if we only want to improve measurement precision. Simplifying the above list is mainly 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 Table 51 and Figure 239 show; the last leg forms the third part of our walk.

---

* Every now and then, researchers provide other lists of open questions. However, they all fall into the list above. The elucidation of dark matter, the quest for unknown elementary particles such as the inflaton field, magnetic monopoles or others, the functioning of cosmic high-energy particle accelerators, the stability of protons, or the origins of the heavy chemical elements are questions that all fall into the table above.

** Several other lists of open questions have been published. For example, astrophysicists in 2002 are interested in the mysteries of dark matter and dark energy, the details of the big bang, the modifications of general relativity by quantum theory, the mass of neutrinos, the working of cosmic accelerators, the stability of protons, the possibility of other matter forms, the possibility of higher spatial dimensions, the origin of the heavy elements and a new theory about interactions between matter and radiation. All these issues can be traced to issues in the table.

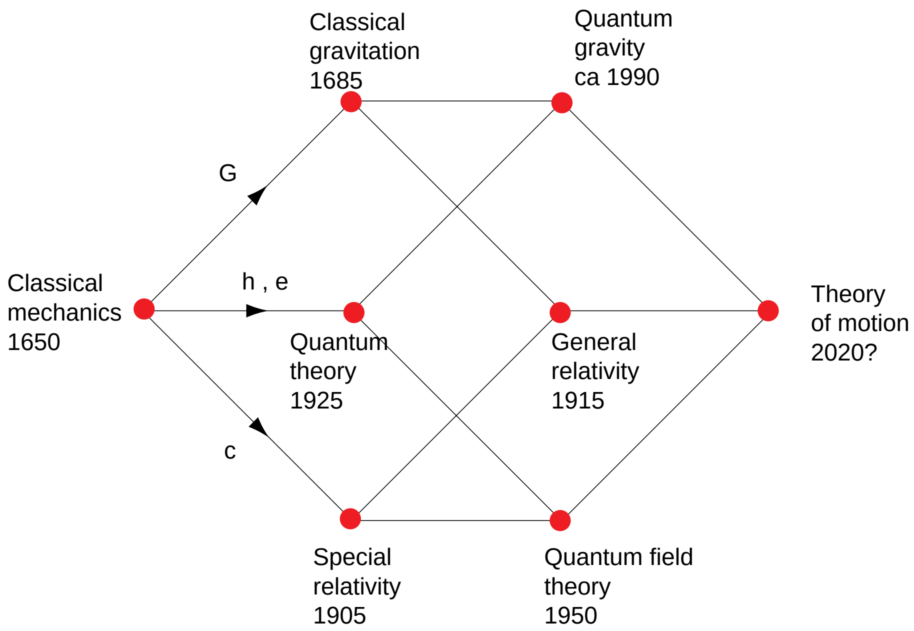


Figure 239 How the description of motion proceeded in the history of physics

How to delude oneself to have reached the top of Motion Mountain

Nowadays is deemed chic to pretend that the adventure is over at this stage.* The reasoning is as follows. If in the previous table on unexplained features of nature we change the values of the constants only ever so slightly, the world would look completely different from what it is.**

Table 52 A tiny selection of the 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

---

* Actually this attitude is not new. Only the arguments have changed. Maybe the greatest physicist ever, James Clerk Maxwell, fought against this attitude: 'The opinion seems to have got abroad 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. ... The history of science shows that even during that phase of her progress in which she devotes herself to improving the accuracy of the numerical measurement of quantities with which she has long been familiar, she is preparing the materials for the subjugation of new regions, which would have remained unknown if she had been contented with the rough methods of her early pioneers.' Ref. 740

** 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.

Observed value	Change	Result
	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 leukaemia
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 and confinement
Global quantities, from general relativity		
horizon size	much smaller:	no people
baryon number	very different:	no smoothness
	much higher:	no solar system
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
Sun’s mass	smaller:	too cold for the evolution of life
Sun’s mass	larger:	sun too short lived for the evolution of life
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 cosmicspeed	+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

Some have condensed Table 52 into a simple sentence: if any parameter is changed, the universe would either have too many or too few black holes.

The table is overwhelming. 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 unexplained numbers and other properties need to be explained, i.e. deduced from more general principles. It is easier to throw in some irrational belief; three fashionable ones are that the universe is *created* or *designed*, that the universe is *designed for people*, and 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 from 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 495

See page 502

Stopping our mountain ascent with a belief at the present point 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 ascent of Motion Mountain while pretending to have reached the top.

That is a pity. In our adventure, accepting the powerful message of Table 52 is one of the most awe-inspiring, touching and motivating moments. There is only one possible implication based on facts: the evidence 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.

Having faced this conclusion, everybody has to make up his own mind on whether to proceed or not with the adventure. Of course, there is no obligation to do so. Challenge 1122 n

What awaits us?

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 storms raging near 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 shall 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 relations of nature fix the famous proton-electron mass ratio of about 1836.2 or the fine structure constant? Like the hero of Douglas Adams, physicists know the answer to the greatest of questions: it is 137.036. But they do not know 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 Democritus *right*? Our adventure has confirmed his statement up to this point; 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 Democritus *wrong*? It is often said 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

$$N = n\, v\, d\, p\, f \qquad (528)$$

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 space-time 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. We thus have

$$N = 10^{92} \cdot 8 \cdot 10^{61} \cdot 10^{183} \cdot f = 10^{336} \cdot f \qquad (529)$$

with the small problem that we know nothing whatsoever about f. Its value could be 0, if all data were interdependent, or 1, if none were. Do good arguments for $f = 0$ really exist? In either case we still need to understand how all the particles get their states assigned from this truly enormous range of possibilities.

See page 117 Were our efforts up to this point *in vain*? Quite at the beginning of our walk we noted that in classical physics, space and time are defined using matter, and matter is defined using space-time. Hundred years of general relativity and of quantum theory, including 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 1123 e

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 unravelling of the essence of motion gets really interesting only from this point onwards!

> That is why Leucippus and Democritus, who say that the atoms move
> always in the void and the unlimited, must say what movement is,
> and in what their natural motion consists.
> Aristotle, Treaty of the Heaven. Ref. 741

References

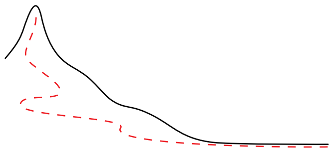
738 An informative account of the world of psychokinesis and the paranormal is given by JAMES RANDI, a professional magician, in *Flim-flam!*, Prometheus Books, Buffalo 1987, as well as in several of his other books. See also the http://www.randi.org web site. Cited on page 698.

739 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 698.

740 JAMES CLERK MAXWELL, *Scientific papers*, **2**, p. 244, October 1871. Cited on page 703.

741 ARISTOTLE, *Treaty of the heaven*, III, II, 300 b 8. See JEAN-PAUL DUMONT, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 392, 1991. Cited on page 707.

Third Part



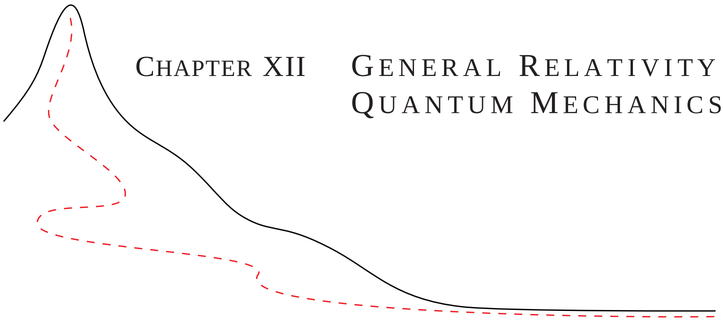
Motion Without Motion

What Are Space, Time and Particles?

Where through the combination of
quantum mechanics and general relativity,
the top of Motion Mountain is reached
and it is discovered
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 a complete description of motion is possible.
(Well, wait a few more years for the last line.)

Typeset in
January 2003

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 have followed up to now, namely that on general relativity and the one on quantum field theory, are both beautiful and successful. Both are confirmed by experiments. We have reached a considerable height in our mountain ascent. The precision we achieved in the description of nature is impressive, and we are now able to describe all known examples of motion. So far we have encountered no exceptions.

However, the most important aspects of any type of motion, the masses of the particles involved and the strength of their coupling, are still unexplained. Furthermore, the origin of the number of particles in the universe, their initial conditions, and the dimensionality of space-time remain hidden from us. Obviously, our adventure is not yet complete.

This last part of our hike will be the most demanding. In the ascent of any high mountain, the head gets dizzy because of the lack of oxygen. The finite energy at our disposal requires that we leave behind all unnecessary baggage and everything that slows us down. In order to determine what is unnecessary, we need to focus on what we want to achieve. Our biggest problem is all those concepts that are at the origin of the *contradictions* between general relativity and quantum theory. To pinpoint this useless baggage, we first list these contradictions.

In classical physics and in general relativity, the vacuum, or empty space-time, is a region with no mass, no energy and no momentum. If matter or gravitational fields are present, space-time is curved. The best way to measure the mass or energy content of space-time is to measure the average curvature of the universe. Cosmology tells us how we can do this; measurements yield an average energy density of the 'vacuum' of

Ref. 766

See page 318

$$E/V \approx 1\,\mathrm{nJ/m^3} \quad . \tag{534}$$

However, quantum field theory tells a different story. Vacuum is a region with zero-point fluctuations. The energy content of vacuum is the sum of the zero-point energies of all the

Ref. 767

---

* 'One needs to replace habits of thought by necessities of thought.'

fields it contains. Indeed, the Casimir effect 'proves' the reality of these zero-point energies. Their energy density is given, within one order of magnitude, by See page ??

$$E/V = \frac{4\pi h}{c^3}\int_0^{\nu_{\max}} \nu^3 d\nu = \frac{\pi h}{c^3}\nu_{\max}^4 \quad . \tag{535}$$

The approximation is valid for the case in which the cut-off frequency $\nu_{\max}$ is much larger than the rest mass m of the particles corresponding to the field under consideration. Indeed, particle physicists argue that the cut-off energy has to be at least the energy of grand unification, about 10^{16} GeV= 1.6 MJ. That would give a vacuum energy density of

$$E/V \approx 10^{99}\,\mathrm{J/m^3} \quad , \tag{536}$$

which is about 10^{108} times higher than the experimental limit deduced from spatial curvature using general relativity estimates. In other words, something is slightly wrong here.

General relativity and quantum theory contradict each other in other ways. Gravity is curved space-time. Extensive research has shown that quantum field theory, the description of electrodynamics and of 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 include gravity consistently and thus to include general relativity. Without the concept of the particle as a countable entity, the ability to perform perturbation calculations is also lost; and these are the only calculations possible in quantum field theory. In short, quantum theory only works because it assumes that gravity does not exist! Indeed, the gravitational constant does not appear in any consistent quantum field theory. Ref. 768

On the other hand, general relativity neglects the commutation rules between physical quantities discovered in experiments on a microscopic scale. General relativity assumes that the position and the momentum of material objects can be given the meaning that they have in classical physics. It thus ignores Planck's constant $\hbar$ and only works by neglecting quantum theory.

Measurements also lead to problems. In general relativity, as in classical physics, it is assumed that infinite precision of measurement is possible, e.g. by using finer and finer ruler marks. In contrast, in quantum mechanics the precision of measurement is limited. The indeterminacy principle gives the limits that result from the mass M of the apparatus. See page 470 See page 610

Time shows the contradictions most clearly. Relativity explains that time is what is read from clocks. Quantum theory says that precise clocks do not exist, especially if the coupling with gravitation is included. What does waiting 10 minutes mean, if the clock goes into a quantum mechanical superposition as a result of its coupling to space-time geometry?

In addition, quantum theory associates mass with an inverse length via the Compton wavelength; general relativity associates mass with length via the Schwarzschild radius.

Similarly, general relativity shows that space and time cannot be distinguished, whereas quantum theory says that matter does make a distinction. 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. See page 336

Most dramatically, the contradiction is shown by the failure of general relativity to describe the pair creation of particles with spin 1/2, a typical and essential quantum process. John Wheeler and others have shown that, in such a case, the topology of space necessarily Ref. 769, 770

Ref. 771, 772 has to *change*; in general relativity, however, the topology of space is fixed. In short, quantum theory says that matter is made of fermions, while general relativity cannot incorporate fermions.

To sum up, general relativity and quantum theory clash. As long as an existing description of nature contains contradictions, it cannot lead to a unified description, to useful explanations, or even to a correct description. In order to proceed, let us take the shortest and fastest path: let us investigate the contradictions in more detail.

33. Does matter differ from vacuum?

There is a simple way to state the origin of *all* contradictions between general relativity and quantum mechanics. Both theories describe motion with objects made up of particles and with space-time made up of events. Let us see how these two concepts are defined. Ref. 773

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.

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, the *length* is defined in terms of objects, either with an old-fashioned ruler or with the help of the motion of light, which in turn is motion of particles. Ref. 774

Modern physics has further sharpened the definitions of particles and space-time. Quantum mechanics assumes that space-time is given (it is included as a symmetry of the Hamiltonian), and studies the properties and the motion of particles, both for matter and radiation. In general relativity and especially in cosmology, the opposite approach is taken: it assumes that the properties of matter and radiation are given, e.g. via their equations of state, and describes in detail the space-time that follows from them, in particular its curvature.

However, one fact remains unchanged throughout all these advances in physics: *the two concepts of particles and of space-time are each defined with the help of the other.* To avoid the contradiction between quantum mechanics and general relativity and to eliminate their incompleteness requires the elimination of this circular definition. As argued in the following, this necessitates a radical change in our description of nature, and in particular of the continuity of space-time.

For a long time, the contradictions between 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 general relativity is valid at large dimensions. However, this artificial separation is not justified; worse, it prevents the solution of the problem. The situation resembles the well-known drawing (Figure 249) by M.C. Escher, where two hands, each holding a pencil, seem to be drawing each other. Taking 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, the drawing gives a description of standard twentieth century physics. The apparent contradiction is solved by recognizing that both concepts (both hands) result from a hidden third concept from which the other two originate. In the picture, this third entity is the hand of the painter.

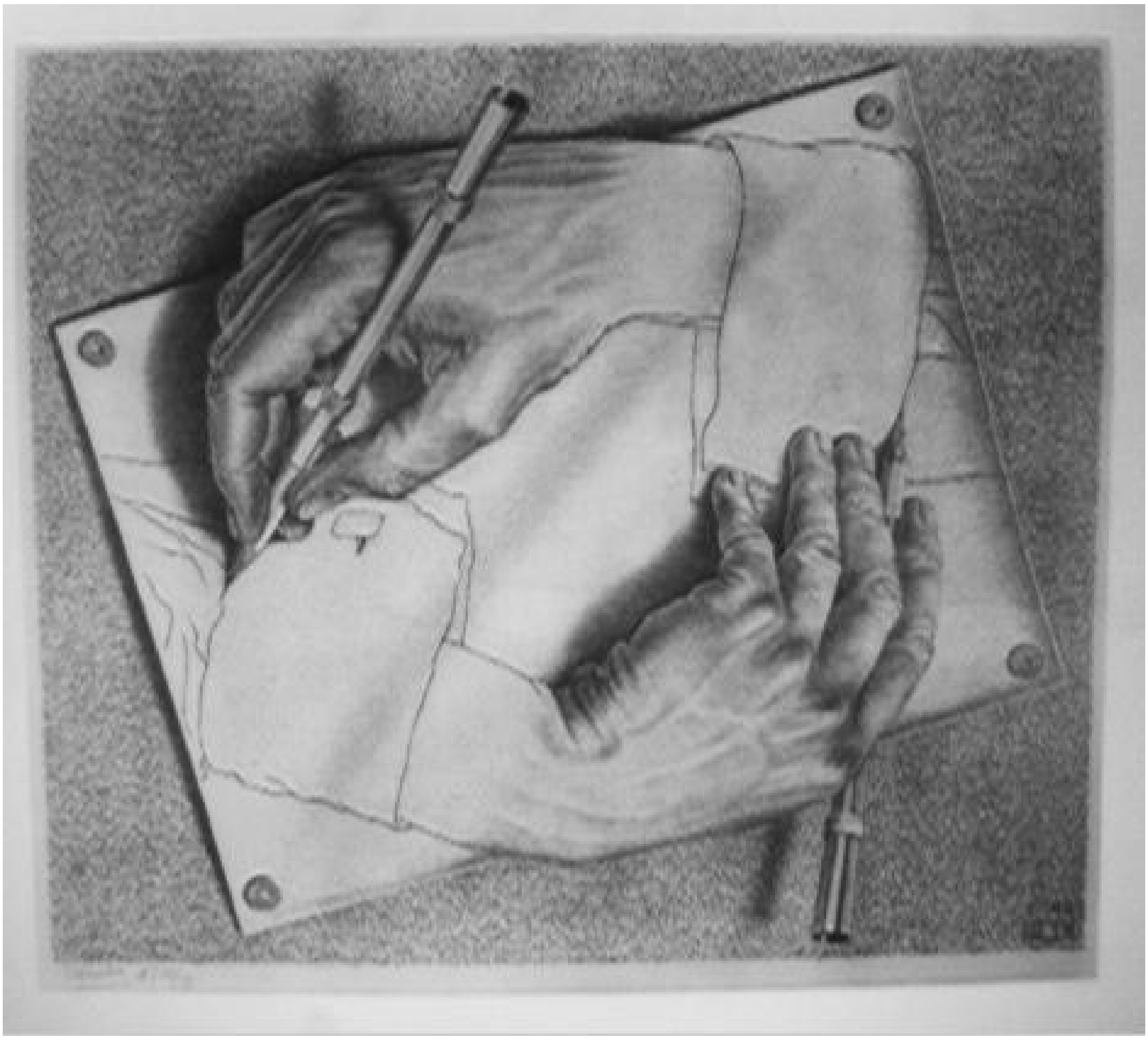


Figure 249 ‘Tekenen’ by M.C. Escher, 1948 – a metaphor for the way in which ‘particles’ and ‘space-time’ are usually defined: each with help of the other

In the case of space-time and matter, the search for the underlying common concept is presently making renewed progress. Ref. 775, 776 The required conceptual changes are so dramatic that they should be of interest to anybody who has an interest in physics. Ref. 777, 778 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 a well-known argument.

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

$$r_S = 2Gm/c^2 \quad . \tag{537}$$

Indeed, 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 its Schwarzschild radius. In general, we are forced to stay away from objects by an even larger multiple of the Schwarzschild radius, as shown in Table 53. 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. 769, 779

Ref. 780

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 53 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.

Similarly, quantum mechanics shows that Galilean physics must be abandoned and quantum effects must be taken into account whenever an object is approached to distances of the order of the Compton wavelength λ_C, given by

$$\lambda_C = \frac{\hbar}{m\,c} \quad . \tag{538}$$

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 or annihilation. Table 53 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.

The *combined* concepts of quantum field theory and general relativity are required in situations 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)

$$\begin{aligned} l_{Pl} &= \sqrt{\hbar G/c^3} = 1.6 \cdot 10^{-35}\,\mathrm{m}, \text{ the } \textit{Planck length,} \\ t_{Pl} &= \sqrt{\hbar G/c^5} = 5.4 \cdot 10^{-44}\,\mathrm{s}\ , \text{ the } \textit{Planck time.} \end{aligned} \tag{539}$$

Whenever we approach objects to these scales, 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, the questions mentioned at the beginning – why do we live in three dimensions or why is the proton 1836.15 times heavier than the electron – require for their answer a precise and complete description of nature. The contradictions between quantum mechanics and general relativity make the search for answers impossible. On the other hand, the unified theory, describing quantum gravity, is not yet finished; however, we can take a few glimpses on its implications already at the present stage.

Note that the Planck scales are one of only two domains of nature where quantum mechanics and general relativity apply at the same time. Planck scales being the most easy to study, they provide the best possible starting point for the following discussion. When Planck discovered them, he was interested in the Planck units mainly as *natural units of measurement*, and that is how he called them. However, their importance in nature is much more pervasive, as we will see now. Ref. 781

Farewell to instants of time

> Time is composed of time atoms ... which in fact are indivisible.
> Moses Maimonides, 12th century.

The appearance of the quantum of action in the description of motion leads to quantum limits to all measurements. These limits have important consequences at Planck dimensions which 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 Ref. 782, 783

$\Delta E \Delta t \geqslant \hbar$ it seems that by making ΔE arbitrary large, we can make Δt arbitrary small.

Any clock is a device with some moving parts. Parts can be mechanical wheels, matter particles in motion, changing electrodynamic fields, i.e. photons, or decaying radioactive particles. For each moving component of a clock, such as the two hands of the dial, the uncertainty principle applies. As discussed most clearly by Michael 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. 784, 785 Ref. 786

In any clock, we need to know both the time and the energy of each hand, in order for it to work. Otherwise it would not be a recording device. Put simply, it must be a classical system. 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 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 uncertainty Δt for its moving components. Thus we have

$$\delta t \geqslant \Delta t \geqslant \frac{\hbar}{\Delta E} \tag{540}$$

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.* 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. $m \leqslant c^2 l/G$, where l is the size of the clock (neglecting 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, since otherwise different parts of the clock could not work together to produce the same time display.** Putting all these conditions together one after the other, we get

$$\delta t \geqslant \frac{\hbar G}{c^5 \delta t} \quad , \tag{541}$$

or

$$\delta t \geqslant \sqrt{\frac{\hbar G}{c^5}} = t_{\mathrm{Pl}} \quad . \tag{542}$$

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 read-outs, we get the general conclusion that *clocks cannot measure time intervals shorter than the Planck time*.

---

* Physically, this condition means to be 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.

** It is amusing to explore how a clock larger than $c\,\delta t$ stops working, due to the loss of rigidity of its components.

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.*

Ref. 790 The same result can also be found in other ways. For example, any clock small enough to measure small time intervals necessarily has a certain energy uncertainty due to the uncertainty relation. 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 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

See page 281

Ref. 779

$$\Delta t = \frac{\Delta E\, G}{l\, c^5} \quad . \tag{543}$$

It determines the precision of the clock. Now the energy uncertainty of the clock is bound by the uncertainty relation for time and energy $\Delta E \geqslant \hbar/\Delta t$, again involving the precision of the clock. Putting this together, we again find the relation $\delta t \geqslant t_{\mathrm{Pl}}$ for the minimum measurable time. 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 impossible to make a meter bar or any other length measuring device able to measure lengths shorter than the Planck length. Obviously, we can deduce this already from $l_{\mathrm{Pl}} = c\,t_{\mathrm{Pl}}$. But a separate proof is also possible.

Ref. 791

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 $\Delta p < mc$; together this gives

$$\delta l \geqslant \Delta l \geqslant \frac{\hbar}{mc} \quad . \tag{544}$$

Furthermore, the measurement cannot be performed if signals cannot leave the objects; thus they may not be black holes. Therefore their masses must be so small that their

---

[*] Note that gravitation is essential here. The present argument differs from the well-known study on the limitations of clocks due to their mass and their measuring time published by Salecker and Wigner, and summarized in pedagogical form by Zimmerman. Here, both quantum mechanics as well as gravity are included, and therefore a different, lower, and much more fundamental limit is found. Note also that the discovery of black hole radiation does not change the argument; black hole radiation notwithstanding, measurement devices cannot be inside black holes.

Ref. 787

Ref. 788

Ref. 789

Schwarzschild radius $r_S = 2Gm/c^2$ is smaller than the distance δl separating them. Dropping again the factor of 2, we get

$$\delta l \geqslant \sqrt{\frac{\hbar G}{c^3}} = l_{Pl} \quad . \tag{545}$$

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. The corresponding energy uncertainty obeys $\Delta E = c(c^2m^2 + (\Delta p)^2)^{1/2} \geqslant c\hbar/\Delta x$. But general relativity shows that a small volume filled with energy changes the curvature, and thus changes 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 we localize 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. 769, 779 Ref. 791, 792, 793 Ref. 794, 795, 796

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. In usual proton-proton collisions we observe both pair production and interference effects. But the Planck limit implies that in everyday, macroscopic situations, such as car-car collisions, embryo-antiembryo pair production or quantum interference effects cannot be observed.

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, we arrive at the conclusion that *lengths smaller than the Planck length cannot be found in measurements*. Whatever the method used, be it a meter bar or time of flight measurement, we cannot overcome this fundamental limit. It follows that *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.

These results are often summarized in the so-called generalized uncertainty principle Ref. 797

$$\Delta p \Delta x \geqslant \hbar/2 + f\frac{G}{c^3}(\Delta p)^2 \tag{546}$$

or

$$\Delta p \Delta x \geqslant \hbar/2 + f\frac{l_{Pl}^2}{\hbar}(\Delta p)^2 \tag{547}$$

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 uncertainty. 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. You easily deduce from (546) that thegeneralized principle automatically implies that Δx can never be smaller than $f^{1/2} l_{Pl}$. Challenge 1132 e

The generalized uncertainty principle is derived in exactly the same way in which Heisenberg derived the original uncertainty principle $\Delta p \Delta x \geqslant \hbar/2$, namely by studying the deflection of light by the object under a microscope. A careful recalculation of the process, not disregarding gravity, yields equation (546). For this reason, *all* approaches which try to Ref. 797 Ref. 798

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. 799 Ref. 800, 801, 802

We remember that quantum mechanics starts when realizing that the classical concept of action makes no sense below the value of $\hbar$; similarly, unified theories start when realizing 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 favour of a more appropriate one.*

Ref. 803 A new uncertainty relation appearing at Planck scales shows that continuity cannot be a good description of space-time. Inserting $c\Delta p \geqslant \Delta E \geqslant \hbar/\Delta t$ into equation (546) we get

$$\Delta x \Delta t \geqslant \hbar G/c^4 = t_{\mathrm{Pl}} l_{\mathrm{Pl}} \ , \qquad (548)$$

which of course has no counterpart in standard quantum mechanics. It shows that space-time events do not exist. 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 $V_{\mathrm{Pl}} = l_{\mathrm{Pl}}^3$.

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: it is impossible to detect, and it is useful to describe observations only until the way to describe nature without it has been found.

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 intervals.

Farewell to the space-time manifold

The consequences of the Planck limits for time and space measurements can be taken much further. It is commonplace to say that given any two points in space or any 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.

All paradoxes resulting from the infinite divisibility of space and time, such as Zeno's argument or the Banach-Tarski paradox, are avoided. We can now dismiss the paradoxes straight-away 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 $l_{\mathrm{Pl}}/c = t_{\mathrm{Pl}}$, 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 time difference between events can only be measured within a Planck time, for two events distant in time by this order of magnitude, it is not possible to say with complete certainty which of the two *precedes* the other! This is an important result. If events cannot be ordered, the concept of time, which is introduced in physics to describe sequences, cannot be defined at all at Planck scales. 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. Therefore, the concept of 'proper time' loses its sense at Planck scales. Ref. 804

It is straightforward to use the same arguments to show that length measurements do not allow us to speak of continuous space, but only of *approximately* continuous space. Due to the lack of measurement precision at Planck scales, the concepts of spatial order, translation invariance, isotropy of the vacuum, and global coordinate systems lack experimental backing.

But there is more to come. The very existence of a minimum length contradicts special relativity, where it is shown lengths undergo Lorentz contraction when switching frame of reference. A minimum length thus cannot exist in special relativity. But we just deduced that there must be such a minimum distance in nature. There is only one conclusion: special relativity cannot be correct at smallest distances. Thus,*space-time is neither Lorentz invariant, nor diffeomorphism invariant, nor dilatation invariant at Planck dimensions.* All 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* 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 equivalent to fluctuations of gravity*. Ref. 791, 805

In addition, even the number of space dimensions makes no sense at Planck scales. Let us remind ourselves how to determine this number experimentally. One possible way is to determine how many points we 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. We recognize 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 way to check for three spatial 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 check whether space has three dimensions or not at Planck scales.

Many other methods to determine space dimensionality exist.[*] All these methods start from the definition of the concept of dimensionality, which is based on a precise definition of the concept of neighbourhood. But at Planck scales, as just mentioned, length measurements do not allow us to say whether a given point is inside or outside a given volume. In short, whatever method we use, 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 we remember Ref. 806 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; 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 *afterwards* with a measure or a metric. Mathematical 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 difficulties distinguishing physical and mathematical space and points arise from the failure to distinguish a mathematical metric from a physical length measurement.[**]

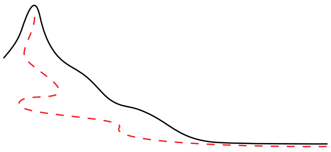
Perhaps the most beautiful way to make this point is the Banach-Tarski theorem. It clearly Ref. 807 shows the limits of the concept of volume. The theorem states that a sphere made of *mathematical points* can be cut into six pieces in such a way that two sets of three pieces can be put together to form two spheres, each of the *same volume* as the original one. However, the necessary cuts are 'infinitely' curved and detailed: they are wildly disconnected. 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 zero-point fluctuations is given by a density times the volume; following the Banach-Tarski theorem, the zero point energy content of a

---

[*] For example, we can determine the dimension using the topological properties of space only. If we draw 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.

In fact, if physical space is not a manifold, the various methods could give different answers for the dimensionality. Indeed, for linear spaces without norm, a unique number of dimensions cannot be defined. The value then depends on the specific definition used and is called e.g. fractal dimension, Lyapunov dimension, etc.

[**] Where does the incorrect idea of continuous space-time have its roots? In everyday life, as well as in physics, space-time is introduced to describe observations. Space-time is a bookkeeping device. Its properties are extracted from the properties of observables. Since observables can be added and multiplied, we extrapolate that they can take continuous values. This extrapolation implies that length and time intervals can take continuous, and in particular arbitrary small values. From this consequence we get the possibility to define points and sets of points. A special fields of mathematics, topology, shows how to start from a set of points to construct, with help of neighbourhood relations and separation properties, first a *topological space*. Then, with help of a metric, a *metric space* can be built, and finally, with the appropriate compactness and connectedness relations, a *manifold*, characterized by its dimension, metric and curvature.



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

single sphere should be equal to the zero point energy of two similar spheres each of the same volume as the original one. The paradox is solved by the Planck length, as it provides a fundamental length scale also for the vacuum, thus making infinitely complex cuts impossible. Therefore, *the concept of volume is only well defined at Planck scales if a minimum length is introduced*.

To sum up, *physical space-time cannot be a set of mathematical points.* But the surprises are not finished. At Planck dimensions, since both temporal 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 addition, it is impossible to state that the topology of space-time is fixed, as general relativity implies. The topology changes – mentioned above – required for particle reactions do become possible. In this way another of the contradictions between general relativity and quantum theory is resolved.

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,* 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.*

> 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 well-defined under the assumption of continuous space and time. The important tool of the evolution equation, based on derivatives, such as the Schrödinger or the Dirac equation, cannot be used any more. Concepts such as ‘derivative’, ‘divergence-free’, ‘source free’, etc., lose their meaning at Planck scales.

In fact, all physical observables are defined using length and time measurements. Any list of physical units shows that each of them is a product 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.) Since time and length are not continuous, observables themselves are not defined, as their value is not fixed. This means that *at Planck scales, observables are not to be described by real numbers.*

---

* A manifold is what *locally* looks like an Euclidean space. The exact definition can be found in Appendix D.

In addition, if time and space are not continuous, the usual expression for an observable field A, namely $A(t,x)$, does not make sense: we have to find a more appropriate description. *Physical fields cannot exist at Planck scales.*

The consequences for quantum mechanics are severe. 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. We recognize that due to measurement errors, we 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 $[x_\mathrm{i}, x_\mathrm{j}] = 0$ for $x_\mathrm{i} \neq x_\mathrm{j}$, 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 conclude 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.*

Finally, also renormalization symmetry is destroyed.

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?

Discretization of space-time has been studied already in the 1940s. More recently, the idea that space-time is described as a lattice has also been explored most notably by David Finkelstein and by Gerard '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. Worse, the lattice idea conflicts with general relativity, in particular with the diffeomorphism invariance of the vacuum. Finally, where would a particle be *during* the jump from one lattice point to the next? In summary, *space-time cannot be a lattice.* 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 at all for this conclusion, and actually there is quite some evidence for the contrary.

Ref. 808
Ref. 809
Ref. 810
Ref. 811
Ref. 812

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. We are 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 about the size of a cloud or of an atom. Measuring water density or electron density, we find non-vanishing values at any distance from the centre 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, we guess 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, a *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 mean time, research efforts in quantum gravity, superstring theory and quantum groups have confirmed independently from each other that a probabilistic and non-local description of space-time at Planck dimensions, resolves the contradictions between general relativity and quantum theory. To clarify the issue, we have to turn to the concept of particle.

Farewell to particles

In every example of motion, some object is involved. One of the important discoveries of the natural sciences was that all objects are composed of small constituents, called *elementary particles*. Quantum theory shows that all composite, non-elementary objects 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, as shown in Table 53. Elementary particles are characterized by their vanishing size, their spin, and their mass.

See page 889

See page 728

Even though the definition of 'elementary' as point particle is all we need 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; for example, it leads to intrinsically unstable components.)

The *size* of an object, such as the one given in Table 53, is defined as the length at which differences from point-like behaviour is observed. 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 how it scatters a beam of probe particles. Also in daily life, when we look at objects, we make use of scattered photons. In general, in order to make use of scattering, the effective wavelength $\lambda = \hbar/mv$ of the probe must be smaller than the object size d to be determined. We thus need $d > \lambda = \hbar/(mv) \geqslant \hbar/(mc)$. In addition, in order to make a scattering experiment possible, the object must not be a black hole, since then it would simply swallow the approaching particle. This means that its mass m must be smaller than that of a black hole of its size; in other words, from equation (537) we must have $m < dc^2/G$. Combining it with the previous condition we get

$$d > \sqrt{\frac{\hbar G}{c^3}} = l_{\mathrm{Pl}} \quad . \tag{549}$$

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. As in the case of time, the lower limit on length results from the combination of quantum mechanics and general relativity.*

The size d of any elementary particle is by definition surely smaller than its own Compton wavelength $\hbar/(mc)$. Moreover, a particle's size is always larger than the Planck length: $d > l_{\mathrm{Pl}}$. Combining these two requirements and eliminating the size d we get a condition for the mass m of any elementary particle, namely

$$m < \frac{\hbar}{c\, l_{\mathrm{Pl}}} = \sqrt{\frac{\hbar c}{G}} = m_{\mathrm{Pl}} = 2.2 \cdot 10^{-8}\,\mathrm{kg} = 1.2 \cdot 10^{19}\,\mathrm{GeV/c^2} \tag{550}$$

This limit, the so-called *Planck mass*, corresponds roughly to the mass of a ten days old human embryo, or equivalently, to 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 Andrei Sakharov in 1968; he 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. We will come back to it.) Ref. 813

---

* Obviously, 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.

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.) Ref. 814 The importance of the Planck mass will become clear shortly.

Another property connected with the size of a particle is its electric dipole moment. It describes the deviation of its charge distribution from spherical shape. Some predictions from the standard model of elementary particles give as *upper* limit for the electron dipole moment d_e a value of Ref. 815

$$|d_e| < 10^{-39}\,\mathrm{m}\,e \quad , \tag{551}$$

where e is the charge of the electron. This value is ten thousand times smaller than $l_{\mathrm{Pl}}\,e$. Since that the Planck length is the smallest possible length, we seem to have a potential contradiction here. However, a more recent prediction from the standard model is more careful and only states Ref. 816

$$|d_e| < 3\cdot 10^{-21}\,\mathrm{m}\,e \quad , \tag{552}$$

which is not in contradiction with a minimal length in nature. The issue is still open. We will see below that the experimental limit is expected to reach the precise predictions from the minimum length idea in the forseeable future.

There are 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, obeying $E^2 = m^2c^4 + p^2c^2$, whereas a virtual particle is off shell, obeying $E^2 \neq m^2c^4 + p^2c^2$. 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. Antimatter can be described as matter moving backwards in time. 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 behaviour under rotations (and if the particle is charged, the behaviour in magnetic fields) and its behaviour under particle exchange. The wave function of spin 1 particles remains 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.

We see directly that both transformations are impossible to study at Planck scales. Given the limit on position measurements, the position of a rotation axis 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 Planck 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

$$\rho_{\mathrm{Pl}} = \frac{c^5}{G^2\hbar} = 5.2 \cdot 10^{96}\,\mathrm{kg/m^3} \tag{553}$$

and is a useful concept in the following. If we want 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 $M < c^2R/G$; since the minimum value for R is the Planck distance, we get (neglecting again factors of order unity) a limit for the mass density, namely Challenge 1134 e

$$\rho < \rho_{\mathrm{Pl}} \quad . \tag{554}$$

In other words, *the Planck density is the maximum possible value for mass density.* Unsurprisingly, a volume of Planck dimensions cannot contain a mass larger than the Planck mass.

Interesting things happen when we start to determine the error ΔM of a mass measurement in a Planck volume. Let us return to the mass measurement by an orbiting probe. From the relation $GM = rv^2$ we deduce by differentiation that $G\Delta M = v^2\Delta r + 2vr\Delta v > 2vr\Delta v = 2GM\Delta v/v$. For the error Δv in the velocity measurement we have 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

$$\Delta M \geqslant \frac{\hbar}{cR} \quad . \tag{555}$$

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. We even 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 (555). Now that we are sure of the result, let us continue.

From equation (555) 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. Therefore, for a box of Planck dimensions, the mass measurement *error* is larger (or at best equal) to the mass contained in it: $\Delta M \geqslant M_{\mathrm{Pl}}$. In other words, if we build 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 only randomly change inclination, staying horizontal on average.

The argument can be rephrased as follows. The largest mass we can put in a box of size R 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. In other words, inside any box of size R we have a mass m whose limits are given by:

$$\text{(full box)}\ \frac{c^2R}{G} > m > \frac{\hbar}{cR}\ \text{(empty box)} \quad . \tag{556}$$

We see directly that for sizes R of the order of the Planck scale, the two limits coincide; in other words, we cannot distinguish a full from an empty box.

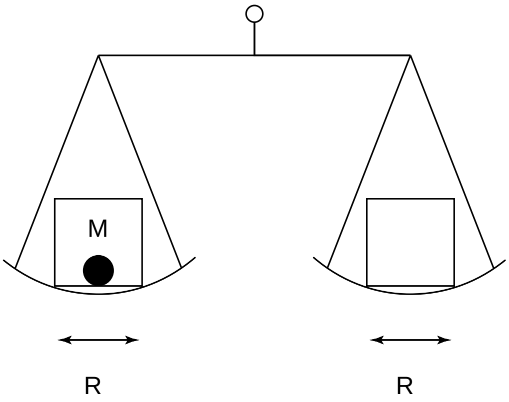


Figure 250 A *Gedankenexperiment* showing that at Planck scales, matter and vacuum cannot be distinguished

To be sure of this strange result, we check whether it also appears 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

$$\frac{\Delta M}{M} = \frac{\Delta\delta\lambda}{\delta\lambda} \quad . \tag{557}$$

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 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 state of affairs 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 rely on a simple fact: whatever definition of mass we use, it is always measured via combined 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.) The error in these measurements makes it *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 find 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, we have confirmed that they are made of the same 'fabric', as suggested above. This approach 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. This is already shown by the discovery of Unruh radiation. Any accelerated observer and any observer in a gravitational field detects particles hitting him, even if he is in vacuum. The effect 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 again follows that it is impossible to say whether it contains particles or not. Ref. 817

Curiosities and challenges

These strange results imply many others; here is a selection.

- We now have a new answer to the old question: why is there anything instead of nothing? Well, there is *no difference* between anything and nothing.
- We now can honestly say about ourselves: we are made of nothing.
- If vacuum and matter or radiation cannot be distinguished, it is incorrect to claim that the universe appeared from nothing. The impossibility of distinction thus makes creation a logical impossibility. Creation is not a description of reality; it is exposed as a lack of imagination.
- 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'.*
- 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 1135 n

On the other side, in everyday life, the Planck energy is rather small. Measured in litres of gasoline, how much fuel does it correspond to? Challenge 1136 n

▪ Do the large mass measurement errors allow us to claim that mass can be negative at Planck energy? Challenge 1137 n

▪ 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. It is sometimes claimed that measurement uncertainties smaller than the Planck values are possible for large enough numbers of particles. Can you show why this cannot be the case for space and time? Ref. 818 Ref. 801 Challenge 1138

▪ Of course, the idea that vacuum is not empty is not new. Already Aristotle argued for a filled vacuum, even though he used incorrect arguments, seen from today's perspective. In the fourteenth century the discussion on whether empty space was composed of indivisible entities was rather common, but died down again later.

▪ Special relativity implies that no length or energy can be invariant. Since we came to the conclusion that the Planck energy and the Planck length are invariant, there must be deviations of Lorentz invariance at high energy. Can you imagine how they could look like? In what experiment could they be measured?If you get an answer, publish it; you will become famous. Ref. 820 Challenge 1139 e

▪ A Planck energy particle falling in a gravitational field would gain energy. However, this is impossible, as the Planck energy is the highest energy in nature. What does this imply for this situation? Challenge 1140 n

▪ 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 idea, developed by Phil Gibbs, provides an additional formulation of the strange behaviour of nature at extreme scales. Ref. 790

See page 349 ▪ Due to a minimum length in nature, naked singularities do not exist. The issue, so hotly debated in the twentieth century, becomes uninteresting, thus ending decades of speculation.

▪ Since mass density and thus energy density is limited, we know that the number of degrees of freedom of any object of finite volume is *finite*. 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. Ref. 819

▪ If vacuum and matter cannot be distinguished, we cannot distinguish objects and environment. However, this was one the starting points of our walk. Some interesting adventures are thus awaiting us! See page 33

▪ 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 are exchanges of particles. This connection can be used to show that either vacuum and See page 489 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.

- Is there a smallest possible momentum? And a smallest momentum error? Challenge 1141

- There is a maximal acceleration in nature. Can you deduce the value of this so-called Planck acceleration? Challenge 1142 n

- Given that time becomes an approximation at Planck scales, can we still say whether nature is deterministic? Let us go back to the beginning. We can define time, because in nature change is not random, but gradual. What is the situation now that we know that time is only approximate? Is non-gradual change possible? Is energy conserved? In other words, are surprises possible?

To say that time is not defined at Planck scales, and that therefore determinism is an undefinable concept is correct, but not a satisfying answer. What happens at daily life scales?

The first answer is that at our scales, the probability for surprises is so small, that the world indeed is effectively deterministic. The second answer is that nature is not deterministic, but that the difference is not measurable, since every measurement and observation, by definition, *implies* a deterministic world. The lack of surprises would the due to the limitations of our human nature, more precisely to the limitations of our senses and brain. The third answer is that the lack of surprises is only apparent, and that we do not grasp them yet.

Can you imagine another possibility?To be honest, there is no answer possible at this point; we will need to keep the alternatives in mind. We have to continue searching. But at every step we are taking, we have to carefully ponder what we are doing. Challenge 1143 e

- If matter and vacuum cannot be distinguished, matter and vacuum each have the properties of the other. For example, space-time being an extended entity, matter and radiation is so as well. Even more so, space-time being an entity which reaches the borders of the system under scrutiny, particles do so as well. This is the first hint for the extension of matter; in the following, we will examine this argument with more detail.

- Vacuum has zero mass density at large scales, but Planck mass density at Planck scales. Cosmological measurements show that the cosmos is flat or almost flat on large scales, i.e. that its energy density is quite low. On the other hand, quantum field theory maintains that the vacuum has a high energy density (or mass density) on small scales. Since mass is scale dependent, both viewpoints are right, providing a hint to the solution of what is usually called the cosmological constant problem. The contradiction is only apparent; more about this issue later on. Ref. 826

- When *can* matter and vacuum be distinguished? At what energy? Challenge 1144 n

- If matter and vacuum cannot be distinguished, a lack of information follows, which in turn produces an intrinsic basic entropy associated with any part of the universe. We will come back to this topic shortly, in the discussion of black hole entropy.

- Can we distinguish between liquids and gases by looking at a single atom? No, only by looking at many. In the same way, we cannot distinguish between matter and vacuum by looking at one point, but only by looking at many. We must always *average*. But even averaging is not completely successful. Distinguishing matter from vacuum is like distinguishing clouds from the clear sky; like clouds, also matter has no defined boundary.

- In our exploration we 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 mountain ascent, we need to leave behind us the 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.*

Farewell to the big bang

A minimum length, or equivalently,* a minimum action, also imply that there is a maximum curvature for space-time. Curvature is an inverse area. A minimum length thus implies a maximum curvature. Within a factor of order one, we get the limit for curvatures

$$K < \frac{c^3}{G\hbar} = 0.39 \cdot 10^{70}\,\mathrm{m}^{-2} \quad . \tag{558}$$

In other words, the universe can never have been a point, can never have had zero age, can never have had infinite density, and can never have had infinite curvature. It is not difficult to get similar limits for temperature or any other physical quantity. In addition, since events do not exist, also the big bang cannot have been an event.

In short, there was no initial singularity and no beginning of the universe.

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, renormalization symmetry, and permutation symmetry. We also have destroyed the foundations of general relativity, namely the existence of the space-time manifold, the field concept, the particle concept, and the concept of mass. It was even shown that matter and space-time cannot be distinguished. It seems that we have lost every concept used for the description of motion, and thus made its description impossible. We naturally ask whether we can save the situation.

First of all, since matter is not distinguishable for vacuum, and since this is correct for all types of particles, be they matter or radiation, we have an 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 are being pursued at the end of the twentieth century. The first, quantum gravity, especially the one using the loop representation and Ashtekar's new variables, starts by generalizing *space-time symmetry*. Ref. 775 The second, string theory, starts by generalizing *gauge symmetries* and interactions, and the Ref. 776 third, the algebraic quantum group approach, looks for generalized *permutation symmetries*. Ref. 777 We will describe them in more detail shortly.

Before we go on however, we should check what we deduced so far.

---

* The big bang section was added in summer 2002.

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 we can define a characteristic energy by setting

$$E_{\mathrm{char}} = \frac{\hbar(\omega_1 - \omega_2)d}{c\Delta t} \quad . \tag{559}$$

This energy value is $8 \cdot 10^{16}$ GeV for the best measurement to date. Ref. 821 The value is not far from the Planck energy, 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 dispersion, different from zero. Ref. 822, 823 This effect would allow to confirm that Lorentz symmetry is not valid at Planck scales.

Another candidate is the direct detection of distance fluctuations between bodies. Ref. 827, 823 Gravitational wave detectors are sensitive to extremely small noise signals in length measurements. There should be a noise signal due to the distance fluctuations induced near Planck energies. The length uncertainty with which a length l can be measured is predicted to be Ref. 828

$$\frac{\delta l}{l} \geqslant \left(\frac{l_{\mathrm{Pl}}}{l}\right)^{2/3} \tag{560}$$

Th expression simply deduced by combining the measurement limit of a ruler in quantum theory with the requirement that the ruler cannot be a black hole. We will discuss this result See page ?? in more detail in the next section. See page 761 The sensitivity to noise of the detectors might reach the required level in the early 21st century. The noise induced by quantum gravity effects is also predicted to lead to detectable quantum decoherence and vacuum fluctuations. Ref. 824

A third candidate for measurable quantum gravity is the detection of the loss of CPT symmetry at high energies. Especially in the case of the decay of certain elementary particles, such as neutral kaons, the experimental measurement precision is approaching the detection of Planck scale effects. Ref. 823

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 Ref. 829 high energy photons or cosmic rays allow to prove that Lorentz invariance is indeed broken near the Planck scale.

It has also been predicted that quantum gravity effects will induce a gravitational Stark effect and a gravitational Lamb shift in atomic transitions. Ref. 824

A few candidates for quantum gravity effects have also been predicted by the author. To get an overview, we summarize and clarify the results found so far.[**] Special relativity

---

[*] 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.

starts 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 limited by 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.

We need to exchange $\hbar$ by $\hbar/2$ and G by $4G$ in all the defining expressions of Planck quantities, in order to find the corresponding measurement limits. In particular, the limit for lengths and times is $\sqrt{2}$ times the Planck value, and the limit for energy is the Planck value divided by $\sqrt{8}$.* 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.

However, we need to correct the argument just presented. The argument is only half the story, because so far, we cheated. The (corrected) Planck values do not seem to be the actual limits to measurements. The actual measurement limits are stricter still.

First of all, for any measurement, we need certain fundamental conditions to be realized. Take the length measurement of an object. We need 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 off distances from the ruler. For a measurement process, we need an interaction, which implies the use of radiation. Note that even the use of particle scattering to determine lengths does not invalidate this general requirement.

Also for the measurement of wavelengths we need to distinguish matter and radiation, as 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 energy of grand unification, in which the electroweak and the strong nuclear interactions are unified. Above this energy, particles of matter and of radiation cannot be distinguished from each other. Ref. 825

If all matter and radiation particles were the same, mass could not be defined. Similarly, spin could not be defined. If all particles were the same, neither charge nor the other quantum numbers can be defined. To sum up, no measurement can be performed at energies at or above the GUT unification energy.

In other words, the particle concept (and thus the matter concept) does not run into trouble at the Planck scale, it does so already earlier, at the unification scale. Only below the unification scale our standard particle and space-time concepts apply. *Only below the unification scale, particles and vacuum can effectively be distinguished.*

As a result, the smallest length in nature is $\sqrt{2}$ times the Planck length reduced by the ratio between the maximal energy $E_{\text{Pl}}/\sqrt{8}$ and the unification energy E_{GUT}. Following present estimates, $E_{\text{GUT}} = 10^{16}$ GeV, implying that Ref. 830

---

* The entropy of a black hole is thus given by the ratio between its horizon and *half* the minimal area. 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 everyday systems, there is no limit.

** To speak in modern high energy concepts, all measurements require broken supersymmetry.

$$L_{\text{min}} = \sqrt{2}\, l_{\text{Pl}} \frac{E_{\text{Pl}}}{\sqrt{8}\, E_{\text{GUT}}} \approx 10^{-32}\,\text{m} \approx 800\, l_{\text{Pl}} \quad . \tag{561}$$

It is unlikely that measurements at these dimensions will ever be possible. Anyway, 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. We thus have to correct our previous statement: *the minimum measurable length cannot be smaller than* L_{min}.

The experimentally determined factor of about 800 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 provide the answer.

In particular, the minimum length puts a bound on the electric dipole moment d of elementary particles, i.e. on any particles without constituents. We get the limit

$$d > d_{\text{min}} = eL_{\text{min}} = 10^{-32}\,\text{m}\, e = 1.5 \cdot 10^{-51}\,\text{Cm} \quad . \tag{562}$$

We saw that this result in contradiction with one of the predictions of the standard model, but not with all of them. More interestingly, the prediction is in the reach of future experiments. This improved limit might be the simplest possible measurement of yet unpredicted quantum gravity effects. Measuring the dipole moment could thus be a way to determine the unification energy (the factor 800) independently of high energy physics experiments, and possibly to higher precision. See page 741 Ref. 832

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 in everyday life, 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 coupling constant as an example. This constant α, also called the fine structure constant, is related to the charge by

$$q = \sqrt{4\pi\varepsilon_o \hbar c \alpha} \tag{563}$$

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, we have

$$\frac{q}{q_{\text{unit}}} = \frac{ma}{m_{\text{unit}} a_{\text{unit}}} \quad . \tag{564}$$

Therefore any error in mass and acceleration measurements implies errors in charge and coupling constant measurements.

We found in the part on quantum theory that 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. See page ?? Ref. 830

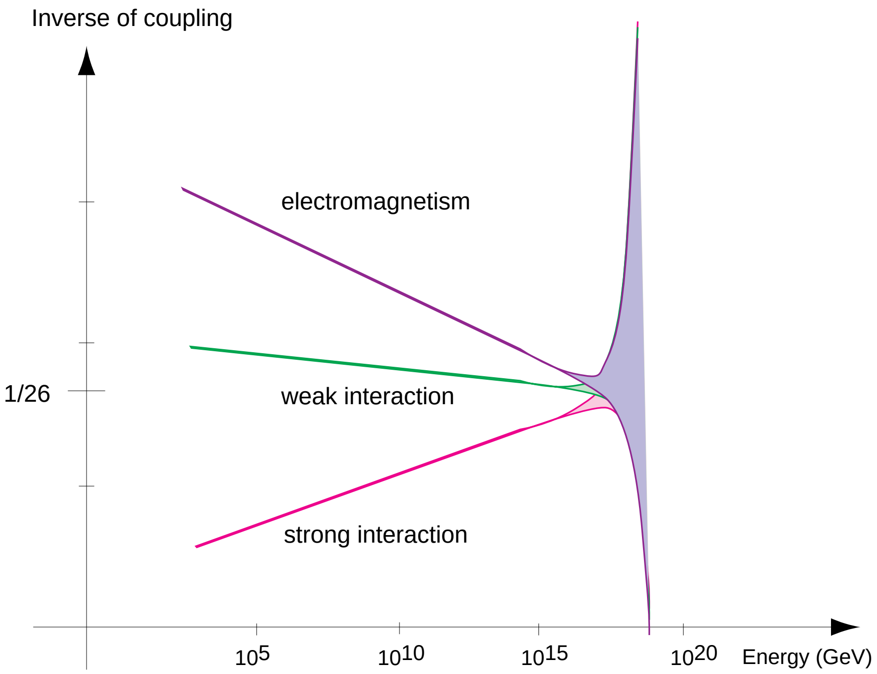


Figure 251 Coupling constants and their spread running with energy

We know from the above discussions that the minimal measurement error for any energy measurement at high energies 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, we get the result shown in Figure 251. 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. But all measurement devices, be they clocks, meter bars, scales or something else, use 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 spreading coupling constant value appears more clearly in the measurements. For example, 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. Also the lifetime of elementary particles could be affected. Can you find another effect? Challenge 1145 n

In summary, the experimental detection of quantum gravity effects should be possible, despite their weakness, during the 21st century. The successful detection of any such effect

---

[*] 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 for the maximum precision.

will be one of the highlights of physics, as it will challenge the usual description of space and time even more than general relativity did.

We now know that the fundamental entity describing space-time and matter 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.

References

766 STEVEN WEINBERG, *The cosmological constant problem*, Reviews of Modern Physics **61**, pp. 1–23, 1989. Cited on page 724.

767 STEVEN WEINBERG, *The quantum theory of fields*, Cambridge University Press, volume I, 1995, and II, 1996. Cited on page 724.

768 The difficulties are summarized by B.S. DE WITT, *Quantum field theory in curved spacetime*, Physics Reports **19**, pp. 295–357, 1975. Cited on page 725.

769 C.W. MISNER & K.S. THORNE, J.A. WHEELER, *Gravitation*, Freeman, 1973. Cited on pages 725, 728 and 732.

770 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 725.

771 J.L. FRIEDMAN & R.D. SORKIN, *Spin 1/2 from gravity*, Physical Review Letters **44**, pp. 1100–1103, 1980. Cited on page 726.

772 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 726.

773 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 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 C. SCHILLER, *Does matter differ from vacuum?* http://www.arxiv.org/abs/gr-qc/9610066. Cited on page 726.

774 See for example R.P. FEYNMAN, *The Feynman lectures of physics*, Addison Wesley, 1963. Cited on page 726.

775 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 pages 727 and 747.

776 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 pages 727 and 747.

777 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 pages 727 and 747.

778 C.J. ISHAM, *Prima facie questions in quantum gravity*, preprint gr-qc/9310031. Cited on page 727.

779 Steven Weinberg, *Gravitation and Cosmology*, (Wiley, 1972). Cited on pages 728, 731 and 732.

780 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 728.

781 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 Planck originally called b. Cited on page 729.

782 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 729.

783 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 729.

784 W. Heisenberg, *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, Zeitschrift für Physik **43**, pp. 172–198, 1927. Cited on page 730.

785 E.H. Kennard, *Zur Quantenmechanik einfacher Bewegungstypen*, Zeitschrift für Physik **44**, pp. 326–352, 1927. Cited on page 730.

786 M.G. Raymer, *Uncertainty principle for joint measurement of noncommuting variables*, American Journal of Physics **62**, pp. 986–993, 1994. Cited on page 730.

787 H. Salecker & E.P. Wigner, *Quantum limitations of the measurement of space-time distances*, Physical Review **109**, pp. 571–577, 1958. Cited on page 731.

788 E.J. Zimmerman, *The macroscopic nature of space-time*, American Journal of Physics **30**, pp. 97–105, 1962. Cited on page 731.

789 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 731.

790 P. Gibbs, *The small structure of space-time: a bibliographical review*, preprint hep-th/9506171. Cited on pages 731 and 745.

791 C.A. Mead, *Possible connection between gravitation and fundamental length*, Physical Review B **135**, pp. 849–862, 1964. Cited on pages 731, 732, 734 and 754.

792 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 732.

793 M.-T. Jaekel & S. Renaud, *Gravitational quantum limit for length measurement*, Physics Letters A **185**, pp. 143–148, 1994. Cited on page 732.

794 D.V. Ahluwalia, *Quantum measurement, gravitation, and locality*, Physics Letters B **339**, pp. 301–303, 1994, also preprint gr-qc/9308007. Cited on page 732.

795 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 732.

796 G. Amelino-Camelia, *Limits on the measurability of space-time distances in (the semi-classical 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 732.

797 The generalized uncertainty relation is an implicit part of reference 791, 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 732.

798 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 732.

799 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 733.

800 M. Maggiore, *A generalized uncertainty principle in quantum mechanics*, Physics Letters B **304**, pp. 65–69, 1993. Cited on page 733.

801 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 pages 733 and 745.

802 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 733.

803 E.J. Hellund & K. Tanaka, *Quantized space-time*, Physical Review **94**, pp. 192–195, 1954. Cited on page 733.

804 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 734.

805 A. Peres & N. Rosen, *Quantum Limitations on the measurement of gravitational fields*, Physical Review **118**, pp. 335–336, 1960. Cited on page 734.

806 It is the first definition in Euclid's *Elements*, ca. 300 BCE. For an English translation see T. Heath, *The thirteen books of the elements*, Dover, 1969. Cited on page 735.

807 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 735.

808 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 738.

809 D. Finkelstein, *'Superconducting' causal nets*, International Journal of Theoretical Physics **27**, pp. 473–519, 1985. Cited on page 738.

810 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 738.

811 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 738.

812 For a discussion, see R. Sorabji, *Time, creation, and the continuum*, Cited on page 738.

813 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 740.

814 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 741.

815 N.F. Ramsey & A. Weis, *Suche nach permanenten elektrischen Dipolmomenten: ein Test der Zeitumkehrinvarianz*, Physikalische Blätter **52**, pp. 859–863, 1996. The paper by E.D. Commins, S.B. Ross, D. DeMille & B.C. Regan, *Improved experimental limit on the electric dipole moment of the electron*, Physical Review A **50**, p. 2960, 1994, gives an upper experimental limit to dipole moment of the electron of $3.4 \cdot 10^{-29}$ e m. W. Bernreuther & M. Suzuki, *The electric dipole moment of the electron*, Reviews of Modern Physics **63**, pp. 313–340, 1991. See also the musings in Hans Dehmelt, *Is the electron a composite particle?*, Hyperfine Interactions **81**, pp. 1–3, 1993. Cited on page 741.

816 K. Akama, T. Hattori & K. Katsuura, *Naturalness bounds on dipole moments from new physics*, http://www.arxiv.org/abs/hep-ph/0111238 Cited on page 741.

817 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 744.

818 W. Jauch, *Heisenberg's uncertainty relation and thermal vibrations in crystals*, American Journal of Physics **61**, pp. 929–932, 1993. Cited on page 745.

819 H.D. Zeh, *On the interpretation of measurement in quantum theory*, Foundations of Physics **1**, pp. 69–76, 1970. Cited on page 745.

820 One way out is proposed in J. Magueijo & L. Smolin, *Lorentz invariance with an invariant energy scale*, Physical Review Letters **88**, p. 190403, 26 April 2002. They propose a modificaion of the mass energy relation of the sort

$$E = \frac{\gamma m c^2}{1 + \frac{\gamma m c^2}{E_{\mathrm{Pl}}}} \quad \text{and} \quad p = \frac{\gamma m v}{1 + \frac{\gamma m c^2}{E_{\mathrm{Pl}}}} \quad . \qquad (565)$$

Cited on page 745.

Another approach of recent years, with a different proposal, is called doubly special relativity. A recent summary is ...

821 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 748.

822 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 748.

823 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. See also his update G. Amelino-Camelia, *Quantum-gravity phenomenology: status and prospects*, preprint http://arxiv.org/gr-qc/0204051, 2002. Cited on page 748.

824 L.J. Garay, *Spacetime foam as a quantum thermal bath*, Physics Review Letters **80**, pp. 2508–2511, 1998, also downloadable as http://arxiv.org/gr-qc/9801024. Cited on page 748.

825 At the unification scale, all particles are of the same type. See for example ... Cited on page 749.

826 STEVEN WEINBERG, *The cosmological constant problem*, Reviews of Modern Physics **61**, pp. 1–23, 1989. Cited on page 746.

827 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 748.

828 F. KAROLYHAZY, Il Nuovo Cimento **A42**, p. 390, 1966, 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. The discussion is neatly summarized in Y.J. NG, H. VAN DAM, *Comment on "Uncertainty in measurements of distance"*, http://www.arxiv.org/gr-qc/0209021. See also Y.J. NG, *Spacetime foam*, http://arxiv.org/gr-qc/0201022. Cited on page 748.

829 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 748.

830 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 pages 749 and 750.

831 J.K. WEBB & al., *Further evidence for cosmological evolution of the fine structure constant*, Physical Review Letters **87**, p. 091301, 2001. They claim to have determined a variation $\Delta\alpha/\alpha$ of $-7.2 \pm 1.8 \cdot 10^{-6}$ for stars with a redshift between 0.5 and 3.5. Cited on page .

832 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 http://www.arxiv.org/hep-ex/0012001. Their upper limit on the dipole moment of mercury atoms is $2.1 \cdot 10^{30}$ em. Research for better measurement methods is progressing; a modern proposal is by S.K. LAMOREAUX, *Solid state systems for electron electric dipole moment and other fundamental measurements*, http://www.arxiv.org/abs/nucl-ex/0109014. Cited on page 750.

34. Nature at large scales – is the universe something or nothing?

Die Grenze ist der Ort der Erkenntnis.*
Paul Tillich

Ref. 833 This strange question is the topic of the present leg of our mountain ascent. We explored the properties of nature in the vicinity of Planck dimensions in the previous section; it is equally fascinating to explore the other limit, namely 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 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. For every sentence of this section you will find at least one physicist disagreeing !

Typeset in January 2003

We have studied the universe several times already. 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, as sum of all matter and space-time, studied the expansion of the universe 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 of the order of the Schwarzschild radius, i.e. whenever

$$d \approx r_{\mathrm{S}} = 2Gm/c^2 \quad . \tag{566}$$

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.We do need general relativity and thus curved space-time when talking about the whole of nature. Challenge 1146

Similarly, quantum theory is required for the description of motion of an object whenever we approach it to distances d of the order of the Compton wavelength, i.e. whenever

$$d \approx \lambda_{\mathrm{C}} = \frac{h}{mc} \quad . \tag{567}$$

---

* The frontier is the place of understanding.

** Written between june and 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. For their description, 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 in space and time, this is indeed the case. We are thus 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 years or 380 Ps, providing an upper limit to time measurements. It is called the 'age' of the universe. It is deduced from two sets of 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. We 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. We thus get a common 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, no 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 1147

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 883

In order to measure the duration of a movement or the age of a system, we need a clock. The clock has to be *independent* of that movement and thus has to be *outside* the system.

---

* 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.

However, there are no clocks outside the universe. Inside it, a clock cannot be independent. In fact we just saw that inside the universe, no clock can run during *all* its history. 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: we can either talk about the age of space-time, as is done in general relativity, by assuming that matter provides suitable clocks; or we can talk about the age of matter, such as stars or galaxies, 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, we *approach* it as much as possible? The best way to clarify the issue is to ask about the measurement *error* we make 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. 835 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

$$m > \frac{\hbar}{c^2} \frac{T}{(\Delta t)^2} \quad . \qquad (568)$$

A simple way to include general relativity into this result was suggested by Ng and Van Ref. 836 Dam. Any clock of mass m has a minimum resolution Δt due to the curvature of space it

---

[*] 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. 834 scenario, i.e. for the matter dominated Einstein-de Sitter model, we have the simple relation $T = (3/2)\, t_o$.

introduces, given by

$$\Delta t > \frac{Gm}{c^3} \quad . \tag{569}$$

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

$$T < \frac{(\Delta t)^3}{t_{\mathrm{Pl}}^2} \quad , \tag{570}$$

where $t_{\mathrm{Pl}} = \sqrt{\hbar G/c^5} = 5.4 \cdot 10^{-44}\,\mathrm{s}$ 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 1148

As a consequence of (570), 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 we try to go beyond limit (570), 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 the conclusion is also valid for clocks made of radiation, such as background radiation. Challenge 1149

In short, measuring an age with a clock always involves some errors; whenever we try to reduce these errors, the clock becomes so imprecise that age measurements become impossible.

Does time exist?

> Time is waste of money.
> Oscar Wilde

From the origins of physics onwards, the concept of ‘time’ has designated what is measured by a clock. Therefore equation (570), 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 or everyday concept. See page 37

Despite this conclusion, time is obviously a useful concept in everyday life. A simple explanation appears when we focus 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 $1\,\mathrm{eV}/10\,\mathrm{GeV} = 10^{-10}$.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; indeed, both its position and speed must Challenge 1150 be measured, while the two measurement errors are related by $\Delta v \Delta x > \hbar/m$. This requires even more mass to screen the pointer from outside influences, thus possibly 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. 833

See page 732

Challenge 1151

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 (570) implies that the highest precision possible for a clock is about 10^{-23} s, orabout the time light takes to move across a proton. The finite age of the universe also yields also a maximal *relative* measurement precision. Expression (570) can be written as

Challenge 1152

$$\frac{\Delta t}{T} > \left(\frac{t_{\rm Pl}}{T}\right)^{2/3} \tag{571}$$

which shows that no time interval can be measured with more than about 40 decimals.

In order to clarify the issue we calculate the measurement error as function of the observation energy. We get two limits. For *small* energies, the error is given by quantum theory as

$$\frac{\Delta t}{T} \sim \frac{1}{E_{\rm meas}} \tag{572}$$

and thus goes down with measurement energy. For high energies, the error is given by gravitational effects by

$$\frac{\Delta t}{T} \sim \frac{E_{\rm meas}}{E_{\rm Pl}} \tag{573}$$

so that the total result is given in Figure 252. In particular, too high energies do not help to reduce measurement errors, as any attempt to reduce the measurement error for the age of the universe below 10^{-23} s would require energies so high that the limits of space-time would be reached, making the measurement itself 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 of this limit, as well as more details, by looking at the methods to determine the age of the universe from space-time.

Imagine you see a tree which, due to some wind storm, fell towards another, touching it at the very top, as shown in Figure 253. 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

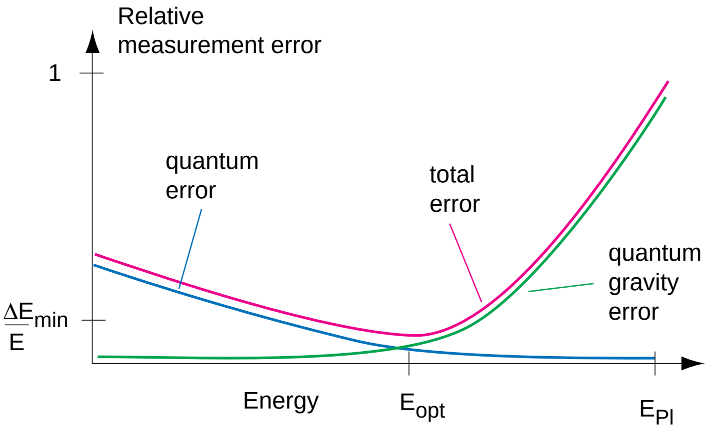


Figure 252 Measurement errors as a function of measurement energy

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

$$T = \frac{d}{v} \quad . \tag{574}$$

This is in simple 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

$$\frac{\Delta T}{T} = \frac{\Delta d}{d} + \frac{\Delta v}{v} \quad ; \tag{575}$$

exploring it in more detail is worthwhile. For any measurement of T we have 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 (575), always taking into account both quantum theory *and* general relativity.

At everyday energies, the result of the determination of the age t_o is about $12 \pm 2 \cdot 10^9$ years. 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.

---

* At higher redshifts, the speed of light as well as the details of the expansion come into play; in the image of inclined trees, we find that the trees are not straight all the way up to the top, and that they grow on a slope, as shown in Figure 254.

What is the smallest possible distance error? Obviously, equation (571) implies

$$\frac{\Delta d}{T} > \frac{l_{\mathrm{Pl}}^{2/3}}{d^{2/3}} \qquad (576)$$

thus giving the same age uncertainty for the universe as found above in the case of material clocks. Challenge 1153

We 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, we need high observation energies. It does not take much to notethat at observation energies near the Planck value, the value of $\Delta T/T$ approaches unity. Challenge 1154 In fact, both terms on the right hand side of expression (575) 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*.

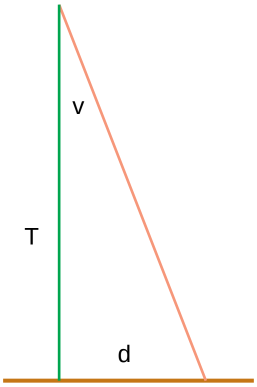


Figure 253 Trees and galaxies

Let us continue with the other extreme, namely objects extremely far away, say with a redshift of $z \gg 1$. Relativistic cosmology requires the diagram of Figure 253 to be replaced by the more realistic diagram of Figure 254. The 'light onion' replaces the familiar light cone of special relativity: light converges near the big bang. Ref. 834

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. We inevitably reach 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.

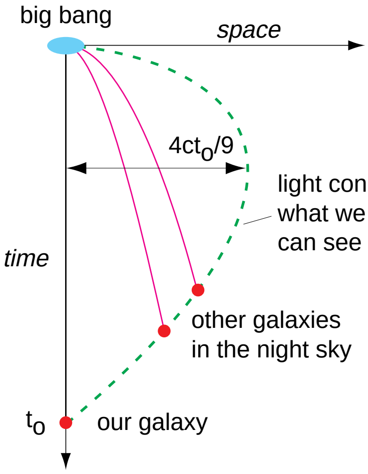


Figure 254 Speed and distance of remote galaxies

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. As a result, for large distances the error on the horizon distance becomes as large as the value to be measured.

In short, even 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, a result we also found at Planck distances. Whenever we aim for perfect precision, we find 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 have to conclude that the anthropomorphic concept of 'age' does not make any sense for the universe as a whole. The usual textbook value 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 issueof 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 1155

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. In a hyperbolic or parabolic universe, even though the size is infinite, the most distant visible events (which form the horizon) are at finite distance.* For elliptical evolution, the total size is finite and depends on the curvature; but also in this case the present measurement limit yields a minimum size for the universe many times larger than the horizon distance. At least, this is what general relativity says. See page 318

On the other hand, quantum field theory is based on flat and infinite space-time. Let us see what happens when both theories are combined. What can we 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 we 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. We should have started installing the meter bar right at the big bang. Ref. 834

For fun, let us assume that we actually managed to do this. How far could we 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, Figure 254 shows that the maximum *spatial distance* an object can be seen away from us is only $(4/9)ct_0$. Obviously, for space-time intervals, the maximum remains ct_0. Ref. 834

In all cases it thus 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 real surprises come now.

---

* In cosmology, we need 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. 834

Is the universe really a big place?

Ref. 837 See page 883 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 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, we have

$$R(t) = C\, t^{2/3} \quad , \tag{577}$$

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 still getting 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 high energies. We saw above that space-time and matter are not distinguishable at Planck scales. Therefore, at Planck energy we cannot state whether objects are *localized* or not. At Planck scales, a basic distinction of our thinking, namely 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, none of the observations in daily life from which we deduce that space is extended are possible at Planck scales. *At Planck scales, the basic distinction between vacuum and matter, namely the opposition between 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.

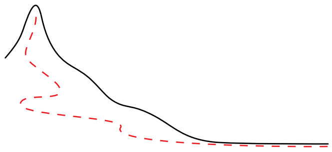
At cosmological distances, the situation is even easier. All arguments given above on the Challenge 1157 measurement errors for the age can be repeated for the distance of the horizon.Essentially, at largest distances and at Planck energies, the measurement errors are of the same magnitude as the measured value. All this happens because length measurements become impossible at nature's limits. This is corroborated by the lack of any standard with which to compare the size of the universe.

Also studying the big bang produces strange results. At Planck energies, whenever we try to determine the size of the big bang, we 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 we 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 Challenge 1158 to discoverthat 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.

Since this also applies when we try 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 from 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. If we start



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

measuring the sky at standard observation energies, trying to increase the measurement precision of the distance to the horizon, the measurement error increases beyond all bounds. 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 the approaches of 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. There is even no way to determine the dimensionality of the horizon, nor the dimensionality of space-time near it.*

Measurements thus 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 1160

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. 833

Finally, there is no way to decide whether the various boundary points are *different* from each other. The distance between two points on the night sky 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?

Ref. 838 One often reads about the quest for the initial conditions of the universe. But before joining the search, we should ask *whether* and *when* such initial conditions make any sense. Obviously, our everyday description of motion requires them. Initial conditions describe the

---

* In addition, the measurement errors imply that no statement can be made about translation symmetry at cosmological scales. Are you able to confirm this?In addition, 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 1159

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, are attributed to a system by an *outside* observer.

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, independently 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 curvature, density or temperature, 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.) Ref. 833

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 requires that two conditions 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 can objects *move* in space. Of course, this is always the case in everyday life.

At this point of our mountain ascent, where time and length are unclearly defined at cosmological scales, it should come as no surprise that the concept of mass has similar difficulties.

Does the universe contain particles and stars?

The number of stars, about $10^{23\pm1}$, is included in every book on cosmology, as it is in the table of Appendix B. A subset of this number can be counted on clear nights. If we ask 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 $10^{90\pm1}$. See page 883

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 we combine them?

In order to define the number of particles in a region, quantum theory first of all requires a vacuum state. Particle number is 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 no 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 are we able to speak of an approximate particle number. See page 561

Comparison between a system and the vacuum is impossible in the case of the universe also for purely practical reasons. The requirement effectively translates into the requirement that the particle counter be outside the system. (Can you confirmthe connection?) In addition, it is impossible to remove particles from the universe. The impossibility to define a vacuum state and thus a particle number for the universe is not surprising. It is an interesting exercise to investigate the measurement errors appearing when we try to determine a particle number despite this fundamental impossibility.

Challenge 1161

Challenge 1162

Can we count 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 we stay away from Planck energies and from the horizon.

Therefore, 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 completely.

This conclusion is so strange that we cannot accept it too easily. Let us try another method to determine the matter content: instead of counting, let us weigh.

Does the universe contain masses and objects?

See page 883

The average density of the universe, of about $10^{-26}\,\mathrm{kg/m^3}$, is frequently cited in texts. 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, we have $E/c = \Delta p > \hbar/\Delta x$, so that in a volume of size R, we have 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

$$\rho_{\mathrm{min}} \approx \frac{m_{min}(R)}{R^3} = \frac{\hbar}{cR^4} \quad . \tag{578}$$

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 (578) does not really provide a clear statement.

Challenge 1163

* In fact, at everyday energies it lies almost exactly between the two values, yielding the strange relation

$$m_0^2/R_0^2 \approx m_{\mathrm{Pl}}^2/R_{\mathrm{Pl}}^2 = c^4/G^2 \quad . \tag{579}$$

Another way to measure the mass of the universe would be to use 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.We would need to produce or to measure a velocity change Δv for the rest of the universe after the collision. To hit all mass in the universe at the same time, we need high energy; but then we are hindered by Planck energy effects. In addition, a properly performed collision measurement would require a mass outside the universe, a rather difficult feat.

See page 221

Challenge 1164

Still another way to measure the mass would be 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.

We are 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 286

$$\kappa = \frac{1}{r^2_{\text{curvature}}} = \frac{3}{4\pi}\frac{4\pi R^2 - S}{R^4} = \frac{15}{4\pi}\frac{4\pi R^3/3 - V}{R^5} \quad . \tag{580}$$

We have to insert the horizon radius R_o and either its surface S_o or its volume V_o. However, given 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 1165

An equivalent method starts with the usual expression for the scalar curvature uncertainty $\Delta\kappa$ for a region of size R provided by Rosenfeld, namely

Ref. 839

$$\Delta\kappa > \frac{16\pi l^2_{\text{Pl}}}{R^4} \quad . \tag{581}$$

But also this expression shows 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 1166

Do symmetries exist in nature?

We have already seen that at the horizon, space-time translation symmetry breaks down. Let us have a quick look at the other symmetries.

See page 636

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 $Gn_b m_p = 1/t_o^2$. The relation is required by several cosmological models.

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. 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 from the 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.For its definition, the concept of gauge field requires a distinction between time, space, and mass; at the horizon this is impossible. We deduce that at the horizon also concepts such as algebras of observables cannot be used to describe nature. Also renormalization breaks down.

Challenge 1167

All symmetries of nature break down at the horizon. The complete vocabulary we use to talk about observations, such as magnetic field, electric field, potential, spin, charge or speed, cannot be used at the horizon. And that is not all.

Is there a boundary of the universe?

It is common to take ‘boundary’ and ‘horizon’ to be synonyms in the case of the universe, as they are the same for all practical purposes. To study it, knowledge of mathematics does not help us; 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. 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!A distinction *is* possible in general relativity; a distinction also possible in quantum theory. 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 no boundary.

Challenge 1168

That is definitely interesting; it suggests that nature might be *symmetric* under transformations which exchange interiors and boundaries. This connection is nowadays called *holography* because it vaguely recalls the working of credit card holograms. It is an busy research field in present high energy physics. However, for the time being we continue with our original theme, which directly leads us to ask: Ref. 840

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 set 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 consists of different initial conditions. But our discussion so far shows that the universe is *not* a set 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, such as between a tooth brush and a mountain, is possible only *approximately*. The approximation is possible because 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. *It is impossible to split the universe into separate entities.*

Another way to reach 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. Only at everyday energies it is approximately possible.

In short, since the universe contains no distinguishable entities, *the universe is not a set*. We envisaged 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.* See page 493

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-space and its generalizations, Fock space or particle space, are based on sets. They all must be abandoned at Planck energy, and thus also in any precise description.*

Also many speculations about unified descriptions do not satisfy the criterion of eliminating sets. In particular, all studies of quantum fluctuations, mathematical categories, posets, complex mathematical spaces, computer programs, Turing machines, Gödel's theorem, creation of any sort, space-time lattices, quantum lattices or Bohm's unbroken wholeness fail to satisfy the requirement. In addition, almost all speculations about the origin of the universe cannot be correct. For example, you might want to check the religious explanations you know against this result.In fact, no approach of theoretical physics in the year 2000 satisfies the requirement to abandon sets; maybe a future version of string or M theory might do so. Challenge 1170

Challenge 1169 * Do you know a concept not based on a set?

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. The difficulties to comply with this 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; 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 failing to be 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 and no cosmological constant. Neither is it thermodynamically closed or open, nor does it contain any information. All thermodynamical quantities, such as entropy, temperature or free energy, 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 are physical systems. The universe is neither.

Curiosities

> Insofern sich die Sätze der Mathematik auf die Wirklichkeit beziehen, sind sie nicht sicher, und sofern sie sicher sind, beziehen sie sich nicht auf die Wirklichkeit.
> Albert Einstein

▪ In mathematics, $2+2=4$. This statement is an idealisation of statements such as “two apples plus two apples are four apples.” However, we now know that at Planck energies this is not a correct statement about nature. At Planck energy, objects cannot be counted. Are you able to confirm this? Challenge 1171

▪ In 2002, Seth Lloyd estimated how much information the information the universe can contain, and how many calculations it has performed since the big bang. His estimate is based on two ideas: that the number of particles in the universe is a well-defined quantity, and that the universe is a computer, i.e. a physical system. We now know that both assumptions are not correct. This example shows the power of the criteria that we deduced for the final description of motion. Ref. 841

▪ People are taking pictures of the cosmic background radiation and its variations. Is it possible that the graphs will show that the spots into one direction of the sky are exactly the same as those in exactly the opposite direction? Challenge 1172

---

* 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.

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 twentieth century. Most problems provided challenges to many mathematicians for decades afterwards. A few are still unsolved, among them the sixth. The sixth problem was challenged mathematicians and physicists to find an *axiomatic* treatment of physics.

See page 472
Ref. 842

Since the universe is not even a set, we 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 contains at least one set.

The impossibility of an axiomatic system for physics is also confirmed in another way. Physics starts with a *circular* definition: space-time is defined with help of objects and objects are defined with help of space-time. Physics thus has *never* been modelled after mathematics. Physicists always had to live with logical problems.

See page 117

The situation is similar to the description of the sky by a child as 'made of air and clouds'. Looking closely, we discover that clouds are made of water droplets. But there is air inside clouds, and there is also water vapour elsewhere in air. When clouds and air are watched through the microscope, there is no clear boundary between the two. We cannot define any of the terms 'cloud' and 'air' without the other. No axiomatic definition is possible.

Objects and vacuum also behave in this way. Virtual particles are found in the vacuum, and vacuum is found inside objects. At Planck scales, there is no clear boundary between the two; we cannot either of them without the other. In both cases, despite the lack of precise definition and despite the logical problems ensuing, the description works well at large, everyday scales.

But then a question arises naturally:

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? Or of particles and vacuum? 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, in the section on Planck scales, and in the intermezzo on brain and language. Table 54 shows the result.

See page 446

Table 54 Physical properties of the universe

- The universe has no age.
- The universe has no beginning.

---

[*] Thus I have devoted myself to magic, [...] that I understand how the innermost of the world is held together.

Table 54 Physical properties of the universe

- The universe has no size.
- The universe has no shape.
- The universe has no mass.
- The universe has no density.
- The universe has no cosmological constant.
- The universe has no state.
- The universe is not a physical system.
- The universe is not isolated.
- The universe has no boundaries.
- The universe cannot be measured.
- The universe cannot be distinguished from nothing.
- The universe contains moments.
- The universe is not a set.
- The universe cannot be described.
- The universe cannot be distinguished from vacuum.
- The universe has no volume.
- The universe has no particle number.
- The universe has no energy.
- The universe contains no matter.
- The universe has no initial conditions.
- The universe has no wave function.
- The universe contains no information.
- The universe is not open.
- The universe does not interact.
- The universe cannot be said to exist.
- The universe cannot be distinguished from a single event.
- The universe is not composite.
- The universe is not a concept.
- There is no plural for ‘universe’.
- The universe was 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 Challenge 1174 to make a single sensible statement. (Can you find one?)We are only able to say which properties it does *not* have. We are unable to find any property the universe *does* have. The universe has no properties! 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.

See page 459 We get a confirmation for this strange conclusion from the first intermezzo. There we found that any concept needs a defined content, defined limits, and a defined domain of application. In this section, we found 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 ‘why’ wrongly suggest that something might exist outside the universe, providing a reason for it, and thus contradicting the definition of the term ‘universe’ itself; most importantly of all, the universe simply does not exist. Any sentence containing the word universe makes no sense. The term ‘universe’ only *seems* to express something, but it doesn’t. 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.

---

[*] There is also a well-known non-physical concept of which nothing positive can be said. Many scholars have Challenge 1173 explored it in detail. Can you spotit?

[**] Of course, the term ‘universe’ still makes sense if it is defined more restrictively, such as ‘everything interacting with a particular human or animal observer in everyday life.’ But such a definition is not useful for our quest, as it lacks the precision required for any description of motion.

No contradictions

By taking into account the limits to length, time, mass and all other quantities we have encountered, we deduced a number of strange conclusions about nature. However, we got something in exchange. All contradictions between general relativity and quantum theory that we mentioned in the beginning of this chapter are resolved. Challenge 1175 e

We had to leave behind us many cherished habits, but in exchange we have the promise of a description of nature without contradictions. But we get even more.

Extremal scales and open questions of physics

In the chapter *Quantum physics in a nutshell* we had listed all the unexplained properties of nature left open either by general relativity or quantum theory. The present conclusions provide a new connection among them. Indeed, many of the cosmological results of this section sound surprisingly familiar; let us compare them systematically with those of the section on Planck scales. Both sections explored topics – some in more details than others – from the list of unexplained properties of nature. See page 701

Table 55 Properties of nature at maximal, everyday and 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
renormalization	does not work	works	does not work
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

First of all, Table 55 shows that *each* of the unexplained properties makes no sense at *both* limits of nature, the small *and* the large. All open questions are open at both extremes.

Secondly and more importantly, nature behaves in the same way at horizon scales and at Planck scales. In fact, we have not found any difference between the two cases. Are you able to discoverone?* We are thus lead to the hypothesis that nature does not distinguish Challenge 1176 between the large and the small. Nature seems to be characterized by *extremal identity*.

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 extremes;
- a description of nature requires both general relativity and quantum theory;
- nature is not a set;
- initial conditions and evolution equations make no sense at nature's limits;
- there is a relation between local and global issues in nature;
- the concept of 'universe' makes no sense.

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 only candidate for the quest of unification. Extremal identity is beautiful in its simplicity, its unexpectedness and its richness of consequences. Just explore it a little for yourself. Challenge 1177

The consequences of extremal identity are presently studied with great intensity in high energy particle physics, though often under different names. Ref. 843 The simplest approach to extremal identity – in fact too simple to be correct – is inversion. It looks as if extremal identity implies a connection such as

$$r \leftrightarrow \frac{l_{\mathrm{Pl}}^2}{r} \quad \text{or} \quad x_\mu \leftrightarrow \frac{l_{\mathrm{Pl}}^2\, x_\mu}{x_\mu x^\mu} \quad . \tag{582}$$

Could this mapping, called *inversion*, be a symmetry of nature? At every point of space? For example, inserting the horizon distance, equation (582) would imply that lengths smaller than $l_{\mathrm{Pl}}/10^{61} \approx 10^{-96}\,\mathrm{m}$ never appear in physics. Is this the case?What would inversion imply for the big bang? Challenge 1178

Numerous fascinating questions are contained in the simple hypothesis of extremal identity. They lead to *two* main directions of investigation.

We have to start by searching for some stronger arguments for the validity of extremal identity. We will discover a number of simple arguments, all showing that extremal identity is indeed a property of nature, and producing many beautiful insights.

The other quest then follows. We need to find the correct version of equation (582). That oversimplified expression is neither sufficient nor correct. It is not sufficient because it does 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 1179

But inversion is also simply wrong. Inversion is not the correct description of extremal identity because it does not realize a central result discovered above: it does not connect *states* and *intrinsic properties*. Inversion keeps them distinct. Among others, this means that inversion does not take into account *interactions*. And most open issues in at this point of our mountain ascent are properties of interactions.

∗ If so, send a message to mm@motionmountain.net; but first of all, publish it!

References

833 Available at http://www.dse.nl/motionmountain/C10-QMGR.pdf. 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 C. Schiller, *Does matter differ from vacuum?* http://www.arxiv.org/abs/gr-qc/9610066. Cited on pages 757, 761, 767 and 768.

834 See the lucid discussion by G.F.R. Ellis & T. Rothman, *Lost horizons*, American Journal of Physics **61**, pp. 883–893, 1993. Cited on pages 759, 763 and 764.

835 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 759.

836 F. Karolyhazy Il Nuovo Cimento **A42**, p. 390, 1966, 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. Newer summaries are Y.J. Ng, H. van Dam, *Comment on "Uncertainty in measurements of distance"*, http://www.arxiv.org/gr-qc/0209021 and Y.J. Ng, *Spacetime foam*, http://arxiv.org/gr-qc/0201022. Cited on page 759.

837 See e.g. the Hollywood movie *Contact*, based on the book by Carl Sagan, *Contact*, 1985. Cited on page 765.

838 For example the international bestseller by S. Hawking, *A Brief History of Time – From the Big Bang to Black Holes*, 1988. Cited on page 767.

839 L. Rosenfeld, in H.-J. Treder, *Entstehung, Entwicklung und Perspektiven der Einsteinschen Gravitationstheorie*, Springer Verlag, 1966. Cited on page 770.

840 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 the often cited paper by L. Susskind, *The world as a hologram*, Journal of Mathematical Physics **36**, pp. 6377–6396, 1995, or as http://arxiv.org/abs/hep-th/9409089 Cited on page 772.

841 S. Lloyd, *Computational capacity of the universe*, Physical Review Letters **88**, p. 237901, 2002. Cited on page 773.

842 See for example *Die Hilbertschen Probleme*, Akademische Verlagsgesellschaft Geest & Portig, 1983, or Ben H. Yandell, *The Honours Class: Hilbert's Problems and their Solvers*, A.K. Peters, 2002. Cited on page 774.

843 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 777.

A classical version of duality is discussed by M.C.B. Abdalla, A.L. Gadelka & I.V. Vancea, *Duality between coordinates and the Dirac field*, hep-th/0002217.

35. 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 head in disbelief, and strongly disapprove. In this section we shall argue that it is possible to learn so much about physics while making love that discussions about their relative beauty can be put aside altogether.

Typeset in January 2003

Imagine to be with your partner on a beautiful tropical island, just after sunset, and to 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.

Imagine being busy enjoying each other's company. The 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 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 animalic.
Sex is holy.
Sex uses motion again.
Sex is private.

Let us see why.

- Sex is *communication*. Communication is possible because nature looks similar from different standpoints and because nature shows 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 implies that the sun will rise tomorrow if the sea level around the island is the usual one. Challenge 1180 n
- 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. 844

* In fact, studying the influences of *sex* on physics is mostly a waste of time. We avoid it. True, maybe one day we shall 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, as shown in this section. In the following, we thus bow to the modern habit of saying 'sex' instead of 'sexuality'.

and find that applying the expression for universal gravity

$$E_{\text{pot}} = -\frac{GMm}{r} \tag{583}$$

to both of you, 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 *makes noise*. That is no news. However, even after sex, even when everybody and everything is quiet, in a completely silent environment, we do hear something. The noises we hear 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. In fact, all proofs for the discreteness of matter, of electric current, of energy, or of light are based on the increase of fluctuations with the smallness of systems under consideration. The persistence of noise thus makes us suspect that matter is made of smallest entities. Making love confirms this suspicion in several ways.

- 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, in other words if the structure can be built from small standard entities. Thus we again suspect ourselves to be made of smallest, discrete entities.

Sex is also a complicated method of reproduction. Mathematics provides a much simpler 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, cut it into six pieces, and rearrange the pieces in such a way that the result are *two* copies of the same size and volume as the original. In fact, even volume increases can be produced in this way, thus realizing growth without any need for food. Mathematics thus provides some interesting methods for growth and reproduction. However, they assume that matter is continuous, without a smallest length scale. The observation that these 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. 845

In summary, sex shows that we are made of some kind of lego bricks: depending on the level of magnification, these bricks are called molecules, atoms, or elementary particles. It is possible to estimate their size using the sea around the tropical island, as well as a bit of oil. Can you imagine how? Challenge 1181 n

- Sex *uses the sense of sight*. Seeing each other is only possible because we are cold whereas the sun is hot. If we and our environment all had the same temperature as the sun, we would not see each other. This can be checked experimentally by looking into a hot

Ref. 846 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, their speed can be measured. Since measurement is a comparison with a standard, there must be a velocity standard in nature, some special velocity standing out. 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. We are thus looking for a maximum value. To estimate the value of the maximum, just take your cellular phone and ring home from the 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, we find that this effect becomes noticeable at energies

$$E_{\text{different time}} \approx mc^2 \quad . \tag{584}$$

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 involved are larger than than those of touching fly legs. In short, people are electric.

In the dark, we observe that discharges emit light. Light is thus related to electricity. In addition, touching 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. Challenge 1182 n

- Sex is *fun*. People like to make love in different ways, such as in a dark room. But rooms get dark when the light is switched off only because we live in a space of odd dimensions. In even dimensions, a lamp would not turn off directly after the switch is flipped, but dim only slowly.

Sex is also fun because with our legs, arms, and bodies we can make knots. Knots are possible only in three dimensions. In short, sex is real fun only because we live in 3 dimensions.

- Sex is *tiring*. The reason is gravity. But what is gravity? A little thinking shows that since there is a maximum speed, gravity is the curvature of space-time. Curved space also means that a *horizon* can appear, i.e. a largest possible visible distance. From equations (583) and (584), we deduce that this happens when distances are of the order of

$$R_{\text{horizon}} \approx Gm/c^2 \quad . \tag{585}$$

For example, only due 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 known that sex happens between your ears. Indeed, biological research has shown that we have a clock inside the brain, due to circulating electrical currents. This clock provides our normal sense of time. Since such a brain clock can be built, there must be a time standard in nature. Again, such a standard must be either a minimum or a maximum time interval. We shall discover it later on. Ref. 850

- 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 provide one. Classical physics only allows for the measurement of speed. Classical physics cannot explain that the measurement of length, time, or mass is possible.* Classically, matter cannot be hard; it should be possible to compress it. But sex shows us that this is not the case. Sex shows 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 when energies are concentrated on small volumes, such as a few aJ per atom. Investigating such situations more precisely, we finds that strange phenomena appear at distances r if energies exceed the value

$$E \approx \frac{\hbar c}{r} \quad ; \tag{586}$$

in particular, energy becomes chunky, things become fuzzy, boxes are no tight, and particles get be confused. These 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 other words, objects tear and break because in nature there is a minimum action, given roughly by $\hbar$.

If even more energy is concentrated in small volumes, such as energies of the order of mc^2 per particle, one even observes transformation of energy into matter, or *pair production*. From equations (584) and (586), we deduce that this happens at distances of

$$r_{\text{pair production}} \approx \frac{\hbar}{m\,c} \quad . \tag{587}$$

At such small distances we cannot avoid using the quantum description of nature.

- Sex is not only *Greek*. The Greeks 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, Democritus 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 Democritus described 2500 years ago.

It seems that physics has not made much progress in the mean time. Take the statement made in 1939 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. 847

Compare it with the version of 2003:

> Baryons in the universe: $10^{81\pm1}$; total charge: near zero.

---

* Note that the classical electron radius is not an exception: it contains the elementary charge e, which contains a length scale, as shown on page 335.

The second is more honest, but which of the two is less sensible? Both sentences show that there are unexplained facts in the Greek description nature, in particular the number of involved particles.

▪ Sex is *animalic*. We have seen that we can learn a lot about nature from the existence of sex. We could be tempted to see this approach of nature as a special case of the so-called *anthropic principle*. However, some care is required here. In fact, we could have learned exactly the same if we had taken as starting point the observation that *apes* or *pigs* have sex. There is no 'law' of nature which distinguishes between them 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, apes, and pigs.

A famous anthropic deduction was drawn by the British astrophysicist Fred Hoyle. While studying stars, he predicted a resonance in the carbon-12 nucleus. If it did not exist, he argued, stars could not have produced the carbon which afterwards was spread out by explosions into interstellar space and collected on earth. Also apes or pigs could reason this way; therefore Hoyle's statement does make sense.

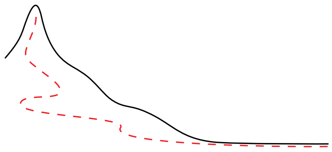
On the other hand, claiming that the universe is made *especially* for people is not sensible: using the same arguments, pigs would say it is made for pigs. The existence of either requires all 'laws' of nature. In summary, the anthropic principle is true only in so far as its consequences are indistinguishable from the porcine or the simian principle. In short, the animalic side of sex puts limits to the philosophy of physics.

▪ Sex is *holy*. Following the famous definition by the theologian Rudolf Otto, holiness Ref. 848 results from a mixture of a *mysterium tremendum* and a *mysterium fascinans*. Tremendum means that it makes one tremble. Indeed, sex produces heat and 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. 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, which brought our parents together, our grandparents, and made life appear on earth.

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 of Table 57, gives all facts still unexplained. It can also be called 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.)



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

Table 57 *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
$0.5\,\mathrm{nJ/m^3}$	the value of the observed vacuum energy density or cosmological constant
$3+1$	the number of space and time dimensions
Global quantities, from general relativity	
$1.2(1) \cdot 10^{26}\,\mathrm{m}$?	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 and stars (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 $\mathrm{R} \times \mathrm{S}^3$?	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 Table 57 were only a tiny bit different. It is not a secret that small changes in nature would lead to completely different observations, as shown in Table 58.

Table 58 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 to realize that we are only a *tiny* part of nature. We are a small droplet shaken around in the ocean of nature. Even the tiniest changes in nature 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 dependent and connected to the rest of the universe.

▪ We said above that sex uses *motion*. It contains 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 – as done in detail in general relativity – and that we define bodies with space-time – as done in detail in quantum theory. This circular reasoning shows that making love is truly 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 approach seems to be M-theory, the modern extension of string theory. But any such attempt has to overcome important difficulties which can also be experienced while making love.

▪ Sex is *private*. But is it? Privacy assumes that a person can separate itself from the rest, without important interactions, at least for a given time, and come back later. This is possible if the person puts enough *empty space* between itself and others. In other words, privacy is based on the idea that objects can be distinguished from vacuum. Let us check whether this is always possible.

What is the smallest measurable distance? 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 Ref. 849

$$\delta l \geqslant \Delta l \geqslant \frac{\hbar}{mc} \quad . \tag{588}$$

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

$$\delta l \geqslant \sqrt{\frac{\hbar G}{c^3}} = l_{\mathrm{Pl}} = 1.6 \cdot 10^{-35}\,\mathrm{m} \quad . \tag{589}$$

This expression defines a minimum length in nature, the so-called *Planck length*. Every other Gedankenexperiment leads 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 800 times the Planck length.

In other words, privacy has its limits. In fact, the issue is even more muddled when we explore the consequences for bodies. A body, also a human one, is something we can touch, throw, hit, carry or weigh. Physicists say that a body is something with energy or momentum. Vacuum has none of it. In addition, vacuum is unbounded, whereas objects are bounded.

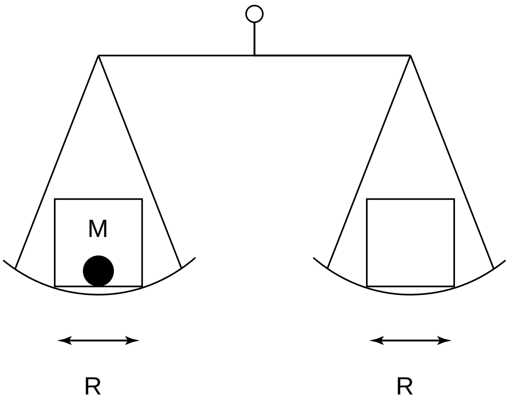


Figure 255 A *Gedankenexperiment* showing that at Planck scales, matter and vacuum cannot be distinguished

What happens if we try to weigh objects at Planck scales? Quantum theory makes a simple prediction. If we put an object of mass M in a box of size R onto a scale, equation (586) implies that there is a minimal mass error ΔM given by

$$\Delta M \approx \frac{\hbar}{cR} \quad . \tag{590}$$

If the box has Planck size, the mass error is the Planck mass

$$\Delta M = M_{\mathrm{Pl}} = \sqrt{\hbar c/G} \approx 22\,\mu\mathrm{g} \quad . \tag{591}$$

How large is the mass we 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

$$\rho \approx \frac{M}{r^3} = \frac{v^2}{Gr^2} \quad . \tag{592}$$

Using equation (588) we find that the maximum mass density in nature, within a factor of order one, is the so-called *Planck density*, given by

$$\rho_{\mathrm{Pl}} = \frac{c^5}{G^2\hbar} = 5.2 \cdot 10^{96}\,\mathrm{kg/m^3} \quad . \tag{593}$$

Therefore the *maximum* mass that can be contained inside 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 every other Gedanken-experiment exploring the issue. Ref. 849

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 1184 n

So, why is there something instead of nothing? Making love shows that there is no difference between the two options!

- Sex makes us *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. Ref. 851

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

844 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 779.

845 The consequences of memory loss in this case are already told by VOLTAIRE, *Aventure de la mémoire*, 1775. Cited on page 780.

846 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 781.

847 The famous quote is found at the beginning of chapter XI, 'The Physical Universe', in ARTHUR EDDINGTON, *The Philosophy of Physical Science*, Cambridge, 1939. Cited on page 782.

848 See the first intermezzo for details and references. Cited on page 783.

849 Details can be found in Chapter XII of this text, a reworked version of the pages 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. Cited on pages 786 and 788.

850 An introduction to the sense of time due to electrical currents circulating in the brain is found in ... Cited on page 781.

851 For details of the arguments leading to duality, see section 34. It also includes suggestions supporting the notion that the universe is not a even a set, thus proposing a solution for Hilbert's sixth problem. Cited on page 788.

36. The shape of points

This section describes a research topic,[*] and as such is *not* a compendium of generally accepted results.

Summary

Typeset in January 2003

Using only the expressions for the Compton wavelength $\lambda = \hbar/mc$ and for the Schwarzschild radius $r_s = 2Gm/c^2$, a number of arguments are presented which lead to the conclusion that at Planck energies, space-time points and point particles must be described, in contrast to their name, by *extended* entities. These arguments point towards a connection between microscopic and macroscopic scales, confirming the present results of string theory and quantum gravity. At the same time, they provide a pedagogical summary of this aspect of present day theoretical physics.

1. It is shown that any experiment trying to measure the size or the shape of an elementary particle with high precision inevitably leads to the result that at least one dimension of the particle is of macroscopic size.

2. It is argued that there is no data showing that space-time is continuous, but enough data showing that it is not. It is then argued that as a consequence, one necessarily needs extended building blocks to build up an entity, such as the vacuum, which is extended in three dimensions.

3. The existence of minimum measurable distances and time intervals is shown to imply the existence of space-time duality, i.e. a symmetry between very large and very small distances. Space-time duality in turn implies that the fundamental entities which make up vacuum and matter are extended.

4. In another, purely logical argument it is argued that the constituents of the universe and thus of space-time, of matter and of radiation cannot form a set, and that as a consequence any useful description of nature must use extended entities.

5. The Bekenstein-Hawking expression for the entropy of black holes is used to argue that its surface dependence confirms that both space-time and of particles are composed of extended entities.

6. A similar argument, based on extending statistical properties to Planck scales, shows that particles behave as braids at high energies, a fact which also requires extended entities.

7. The minimum distance relation implies that near Planck scales, braid statistics becomes important for space-time points, which in turn means that they are extended.

8. It is finally argued that even at low energy all fermions intrinsically are extended entities and that the Dirac construction for spin provides a model for them, without contradiction with experiments.

An overview of other arguments in favour of extended entities provided by present research efforts is also listed. To complete the discussion, a number of experimental and theoretical checks for the idea of extended building blocks of nature are presented.

---

[*] It was written between december 2001 and may 2002.

Introduction: the mass of vacuum

Nihil tam difficile est, quin quaerendo investigari potest.
Terentius*

The separation of the description of nature into general relativity and quantum theory has many strange consequences. Most of all, what we use to call 'point' turns out to have quite counterintuitive properties, some of which we will explore in the following. So far, we have given the standard argument showing that points do not exist in nature. The Compton wavelength and the Schwarzschild radius together determine a *minimal* length and a *minimal* time interval in nature. They are given (within a factor of order one, usually omitted in the following) by the Planck length and the Planck time, with the values Ref. 852, 853

$$\begin{aligned} l_{\mathrm{Pl}} &= \sqrt{\hbar G/c^3} = 1.6\cdot 10^{-35}\,\mathrm{m} \\ t_{\mathrm{Pl}} &= \sqrt{\hbar G/c^5} = 5.4\cdot 10^{-44}\,\mathrm{s} \quad . \end{aligned} \tag{594}$$

The existence of a minimal length and space interval in nature implies that points in space, in time, or in space-time have no experimental backing, and that we are forced to part from the traditional idea of continuous space and time. Even though properly speaking, points do not exist, and thus space points, events, or point particles do neither, we can still ask what happens when we study these entities in detail. The results provide many fascinating surprises.

Using a simple *Gedankenexperiment*, we have found that particles and space-time cannot be distinguished from each other at Planck scales. The argument was the following. The largest mass that can be put in a box of size R is a black hole with a Schwarzschild radius of the same value. But any piece of vacuum also has a minimum mass; it is given by the smallest possible mass measurement uncertainty.

The issue is important and merits a special focus. Mass measurement errors always prevent humans to state that a region of space has zero mass. In exactly the same way, also nature cannot 'know' that the mass of a region is zero, provided that this error is due to quantum uncertainty. Otherwise nature would circumvent quantum theory itself. Energy, and thus mass, is always unsharp in quantum theory. We are not used to apply this knowledge to vacuum itself, but now we have to. We remember from quantum field theory that the vacuum, like any other system, has a mass; of course, its value is zero for long time averages. But for finite measuring times, its value will be uncertain and thus different from zero. In the present discussion we show that not only limitations in time, but also limitations in space lead to mass uncertainty for the vacuum. These uncertainties in turn lead to a minimum mass for vacuum regions of finite size. Quantum theory implies that nobody, not even nature, *knows* the mass value of a region of empty space.

A box is empty if it does not contain anything. But emptiness is not well defined for photons with wavelength of the size R of the box or larger. Thus the mass that is present in such a box – corresponding to what we call vacuum – is due to the uncertainty relation

---

* 'Nothing is so difficult that it could not be investigated.' Publius Terentius Afer (ca. 190–159 BCE), roman poet, in his *Heautonimorumenos*, 675.

and is given by that mass whose Compton wavelength matches the size of the box. As shown above, the same mass value is found by every other Gedankenexperiment: trying to determine the gravitational mass by weighing the 'piece' of vacuum or by measuring its kinetic energy gives the same result, whatever the method chosen. Another, but in the end equivalent way to show that a region of vacuum has a finite mass is to study how the vacuum energy depends on the position uncertainty of the border of the region. Any region is defined through its border. The position uncertainty of the border will induce a mass error for the contents of the box, in the same way that a time limit does. Again, the resulting mass value for a region of vacuum is the one for which the box size is the Compton wavelength.

See page 742

Ref. 852

Summarizing, in a box of size R, nature allows only mass values m between two limits:

$$\text{(full box)}\ \frac{c^2R}{G} \geqslant m \geqslant \frac{\hbar}{cR}\ \text{(empty box)} \quad . \tag{595}$$

We see directly that for sizes R of the order of the Planck length, the two limits coincide; they both give the Planck mass

$$M_{\mathrm{Pl}} = \frac{\hbar}{c\,l_{\mathrm{Pl}}} = \sqrt{\frac{\hbar c}{G}} \approx 10^{-8}\,\mathrm{kg} \approx 10^{19}\,\mathrm{GeV/c^2} \quad . \tag{596}$$

In other words, at Planck scales we cannot distinguish a full box from an empty one. That means that there is no difference between vacuum and matter at Planck scales. Of course, a similar statement holds for the distinction between vacuum and radiation.

How else can we show that matter and vacuum cannot be distinguished?

A strong statements needs additional proof. Mass can also be measured by probing its inertial aspect, i.e. by colliding the unknown mass M with known velocity V with a known probe particle of mass m and momentum p. We then have

$$M = \frac{\Delta p}{\Delta V} \quad , \tag{597}$$

where the differences are taken between the values before and after the collision. The error δM of such a measurement is simply given by

$$\frac{\delta M}{M} = \frac{\delta \Delta v}{\Delta v} + \frac{\delta m}{m} + \frac{\delta \Delta V}{\Delta V} \quad . \tag{598}$$

At Planck scales we have $\delta\Delta v/\Delta v \approx 1$, because the velocity error is always, like the velocities themselves, of the order of the speed of light. In other words, at Planck scales the mass measurement error is so large that we cannot determine whether a mass is different from zero: vacuum is indistinguishable from matter.

The same happens if we take light with a wavelength λ as the probe particle. In this case, expression (597) leads to a mass error

$$\frac{\delta M}{M} = \frac{\delta \Delta\lambda}{\Delta\lambda} + \frac{\delta \Delta V}{\Delta V} \quad . \tag{599}$$

In order that photon scattering can probe Planck dimensions, we need a wavelength of the order of the Planck value; but in this case the first term is approximately 1. Again we find that at Planck scales the energy uncertainty is always of the *same* size as the energy value to be measured. Measurements cannot distinguish between vanishing and non-vanishing mass M at Planck scales. This result appears for all methods of mass measurement that can be imagined.At Planck scales, matter cannot be distinguished from vacuum. Challenge 1185

We note that the same arguments are valid if instead of the mass of matter we use the energy of radiation. In other words, it is also impossible to distinguish radiation from vacuum at high energies. In short, no type of particle differs from vacuum at high energies.

The indistinguishability of particles and vacuum, together with the existence of minimum space-time intervals, suggest that space, time, radiation and matter are macroscopic approximations of an underlying, common and discrete structure. How do the common constituents of the two look like? We will argue that these constituents are *extended* and *fluctuating*. Ref. 857

> Unsere Aufgabe ist nicht zu sehen was noch niemand gesehen hat,
> sondern über dasjenige was jeder schon gesehen hat
> zu denken was noch nie jemand gedacht hat.*
> Erwin Schrödinger

Argument 1: The size and shape of elementary particles

Size is the length of vacuum taken by an object. This definition comes natural in everyday life, quantum theory and relativity. However, approaching Planck energy, vacuum and matter cannot be distinguished: it is impossible to define the boundary between the two, and thus it is impossible to define the size of an object: every object becomes as extended as the vacuum!

So we need to be careful. What happens if Planck energy is *approached*, advancing step by step to higher energy? Every measurement requires comparison with a standard. A standard is made of matter, and comparison is performed using radiation. Thus any measurement requires to distinguish between matter and radiation. However, the distinction between matter and radiation is possible only up to the (grand) unification energy, which is measured to be about 800 times lower than the Planck energy. Let us try to measure the size of a particle at those energies.

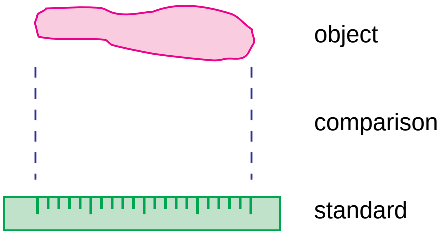


Figure 256 Measurement requires matter and radiation

---

* Our task is not to see what nobody has seen yet, but to think what nobody has thought yet about that which everybody has seen already.

Do boxes exist?

The first and simplest way to determine the size of a particle is to measure the size of a *box* it fits in. To be sure that the particle fits inside, we first of all must be sure that the box is tight. This is done by checking whether something, such as matter or radiation, can leave the box. However, there is no way to be sure that a box has no holes. Potential hills cannot get higher than the maximum energy, namely the Planck energy. In short, at high energies there is no way to make tight boxes.

In addition, at the latest at the unification energy there is no way to distinguish between the box and the object enclosed in it, as all matter has the same properties. At highest energies, there is thus no way to say that a particle is compact using a box.

Let us cross-check this result. In everyday life, we call particles ‘small’ because they can be enclosed. Enclosure is possible because in daily life walls are impenetrable. However, walls are impenetrable for matter particles only up to roughly 10 MeV, and only up to 10 keV for photons. In fact, boxes do not even exist at medium energies. We thus cannot extend the idea of ‘box’ to high energies at all.

In summary, we cannot state that particles are compact or of finite size using boxes. Thus we try to achieve this in other ways.

Can the Greeks help?

The Greeks deduced the existence of atoms by noting that division of matter must end. In contrast, whenever we think of space (or space-time) as made of points, we assume that it can be subdivided without end. Zeno noted this already long time ago and strongly criticized this assumption. He was right. At Planck energy, infinite subdivision is impossible. Any attempt to divide space stops at Planck dimensions at the latest. We have thus no possibility to deduce the existence of points using any method of dividing space.

Are cross-sections finite?

To determine size, we can take the opposite approach and determine the departure from point-likeness for a particle. At high energy, detecting this departure requires scattering. For example, we can suspend the particle in some trap and then shoot some probe at it. What happens in a scattering experiment at high energies? The question has been studied already by Leonard Susskind. When shooting at the particle with a high energy probe, the scattering process is characterized by an interaction time. But extremely short interaction times imply sensitivity to the size and shape fluctuations due to the quantum of action. A short interaction time provides a cut-off for high energy shape and size fluctuations and thus determines the measured size. As a result, the size measured for any microscopic, but extended object increases when the probe energy is increased towards the Planck value. Ref. 854

In summary, even though at experimentally achievable energies the size is always smaller than measurable, when approaching the Planck energy, size would increase above all bounds. As a result, at high energies we cannot give a limit to sizes of particles! In other words, since particles are not point-like at everyday energy, at Planck energies they are enormous: particles are extended.

Duration	blur	observation possibilities
1 h	high	Ability to see faint quasars at night
1 s	high	Everyday motion is completely blurred
20 ms	lower	Interruption by eyelids
10 ms	lower	Effective eye/brain shutter time – impossibility to see tennis ball when hitting it
0.25 ms	lower	Shortest commercial photographic camera shutter time, for photographing fast cars
1 μs	very low	Ability to photograph flying bullets; requires strong flashlights
ca. 10 ps	lowest	Study of molecular processes; ability to photograph flying light pulses; requires laser light to get sufficient illumination
10 fs	higher	Light photography becomes impossible due to wave effects
100 zs	high	X-ray photography and becomes impossible; only γ-ray imaging is left over
shorter times	very high	Photographs get darker as illumination gets dimmer; gravitational effects start playing a role
10^{-41} s	highest	imaging makes no sense

Table 59 Effects of various camera shutter times

That is quite a deduction. Right at the start of our mountain ascent, we distinguished objects from their environment. Objects are by definition localized, bounded and compact. All objects have a boundary, which is a surface which itself does not have a boundary. Objects are also bounded in abstract ways; boundedness is also a property of the symmetries of any object, such as its gauge group. In contrast, the environment is not localized, but extended and unbounded. All these basic assumptions disappear at Planck scales. At Planck length, matter cannot be distinguished from vacuum. In fact our discussion above showed that it is impossible to determine whether something is bounded or compact at Planck energy. Compactness is only an approximate property; it is not correct at high energies. As a result, the idea of *a point particle is a low energy, approximated concept.*

A final check can be performed studying the following question:

Can we take a photograph of a point?

Humans or any other types of observers can only observe a part of the world with finite resolution in time and in space. In this they resemble a film camera. The magnitude of the resolution has (almost) been discovered in 1899: the Planck time and the Planck length. No measurement apparatus can measure space or time intervals smaller than the Planck values. But what would happen if we would take a photograph with a shutter time approaching the Planck time? Ref. 855 Ref. 852, 856

Imagine having perfect a shutter and taking photographs at increasingly shorter times. Table gives a rough overview of the possibilities. For shorter and shorter shutter times, photographs get darker and darker. Once the shutter time reaches the oscillation time of light, strange things happen: light has no chance to pass undisturbed; signal and noise become impossible to distinguish; in addition, the moving shutter will produce colour shifts.

In contrast to our intuition, the picture would get *blurred* at extremely short shutter times. Photography is not only impossible at long but also at short shutter times.

The difficulty of taking photographs is independent of the used wavelength. The limits move, but remain. A short shutter time τ does not allow photons of energy lower than $\hbar/\tau$ to pass undisturbed. The blur is small when shutter times are those of everyday life, but *increases* when shutter times are shortened towards Planck times. As a result, there is no way to detect or confirm the existence of points by taking pictures. Points in space, as well as instants of time, are *imagined* concepts; they do not allow a precise description of nature.

At Planck shutter times, only signals with Planck energy can pass through the shutter. Since at these energies matter cannot be distinguished from radiation or from empty space, all objects, light and vacuum look the same. As a result, it becomes impossible to say how nature looks at shortest times.

But the situation is much worse: a Planck shutter does not exist at all, as it would need to be as small as a Planck length. A camera using it could not be built, as lenses do not work at this energy. Not even a *camera obscura* – without any lens – would work, as the diffraction effects would make image production impossible. In other words, the idea that at short shutter times, a photograph of nature shows a frozen version of everyday life, like a stopped movie, is completely wrong! Zeno criticized this image already in ancient Greece, in his discussions about motion, though not so clearly as we can do now. Indeed, at a single instant of time nature is not frozen at all.* At short times, nature is blurred and fuzzy. This is also the case for point particles. Whatever the intrinsic shape of what we call a ‘point’ might be, we know that being always blurred, it is first of all a cloud.

Any photograph of a point particle would thus show an extended entity. Let us study this issue in more detail.

What is the shape of an electron?

Since particles are not point-like, they have a shape. How can we determine it? Everyday object shape determination is performed by *touching* the object from all sides. This works with plants, people or machines. It works with molecules, such as water molecules. We can put them at rest, e.g. in ice, and then scatter small particles off them. Scattering is just a higher energy version of touching. However, scattering cannot determine shapes smaller than the wavelength of the used probes. To determine a size as small as that of an electron, we need highest energies. But we already know what happens when approaching Planck scales: the shape of a particle becomes the shape of all the space surrounding it. Shape cannot be determined in this way.

Another method to determine the shape is to build a tight box filled of wax around the system under investigation. We let the wax cool and then open the box and observe the hollow part. However, near Planck energies boxes do not exist. We are unable to determine the shape in this way.

A third way to measure shapes is cutting something into pieces and then study the pieces. But cutting is just a low-energy version of a scattering experiment. It does not work at

---

* In fact, a shutter does not exist even at medium energy, as shutters, like walls, stop existing at around 10 MeV.

high energies. Since the term 'atom' means 'uncuttable' or 'indivisible', we have just found out that neither atoms nor indivisible particles can exist, as there is no way to prove this property. Our everyday intuition leads us completely astray at Planck energies.

A fourth way to measure shapes could appear by distinguishing transverse and longitudinal shape, with respect to the direction of motion. However, for transverse shape we get the same issues as for scattering; transverse shape diverges for high energy. To determine longitudinal shape, we need at least two infinitely high potential walls. Again, we already know that this is impossible.

A further, indirect way of measuring shapes is the measurement of the moment of inertia. A finite moment of inertia means a compact, finite shape. However, when the measurement
Ref. 852 energy is increased, rotation, linear motion and exchange become mixed up. We do not get meaningful results.

Still another way to determine shapes is to measure the *entropy* of a collection of particles we want to study. This allows to determine the dimensionality and the number of internal degrees of freedom. At high energies, a collection of electrons would become a black hole. We study the issue separately below, but again we find no new information.

Are these arguments water-tight? We assumed three dimensions at all scales, and assumed that the shape of the particle itself is fixed. Maybe these assumptions are not valid at Planck scales. Let us check the alternatives. We have already shown above that due to the fundamental measurement limits, the dimensionality of space-time cannot be determined at
Ref. 852 Planck scales. Even if we could build perfect three-dimensional boxes, holes could remain in other dimensions. It does not take long to see that all the arguments against compactness work even if space-time has additional dimensions.

Is the shape of an electron fixed?

Only an object composed of localized entities, such as a house or a molecule, can have a fixed shape. The smaller a system gets, the more quantum fluctuations play a role. Any entity with a finite size, thus also an elementary particle, cannot have a fixed shape. Every Gedankenexperiment leading to finite shape also implies that the shape itself fluctuates. But we can say more.

The distinction between particles and environment resides in the idea that particles have *intrinsic* properties. In fact, all intrinsic properties, such as spin, mass, charge, and parity, are localized. But we saw that no intrinsic property is measurable or definable at Planck scales. Thus it is impossible to distinguish particles from the environment, as we know already. In addition, at Planck energy particles have all properties the environment also has. In particular, particles are extended.

In short, one cannot prove by experiments that at Planck energy elementary particles are finite in size in all directions. In fact, *all experiments one can think of are compatible with extended particles*, with infinite size. We can also say that particles have *tails*.

Not only are particles extended; in addition, their shape cannot be determined by the methods just explored. The only possibility left over is also suggested by quantum theory: *The shape of particles is fluctuating.*

We note that for radiation particles we reach the same conclusions. The box argument shows that also radiation particles are extended and fluctuating.

In our enthusiasm we have also settled an important detail about elementary particles. We saw above that any particle which is smaller than its own Compton wavelength must be 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.* See page 552

However, an elementary particle can have constituents if they are not compact. The difficulties of compact constituents were already described by Sakharov in the nineteen sixties. Ref. 861 But if the constituents are extended, they do not fall under the argument, as extended entities have no localized mass. In particular, a flying arrow, Zeno's famous example, cannot be said to be at a given position at a given time. Shortening the observation time towards the Planck time makes an arrow disappear in the same cloud that also makes up space-time.**

In summary, only the idea of points leads to problems at Planck scales. If space-time and matter are imagined to be made, at Planck scales, of *extended and fluctuating* entities, all problems disappear. We note directly that for extended entities the requirement of a non-local description is realized. Similarly, the entities being fluctuating, the requirement of a statistical description of vacuum is realized. Ref. 852 Finally, the argument forbidding composition of elementary particles is circumvented, as extended entities have no clearly defined mass. Thus the concept of Compton wavelength cannot be defined or applied. Elementary particle can thus have constituents if they are extended. But if the components are extended, how can compact 'point' particles be formed with them?A few options will be studied shortly. Challenge 1186 e

Argument 2: The shape of vacuum

> Thus, since there is an impossibility that [finite] quantities are built
> from contacts and points, it is necessary that there be
> indivisible material elements and [finite] quantities.
> Aristotle, Of generation and corruption. Ref. 863

We are used to think that empty space is made of spatial points. Let us check whether this is true at high energy. At Planck scales no measurement can give zero length, zero mass, zero area, zero volume. There is no way to state that something in nature is a point without contradicting experimental results. In addition, the idea of a point is an extrapolation of what is found in small empty boxes getting smaller and smaller. However, we just saw that at high energies small boxes cannot be said to be empty. In fact, boxes do not exist at all, as they are never tight and do not have impenetrable walls at high energies.

Also the idea of a point as a continuous subdivision of empty space is untenable. At small distances, space cannot be subdivided, as division requires some sort of dividing wall, which does not exist.

---

[*] Examples are the neutron, positronium or the atoms. Note that the argument does not change when the elementary particle itself is unstable, such as the muon. Note also that the possibility that all components be heavier than the composite, which would avoid this argument, does not seem to lead to satisfying physical properties; e.g. it leads to intrinsically unstable composites.

Ref. 862 [**] Thus at Planck scales there is no quantum Zeno effect any more.

Even the idea of repeatedly putting a point between two others cannot be applied. At high energy, it is impossible to say whether a point is exactly on the line connecting the outer two points; and near Planck energy, there is no way to find a point between them at all. In fact, the term 'in between' makes no sense at Planck scales.

We thus find that space points do not exist. But there is more; space cannot be made of points for additional reasons. Common sense tells us that points need to be kept *apart* somehow, in order to form space. Indeed, mathematicians have found a strong argument that physical space cannot be made of mathematical points: the behaviour described by the Ref. 852 Banach-Tarski paradox for mathematical spaces is quite different from what is found for physical vacuum. The Banach-tarski paradox states states that a sphere made of mathematical points ca be cut into 5 pieces which can be reassembled into two spheres each of the same volume as the original sphere. Mathematically, volume makes no sense. Physically speaking, we can say that the concept of volume does not exist for continuous space; it is only definable if an intrinsic length exists. This is the case for matter; it must also be the case for vacuum. As a result, any concept with an intrinsic length, like the vacuum, must be described by one or several extended components.[*] In summary, we need *extended* entities to make up space-time!

Not only is it impossible to generate a volume with mathematical points; it is also impossible to generate exactly three physical dimensions with mathematical points. Mathematics shows that any compact one-dimensional set has as many points as any compact three-dimensional set. And the same is true for any other pair of dimension values. To build up the physical three-dimensional vacuum we need entities which organize their neighbourhood. This cannot be done with points. Entities must possess some sort of bond forming ability. Bonds are needed to construct or fill three dimensions and not any other number. Bonds are extended entities. Also a collection of tangled entities extending to the maximum scale of the region under consideration, would work perfectly. Of course the precise shape is not known at this point in time. In any case we again find that any constituents of physical three-dimensional space must be extended. We need extension to define dimensionality and to define volume.

We are not surprised. Above we deduced that the constituents of particles are extended. Since vacuum is not distinguishable from matter, we expect that the constituents of vacuum to be extended as well. Stated simply, if elementary particles are not point-like, then space-time points cannot be either.

Measuring the void

To check whether space-time constituents are extended, we perform some additional Gedankenexperiments. Let us measure the size of a point of space-time. The clearest definition of size is through the cross section. How can we determine the cross section of a point? We can determine the cross section of a piece of vacuum and determine the number

---

Ref. 864
Challenge 1187 n

[*] A a collection of entities with Planck size in all directions, such as spheres, would avoid the Banach-Tarski paradox, but would not allow to deduce the numbers of dimensions of space and time. It would also contradict the other results of this section. We therefore do not explore it further.

of points inside it. From the two results we can deduce the cross section of a single point. At Planck energies however, we get a simple result: the cross section of a volume of empty space is depth independent. At Planck energies, vacuum has a surface, but no depth. In other words, at Planck energy we can only state that a Planck layer covers the surface of a volume. We cannot say anything about its interior. One way to picture the result is to say that space points are long tubes.

Another way to determine the size of a point is to count the points found in a given volume of space-time. The first approach is to count the possible positions of a point particle in a volume. However, point particles are extended at Planck energies, and indistinguishable from vacuum. At Planck energy, this number is given by surface area of the volume divided by the Planck area.

The second approach to count the number of points in a volume is to fill a piece of vacuum with point particles.

What is the maximum number of particles that fits inside a piece of vacuum?

The maximum mass that fits into a piece of vacuum is a black hole. This maximum mass depends only on the *surface* of the given vacuum. The maximum mass increases less rapidly than the volume. In other words, the number of points in a volume is only proportional to the surface area of that volume. There is only one solution: vacuum must be made of extended entities crossing the whole volume, independently of the shape of the volume.

Two thousand years ago, the Greek argued that matter must be made of particles because salt can be dissolved in water and because fish can swim through water. Knowing more about Planck scales, we have to reconsider the argument. Like fish through water, particles can move through vacuum; but since vacuum has no bounds and since it cannot be distinguished from matter, vacuum cannot be made of particles. However, there is another possibility that allows for motion of particles through vacuum: both vacuum and particles can be made of a web of extended entities. Let us study this option in more detail. Ref. 865

Argument 3: The large and the small

> I could be bounded in a nutshell and count myself a king of infinite space.
> William Shakespeare, *Hamlet*.

If two observables cannot be distinguished, there is a symmetry transformation connecting them. For example, when switching observation frame, an electric field may change into a magnetic one. A symmetry transformation means that we can change the viewpoint (i.e. the frame of observation) with the consequence that the same observation is described by one quantity from one viewpoint and by the other quantity from the other viewpoint.

When measuring length at Planck scales it is impossible to say whether we are measuring the length of a piece of vacuum, the Compton wavelength of a body, or the Schwarzschild diameter of a body. For example, the maximum size for an elementary object is its Compton wavelength. The minimum size for an elementary object is its Schwarzschild radius. The actual size of an elementary object is somewhere in between. If we want to measure the size precisely, we have to go to Planck energy: but then all these quantities are the same. In other words, at Planck scales, there is a *symmetry* transformation between Compton wavelength and Schwarzschild radius. In short, *at Planck scales there is a symmetry between mass and inverse mass.*

As a further consequence, at Planck scales there is a symmetry between size and inverse size. *Matter-vacuum indistinguishability means that there is a symmetry between length and inverse length* at Planck energies. This symmetry is called *space-time duality* or *T-duality* in the literature of superstrings.* Space-time duality is a symmetry between situations at scale $n\,l_{\mathrm{Pl}}$ and at scale $f\,l_{\mathrm{Pl}}/n$, or, in other words, between R and $(f\,l_{\mathrm{Pl}})^2/R$, where the experimental number f has a value somewhere between 1 and 1000. Ref. 859

Duality is a genuine non-perturbative effect. It does not exist at low energy, since duality automatically also relates energy E and energy $E_{\mathrm{Pl}}^2/E = \hbar c^3/GE$, i.e. it relates energies below and above Planck scale. Duality is a quantum symmetry. It does not exist in everyday life, as Planck's constant appears in its definition. In addition, it is a general relativistic effect, as it includes the gravitational constant and the speed of light. Let us study duality in more detail.

Small is large?

> [Zeno of Elea maintained:] If the existing are many,
> it is necessary that they are at the same time small and large,
> so small to have no size, and so large to be without limits.
> Simplicius, Commentary on the Physics of Aristotle, 140, 34. Ref. 897

To explore the consequences of duality, we can compare it to the 2π rotation symmetry in everyday life. Every object in daily life is symmetrical under a full rotation. For the rotation of an observer, angles make sense only as long as they are smaller than 2π. If a rotating observer would insist on distinguishing angles of $0, 2\pi, 4\pi$ etc., he would get a new copy of the universe at each full turn.

Similarly, in nature, scales R and l_{Pl}^2/R cannot be distinguished. Lengths make no sense when they are smaller than l_{Pl}. If however, we insist on using even smaller values and insist on distinguishing them from large ones, we get a new copy of the universe at those small scales. Such an insistence is part of the standard continuum description of motion, where it is assumed that space and time are described by the real numbers, which are defined for arbitrary small intervals. Whenever the (approximate) continuum description with infinite extension is used, the $R \leftrightarrow l_{\mathrm{Pl}}^2/R$ symmetry pops up.

Duality implies that diffeomorphism invariance is only valid at medium scales, not at extremal ones. At extremal scales, quantum theory has to be taken into account in the proper manner. We do not know yet how to do this.

Space-time duality means that introducing lengths smaller than the Planck length (like when one defines space points, which have size zero) means at the same time introducing things with very large ('infinite') value. Space-time duality means that for every small enough sphere the inside equals outside.

---

* There is also an *S-duality*, which connects large and small coupling constants, and a *U-duality*, which is the combination of S- and T-duality.

Duality means that if a system has a small dimension, it also has a large one. And vice versa. There are thus no small objects in nature. As a result, *space-time duality is consistent with the idea that the basic entities are extended.*

Unification and total symmetry

So far, we have shown that at Planck energy, time and length cannot be distinguished. Duality has shown that mass and inverse mass cannot be distinguished. As a consequence, length, time, and mass cannot be distinguished from each other. Since every observable is a combination of length, mass and time, *space-time duality means that there is a symmetry between all observables*. We call it the *total symmetry*.*

Total symmetry implies that there are many types of specific dualities, one for each pair of quantities under investigation. Indeed, in string theory, the number of duality types discovered is increasing every year. It includes, among others, the famous electric-magnetic duality we first encountered in the chapter on electrodynamics, coupling constant duality, surface-volume duality, space-time duality and many more. All this confirms that there is an enormous symmetry at Planck scales. Similar statements are also well-known right from the beginning of string theory. See page 390 Ref. 860 Ref. 859

Most importantly, total symmetry implies that gravity can be seen as equivalent to all other forces. Space-time duality shows that unification is possible. Physicist have always dreamt about unification. Duality tells us that this dream can indeed be realized.

It may seem that total symmetry is in complete contrast with what was said in the previous section, where we argued that all symmetries are lost at Planck scales. Which result is correct? Obviously, both of them are.

At Planck scales, all low energy symmetries are indeed lost. In fact, all symmetries that imply a *fixed* energy are lost. Duality and its generalizations however, combine both small and large dimensions, or large and small energies. Most symmetries of usual physics, such as gauge, permutation and space-time symmetries, are valid at each fixed energy separately. But nature is not made this way. The precise description of nature requires to take into consideration large and small energies at the same time. In everyday life, we do not do that. Everyday life is a low and fixed energy approximation of nature. For most of the 20th century, physicists aimed to reach higher and higher energies. We believed that precision increases with increasing energy. But when we combine quantum theory and gravity we are forced to change this approach; to achieve high precision, we must take both high and low energy into account at the same time.**

The large differences in phenomena at low and high energies are the main reason why unification is so difficult. So far, we were used to divide nature along the energy scale. We

---

* A symmetry between size and Schwarzschild radius, i.e. a symmetry between length and mass, will lead to general relativity. Additionally, at Planck energy there is a symmetry between size and Compton wavelength. In other words, there is a symmetry between length and 1/mass. It means that there is a symmetry between coordinates and wavefunctions. Note that this is a symmetry between states and observables. It leads to quantum theory.

** Renormalization energy does connect different energies, but not in the correct way; in particular, it does not include duality.

thought about high energy physics, atomic physics, chemistry, biology, etc. The differences between these sciences is the energy of the processes involved. We are not allowed to think in this way any more. We have to take all energies into account at the same time. That is not easy, but we do not have to despair. Important conceptual progress has been achieved in the last decade of the twentieth century. In particular, we now know that we need only *one single concept* for all things which can be measured. Since there is only one concept, there are many ways to study it. We can start from any (low-energy) concept in physics and explore how it looks and behaves when we approach Planck scales. In the present section, we are looking at the concept of point. Obviously, the conclusions must be the same, independently of the concept we start with, be it electric field, spin, or any other. Such studies thus provide a check for the results in this section. Challenge1188 h

Unification thus implies to think using duality and using concepts which follow from it. In particular, we need to understand what exactly happens to duality when we restrict ourselves to low energy only, as we do in everyday life.This question is left for the next section. Challenge 1189 e

Argument 4: Does nature have parts?

> Pluralitas non est ponenda sine necessitate.*
> William of Occam

Another argument, independent of the ones given above, underlines the correctness of a model of nature made of extended entities. Let us take a little broader view. Any concept for which we can distinguish parts is described by a set. We usually describe nature as a set of objects, positions, instants, etc. The most famous set description of nature is the oldest known, by Democritus: ‘The world is made of indivisible particles and void.’ This description was extremely successful in the past; there were no discrepancies with observations yet. However, after 2500 years, the conceptual difficulties of this approach are obvious.

We now know that Democritus was wrong, first of all, because vacuum and matter cannot be distinguished at Planck scales. Thus the word ‘and’ in his sentence is already mistaken. Secondly, due to the existence of minimal scales, the void cannot be made of ‘points,’ as we usually assume nowadays. Thirdly, the description fails because particles are not compact objects. Fourth, the total symmetry implies that we cannot distinguish parts in nature; nothing can really be distinguished from anything else with complete precision, and thus the particles or points in space making up the naive model of void cannot exist.

In summary, quantum theory and general relativity together show that in nature, *all differences are only approximate*. Nothing can really be distinguished from anything else with complete precision. In other words, there is no way to define a ‘part’ of nature, neither for matter, nor for space, nor for time, nor for radiation. *Nature cannot be a set.*

---

* ‘Multitude should not be introduced without necessity.’ This famous principle is commonly called *Occam’s razor*. William of Ockham (1285/1295, Ockham–1349/50, München), or Occam in the common latin spelling, was one of the great thinkers of his time. In his famous statement he expresses that only those concepts which are strictly necessary should be introduced to explain observations. It can be seen as the requirement to abandon beliefs when talking about nature. In addition, at this stage of our mountain ascent it has an even more direct interpretation.

The conclusion that nature is not a set does not come as a surprise. We have already encountered another reason to doubt that nature is a set. Whatever definition we use for the term 'particle', Democritus cannot be correct for a purely logical reason. The description he provided is *not complete*. Every description of nature defining nature as a set of parts necessarily misses certain aspects. Most importantly, it misses the *number* of these parts. In particular, the number of particles and the number of dimensions of space-time must be specified if we describe nature as made from particles and vacuum. Above we saw that it is rather dangerous to make fun of the famous statement by Arthur Eddington See page 493

> 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. 867

In fact, practically all physicists share this belief; usually they either pretend to favour some other number, or worse, they keep the number unspecified. We have seen during our walk that in modern physics many specialized sets are used to describe nature. We have used vector spaces, linear spaces, topological spaces and Hilbert spaces. But very consistently we refrained, like all physicists, from asking about the origin of their sizes (mathematically speaking of their dimensionality or their cardinality). In fact, it is equally unsatisfying to say that the universe contains some specific number of atoms as it is to say that space-time is made of point-like events arranged in $3+1$ dimensions. Both statements are about set sizes in the widest sense. In a complete, i.e. in a unified description of nature the number of smallest particles and the number of space-time points must not be added to the description, but must *result* from the description. Only in this case is unification achieved.

Requiring a complete explanation of nature leads to a simple consequence. Any part of nature is by definition smaller than the whole of nature and different from other parts. As a result, any description of nature by a set *cannot* possibly yield the number of particles nor space-time dimensionality. As long as we insist in using space-time or Hilbert spaces for the description of nature, we *cannot* understand the number of dimensions or the number of particles.

Well, that is not too bad, as we know already that nature is *not* made of parts. We know that parts are only approximate concepts. In short, if nature were made of parts, it could not be a unity, or a 'one.' If however, nature is a unity, a one, it cannot have parts.[*] Nature cannot be separable exactly. It cannot be made of particles.

To sum up, nature cannot be a set. Sets are lists of distinguishable elements. When general relativity and quantum theory are unified, nature shows no elements: *nature stops being a set at Planck scales*. The result confirms and clarifies a discussion we have started in classical physics. There we had discovered that matter objects were defined using space and time, and that space and time were defined using objects. Including the results of quantum theory,

---

[*] As a curiosity, practically the same discussion can already be found, in Plato's *Parmenides*, written in the 4th century BCE. There, Plato musically ponders different arguments on whether nature is or can be a *unity* or a *multiplicity*, i.e. a set. It seems that the text is based on the real visit by Parmenides and Zeno in Athens, where they had arrived from their home city Elea, which lies near Naples. Plato does not reach a conclusion. modern physics however, does. Ref. 866

this implies that in modern physics particles are defined with help of the vacuum, and the vacuum with particles. That is not a good idea. We have just seen that since the two concepts are not distinguishable from each other, we cannot define them with each other. Everything is the same; in fact, there is no 'every' and no 'thing.' Since nature is not a set, the circular reasoning is dissolved.

Space-time duality also implies that space is not a set. Duality implies that events cannot be distinguished from each other. They thus do not form elements of some space. Phil Gibbs Ref. 868 has given the name *event symmetry* to this property of nature. This thought-provoking term, even though still containing the term 'event', underlines that it is impossible to use a set to describe space-time.

In summary, nature cannot be made of vacuum and particles. That is bizarre. People propagating this idea have been persecuted for 2000 years. This happened to the atomists from Democritus to Galileo. Was it all in vain? Let us continue to clarify our thoughts.

Does the universe contain anything?

Stating that the universe contains something implies that we are able to distinguish the contents from the universe. However, we now know that precise distinctions are impossible. If nature is not made of parts, it is wrong to say that the universe *contains* something.

Let us go further. As nothing can be distinguished, we need a description of nature which allows to state that at Planck energies nothing can be distinguished from anything else. For example, it must be impossible to distinguish particles from each other or from the vacuum. There is only one solution: everything, or at least what we call 'everything' in everyday life, must be made of the same single entity. All particles are made of one same 'piece.' Every point in space, every event, every particle and every instant of time must be made of the same single entity.

An amoeba

> A theory of nothing describing everything is better
> than a theory of everything describing nothing.

We found that parts are approximate concepts. The parts of nature are not strictly smaller than nature itself. As a result, any 'part' must be extended. Let us try to extract more information about the constituents of nature.

The search for a unified theory is the search for a description in which all concepts appearing are only *approximately* parts of the whole. Thus we need an entity Ω, describing nature, which is not a set but which can be approximated by one. This is unusual. We all are convinced very early in our life that we are a *part* of nature. Our senses provide us with this information. We are not used to think otherwise. But now we have to.

Let us eliminate straight away a few options for Ω. One concept without parts is the empty set. Perhaps we need to construct a description of nature from the empty set? We could be See page 467 inspired by the usual construction of the natural numbers from the empty set. However, the empty set makes only sense as the opposite of some full set. That is not the case here. The empty set is not a candidate for Ω.

Another possibility to define approximate parts is to construct them from multiple copies of Ω. But in this way we would introduce a new set through the back door. In addition, new concepts defined in this way would not be approximate.

We need to be more imaginative. How can we describe a whole which has no parts, but which has parts approximately? Let us recapitulate. The world must be described by a single entity, sharing all properties of the world, but which can be approximated into a set of parts. For example, the approximation should yield a set of space-points and a set of particles. We also saw that whenever we look at any 'part' of nature without any approximation, we should not be able to distinguish it from the whole world. In other words, composed entities are not always larger than constituents.

Composed entities must usually appear to be smaller than their constituents. For example, space 'points' or 'point' particles are tiny, even though they are only approximations. Which concept without boundaries can be at their origin? Using usual concepts the world is everywhere at the same time; if nature is to be described by a single constituent, this entity must be extended.

The entity has to be a single one, but it must *seem* to be multiple, i.e. it has to be multiple approximately, as nature shows multiple aspects. The entity must be something folded. It must be possible to count the folds, but only approximately. (An analogy is the question of how many tracks are found on an LP or a CD; depending on the point of view, local or global, one gets different answers.) Counting folds would correspond to a length measurement.

The simplest model would be the use of a single entity which is extended, fluctuating, going to infinity, and allowing approximate localization, thus allowing approximate definition of parts and points.* In more vivid imagery, nature could be described by some deformable, folded and tangled up entity: a giant, knotted amoeba. An amoeba slides between the fingers whenever one tries to grab a part of it. A perfect amoeba flows around any knife trying to cut it. The only way to hold it would be to grab it in its entirety. However, for an actor himself made of amoeba strands this is impossible. He can only grab it approximately, by catching part of it and approximately blocking it, e.g. using a small hole so that the escape takes a long time.

To sum up, nature is modelled by an entity which is *a single unity* (to eliminate distinguishability), *extended* (to eliminate localizability) and *fluctuating* (to ensure approximate continuity). A far-reaching, fluctuating fold, like an amoeba. The tangled branches of the amoeba allow a definition of length via counting of the folds. In this way, *discreteness* of space, time, and particles could also be realized; the quantization of space-time, matter and radiation thus follows. Any flexible and deformable entity is also a perfect candidate for the realization of diffeomorphism invariance, as required by general relativity.

A simple candidate for the extended fold is the image of a fluctuating, flexible *tube*. Counting tubes implies to determine distances or areas. The minimum possible count of one gives the minimum distance, and thus allows us to deduce quantum theory. In fact we can use as model any object which has flexibility and a small dimension, such as a tube, a

---

Challenge 1190

* Is this the only method to describe nature?Is it possible to find another description, in particular if space and time are not used as background? The answers are unclear at present.

thin sheet, a ball chain or a woven collection of rings. These options give the different but probably equivalent models presently explored in simplicial quantum gravity, in Ashtekar's variables and in superstrings.

Argument 5: The entropy of black holes

We are still collecting arguments to determine particle shape. A completely different way to explore the shape of particles, it is useful to study situations where they appear in large numbers. Collections of high numbers of constituents behave differently if they are point-like or extended. In particular, their entropy is different. Studying entropy thus allows to determine shape. The best approach is to study extreme situations, when large numbers of particles are in small volumes. This leads to study the entropy of black holes. A black hole is a body whose gravity is so strong that even light cannot escape. Black holes tell us a lot about the fundamental entities of nature. It is easily deduced in general relativity that any body whose mass m fits inside the so-called Schwarzschild radius

$$r_S = 2Gm/c^2 \qquad (600)$$

is a black hole. A black hole can be formed when a whole star collapses under its own weight. A black hole is thus a macroscopic body with a large number of constituents. As for every macroscopic body, an entropy can be defined. The entropy S of a macroscopic black hole was determined by Bekenstein and Hawking and is given by Ref. 870, 871

$$S = \frac{k}{l_{Pl}^2}\frac{A}{4} \quad \text{or} \quad S = k\frac{4\pi G m^2}{\hbar c} \qquad (601)$$

where k is the Boltzmann constant and $A = 4\pi r_S^2$ is the surface of the black hole horizon. This deep result has been derived in many different ways. The various derivations also confirm that space-time and matter are equivalent, as the entropy value can be seen both as an entropy of matter as one of space-time. In the present context, the two main points of interest are that the entropy is *finite*, and that it is *proportional to the area* of the black hole horizon. Ref. 872

In view of the existence of minimum lengths and times, the finiteness of entropy is not surprising any more. A finite black hole entropy confirms the idea that matter is made of a finite number of discrete entities per volume. The existence of an entropy also shows that these entities behave statistically; they fluctuate. In fact, quantum gravity leads to a finite entropy for any object, not only for black holes; Bekenstein has shown that the entropy of any object is always smaller than the entropy of a (certain type of) black hole of the same mass. Ref. 873

The entropy of a black hole is also proportional to its horizon area. Why? This question is the topic of a stream of publications up to this day.* A simple way to understand the entropy-surface proportionality is to look for other systems in nature with the property that Ref. 875

Ref. 874 * The result can be derived from quantum statistics alone. But this derivation does not yield the proportionality coefficient.

entropy is proportional to system surface instead of system volume. In general, the entropy of any collection of one-dimensional flexible objects, such as polymer chains, shows this property. The expression for the entropy of a polymer chain made of N monomers, each of length a, whose ends are kept a distance r apart, is given by Ref. 877 Ref. 876

$$S(r) = k\frac{3r^2}{2Na^2} \quad \text{for} \quad Na \gg \sqrt{Na} \gg r \quad . \tag{602}$$

The formula is derived in a few lines from the properties of a random walk on a lattice, using only two assumptions: the chains are extended, and they have a characteristic internal length a given by the smallest straight segment. Expression (602) is only valid if the polymers are effectively infinite, i.e. if the length Na of the chain and their effective average size, the *elongation* $a\sqrt{N}$, are much larger than the radius r of the region of interest; if the chain length is comparable or smaller than the region of interest, one gets the usual extensive entropy, fulfilling $S \sim r^3$. Thus *only flexible extended entities yield a* $S \sim r^2$ *dependence.*

However, there is a difficulty. From the entropy expression of a black hole we deduce that the elongation $a\sqrt{N}$ is given by $a\sqrt{N} \approx l_{\mathrm{Pl}}$; thus it is much smaller than the radius of a general, macroscopic black hole which can have diameters of several kilometres. On the other hand, the formula for long entities is only valid when the chains are longer than the distance r between the end points.

This difficulty disappears once we remember that space near a black hole is strongly curved. All lengths have to be measured in the same coordinate system. It is well known that for an outside observer, any object of finite size falling into a black hole seems to cover the complete horizon for long times (whereas it falls into the hole in its original size for an observer attached to the object). In short, an extended entity can have a proper length of Planck size but still, when seen by an outside observer, be as long as the horizon of the black hole in question. We thus find that black holes are made of extended entities. Ref. 854

Another viewpoint can confirm the result. Entropy is (proportional to) the number of yes/no questions needed to know the exact state of the system. This view of black holes has been introduced by Gerard ’t Hooft. But if a system is defined by its surface, like a black hole is, its components must be extended.

Finally, imagining black holes as made of extended entities is also consistent with the so-called no-hair theorem: black holes properties do not depend on what material falls into them, as all matter and radiation particles are made of the same elongated components. The final state only depends on the number of entities, and on nothing else. In short, the entropy of a black hole is consistent with the idea that it is made of a big tangle of elongated entities, fluctuating in shape.

Argument 6: Exchanging space points

We are still collecting arguments for the extension of fundamental entities in nature. Let us focus on their exchange behaviour. We saw above that points in space have to be eliminated in favour of continuous, fluctuating entities common to space, time, and matter. Is such a space ‘point’ or space entity a boson or a fermion? If we exchange two points of empty space in everyday life, nothing happens. Indeed, quantum field theory is based – among others – on the relation

$$[x,y] = xy - yx = 0 \qquad (603)$$

between any two points with coordinates x and y, making them bosons. But at Planck scale, due to the existence of minimal distances and areas, this relation is at least changed to

$$[x,y] = l_{\mathrm{Pl}}^2 + \dots \quad . \qquad (604)$$

This means that 'points' are neither bosons nor fermions.* They have more complex exchange properties. In fact, the term on the right hand side will be energy dependent, with Ref. 852 an effect increasing towards Planck scales. In particular, we saw that gravity implies that double exchange does not lead back to the original situation at Planck scales. Entities following Ref. 869 this or similar relations have been studied in mathematics for many decades: braids. In summary, at Planck scales space-time is not made of points, but of braids or some of their generalizations. Thus quantum theory and general relativity taken together again show that vacuum must be made of extended entities.

Argument 7: Are particles identical?

We know that at low, everyday energies, particles of the same type are *identical*. Experiments sensitive to quantum effects show that there is no way to distinguish them: any system of several identical particles obeys permutation symmetry. On the other hand we know that at Planck energy all low-energy symmetries disappear. We also know that at Planck energy, permutation cannot be carried out, as it implies exchanging positions of two particles. At Planck energy, nothing can be distinguished from vacuum; thus no two entities can be shown to have identical properties. Indeed, no two particles can be shown to be indistinguishable, as they cannot even be shown to be separate.

What happens when we slowly approach Planck energy? At everyday energies, permutation symmetry is defined by commutation or anticommutation relations of the particle creation operators

$$a^\dagger b^\dagger \pm b^\dagger a^\dagger = 0 \quad . \qquad (605)$$

At Planck energies this cannot be correct. At those energies, quantum gravity effects appear and modify the right hand side; they add an energy dependent term that is negligible at most experimentally accessible energies, but which becomes be important at Planck energy. Ref. 852 We know from our experience with Planck scales that exchanging particles twice cannot lead back to the original situation, in contrast to everyday life, as such statements cannot be correct at Planck energy. The simplest extension of the commutation relation satisfying Ref. 869 this requirement is again *braid symmetry*. Thus braid symmetry suggests that also particles themselves are made of extended entities.

---

* The same reasoning destroys the fermionic or Grassmann coordinates used in supersymmetry.

Argument 8: The meaning of spin

In our last argument, we will show that even at everyday energy, the extensions of particles makes sense. Any particle is a part of the universe. A part is something which is different from anything else. Being 'different' means that exchange has some effect. *Distinction means possibility of exchange.* In other words, any part of the universe is described by its exchange behaviour. Everyday life tells us that exchange can be seen as composed of rotation. In short, distinguishing parts are described by their rotation behaviour. For this reason, for microscopic particles, exchange behaviour is specified by spin. *Spin distinguishes particles from vacuum.*[*]

We note that volume does not distinguish vacuum from particles; neither does rest mass: there are particles without mass, both fermions (neutrinos) – even though recent results cast some doubt on this statement – and bosons (photons); same for charge. The only candidate observables to distinguish particles from vacuum are spin and momentum. But linear momentum after all is only a limiting case of angular momentum. So we confirm that *rotation behaviour is the basic aspect distinguishing particles from vacuum.*

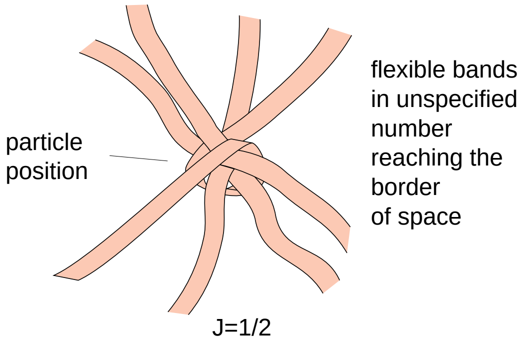


Figure 257 A possible model for a spin 1/2 particle

If spin is the central property distinguishing particles from vacuum, finding a model for spin is of central importance. But we do not have to search for long. An well-known model for spin 1/2 is part of physics folklore. Any belt provides an example, as we discussed in detail in chapter VII on permutation symmetry. Any localized structure with any number of long tails attached to it – and reaching the border of the region of space under consideration – has the same properties as a spin 1/2 particle. It is a famous exercise to show that such a model indeed is invariant under 4π rotation but not under 2π rotations, that two such particles get entangled when exchanged, but get untangled when exchanged twice. The model has all properties of spin 1/2 particles, independently of the precise structure of the central region, which remains unknown at this point. The tail model even has the same problems with highly curved space as real spin $1/2$ particles have.

See page 559

The tail model thus confirms that rotation is partial exchange. More interestingly, it shows that rotation implies connection with the border of space-time. Extended particles can be rotating. Particles can have spin $1/2$ provided that they have tails going to the border of space-time. If the tails do not reach the border, the model does not work. Spin $1/2$ thus even

Ref. 890

---

[*] With a flat (or other) background, it is possible to define a *local* energy-momentum tensor. Thus particles can be defined. Without background, this is not possible, and only global quantities can be defined. Without background, particles then cannot be defined! Therefore, we assume that we have a slowly varying space-time background in this section.

seems to *require* extension. We again reach the conclusion that extended entities are a good description for particles.

Present research

Ref. 865 The Greek deduced the existence of atoms because fish can swim through water. They argued that only if water is made of atoms, a fish can find its way through it by pushing the atoms aside. When a particle flies through vacuum, we can ask a similar question: why can particles move through vacuum? Vacuum cannot be a fluid or a solid of small entities, as this would not fix its dimensionality. Only one possibility remains: both vacuum and particles are made of a web of extended entities. Describing nature as composed of extended entities is an idea from the 1960s. Describing it as composed of 'infinitely' extended entities seems to be an idea from the 1990s. Indeed, in addition to the arguments presented so far, present research provides several others approaches that arrive at the same conclusion.

- Bosonization, the construction of fermions using an infinite number of bosons, is a central property of modern unification attempts. Ref. 893 It also implies coupling duality, and thus the extension of fundamental constituents. Through bosonization, also supersymmetry is connected to extension of fundamental particles.
- String theory and in particular its generalization to membranes are explicitly based on extended entities, as the name already states. Ref. 891 The fundamental entities are indeed assumed to reach the limits of space-time.
- Ref. 885 Research into quantum gravity and the study of spin networks and spin foams applied to space-time have shown that the vacuum must be thought as a collection of extended entities.
- In the 1990s, Dirk Kreimer has shown that high-order QED diagrams are related to knot theory. Ref. 892 He thus proved that extension appears through the back door even when electromagnetism is described using point particles.
- In 1995, Ted Jacobson showed that thermodynamics of space-time yields general relativity. Ref. 882 The equations of motion of general relativity follow from the idea of fluctuating entities of space-time.
- A recent development in particle physics, 'holography', studies the connection of surface and volume of physical systems at high energy. Ref. 894 Even if it were not part of string theory, it would still imply extended entities.
- Any other fundamental nonlocalities in the description of nature obviously also imply extended entities.

Testing the model

Is nature really described by extended entities? The idea is taken for granted by all present approaches in theoretical physics. How can we be sure about this result? The arguments presented above provide several possible checks. Challenge 1191

Conceptual checks of extension

- Any model of nature must be supersymmetric, dual and high dimensional. The idea of extended entities would be dead as soon as it is shown not to be compatible with these three requirements.

- Any model of nature must be easily extendible to a model for black holes. If not, it cannot be taken seriously.
- Showing that the results on quantum gravity, such as the results on the area and volume quantization, are in contradiction with extended entities would directly invalidate the model.
- One has to reach the same conclusion of extended entities if one starts from *any* physical (low-energy) concept – not only from length measurements – and continues to study how it behaves at Planck scales. If not, the idea is not consistent and thus incorrect.
- Showing that any conclusion of the idea of extension is in contrast with string theory or with M theory would lead to strong doubts.

Possibilities for falsification of extension models

- Observing a single particle in cosmic rays with energy above the Planck energy would provide the end of this approach.
- Showing that the measurement of length cannot be related to the counting of folds would invalidate the model.
- Finding a single *Gedankenexperiment* invalidating the extended entity idea would prove it wrong.
- Finding any property of nature not consistent with extended entities would spell the end of the idea.
- Finding an elementary particle of spin 0 would invalidate the idea. In particular, finding the Higgs boson and showing that it is elementary, i.e. that its size is smaller than its own Compton wavelength, would invalidate the model.

Note that many of these possibilities probably would also spell the death penalty for most present unification attempts.

Possibilities for confirmation of extension models

- The best confirmation would be to find a concrete model for the electron, muon and tau, and for their neutrinos. In addition, a concrete model for photons and gravitons is needed. With these models, finding a knot-based definition for the electrical charge and the lepton number would be a big step ahead. All quantum numbers should be topological quantities deduced from these models and should behave as observed.
- In July 2002, the Italian physicist Andrea Gregori has made a surprising prediction valid for any model using extended entities that reach the border of the universe: if particles are extended in this way, their mass should depend on the size of the universe. Particle masses should thus change with time, especially around the big bang. This completely new point is still a topic of research. Ref. 895
- Estimating the coupling constants and comparing them with the experimental values is of course the main dream of modern theoretical physics.
- Proving in full detail that extended entities imply exactly three plus one space-time dimensions is still necessary.
- Estimating the total number of particles in the visible universe would provide the final check of any extended entity model.

Generally speaking, the *only* possible confirmations are those from the one-page table of unexplained properties of nature given in Chapter XI. No other confirmations are possible. The ones mentioned here are the main ones. See page 701

A world of worms?

Instead of amoebas, we can take another analogy to describe nature. Flat, empty space is similar to a heap of interwoven earthworms. The worms can stretch, contract, and move. They represent the extended constituents of nature. Seen from far away, i.e. at small energies, the heap looks smooth and continuous. But seen from a shorter distance, the heap is not smooth; it is made of elongated discrete entities, wriggling, knotting and moving around in a disordered way which can only be described by statistical methods. Length and area values are deduced by counting the wriggles. Despite its limits, such a simple model of space and time can guide the imagination in a number of ways. However, we stress again that this way to describe nature is still speculative.

▪ The picture described that space is (at least) three-dimensional; three dimensions are what is left over in a heap of worms in any dimension, as windings and knots only appear in three spatial dimensions. This relation between space dimensionality and extension, though suggestive, still needs to be proven.

▪ The knot image allows to say why all electrons are the same. Every electron is the same knot. Different knots are different particles. All particles *seem* point-like at low energies. When the worms are averaged out over time, the tails disappear, and the knotted regions remain. Different types of particles correspond to different types of knots.

Ref. 878 ▪ Present research in quantum gravity suggests that the lowest energy level of free, empty space-time could be a regular arrangement of worms, similar to a weave, in which the wriggling is reduced to a bare minimum. One then can picture the first excitations of this lowest energy state as gravitons. At high excitations however, i.e. at high worm disorder, the concept of graviton loses its meaning, similar to the way in which phonons lose their meaning at high temperatures. This is the reason for what mathematicians call the non-renormalizability of gravity.

▪ Space-time curvature could be a distribution of tails skewed in a specific way; along certain directions there would be more tails than in the opposite one.

▪ Real worms have no side branches nor can they be toroidal, whereas for the basic entities of nature this can be the case.

▪ In the naive picture, time is tacitly assumed to be continuous; of course, in a precise description this is not the case. The issue still needs clarification; the precise limitations of the use of continuous time as well as the conditions for its use need to be worked out.

▪ At low energies, particles of the same type are *identical*, like knots knotted in the same way. Thus, when the worms are averaged out, they obey permutation symmetry. However, particles are not completely identical. Any two particles have some strands of the space-time weave in common; therefore, at high energies, they cannot obey permutation symmetry exactly, which is only true for point-like particles. Permutation symmetry thus becomes a low energy approximation of a more general symmetry: braid symmetry. This correspondence was already suggested above.

- The more precisely we try to localize matter, the more it fluctuates over large distances. A string has large size if observed with high temporal resolution. Matter appears to localized only because observation at finite energy implies finite time resolution.
- The energy density of vacuum at large scales is zero, but at short scales it is indistinguishable from that of particles. The model thus naturally incorporates the idea that particles and vacuum cannot be distinguished at Planck energies. The problems of defining precisely the vacuum and the particle concept in quantum field theory for strong space curvature – also mentioned in the introduction – are avoided by the worm picture: at Planck scales, the two cannot be distinguished anyway.
- Finally, no infinities appear for any quantum mechanical description, because a cut-off is introduced for all integrations. The problems of quantum field theory could disappear for good.

The big bang and the rest of the universe

Extended entities lead us to some additional consequences. We find that the beginning of the big bang does not exist, but is given by that piece of continuous entity which is encountered when going backwards in time as much as possible. This has several implications.

- Going backwards in time as far as possible – towards the 'beginning' of time – is the same as zooming to smallest distances: we find a single strand of the amoeba.
- In other words, we speculate that the whole world is one single piece, knotted, branched and fluctuating.
- Going far away into space – to the border of the universe – is like taking a snapshot with a short shutter time: strands everywhere.
- From duality, the extreme past of the universe must be the same as its future. In this sentence, 'same' must mean identical – it does not mean a repetition. (Otherwise problems with the increase of entropy would appear.) As a result the small scale of nature is the same as its early part.
- Whenever we sloppily say that extended entities are 'infinite' in size, we only mean that they reach the horizon of the universe.

In summary, no starting point of the big bang exists, because time does not exist there. For the same reason, no initial conditions for particles or space-time exist. In addition, this shows there was no creation involved, since without time and without possibility of choice, the term 'creation' makes no sense.

On the other side, we need a description for the expansion of the universe in terms of extended entities. First approaches are being explored; no final conclusions can be drawn yet. Ref. 895, 896

Curiosities and challenges

Even though this section already provided sufficient food for thought, here is some more.

- If measurements become impossible near Planck energy, we cannot even draw an energy axis reaching that value. Is this valid in allcases? Challenge 1192

- Boxes do not exit. Does this mean that the we cannot paradox with which we deduced Quantum theory implies that even if tight walls would exist, the lid of such a box can never be tightly shut. Can you provide the argument? Challenge 1193 n
- Quantum theory implies that even if tight walls would exist, the lid of such a box can never be tightly shut. Can you provide the argument? Challenge 1194 n
- Is it correct that a detector able to detect Planck mass particles would be of infinite size? What about a detector to detect a particle moving with Planck energy? Challenge 1195 n
- Can you provide an argument against theidea of extended entities?[*] Challenge 1196 e

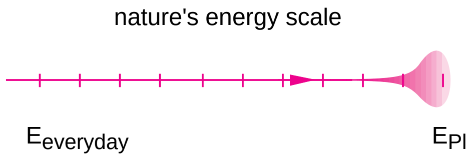


Figure 258 Planck effects make the energy axis an approximation

- Does duality imply that the cosmic background fluctuations (at the origin of galaxies and clusters) are the same as vacuumfluctuations? Challenge 1197
- Does duality imply that a system with two small masses colliding is the same as one with two large masses gravitating? Challenge 1198
- It seems that in all arguments so far we assumed and used continuous time, even though we know it is not. Does this change the conclusions so far? Challenge 1199 d
- Duality also implies that large and small masses are equivalent in some sense. A mass m in a radius r is equivalent to a mass $m_{\rm Pl}^2/m$ in a radius $l_{\rm Pl}^2/r$. In other words, duality transforms mass density from ρ to $\rho_{\rm Pl}^2/\rho$. Vacuum and maximum density are equivalent. Vacuum is thus dual to black holes.
- Duality implies that there are no initial conditions for the big bang make no sense. Duality again shows the uselessness of the idea, as minimal distance did before. As duality implies a symmetry between large an small energies, the big bang itself becomes an unclearly defined concept.
- The total symmetry, as well as space-time duality, imply that there is a symmetry between all values an observable can take. Do you agree? Challenge 1200 n
- Can supersymmetry be an aspect or special case of total symmetry or is it something else? Challenge 1201 n
- Any description is a mapping from nature to mathematics, i.e. from observed differences (and relations) to thought differences (and relations). How can we do this accurately, if differences are only approximate? Is this the end of physics? Challenge 1202 n
- Can you show that going to high energies or selecting a Planck size region of space-time is equivalent to visiting the big-bang? Challenge 1203 d

An intermediate status report

Wir müssen wissen, wir werden wissen.[**]

---

[*] If so, please email it to mm@motionmountain.net

[**] 'We must know, we will know.'

David Hilbert, 1930.

Many efforts for unification advance by digging deeper and deeper into details of quantum field theory and general relativity. Here we took the opposite approach: we took a step back and looked at the general picture. Guided by this idea we found eight arguments, all leading to the same conclusion: space-time points and point particles are made of *extended* entities.

Somehow it seems that the universe is best described by a fluctuating, multi-branched entity, a crossing between a giant amoeba and a heap of worms. Another analogy is a big pot of boiling and branched spaghetti. Such an extended model of nature is beautiful and simple, and these two criteria are often taken as indication, before any experimental tests, of the correctness of a description. We toured topics such as the existence of Planck limits, 3-dimensionality, curvature, renormalization, spin, bosonization, the cosmological constant problem, as well as a few specialized research topics, such as the search for a *background free* description of nature. We will study and test specific details in the next section. All these tests concern one of only two possible topics: the construction of the universe and the construction of particles. These are the only two issues remaining on our mountain ascent of Motion Mountain.

Ref. 881

To 'show' that we are not far from the top of Motion Mountain, we give a less serious argument as final curiosity. Salecker, Wigner and Zimmerman formulated the fundamental limit for the measurement precision τ attainable by a clock of mass M. It is given by $\tau = \sqrt{\hbar T/Mc^2}$, where T is the time to be measured. We can ask what time T can be measured with a precision of a Planck time $t_{\rm Pl}$, given a clock of the mass of the whole universe. We get a maximum time of

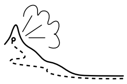
Ref. 879, 880

$$T = t_{\rm Pl}^2 Mc^2/\hbar \quad . \tag{606}$$

Inserting numbers, we find rather precisely the present age $t_{\rm o}$ of the universe.In other words, only the universe knows its own age with precision. However, soon even the universe will lose this ability. Being in this transition region is a real coincidence, because it hinges on the fact that we live in just the present stage of the universe. In fact, we quickly see that this 'coincidence' is still another rephrasing of a famous coincidence already encountered, namely Dirac's large number hypothesis. With the right dose of humour we can see the result as an omen for the fact that time is now ripe, after so much waiting, to understand the universe down to the Planck scale. We are getting nearer to the top of Motion Mountain. Be prepared for even more fun.

Challenge 1204 e

See page 636



References

852 C. Schiller, *Does matter differ from vacuum?* http://www.arxiv.org/abs/gr-qc/9610066. A French language version has been published 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. Cited on pages 791, 792, 795, 797, 798, 799, 809, 825, 826 and 835.

853 The main arguments for minimal temporal and spatial intervals are already given in 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 791.

854 See L. SUSSKIND, *String theory and the principle of black hole complementarity*, Physical Review Letters **71**, pp. 2367–2368, 1993, and M. KARLINER, I. KLEBANOV & L. SUSSKIND, *Size and shape of strings*, International Journal of Modern Physics A **3**, pp. 1981–1996, 1988, as well as L. SUSSKIND, *Structure of hadrons implied by duality*, Physical Review D **1**, pp. 1182–1186, 1970. Cited on pages 794 and 808.

855 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 795.

856 C.A. MEAD, *Possible connection between gravitation and fundamental length*, Physical Review B **135**, pp. 849–862, 1964. Cited on page 795.

857 A similar point of view, often called monism, was proposed by BARUCH SPINOZA, *Ethics demonstrated in geometrical order*, 1677, originally in Latin; an affordable French edition is B. SPINOZA, *L'Ethique*, Folio-Gallimard, 1954. For a discussion of his ideas, especially his monism, see DON GARRET ED., *The Cambridge Companion to Spinoza*, Cambridge University Press, 1996, or any general text on the history of philosophy. Cited on page 793.

858 J.H. SCHWARZ, his paper(s) about the enormous symmetry of nature. Cited on page 827.

859 J.H. SCHWARZ, *The second superstring revolution*, http://www.arxiv.org/hep-th/9607067. More on dualities in ... Cited on pages 801 and 802.

860 D. OLIVE & C. MONTONEN, *Magnetic monopoles as gauge particles*, Physics Letters **72B**, pp. 117–120, 1977. Cited on page 802.

861 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 798.

862 P. FACCHI & S. PASCAZIO, *Quantum Zeno and inverse quantum Zeno effects*, pp. 147–222, in E. WOLF, editor, *Progress in Optics*, **42**, 2001. Cited on page 798.

863 ARISTOTLE, *Of Generation and Corruption*, I, II, 316 a 13. See JEAN-PAUL DUMONT, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 427, 1991. Cited on page 798.

864 See for example the speculative model of vacuum as composed of Planck-size spheres proposed by F. WINTERBERG, Zeitschrift für Naturforschung **52a**, p. 183, 1997. Cited on page 799.

865 The Greek salt and fish arguments are given by Lucrece, or Titus Lucretius Carus, *De natura rerum*, around 50 BCE. Cited on pages 800 and 811.

866 PLATO, *Parmenides*, ca. 370 BCE. It has been translated into most languages. Reading it aloud, like a song, is a beautiful experience. A pale reflection of these ideas is Bohm's concept of 'unbroken wholeness'. Cited on page 804.

867 The famous quote is found at the beginning of chapter XI, 'The Physical Universe', in ARTHUR EDDINGTON, *The Philosophy of Physical Science*, Cambridge, 1939. Cited on page 804.

868 P. GIBBS, *Event-symmetric physics*, http://www.arXiv.org/hep-th/9505089; see also his http://www.weburbia.com/pg/contents.htm web site. Cited on page 805.

869 See for example S. MAJID, *Introduction to braided geometry and q-Minkowski space*, http://www.arxiv.org/hep-th/9410241 or S. MAJID, *Duality principle and braided geometry*, preprint gr-qc/940957. Cited on page 809.

870 J.D. Bekenstein, *Black holes and entropy*, Phys. Rev. D7, pp. 2333–2346, 1973. Cited on page 807.

871 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 807.

872 J.B. Hartle & S.W. Hawking, *Path integral derivation of black hole radiance*, Physical Review D **13**, pp. 2188–2203, 1976. A. Strominger & C. Vafa, *Microscopic origin of Bekenstein-Hawking entropy*, Physics Letters **B379**, pp. 99–104, 1996. The e-print is http://www.arxiv.org/hep-th/9601029. For other derivations of black hole entropy, see ... Cited on page 807.

873 J.D. Bekenstein, *Entropy bounds and black hole remnants*, Physical Review D **49**, pp. 1912–1921, 1994. Cited on page 807.

874 L. Bombelli, R.K. Koul, J. Lee & R.D. Sorkin, *Quantum source of entropy of black holes*, Physical Review D **34**, pp. 373–383, 1986. Cited on page 807.

875 J. Maddox, *When entropy does not seem extensive*, Nature **365**, p. 103, 1993. Other references on the proportionality of entropy and surface are ... Cited on page 807.

876 P.-G. De Gennes, *Scaling concepts in polymer physics*, Cornell University Press, 1979. Cited on page 808.

877 The analogy between polymers and black holes is by G. Weber, *Thermodynamics at boundaries*, Nature **365**, p. 792, 1993. Cited on page 808.

878 The weave model of space-time appears in certain approaches of quantum gravity, as told by A. Ashtekar, *Quantum geometry and gravity: recent advances*, December 2001, http://arxiv.org/abs/gr-qc/0112038. Cited on page 813.

879 H. Salecker & E.P. Wigner, *Quantum limitations of the measurement of space-time distances*, Physical Review **109**, pp. 571–577, 1958. Cited on page 816.

880 E.J. Zimmerman, *The macroscopic nature of space-time*, American Journal of Physics **30**, pp. 97–105, 1962. Cited on page 816.

881 Searches for background free approaches are described by E. Witten, ... Cited on pages 816 and 833.

882 T. Jacobson, *Thermodynamics of spacetime: the Einstein equation of state*, preprint http://arxiv.org/abs/gr-qc/9504004 or Physical Review Letters **75**, pp. 1260–1263, 1995. Cited on pages 811, 826 and 828.

883 See for example C. Rovelli, *Notes for a brief history of quantum gravity*, http://www.arxiv.org/gr-qc/0006061. Cited on pages 825, 826 and 828.

884 The papers on braid symmetry cannot be counted any more; see for an overview S. Majid, *Duality principle and braided geometry*, preprint gr-qc/940957. Cited on page 834.

885 A. Ashtekar, *Quantum geometry and gravity: recent advances*, http://www.arxiv.org/gr-qc/0112038. A. Ashtekar, *Quantum geometry in action: big bang and black holes*, http://www.arxiv.org/math-ph/0202008. On a slightly different topic, see also S.A. Major, *A spin network primer*, http://www.arxiv.org/gr-qc/9905020. Cited on pages 811, 824, 825 and 828.

886 For an up-to date list of string theory and M theory reviews, see the http://www.nuclecu.unam.mx/ alberto/physics/stringrev.html web site. Cited on page 824.

887 On quantum groups, see ... Cited on page 824.

888 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 832.

889 The most prominet proposer of the idea that particles maight be knots was William Thomson, ... Cited on page 834.

890 The relation between spin and statistics has been studied recently by M.V. Berry & J.M. Robbins, *Quantum indistinguishability: spin-statistics without relativity or field theory?*, in R.C. Hilborn & G.M. Tino, editors, *Spin-Statistics Connection and Commutation Relations*, American Institute of Physics CP545 2000. Cited on page 810.

891 A popular account making the point is the bestseller by the US-american string theorist Brian Greene, *The Elegant Universe*, ... Cited on page 811.

892 A good introduction into his work is the paper D. Kreimer, *New mathematical structures in renormalizable quantum field theories*, http://www.arxiv.org/hep-th/0211136 to appear in Annals of Physics Cited on page 811.

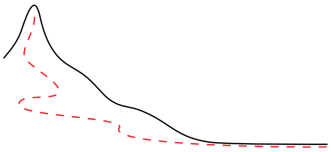
893 An example is given by A.A. Slavnov, *Fermi-Bose duality via extra dimension*, preprint http://www.arxiv.org/hep-th/9512101. See also the standard work by Michael Stone, editor, *Bosonization*, World Scientific, 1994. Cited on page 811.

894 Introductions to holography are E. Alvarez, J. Conde & L. Hernandez, *Rudiments of Holography*, http://www.arxiv.org/hep-th/0205075 or R. Bousso, *The holographic principle*, http://www.arxiv.org/hep-th/02003101. The original idea is from J. Maldacena, *The large N limit of superconformal field theories and supergravity*, http://www.arxiv.org/hep-th/9711200. Cited on page 811.

895 Andrea Gregori, *Entropy, string theory, and our world*, http://arxiv.org/hep-th/0207195 preprint. Cited on pages 812 and 814.

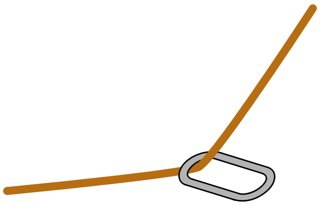
896 String cosmology is a passtime for many. Examples are N.E. Mavromatos, *String cosmology*, http://arxiv.org/hep-th/0111275 and N.G. Sanchez, *New developments in string gravity and string cosmology - a summary report*, http://arxiv.org/hep-th/0209016. Cited on page 814.

897 Cited in Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, p. 379, 1991. Cited on page 801.



APPENDICES

Where the reference information necessary
for mountain ascents is given, allowing
everybody to be prepared for
any other adventure as well.



Typeset in
January 2003

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 by italic page numbers. Throughout the text SI units are used; they are defined in Appendix B. Experimental results are cited with limited precision, usually only two digits, as this is usually sufficient for discussion. Precise reference values can be found in Appendices B and C.

In relativity we use the time convention, where the metric has the signature $(+---)$, as used by about 70% of the literature worldwide. We use indices i, j, k for three-vectors, and Ref. 899 indices a, b, c, etc. for four-three-vectors. Other conventions specific to general relativity are explained in the corresponding chapter.

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*

Books are collections of symbols. Most symbols have been developed over hundreds of years; only the clearest and simplest are now in use. In this mountain ascent, the symbols used as abbreviations for *physical* quantities are all taken from the Latin or Greek alphabets. They are always defined in the context where they are used. The symbols designating units, constants, and particles are defined in Appendices B and C. All conform as much as possible to the ISO standard. Ref. 901

Mathematical symbols used in this text, in particular those for operations and relations, are given in the following list, together with their origin. Ref. 900

$+, -$	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

---

* Robert Recorde (ca.1510–1558), English mathematician and physician; he died in prison, though not for his Ref. 900 false pretention to be the inventor of the equal sign, which he simply took from his Italian colleagues, but for a smaller crime, namely debth. The quotation is from his *The Whetstone of Witte*, 1557.

$=$	equal to — Italian mathematicians, early 16th century, then brought to England by R. Recorde
$\{\,\}, [\,], (\,)$	grouping symbols — use starts in the 16th century
$>, <$	larger than, smaller than — T. Harriot 1631
$\times$	multiplied with, times — W. Oughtred 1631
$:$	divided by — G. Leibniz 1684
$\cdot$	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
$\mathrm{d}/\mathrm{d}x$, d^2x, $\int y\,\mathrm{d}x$	derivative, differential, integral — G. Leibniz 1675
φx	function of x — J. Bernoulli 1718
$fx, f(x)$	function of x — L. Euler 1734
$\Delta x, \sum$	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 by A. Legendre in 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	$\sum_{n=0}^{\infty} = \lim_{n\to\infty}(1+1/n)^n$ — 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 used here 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. 902

The section sign § dates from the 13th century in northern Italy, as was shown by the German palaeographer 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. 903

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 , . ; : ! ? ‘ ’ » « – ( ) ... , each 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.[*] 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 have become popular. The smiley 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. 904 Ref. 905

The most important sign of all, the white space separating words, was due to Celtic 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. Ref. 906

The Latin alphabet

> What is written without effort is in general written without pleasure.
> Samuel Johnson (1709–1784)

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 1208 n

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 other signs. Ref. 908

---

[*] On the parenthesis see the beautiful book by J. Lennard, *But I disgress*, Oxford University Press, 1991.

[**] 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.'

Similarly, lower case letters are not classical Latin; they date only from the middle ages. Like most accents such as ê or ä, who were also defined in the middle ages, they were introduced to save the then expensive paper surface by shortening printed words.

> Outside a dog, a book is a man's best friend.
> Inside a dog, it's too dark to read.
> Groucho Marx

The Greek alphabet

Ref. 907

The Greek alphabet in turn was derived from the Phoenician or a similar northern 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 derivatives. The Greek letter names of course are the origin of the term *alphabet* itself.

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, a few new ones and the lower case versions were added, giving the *classic* Greek 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; thus they also have a lower case form.

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
Ϝ ϛ		Digamma[b]	w	6	Ϙ ϙ		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

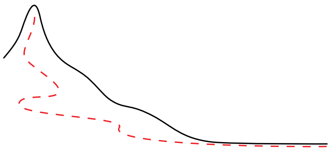
a. Only if before velars, i.e. before kappa, gamma, xi and chi.

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 pronunciation, 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



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

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 'iː'.) Obviously, 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, omicron 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, interpreted 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 exists no standard transcription from cyrillic to Latin, so that often the same author is spelled differently in different countries and even in different occasions.

The Hebrew alphabet and other scripts

Ref. 901 The phoenician alphabet is also at the origin of the Hebrew alphabet, which begins

letter	name	corr.
א	aleph	a
ב	beth	b
ג	gimel	g
ד	daleth	d
etc.		

See page 463 Only the first of these letters is commonly 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 *syllable*, such as Maya, Korean, or Japanese. In addition there are non-alphabetic writing systems, having signs for each *word*, such as Chinese. Even though there are about 7000 languages on earth,

---

[*] 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.

The Gothic alphabet 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 onwards. In Latin countries, Gothic letters were replaced in the 16th century by the *antiqua*, the ancestor of the type in which this text is set. In other countries, Gothic letters remained in use much later. The were used in type and handwriting in Germany until in 1941 the national-socialist government suddenly abolished them. They remain in sporadic use across Europe. In many physics and mathematics books, gothic letters are used to denote vector quantities instead of bold letters.

there are only about *two dozen* writing systems. For physical and mathematical formulas though, the sign system presented here, based on Latin and Greek letters, is a standard the world over. It is used independently of the writing system of the text containing it.

Digits and numbers

Both the digits and the method used in this text to write numbers stem from India. They were 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 resemble closely those used in arabic writing,[*] if they are turned clockwise by 90 °. 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, and to perform calculations with them. The Indian-Arabic method brought two innovations: 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 MCMXCVIII, as well as the *Greek number system*, in which the Greek letters were used for numbers, as shown above. In short, compared to the previous systems the 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 lunar time keeping is easily organized locally. This lead to the use of the month as calendar unit. Centralized states imposed *solar* calendars, based on the year. Solar systems require astronomers and thus a central authority to finance them. For various reasons, farmers, politicians, tax collectors, astronomers, and some, but not all, religious

---

[*] 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. However, the book contains many factual errors in the text, as explained in the http://www.ams.org/notices/200201/rev-dauben.pdf and http://www.ams.org/notices/200202/rev-dauben.pdf review.

[**] About the stories and the methods of calculating prodigies, 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. One can even learn from the book how to emulate them.

groups wanted the calendar to follow the solar year as precisely as possible. The compromises necessary between months and years are the origin of leap days. The compromises require that different months in a year have different length; in addition, their length is different in different calendars. The most commonly used year-month structure was organized over 2000 years ago by Gaius Julius Ceasar, and is thus 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 included in the calendar only around the year 400, towards the end of the Roman empire. The final change took place between 1582 and 1917 (depending on the country), when more precise measurements of the solar year were used to set a new method to determine leap days, a method still in use today. Together with a reset of the date and the fixation of the week rhythm, this standard is called the *gregorian calendar* or simply the *modern calendar*. It is used by a majority of the world's population.

Despite this complexity, the modern calendar allows you 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 or February of a leap year,
- 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

---

[*] 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 explain 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 and temperature 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, then for a short time to Latin with German grammar, after which it changed to Latin vocabulary and British/American grammar.

worth considering?

Abbreviations and 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. First of all, it uses abbreviations, which is 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 number of people simply reflects an increasingly ineffective drive to fame.

Eponyms are a lasting proof of the lack of imagination of scientists. Eponyms are avoided as much as possible in our walk; mathematical equations or entities are given *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's equation of motion' in place of 'Heisenberg's equation'.

However, some 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 effect, Lie groups and the Virasoro algebra are examples. In compensation, it is made sure that the definitions can be looked up using the index.

* * *

References

899 For a clear overview of the various sign conventions in general relativity, see the front cover of CHARLES W. MISNER, KIP S. THORNE & JOHN A. WHEELER, *Gravitation*, Freeman, 1973. We use the gravitational sign conventions of HANS C. OHANIAN & REMO RUFFINI, *Gravitazione e spazio-tempo*, Zanichelli, 1997. Cited on page 862.

900 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 862.

901 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

---

Many units of measurement also date from Roman times, as explained in the next appendix. Even the infatuation with Greek technical terms, as shown in coinages such as 'gyroscope', 'entropy', or 'proton', dates from Roman times. Ref. 909

is given on page 2-90. The list of abbreviations of physical quantities for use in formulas approved by ISO, IUPAP and IUPAC can also be found there. Cited on pages 862 and 867.

902 JAN TSCHICHOLD, *Vormveranderingen van het &-teken*, Uitgeverij De Buitenkant, 1994. Cited on page 863.

903 PAUL LEHMANN, *Erforschung des Mittelalters – Ausgewählte Abhandlungen und Aufsätze*, Anton Hiersemann, Stuttgart, 1961, pp. 4–21. Cited on page 863.

904 M.B. PARKES, *Pause and Effect: An Introduction to the History of Punctuation in the West*, University of California Press, 1993. Cited on page 864.

905 This is explained by BERTHOLD LOUIS ULLMAN, *Ancient Writing and its Influence*, 1932. Cited on page 864.

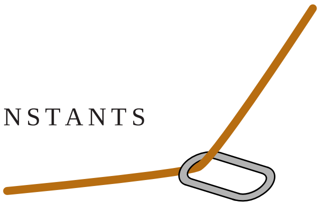
906 BERNARD BISCHOFF, *Paläographie des römischen Altertums und des abendländischen Mittelalters*, Erich Schmidt Verlag, 1979, pp. 215–219. Cited on page 864.

907 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 865.

908 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 864.

909 The connections between Greek roots and many French words, and thus many English ones, can be used to rapidly build up a 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 870.

Typeset in
January 2003



Appendix B Units, Measurements and Constants

Measurements are comparisons. The standard used for the comparison is called a *unit*. Many different systems of units have been used throughout the world. Unit systems are standards, and 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 applied to 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. 899

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 caesium 133 atom.' (1967)*
- 'The *metre* is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second.' (1983)
- 'The *kilogram* is the unit of mass; it is equal to the mass of the international prototype of the kilogram.' (1901)*
- 'The *ampere* is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 metre apart in vacuum, would produce between these conductors a force equal to $2 \cdot 10^{-7}$ newton per metre of length.' (1948)
- 'The *kelvin*, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.' (1967)*
- 'The *mole* is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12.' (1971)*
- 'The *candela* is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency $540 \cdot 10^{12}$ hertz and has a radiant intensity in that

direction of (1/683) watt per steradian.’ (1979)[*]

Note that both time and length units are defined as certain properties of a standard example of motion, namely light. 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 artefact, is the last exception to this requirement; extensive research is under way to eliminate this artefact from the definition – an international race that will take a few more years. A definition can be based only on two ways: counting particles or fixing $\hbar$. The former can be achieved in crystals, the latter using any formula where $\hbar$ appears, such as the de Broglie wavelength, Josephson junctions, etc.)
- 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 derived, the supplementary and the admitted units:

The *derived* units with special names, in their official English spelling, i.e. without capital letters and accents, are:

name	abbreviation & definition	name	abbreviation & 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

---

[*] Ref. 900 The international prototype of the kilogram is a platinum-iridium cylinder kept at the BIPM in Sèvres, in France. For more details on the levels of the caesium atom, consult a book on atomic physics. The Celsius scale of temperature 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.

[*] Jacques Babinet (1794–1874), French physicist who published important work in optics.

name	abbreviation & definition	name	abbreviation & definition
lumen	$\mathrm{lm} = \mathrm{cd\,sr}$	lux	$\mathrm{lx} = \mathrm{lm}/\mathrm{m}^2 = \mathrm{cd\,sr}/\mathrm{m}^2$
becquerel	$\mathrm{Bq} = 1/\mathrm{s}$	gray	$\mathrm{Gy} = \mathrm{J}/\mathrm{kg} = \mathrm{m}^2/\mathrm{s}^2$
sievert	$\mathrm{Sv} = \mathrm{J}/\mathrm{kg} = \mathrm{m}^2/\mathrm{s}^2$	katal	$\mathrm{kat} = \mathrm{mol}/\mathrm{s}$

We note that in all definitions of units, the kilogram only appears to the powers of 1, 0 and -1. The final explanation for this fact appeared only recently.

See page 839

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 *admitted* non-SI units are *minute*, *hour*, *day* (for time), *degree* $1° = \pi/180\,\mathrm{rad}$, *minute* $1' = \pi/10\,800\,\mathrm{rad}$, *second* $1'' = \pi/648\,000\,\mathrm{rad}$ (for angles), *litre*, and *tonne*.

All other units are to be avoided.

All SI units are made more practical by the introduction of standard names and abbreviations for the powers of ten, the so-called *prefixes*:[*]

	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	unofficial:				Ref. 901	
10^{9}	Giga	G	10^{-9}	nano	n	10^{27}	Xenta	X	10^{-27}	xenno	x
10^{12}	Tera	T	10^{-12}	pico	p	10^{30}	Wekta	W	10^{-30}	weko	w
10^{15}	Peta	P	10^{-15}	femto	f	10^{33}	Vendekta	V	10^{-33}	vendeko	v
						10^{36}	Udekta	U	10^{-36}	udeko	u

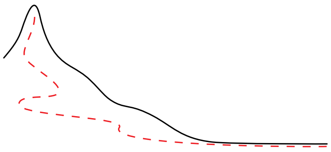
- SI units form a *complete* system; they cover in a systematic way the complete set of observables of physics. Moreover, they fix the units of measurements for physics and for all other sciences as well.
- They form a *universal* system; they can be used in trade, in industry, in commerce, at home, 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.

---

[*] 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, the unofficial ones to sound similar to the Greek words for nine, ten, eleven, and twelve), some are from Danish/Norwegian (atto from atten 'eighteen', femto from femten 'fifteen'), some are from Latin (from mille 'thousand', from centum 'hundred', from decem 'ten', from nanus 'dwarf'), some are from Italian (from piccolo 'small'), some are Greek (micro is from μικρός 'small', deca/deka from δέκα 'ten', hecto from ἑκατόν 'hundred', kilo from χίλιοι 'thousand', mega from μέγας 'large', giga from γίγας 'giant', tera from τέρας 'monster').

Translate: I was caught in such a traffic jam that I needed a microcentury for a picoparsec and that my car's fuel consumption was two tenths of a square millimetre.

Challenge 1210 e



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

The SI units are not the only possible set that fulfils all these requirements, but they form the only existing system doing so.[*]

We remind that since every measurement is a comparison with a standard, any measurement requires matter to realize the standard (yes, even for the speed standard) and radiation to achieve the comparison. Our concept of measurement thus assumes that matter and radiation exist and can be clearly separated from each other. See page 793

Planck's (improved) natural units

Since the exact form of many equations depends on the used system of units, theoretical physicists often use unit 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 $c = 1$, $\hbar = 1$, $G = 1$, $k = 1$, $\varepsilon_o = 1/4\pi$ and $\mu_o = 4\pi$ in equations written in SI units. Planck units are thus defined from combinations of fundamental constants; those 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; every quantity X is substituted by $X/X_{\rm Pl}$.

Table 63 Planck's natural units

Name	definition	value
Basic units		
the Planck length	$l_{\rm Pl} = \sqrt{\hbar G/c^3}$	$= 1.616\,0(12) \cdot 10^{-35}$ m
the Planck time	$t_{\rm Pl} = \sqrt{\hbar G/c^5}$	$= 5.390\,6(40) \cdot 10^{-44}$ s
the Planck mass	$m_{\rm Pl} = \sqrt{\hbar c/G}$	$= 21.767(16)$ µg
the Planck current	$I_{\rm Pl} = \sqrt{4\pi\varepsilon_o c^6/G}$	$= 3.479\,3(22) \cdot 10^{25}$ A
the Planck temperature	$T_{\rm Pl} = \sqrt{\hbar c^5/Gk^2}$	$= 1.417\,1(91) \cdot 10^{32}$ K
Trivial units		
the Planck velocity	$v_{\rm Pl} = c$	$= 0.3$ Gm/s
the Planck angular momentum	$L_{\rm Pl} = \hbar$	$= 1.1 \cdot 10^{-34}$ Js
the Planck action	$S_{\rm aPl} = \hbar$	$= 1.1 \cdot 10^{-34}$ Js

---

[*] 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, after having been defined as minute of arc, is exactly 1852 m), 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 furlong/fortnight as unit for velocity – are now officially defined as multiples of SI units.

[**] 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 $4\pi\varepsilon_o$ in the expressions; for example, using $4\pi\varepsilon_o\alpha$, with the *fine structure constant* α, gives $q_{\rm Pl} = e$. For the explanation of the numbers between brackets, the standard deviations, see page 881.

Name	definition	value
the Planck entropy	$S_{\mathrm{ePl}} = k$	$= 13.8\ \mathrm{yJ/K}$
Composed units		
the Planck mass density	$\rho_{\mathrm{Pl}} = c^5/G^2\hbar$	$= 5.2 \cdot 10^{96}\ \mathrm{kg/m^3}$
the Planck energy	$E_{\mathrm{Pl}} = \sqrt{\hbar c^5/G}$	$= 2.0\ \mathrm{GJ} = 1.2 \cdot 10^{28}\ \mathrm{eV}$
the Planck momentum	$p_{\mathrm{Pl}} = \sqrt{\hbar c^3/G}$	$= 6.5\ \mathrm{Nm}$
the Planck force	$F_{\mathrm{Pl}} = c^4/G$	$= 1.2 \cdot 10^{44}\ \mathrm{N}$
the Planck power	$P_{\mathrm{Pl}} = c^5/G$	$= 3.6 \cdot 10^{52}\ \mathrm{W}$
the Planck acceleration	$a_{\mathrm{Pl}} = \sqrt{c^7/\hbar G}$	$= 5.6 \cdot 10^{51}\ \mathrm{m/s^2}$
the Planck frequency	$f_{\mathrm{Pl}} = \sqrt{c^5/\hbar G}$	$= 1.9 \cdot 10^{43}\ \mathrm{Hz}$
the Planck electric charge	$q_{\mathrm{Pl}} = \sqrt{4\pi\varepsilon_o c\hbar}$	$= 1.9\ \mathrm{aC} = 11.7\ \mathrm{e}$
the Planck voltage	$U_{\mathrm{Pl}} = \sqrt{c^4/4\pi\varepsilon_o G}$	$= 1.0 \cdot 10^{27}\ \mathrm{V}$
the Planck resistance	$R_{\mathrm{Pl}} = 1/4\pi\varepsilon_o c$	$= 30.0\ \Omega$
the Planck capacitance	$C_{\mathrm{Pl}} = 4\pi\varepsilon_o\sqrt{\hbar G/c^3}$	$= 1.8 \cdot 10^{-45}\ \mathrm{F}$
the Planck inductance	$L_{\mathrm{Pl}} = (1/4\pi\varepsilon_o)\sqrt{\hbar G/c^7}$	$= 1.6 \cdot 10^{-42}\ \mathrm{H}$
the Planck electric field	$E_{\mathrm{Pl}} = \sqrt{c^7/4\pi\varepsilon_o\hbar G^2}$	$= 6.5 \cdot 10^{61}\ \mathrm{V/m}$
the Planck magnetic flux density	$B_{\mathrm{Pl}} = \sqrt{c^5/4\pi\varepsilon_o\hbar G^2}$	$= 2.2 \cdot 10^{53}\ \mathrm{T}$

The natural units are important for another reason: whenever a quantity is sloppily called ‘infinitely small (or large)’, the correct expression is ‘small (or large) as the corresponding Planck unit’. As explained in special relativity, general relativity, and quantum theory, the third part, this substitution is correct because almost all Planck units provide, within a factor of the order 1, the extreme value for the corresponding observable. Unfortunately, these factors have not entered the mainstream yet; if G is substituted by $4G$, $\hbar$ by $\hbar/2$ and $4\pi\varepsilon_o$ by $4\pi\varepsilon_o\alpha$ in all formulas, the exact extremal value for each observable in nature are obtained. Exceptions are possible only for extensive quantities, i.e. for those quantities for which many particle systems can exceed single particle limits, such as mass or electrical resistance.

See page 726

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 only sets $c = \hbar = k = 1$ and $\mu_o = 1/\varepsilon_o = 4\pi$,* leaving

---

* 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 JOHN DAVID JACKSON, *Classical Electrodynamics*, 3rd edition, Wiley, 1998.

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. The most important physical observables are related by

$$\begin{aligned}
[l] &= 1/[E] = [t] = [C] = [L] \quad , \\
1/[l] &= [E] = [m] = [p] = [a] = [f] = [I] = [U] = [T] \quad , \\
[l]^2 &= 1/[E]^2 = [G] = [P] = 1/[B] = 1/[E_{\text{el.}}] \quad \text{and} \\
1 &= [v] = [q] = [e] = [R] = [S_{\text{action}}] = [S_{\text{entropy}}] = \hbar = c = k = [\alpha]
\end{aligned}$$

with the usual convention to write $[x]$ for the 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 an electron when accelerated by an electrical potential difference of 1 Volt ('proton Volt' would be a better name). Therefore one has 1 eV=$1.6 \cdot 10^{-19}$ J, or roughly

$$1\,\text{eV} \approx \tfrac{1}{6}\,\text{aJ}$$

which is easily remembered. The simplification $c = \hbar = 1$ yields $G = 6.9 \cdot 10^{-57}\,\text{eV}^{-2}$ and allows to use the unit eV also for mass, momentum, temperature, frequency, time and length, with the respective correspondences $1\,\text{eV} \,\widehat{=}\, 1.8 \cdot 10^{-36}\,\text{kg} \,\widehat{=}\, 5.4 \cdot 10^{-28}\,\text{Nm} \,\widehat{=}\, 242\,\text{THz}$ Challenge 1212 e
$\widehat{=}\, 11.6\,\text{kK}$ and $1\,\text{eV}^{-1} \,\widehat{=}\, 4.1\,\text{fs} \,\widehat{=}\, 1.2\,\mu\text{m}$.

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 0.025 eV or 4.0 zJ. The highest particle energy measured so far is a cosmic ray of energy of $3 \cdot 10^{20}$ eV
or 48 J. Down here on the earth, an accelerator with an energy of about 105 GeV or 17 nJ Ref. 902
for electrons and antielectrons has been built, and one with an energy of 10 TeV or 1.6 μJ for protons will be built. Both are owned by CERN in Geneva and have a circumference of 27 km.

The lowest temperature measured up to now is 280 pK, in a system of Rhodium nuclei in-
side a special cooling system. The interior of that cryostat possibly is the coolest point in the Ref. 903
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 vJ=$3.8 \cdot 10^{-33}$ 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.

---

* The web page http://www.chemie.fu-berlin.de/chemistry/general/units-en.html allows to convert various units into each other.

In general relativity still another system is sometimes used, 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.

Curiosities

Here are a few facts making the concept of unit more vivid.

- A gray is the amount of radioactivity that deposes 1 J on 1 kg of matter. A sievert is a unit adjusted to human scale, where the different types of human tissues are weighted with a factor describing the effectiveness of radiation deposition. Four to five sievert are a lethal dose to humans. In comparison, the natural radioactivity present inside human bodies leads to a dose of 0.2 mSv per year. An average X-ray image is an irradiation of 1 mSv; a CAT scan 8 mSv. Ref. 904

- Are you confused by the candela? The definition simply says that $683\,\mathrm{cd} = 683\,\mathrm{lm}/\mathrm{sr}$ correspond to $1\,\mathrm{W}/\mathrm{sr}$. 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 indeed emits a luminous intensity of about a candela. Therefore, at night, a candle can be seen up to a distance of one or two dozen kilometres. Challenge 1213 e A 100 W incandescent light bulb produces 1700 lm, and the brightest light emitting diodes about 5 lm.

The *irradiance* of sunlight is about $1300\,\mathrm{W}/\mathrm{m}^2$ on a sunny day; the *illuminance* is $120\,\mathrm{klm}/\mathrm{m}^2$ or $170\,\mathrm{W}/\mathrm{m}^2$, 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 $10^{18}\,\mathrm{W}/\mathrm{m}^2$, 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 atoms; such a beam ionizes all matter it encounters. Ref. 905

- The Planck length is roughly the de Broglie wavelength $\lambda_B = h/mv$ of a man walking comfortably ($m = 80$ kg, $v = 0.5\,\mathrm{m/s}$); this motion is therefore aptly called the 'Planck stroll.' Ref. 906

- 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 1/86 400th of the day any more (it did so in the year 1900); the earth now takes about 86 400.002 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; the average rotation speed even changes from winter to summer due to the change in polar ice caps, and in addition that average decreases over time, due to the

---

* 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.

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 usually lies between 1 ms for 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. 907 Ref. 908
- 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, where 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. 909 Challenge 1214 n
- 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. 910
- 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 883
- 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? Challenge 1215 n
- 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. 911 Ref. 912
- The swedish astronomer Anders Celsius (1701–1744) originally set the freezing point at 100 degrees and the boiling point of water at 0 degrees. Then the numbers were switched to get today's scale, with a small detail though. 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 100 °C any more? Ref. 913 Challenge 1216 n
- 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.'

---

[*] 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 table of SI prefixes covers seventy-two decades. How many more will ever be needed? Even this extended list will include only a small part of the infinite number of possible decades. Why is this the case? Challenge 1217 n

- 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, visitors can admire the standard metre, the standard kilogram, and the standard apple.*

Precision and accuracy of measurements

As explained on page 193, *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

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (616)$$

where $\bar{x}$ is the average of the measurements x_i. (Can you imagine why $n-1$ is used in the formula instead of n?) By the way, for a Gaussian distribution, 2.35 σ is the full width at half maximum. Challenge 1218 n

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 also called *total uncertainty*. Ref. 915

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 references. Ref. 916 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.**

What are the limits to accuracy and precision? First of all, there is no way, even in principle, to measure a quantity x to a *precision* higher than about 61 digits, because

---

[*] It is not a joke however, that owners of several apple trees in Britain and in the US claim descent, by rerooting, from the original tree under which Newton had his insight. Ref. 914 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, 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 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.

$\Delta x/x \gtrsim l_{\text{Pl}}/d_{\text{horizon}} = 10^{-61}$. In the third part of our text, studies of clocks and meter bars will further reduce this theoretical limit. See page 760

But it is not difficult to deduce more stringent practical limits. No reasonable machine can measure quantities with a higher precision than measuring the diameter of the earth within the smallest length ever measured, about 10^{-19} m; that 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. Indeed, an additional digit in measurement precision means often an additional digit in equipment cost.

Basic physical constants

In principle, all experimental measurements of matter properties, such as colour, density, or elastic properties, can be predicted using the values of the following constants, using them in quantum theory calculations. Specifically, this is possible using the equations of the standard model of high energy physics. Ref. 916 See page 674

Table 64 Basic physical constants

Quantity	name	value in SI units	uncertainty
vacuum speed of light[a]	c	$299\,792\,458$ 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	$4\pi \cdot 10^{-7}$ H/m	0
		$= 1.256\,637\,061\,435\,917\,295\,385\ldots\ \mu$H/m	
vacuum permittivity[a]	$\varepsilon_o = 1/\mu_o c^2$	$8.854\,187\,817\,620\ldots$ pF/m	0
Planck constant	h	$6.626\,068\,76(52) \cdot 10^{-34}$ Js	$7.8 \cdot 10^{-8}$
reduced Planck constant	$\hbar$	$1.054\,571\,596(82) \cdot 10^{-34}$ Js	$7.8 \cdot 10^{-8}$
positron charge	e	$0.160\,217\,646\,2(63)$ aC	$3.9 \cdot 10^{-8}$
Boltzmann constant	k	$1.380\,650\,3(24) \cdot 10^{-23}$ J/K	$1.7 \cdot 10^{-6}$
gravitational constant	G	$6.673(10) \cdot 10^{-11}$ Nm^2/kg^2	$1.5 \cdot 10^{-3}$
gravitational coupling constant	$\kappa = 8\pi G/c^4$	$2.076(3) \cdot 10^{-43}$ s^2/kg m	$1.5 \cdot 10^{-3}$
fine structure constant,[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	$= g_{\text{em}}(m_e^2 c^2)$	$= 0.007\,297\,352\,533(27)$	$3.7 \cdot 10^{-9}$
Fermi coupling constant,[b]	$G_F/(\hbar c)^3$	$1.166\,39(1) \cdot 10^{-5}$ GeV^{-2}	$8.6 \cdot 10^{-6}$
weak coupling constant	$\alpha_w(M_Z) = g_w^2/4\pi$	$1/30.1(3)$	
weak mixing angle	$\sin^2\theta_W(\overline{MS})$	$0.231\,24(24)$	$1.0 \cdot 10^{-3}$
weak mixing angle	$\sin^2\theta_W$ (on shell)	$0.2224(19)$	$8.7 \cdot 10^{-3}$
	$= 1 - (m_W/m_Z)^2$		
strong coupling constant[b]	$\alpha_s(M_Z) = g^2 s/4\pi$	$0.118(3)$	$25 \cdot 10^{-3}$

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 for the electromagnetic coupling 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 See page **??**

strong coupling constant has higher values at lower momentum transfers; e.g. one has $\alpha_s(34\,\mathrm{GeV}) = 0.14(2)$.

Why do all these constants have the values they have? The answer depends on the constant. For any constant having a unit, such as the quantum of action $\hbar$, the numerical value has no intrinsic meaning. It is $1.054 \cdot 10^{-34}$ Js because of the SI definition of the joule and the second.

However, the question why the value of a constant with units is not larger or smaller always requires 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 if we asks about the size of atoms, people, trees and stars, about the duration of molecular and atomic processes, or about the mass of nuclei and mountains. Understanding the values of all dimensionless constants is thus the key to understanding nature.

The basic constants yield the following useful high-precision observations.

Table 65 Derived physical 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]	$R_\infty = m_e c \alpha^2/2h$	$10\,973\,731.568\,549(83)\ \mathrm{m}^{-1}$	$7.6 \cdot 10^{-12}$
mag. flux quantum	$\varphi_o = h/2e$	$2.067\,833\,636(81)$ pWb	$3.9 \cdot 10^{-8}$
Josephson freq. ratio	$2e/h$	$483.597\,898(19)$ 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	$\mu_B = e\hbar/2m_e$	$9.274\,008\,99(37) \cdot 10^{-24}$ J/T	$4.0 \cdot 10^{-8}$
classical electron radius	$r_e = e^2/4\pi\varepsilon_o m_e c^2$	$2.817\,940\,285(31)$ fm	$1.1 \cdot 10^{-8}$
Compton wavelength	$\lambda_c = h/m_e c$	$2.426\,310\,215(18)$ pm	$7.3 \cdot 10^{-9}$
of the electron	$\lambda\!\!\!^{-}_c = \hbar/m_e c = r_e/\alpha$	$0.386\,159\,264\,2(28)$ pm	$7.3 \cdot 10^{-9}$
Bohr radius [a]	$a_\infty = r_e/\alpha^2$	$52.917\,720\,83(19)$ pm	$3.7 \cdot 10^{-9}$
cyclotron frequency of the electron	$f_c/B = e/2\pi m_e$	$27.992\,4925(11)$ GHz/T	$4.0 \cdot 10^{-8}$
nuclear magneton	$\mu_N = e\hbar/2m_p$	$5.050\,783\,17(20) \cdot 10^{-27}$ J/T	$4.0 \cdot 10^{-8}$
proton electron mass ratio	m_p/m_e	$1\,836.152\,667\,5(39)$	$2.1 \cdot 10^{-9}$
Stephan-Boltzmann constant	$\sigma = \pi^2 k^4/60\hbar^3 c^2$	$5.670\,400(40) \cdot 10^{-8}\ \mathrm{W/m^2K^4}$	$7.0 \cdot 10^{-6}$
Wien displacement law constant	$b = \lambda_{max} T$	$2.897\,7686(51)$ mmK	$1.7 \cdot 10^{-6}$
bits to entropy conv. const.		10^{23} bit $= 0.956\,994\,5(17)$ J/K	
TNT energy content		3.7 to 4.0 MJ/kg=$4 \cdot 10^3\ \mathrm{m^2/s^2}$	

[a]. For infinite mass of the nucleus.

Some properties of the universe as a whole are listed in the following.

Table 66 Astrophysical constants

Quantity	name	value
gravitational constant	G	$6.672\,59(85) \cdot 10^{-11}\ \mathrm{m^3/kg\,s^2}$
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	$23^h 56' 4.090\,53''$
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	$> 3.5(4) \cdot 10^{17}$ s or $> 11.5(1.5) \cdot 10^9$ 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 space-time, 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	$\rho_\gamma = \pi^2 k^4/15 T_o^4$	$4.6 \cdot 10^{-31}\ \mathrm{kg/m^3}$
photon number density		$400\,/\mathrm{cm^3}(T_o/2.7\,\mathrm{K})^3$, 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 ms/a.There is even an empirical formula available for the change of the length of the year over time.

Challenge 1219 e

Ref. 917

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 $16 \pm 4\,\mu\mathrm{K} = 6(2) \cdot 10^{-6}\, T_{\mathrm{o}}$.

See page 311

Attention: in the third part of this text it is shown that many constants in Table 66 are *not* physically sensible quantities. They have to be taken with lots of grains of salt. The more specific constants given in the following table are all sensible though.

Table 67 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}	$6378.1367(1)$ km
earth radius, polar [a]	R_{p}	$6356.7517(1)$ km
equator pole distance [a]		10 001.966 km (average)
earth flattening [a]	e	$1/298.25231(1)$
earth's av. density	ρ	$5.5\ \mathrm{Mg/m^3}$
moon's radius	R_{mv}	1738 km in direction of earth
moon's radius	R_{mh}	17.. km in other two directions
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]		avg. $0.5181° = 31.08'$, min. $0.49°$, max. $0.55°$
moon's av. density	ρ	$3.3\ \mathrm{Mg/m^3}$
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; minimum on 4th of July (aphelion) $1888''$, maximum on 4th of January (perihelion) $1952''$
suns's av. density	ρ	$1.4\ \mathrm{Mg/m^3}$
sun's distance, average	AU	$149\,597\,870.691(30)$ km
solar velocity around centre of galaxy	$v_{\odot \mathrm{g}}$	$220(20)$ km/s
solar velocity against cosmic background	$v_{\odot \mathrm{b}}$	$370.6(5)$ 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. The International Geodesic Union has refined the data in 2000. The radii and the flattening given here are those for the 'mean tide system'. They differ from those of the 'zero tide system' and other systems by about 0.7 m. The details are a science by its own.

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 $= 1^\circ = \pi/180\,\mathrm{rad}$, 1 (first) minute $= 1' = 1^\circ/60$, 1 second (minute) $= 1'' = 1'/60$. The ancient units 'third minute' and 'fourth minute', each $1/60$th 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. 918

If the number π were *normal*, i.e. if all digits and digit combinations would appear with the same probability, then every text written or to be written, as well as every word spoken or to be spoken, can be found coded in its sequence. The property of normality has not yet been proven, even though it is suspected to be true. What is the significance? Is all wisdom encoded in the simple circle? No. The property is nothing special, as it also applies to the number 0.123456789101112131415161718192021... and many others. Can you specify a few?

Challenge 1220 n

By the way, in the graph of the exponential function e^x, the point $(0,1)$ is the only one with two rational coordinates. If you imagine to paint in blue all points on the plane with two rational coordinates, the plane would look quite bluish. Nevertheless, the graph goes only through one of these points and manages to avoid all the others.

References

899 *Le Système International d'Unités*, Bureau International des Poids et Mesures, Pavillon de Breteuil, Parc de Saint Cloud, 92 310 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 872.

900 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 Tony Jones, *Splitting the Second*, Institute of Physics Publishing, 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 225.

On electric current measurements, see ... On precision temperature measurements, see page 196. Cited on page 873.

901 The unofficial prefixes have been originally proposed in the 1990s by Jeff K. Aronson, professor at the University of Oxford, and are slowly coming into general usage. Cited on page 874.

902 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 878.

903 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 878.

904 G. CHARPAK & R.L. GARWIN, *The DARI*, Europhysics News **33**, pp. 14–17, January/February 2002. Cited on page 879.

The answer should lie between one or two dozen kilometers, assuming ideal atmospheric circumstances.

905 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 879.

906 A. ZEILINGER, *The Planck stroll*, American Journal of Physics **58**, p. 103, 1990. Cited on page 879.

907 The most precise clock ever built is ... Cited on page 880.

908 J. BERGQUIST, editor, *Proceedings of the Fifth Symposium on Frequency Standards and Metrology*, World Scientific, 1997. Cited on page 880.

909 About short lifetime measurements, see e.g. the paper on D particle lifetime ... Cited on page 880.

910 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 880.

911 About the detection of gravitational waves, see ... Cited on page 880.

912 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 880.

913 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 196. Cited on page 880.

914 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 881.

915 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 881.

916 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 pages 881 and 882.

917 The details are given in the well-known astronomical reference, P. KENNETH SEIDELMANN, *Explanatory Supplement to the Astronomical Almanac*, 1992. Cited on page 884.

918 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

$$\pi + 3 = \sum_{n=1}^{\infty} \frac{n\,2^n}{\binom{2n}{n}} \qquad (617)$$

or the beautiful formula discovered in 1996 by Bailey, Borwein, and Plouffe

$$\pi = \sum_{n=0}^{\infty} \frac{1}{16n} \left(\frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6} \right) \quad . \qquad (618)$$

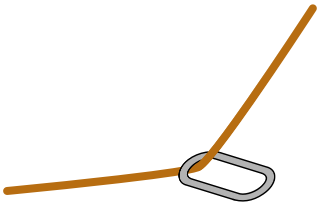
The site also explains the newly discovered methods to calculate specific binary digits of π without having to calculate all the preceding ones. By the way, the number of (consecutive) digits known in 1999 was over 1.2 million million, as told in Science News **162**, 14 December 2002. They pass all tests for a random string of numbers, as the http://www.ast.univie.ac.at/~wasi/PI/pi_normal.html web site explains. However, this property, called *normality*, has never been proven; it is the biggest open question about π. It is possible that the theory of chaotic dynamics will lead to a solution of this puzzle in the coming years.

Another method to calculate π and other constants was discovered and published by 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 an industrial supercomputer. New formulas to calculate π are still irregularly discovered.

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 it is not.) Do you want to become a mathematician? Challenge1221 h Challenge 1222 n

Cited on page 886.



Typeset in
January 2003

Appendix C Particle Properties

The following table gives the overview of the known and predicted elementary particles. There have been no changes in it since the mid 1970s.

Radiation	electromagnetic interaction	weak interaction	strong interaction
	$\boxed{\gamma}$	$\boxed{W^{+}}$, $\boxed{W^{-}}$ $\boxed{Z^{o}}$	$\boxed{g_1}$ .. $\boxed{g_8}$
	photon	intermediate vector bosons	gluons
	Radiation particles are bosons with spin 1. W^{-} is antiparticle of W^{+}; all others are their own antiparticles.		
Matter	generation 1	generation 2	generation 3
Leptons	$\boxed{\mathrm{e}}$ $\boxed{\nu_e}$	$\boxed{\mu}$ $\boxed{\nu_\mu}$	$\boxed{\tau}$ $\boxed{\nu_\tau}$
Quarks (each in three colours)	$\boxed{\mathrm{u}}$ $\boxed{\mathrm{d}}$	$\boxed{\mathrm{s}}$ $\boxed{\mathrm{c}}$	$\boxed{\mathrm{t}}$ $\boxed{\mathrm{b}}$
	Matter particles are fermions with spin $1/2$; all have a corresponding antiparticle.		
Hypothetical matter and radiation			
Higgs boson	$\boxed{\mathrm{H}}$	predicted to have spin 0 and to be elementary	
Supersymmetric partners	$\boxed{\mathrm{se}}$ ... $\boxed{\mathrm{h}}$	one partner for each of the above	

Table 68 The elementary particles

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 also allows to deduce a complete characterization of the intrinsic properties of any *composed* moving entity, be it an object or an image.

Table 69 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	$0\ (< 6\cdot10^{-16}\,\mathrm{eV}/c^2)$	stable	$I(J^{PC}) = 0,1(1^{--})$	000 000	$0,0,1$
$\mathrm{W}^{\pm}$	$80.75(64)\ \mathrm{GeV}/c^2$	$2.06(6)$ GeV 67.8(1.0)% hadrons, 32.1(2,0)% $l^+\nu$	$J=1$	$\pm$100 000	$0,0,1$
Z	$91.187(7)\ \mathrm{GeV}/c^2$	$2.65(1)\cdot10^{-25}\,\mathrm{s}$ 69.90(15)% hadrons	$J=1$	000 000	$0,0,1$
gluon	0	stable	$I(J^P) = 0(1^-)$	000 000	$0,0,1$
elementary matter (fermions): leptons					
electron e	$9.109\,381\,88(72)\cdot10^{-31}$ kg $= 81.871\,0414(64)\ \mathrm{pJ}/c^2$ $= 0.510\,998\,902(21)\ \mathrm{MeV}/c^2 = 0.000\,548\,579\,9110(12)$ u gyromagnetic ratio $g = \mu_e/\mu_B = -1.001\,159\,652\,1883(42)$ electric dipole moment $d = (-0.3\pm0.8)\cdot10^{-29}e$ m	$> 13\cdot10^{30}$ s	$J=\frac{1}{2}$	$-100\,000$	$1,0,1$
muon μ	$0.188\,353\,109(16)$ yg $= 105.658\,3568(52)\ \mathrm{MeV}/c^2 = 0.113\,428\,9168(34)$ u gyromagnetic ratio $g = \mu_\mu/(e\hbar/2m_\mu) = -1.001\,165\,916\,02(64)$ electric dipole moment $d = (3.7\pm3.4)\cdot10^{-22}e$ m	$2.197\,03(4)\ \mu$s 99% $e^-\bar{\nu}_e\nu_\mu$	$J=\frac{1}{2}$	$-100\,000$	$1,0,1$
tau τ	$1.777\,05(29)\ \mathrm{GeV}/c^2$	$290.0(1.2)$ fs	$J=\frac{1}{2}$	$-100\,000$	$1,0,1$
el. neutrino ν_e	$< 7.2\ \mathrm{eV}/c^2$		$J=\frac{1}{2}$		$1,0,1$
muon neutrino ν_μ	$< 0.17\ \mathrm{MeV}/c^2$		$J=\frac{1}{2}$		$1,0,1$

---

* The official reference for all this data, worth a look by every physicist, is the massive collection by the particle data group, with the web site http://pdg.web.cern.ch/pdg containing the most recent information.

A printed review is published about every two years with updated data in one of the large journals on elementary particle physics. See for example C. CASO & al., The European Physical Journal C **3**, p. 1, 1998. For measured properties of these particles, the official reference is the set of CODATA values. The most recent list was 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
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.
- Ref. 899 The electron radius is below 10^{-22} m.
- Ref. 900 It is possible to store single electrons in traps for many months.
- See also the table of SI prefixes on page 873. About the eV/c^2 mass unit, see page 878.
- 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 non-vanishing 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 $1\,\mathrm{u} = \frac{1}{12} m(^{12}\mathrm{C}) = 1.660\,5402(10)\,\mathrm{yg}$.
- See page 662 for the precise definition and meaning of the quark masses.
- The electric polarizability is defined on page 410; it is predicted to vanish for all elementary particles.

Using the table of elementary particle properties, together with the standard model and the fundamental constants, in principle *all* properties of composite matter and radiation can be deduced, including all those encountered in everyday life. (Can you explain how the size of an object follows from them?) In a sense, this table contains the complete results of the study of matter and radiation, such as material science, chemistry, and biology.

Challenge 1224 n

The most important examples examples of composites are grouped in the following table.

Table 70 Properties of selected composites

Composite	mass m, quantum numbers	lifetime τ, main decay modes	size (diameter)
mesons (hadrons, bosons) out of the over 130 types known			
pion $\pi^0 (u\bar{u} - d\bar{d})/\sqrt{2}$	$134.976\,4(6)\,\mathrm{MeV}/c^2$ $I^G(J^{PC}) = 1^-(0^{-+})$, $S = C = B = 0$	$84(6)$ as, 2γ 98.798(32)%	~ 1 fm
pion $\pi^+ (u\bar{d})$	$139.569\,95(35)\,\mathrm{MeV}/c^2$ $I^G(J^P) = 1^-(0^-)$, $S = C = B = 0$	$26.030(5)$ ns, $\mu^+\nu_\mu$ 99.987 7(4)%	~ 1 fm
kaon K^o_S	$m_{K^o_S}$	$89.27(9)$ ps	
kaon K^o_L	$m_{K^o_S} + 3.491(9)\,\mu\mathrm{eV}/c^2$	$51.7(4)$ ns	
kaon $K^\pm$ $(u\bar{s}, \bar{u}s)$	$493.677(16)\,\mathrm{MeV}/c^2$	$12.386(24)$ ns, $\mu^+\nu_\mu$ 63.51(18)% 21.16(14)% $\pi^+\pi^o$	
kaon K^o $(d\bar{s})$ (50%K_S, 50% K_L)	$497.672(31)\,\mathrm{MeV}/c^2$	n.a.	~ 1 fm
kaons $K^\pm, K^o, K^o_S, K^o_L$	$I(J^P) = \frac{1}{2}(0^-)$, $S = \pm 1$, $B = C = 0$		
baryons (hadrons, fermions) out of the over 100 types known			
proton p or N^+ (uud)	$1.672\,621\,58(13)$ yg $= 1.007\,276\,466\,88(13)$ u $= 938.271\,998(38)\,\mathrm{MeV}/c^2$ $I(J^P) = \frac{1}{2}(\frac{1}{2}^+)$, $S = 0$ gyromagnetic ratio $\mu_p/\mu_N = 2.792\,847\,337(29)$ electric dipole moment $d = (-4 \pm 6) \cdot 10^{-26} e$ m electric polarizability $\alpha = 12.1(0.9) \cdot 10^{-4}\,\mathrm{fm}^3$ magnetic polarizability $\alpha = 2.1(0.9) \cdot 10^{-4}\,\mathrm{fm}^3$	$\tau_{\mathrm{total}} > 1.6 \cdot 10^{25}$ a, $\tau(p \rightarrow e^+\pi^o) > 5.5 \cdot 10^{32}$ a	$0.89(1)$ fm Ref. 901
neutron n or N^o (udd)	$1.674\,927\,16(13)$ yg $= 1.008\,664\,915\,78(55)\,\mathrm{u} = 939.565\,330(38)\,\mathrm{MeV}/c^2$	$887.0(2.0)$ s, $pe^-\bar{\nu}_e$ 100 %	~ 1 fm

Composite	mass m, quantum numbers	lifetime τ, main decay modes	size (diameter)
	$I(J^P)=\frac{1}{2}(\frac{1}{2}^+)$, $S=0$		
	gyromagnetic ratio $\mu_n/\mu_N=-1.913\,042\,72(45)$		
	electric dipole moment $d_n=(-3.3\pm4.3)\cdot10^{-28}e$ m		
	electric polarizability $\alpha=0.98(23)\cdot10^{-3}\,\mathrm{fm}^3$		
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 known isotopes Ref. 902			
hydrogen (^{1}H) [lightest]	$1.007\,825\,032(1)\,\mathrm{u}=1.6735\,\mathrm{yg}$		$2\cdot30\,\mathrm{pm}$
antihydrogen			
helium (^{4}He) [smallest]	$4.002\,603250(1)\,\mathrm{u}=6.6465\,\mathrm{yg}$		$2\cdot32\,\mathrm{pm}$
carbon (^{12}C)	$12\,\mathrm{u}=19.926\,482(12)\,\mathrm{yg}$		$2\cdot77\,\mathrm{pm}$
bismuth $^{209}_{83}$Bi [shortest living and rarest]	209 u	0.1 ps	
tantalum ^{180}Ta [longest living radioactive]	180 u	$>10^{15}$ a Ref. 903	
francium [largest of all]	223 u	...	$2\cdot0.28\,\mathrm{nm}$
atom 116 [heaviest of all]	289 u	...	
molecules out of the over 10^7 known types			
hydrogen (H_2)	~2 u	$>10^{25}$ a	
water (H_2O)	~18 u	$>10^{25}$ a	
ATP (adenosinetriphosphate)	360 u	$>10^{10}$ a	ca. 3 nm
human Y chromosome	… ag	$>10^6$ a	ca. 50 mm
other composites			
whale nerve cell	~100 g	~50 a	20 m
cell (red blood)	1 ng	4-100 days	$\sim10\,\mu$m
cell (sperm)	10 pg	not fecundated: ~5 d	length 60 μm, head 3 μm times 5 μm
cell (ovule)	1 μg	fecundated: over 4000 million years	$\sim120\,\mu$m
cell (E. Coli)	1 pg	4000 million years	body: 2 μm
adult human	$35\,\mathrm{kg}<m<350\,\mathrm{kg}$	$\tau\approx2.5\cdot10^9$ s Ref. 904 $\approx$ 600 million breaths $\approx$ 2 500 million heartbeats $\lesssim122$ a, 60% H_2O and 40% dust	~1.7 m
largest living thing	10^5 kg	ca. 1000 a	~1 km

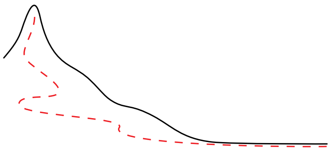
Composite	mass m, quantum numbers	lifetime τ, main decay modes	size (diameter)
larger composites see table on page 140.			

Notes (see also those of the previous table)

- G parity is defined only for mesons and given by $G = (-1)^{L+S+I} = (-1)^I \cdot C$.
- Neutrons bound in nuclei have a lifetime of at least 10^{20} years.
- The $f_o(1500)$ resonance is now accepted as a glueball and thus as a *radiation composite* by the high energy community, as announced at the HEO conference in Warsaw in 1996.
- In 2002, first evidence for the existence of tetra-neutrons was published by a French group.
- The number of existing molecules is several orders of magnitude larger than the number of analysed and listed molecules.
- Some nuclei are not yet discovered; in 2002 the known nuclei range from 1 to 116, but 113, 115 and 117 are still missing.
- The first *anti-atoms*, made of antielectrons and antiprotons, have been made in January 1996 at CERN in Geneva. All properties for antimatter checked so far are consistent with the predictions. Ref. 905
- 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 $\mathbf{r} \to -\mathbf{r}$. For mesons, it is connected through $P = (-1)^{L+1}$ with the orbital angular momentum L.

The most important matter composites are the *atoms*. Their size, structure and interactions determine the properties and colour of everyday objects. Atom types, also called *elements* in chemistry, are most efficiently grouped in the so-called *periodic table*, in which atoms behaving similarly are neighbours. The table results from the various ways that protons, neutrons and electrons can combine to form aggregates.

There are similar tables also for the mesons (made of any two quarks) and the baryons (made of three quarks). Neither the meson nor the baryon table is included here; they can be found in the cited *Review of particle physics*. In fact, the baryon table still has a number of vacant spots. However, the missing particles are extremely heavy and short lived (which means expensive to make and detect) and their discovery is not expected to yield deep new insights.



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	I	II	IIIa	IVa	Va	VIa	VIIa		VIIIa		Ia	IIa	III	IV	V	VI	VII	VIII
Period																		
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	∗	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	∗∗	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Uun	111 Uuu	112 Uub	113	114 Uuq	115	116 Uuh	117	118
Lanthanoids			∗	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
Actinoids			∗∗	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr

Table 71 Periodic table of the elements known in 2002, with their atomic numbers

The atomic number gives the number of protons and of electrons found in an atom of a given element. This number also specifies each element in its chemical behaviour. Most – but not all – elements up to 92 are found in nature; the others are artificially produced in the laboratories. The last element discovered is element 116.

Elements in the same *group* behave similarly in chemical reactions. The *periods* define the repetition of these similarities. Extensive physical and chemical data is available for every element. More elaborate periodic tables can be found on the chemlab.pc.maricopa.edu/periodic/stowetable.html web site. Ref. 908

Group 1 are the alkali metals, group 2 the earth-alkali metals. Actinoids, lanthanoids, and groups 3 to 13 are metals; in particular, groups 3 to 12 are heavy metals. The elements of group 16 are called *chalkogens*, i.e. ore-formers; group 17 are the *halogens*, i.e. the salt-formers, and group 18 are the *noble gases*, which do not form (almost) any chemical compounds. Groups 13, 14 and 15 contain metals, semimetals, liquids, and gases; they are of importance for the appearance of life.

Many elements exist in versions with different numbers of neutrons in their nucleus and thus with different mass; these various *isotopes* – called this way because they are found on the *same place* in the periodic table – behave identically in chemical reactions. There are over 2000 of them. Ref. 902, 906

Table 72 The elements and their main properties

Name	symbol	at. num.	average mass[a] in u (with error) and longest lifetime	atomic[e] radius in pm	main properties (naming) use, and discovery date
Actinium[b]	Ac	89	(227.0277(1)) 21.77(2) a	(188)	radioactive metal (Greek ‘aktis’ ray) 1899
Aluminium	Al	13	26.981 538(2) stable	118c, 143m	light metal (Latin ‘alumen’ ...) 1827
Americium[b]	Am	95	(243.0614(1)) 7.37(2) ka	(184)	radioactive metal (Italian ‘America’) 1945
Antimony	Sb	51	121.760(1)[f] stable	137c, 159m, 205v	semimetal (via Arabic from Latin stibium, itself from Greek, Egyptian for one of its minerals) colours rubber, used in medicines, constituent of enzymes
Argon	Ar	18	39.948(1)[f] stable	(71n)	noble gas, (Greek ‘argos’ inactive from ‘anergos’ without energy) 1894, third component of air, used for welding, in lasers
Arsenic	As	33	74.921 60(2) stable	120c, 185v	poisonous semimetal (Greek ‘arsenikon’ tamer of males) antiquity, for poisoning pigeons and doping semiconductors
Astatine[b]	At	85	(209.9871(1)) 8.1(4) h	(140)	radioactive halogen, (Greek ‘astatos’ unstable) 1940, no use
Barium	Ba	56	137.327(7) stable	224m	(Greek ‘bary’ heavy) 1808
Berkelium[b]	Bk	97	(247.0703(1)) 1.4(3) ka	n.a.	(Berkeley, US town) 1949
Beryllium	Be	4	9.012 182(3) stable	106c, 113m	(Greek ‘beryllos’, a mineral) 1797
Bismuth	Bi	83	208.980 38(2) stable	170m, 215v	(Latin via German ‘weisse Masse’ white mass) 1753
Bohrium[b]	Bh	107	(264.12(1)) 0.44 s[g]	n.a.	(after Niels Bohr) 1981
Boron	B	5	10.811(7)[f] stable	83c	(Latin borax, from Arabic and Persian for brilliant) 1808
Bromine	Br	35	79.904(1) stable	120c, 185v	(Greek ‘bromos’ strong odour) 1826
Cadmium	Cd	48	112.411(8)[f] stable	157m	(Greek kadmeia, a mineral) 1817
Caesium	Cs	55	132.905 45(2) stable	273m	(Latin ‘caesius’ sky blue) 860
Calcium	Ca	20	40.078(4)[f] stable	197m	(Latin ‘calcis’ chalk)
Californium[b]	Cf	98	(251.0796(1)) 0.90(5) ka	n.a.	(Latin ‘calor’ heat and ‘fornicare’ have sex, the land of hot sex) 1950
Carbon	C	6	12.0107(8)[f] stable	77c	makes up coal and diamond (Latin ‘carbo’ coal)

Name	sym-bol	at. num.	average mass[a] in u (with error) and longest lifetime	atomic[e] radius in pm	main properties (naming) use, and discovery date
Cerium	Ce	58	140.116(1)[f] stable	183m	(after asteroid Ceres, roman goddess) 1803
Chlorine	Cl	17	35.453(2)[f] stable	102c, 175v	green gas (Greek 'chloros' yellow-green) 1774
Chromium	Cr	24	51.9961(6) stable	128m	(Greek 'chromos' colour) 1797
Cobalt	Co	27	58.933 200(9) stable	125m	metal (German 'Kobold' goblin) 1694; part of vitamin B_{12}
Copper	Cu	29	63.546(3)[f] stable	128m	red metal (Latin cuprum from Cyprus island) antiquity, part of many enzymes
Curium[b]	Cm	96	(247.0704(1)) 15.6(5) Ma	n.a.	(after Pierre and Marie Curie) 1944
Dubnium[b]	Db	105	(262.1141(1)) 34(5) s	n.a.	('Dubna' Russian city) 1967
Dysprosium	Dy	66	162.50(3)[f] stable	177m	(Greek 'dysprositos' difficult to obtain) 1886
Einsteinium[b]	Es	99	(252.0830(1)) 472(2) d	n.a.	(after Albert Einstein) 1952
Erbium	Er	68	167.259(3)[f] stable	176m	('Ytterby' Swedish town) 1843
Europium	Eu	63	151.964(1)[f] stable	204m	1901, used in red tv screen colour
Fermium[b]	Fm	100	(257.0901(1)) 100.5(2) d	n.a.	(after Enrico Fermi) 1952
Fluorine	F	9	18.998 4032(5) stable	62c, 147v	(from flourine, a mineral, from Greek 'fluo' flow) 1886
Francium[b]	Fr	87	(223.0197(1)) 22.0(1) min	(278)	(from France) 1939
Gadolinium	Gd	64	157.25(3)[f] stable	180m	(after Johan Gadolin) 1880
Gallium	Ga	31	69.723(1) stable	125c, 141m	(Latin for both the discoverer's name and his nation, France) 1875
Germanium	Ge	32	72.64(1) stable	122c, 195v	(from Germania, as opposed to gallium) 1886
Gold	Au	79	196.966 55(2) stable	144m	heavy noble metal (Latin aurum) antiquity, electronics, jewels
Hafnium	Hf	72	178.49(2)[c] stable	158m	(Latin for Copenhagen) 1923
Hassium[b]	Hs	108	(277) 16.5 min[g]	n.a.	(Latin form of German state Hessen) 1984
Helium	He	2	4.002 602(2)[f] stable	(31n)	noble gas (Greek 'helios' sun) where it was discovered 1895
Holmium	Ho	67	164.930 32(2) stable	177m	(Stockholm, Swedish capital) 1878

Name	sym-bol	at. num.	average mass[a] in u (with error) and longest lifetime	atomic[e] radius in pm	main properties (naming) use, and discovery date
Hydrogen	H	1	1.007 94(7)[f] stable	30c	gas (Greek for water-former) 1766
Indium	In	49	114.818(3) stable	141c, 166m	soft metal (Greek 'indikon' indigo) 1863
Iodine	I	53	126.904 47(3) stable	140c, 198v	(Greek 'iodes' violet) 1811
Iridium	Ir	77	192.217(3) stable	136m	(Greek 'iris' rainbow) 1804
Iron	Fe	26	55.845(2) stable	127m	(Latin ferrum) antiquity
Krypton	Kr	36	83.80(1)[f] stable	(88n)	noble gas (Greek 'kryptos' hidden) 1898, used in lasers
Lanthanum	La	57	138.9055(2)[c,f] stable	188m	(Greek 'lanthanein' to be hidden) 1839
Lawrencium[b]	Lr	103	(262.110 97(1)) 3.6(3) h	n.a.	(after Ernest Lawrence) 1961
Lead	Pb	82	207.2(1)[c,f] stable	175m	poisonous heavy metal (Latin plumbum) antiquity, used in car batteries
Lithium	Li	3	6.941(2)[f] stable	156m	(Greek 'lithos' stone) 1817
Lutetium	Lu	71	174.967(1)[f] stable	173m	(Latin 'Lutetia' , old name of Paris) 1907
Magnesium	Mg	12	24.3050(6) stable	160m	light metal, (from Magnesia, a Greek district in Thessalia) 1755
Manganese	Mn	25	54.938 049(9) stable	126m	(Italian 'Manganese', a mineral) 1774
Meitnerium[b]	Mt	109	(268.1388(1)) 0.070 s[g]	n.a.	(after Lise Meitner) 1982
Mendelevium[b]	Md	101	(258.0984(1)) 51.5(3) d	n.a.	(after Dimitri Ivanovitch Mendeleiev) 1955
Mercury	Hg	80	200.59(2) stable	157m	liquid metal (Greek 'hydrargyrum' liquid silver) antiquity
Molybdenum	Mo	42	95.94(1)[f] stable	140m	(Greek 'molybdos' lead) 1788
Neodymium	Nd	60	144.24(3)[c,f] stable	182m	(Greek 'neos' and 'didymos' new twin) 1885
Neon	Ne	10	20.1797(6)[f] stable	(36n)	noble gas (Greek 'neos' new) 1898
Neptunium[b]	Np	93	(237.0482(1)) 2.14(1) Ma	n.a.	(planet Neptune, after Uranus) 1940
Nickel	Ni	28	58.6934(2) stable	125m	metal (German 'Nickel' goblin) 1751
Niobium	Nb	41	92.906 38(2) stable	147m	(Greek 'Niobe', mythical daughter of Tantalos) 1801

Name	sym-bol	at. num.	average mass[a] in u (with error) and longest lifetime	atomic[e] radius in pm	main properties (naming) use, and discovery date
Nitrogen	N	7	14.0067(2)[f] stable	70c, 155v	gas (Greek for nitre former) 1772
Nobelium[b]	No	102	(259.1010(1)) 58(5) min	n.a.	(after Alfred Nobel) 1958
Osmium	Os	76	190.23(3)[f] stable	135m	heavy metal (from Greek 'osme' odour) 1804
Oxygen	O	8	15.9994(3)[f] stable	66c, 152v	gas (formed from Greek to mean 'acid former') 1774
Palladium	Pd	46	106.42(1)[f] stable	138m	heavy metal (from asteroid 'Pallas' after the Greek goddess) 1802
Phosphorus	P	15	30.973 761(2) stable	109c, 180v	(Greek 'phosphoros' light bearer) 1669
Platinum	Pt	78	195.078(2) stable	139m	heavy metal (Spanish 'platina' little silver) 1735
Plutonium	Pu	94	(244.0642(1)) 80.0(9) Ma	n.a.	(after the planet) 1940
Polonium	Po	84	(208.9824(1)) 102(5) a	(140)	(from Poland) 1898
Potassium	K	19	39.0983(1) stable	238m	(Latin 'kalium' from Arabic 'quilyi', a plant used to produce potash, German 'Pottasche') 1807
Praseodymium	Pr	59	140.907 65(2) stable	183m	(Greek 'praesos didymos' green twin) 1885
Promethium[b]	Pm	61	(144.9127(1)) 17.7(4) a	181m	(from the Greek mythical figure of Prometheus) 1945
Protactinium	Pa	91	(231.035 88(2)) 32.5(1) ka	n.a.	radioactive (Greek 'protos' first, as it decays into Actinium) 1917, found in nature
Radium	Ra	88	(226.0254(1)) 1599(4) a	(223)	(Latin 'radius' ray) 1898
Radon	Rn	86	(222.0176(1)) 3.823(4) d	(130n)	radioactive noble gas
Rhenium	Re	75	186.207(1)[c] stable	138m	(Latin 'rhenus' for Rhine river) 1925
Rhodium	Rh	45	102.905 50(2) stable	135m	metal (Greek 'rhodon' rose) 1803
Rubidium	Rb	37	85.4678(3)[f] stable	255m	(Latin 'rubidus' red) 1861
Ruthenium	Ru	44	101.107(2)[f] stable	134m	(Latin 'Rhuthenia' for Russia) 1844
Rutherfordium[b]	Rf	104	(261.1088(1)) 1.3 min[g]	n.a.	(after Ernest Rutherford) 1964
Samarium	Sm	62	150.36(3)[c,f] stable	180m	(from the mineral Samarskite, after Wassily Samarski) 1879, used in magnets

Name	sym-bol	at. num.	average mass^a in u (with error) and longest lifetime	atomic^e radius in pm	main properties (naming) use, and discovery date
Scandium	Sc	21	44.955 910(8) stable	164m	(from Latin ‘Scansia’ Sweden) 1879
Seaborgium^b	Sg	106	266.1219(1) 21 s^g	n.a.	(after Glenn Seaborg) 1974
Selenium	Se	34	78.96(3)^f stable	120c, 190v	(Greek ‘selene’ moon) 1818
Silicon	Si	14	28.0855(3)^f stable	105c, 210v	semiconductor (Latin ‘silex’ pebble) 1823, earth’s crust and electronics
Silver	Ag	47	107.8682(2)^f stable	145m	metal, (Latin argentum, Greek ‘argyros’) antiquity, used in photography
Sodium	Na	11	22.989 770(2) stable	191m	light metal (Egyptian, Arabic ‘natrium’ and Arabic ‘souwad’ soda) component of many salts
Strontium	Sr	38	87.62(1)^f stable	215m	(Strontian, Scottish town) 1790
Sulphur	S	16	32.065(5)^f stable	105c, 180v	yellow solid (Latin) antiquity
Tantalum	Ta	73	180.9479(1) stable	147m	metal (Greek Tantalos, a mythical figure) 1802
Technetium^b	Tc	43	(97.9072(1)) $6.6(10)$ Ma	136m	(Greek ‘technetos’ artificial) 1939
Tellurium	Te	52	127.60(3)^f stable	139c, 206v	(Latin ‘tellus’ earth) 1783
Terbium	Tb	65	158.925 34(2) stable	178m	(’Ytterby’ Swedish town) 1843
Thallium	Tl	81	204.3833(2) stable	172m	poisonous heavy metal (Greek ‘thallos’ branch) 1861
Thorium	Th	90	232.0381(1)^{d,f} $14.0(1)$ Ga	180m	radioactive (nordic god Thor, as in thursday) 1828, found in nature
Thulium	Tm	69	168.934 21(2) stable	175m	(Thule, mythical name for Scandinavia) 1879
Tin	Sn	50	118.710(7)^f stable	139c, 210v, 162m	(Latin stannum)
Titanium	Ti	22	47.867(1) stable	146m	(Greek Titanos) 1791
Tungsten	W	74	183.84(1) stable	141m	highest melting, heavy metal (German Wolfram, Swedish ‘tung sten’ heavy stone) 1783, light bulbs
Ununnilium^b	Uun	110	(281) 1.6 min^g	n.a.	1994, no use
Unununium^b	Uuu	111	(272.1535(1)) 1.5 ms^g	n.a.	1994 no use
Ununbium^b	Uub	112	(285) 15.4 min^g	n.a.	1996, no use
Ununtrium	Uut	113		n.a.	not yet observed
Ununquadium^b	Uuq	114	(289) 30.4 s^g	n.a.	1999, no use
Ununpentium	Uup	115		n.a.	not yet observed
Ununhexium^b	Uuh	116	(289) 0.6 ms^g	n.a.	1999, no use

Name	sym-bol	at. num.	average mass[a] in u (with error) and longest lifetime	atomic[e] radius in pm	main properties (naming) use, and discovery date
Ununseptium	Uus	117		n.a.	not yet observed
Ununoctium	Uuo	118		n.a.	not yet observed, but false claim in 1999
Uranium	U	92	238.028 91(3)[d,f] $4.468(3) \cdot 10^9$ a	156m	radioactive (Planet Uranos, the Greek sky god) 1789, found in nature, used for nuclear energy
Vanadium	V	23	50.9415(1) stable	135m	metal (‘Vanadis’ scandinavian goddess of beauty) 1830, used in steel
Xenon	Xe	54	131.293(6)[f] stable	(103n) 200v	noble gas (Greek ’xenos’ foreign) 1898, used in lamps and lasers
Ytterbium	Yb	70	173.04(3)[f] stable	174m	malleable heavy metal (’Ytterby’ Swedish town) 1878, used in superconductors
Yttrium	Y	39	88.905 85(2) stable	180m	malleable light metal (’Ytterby’ Swedish town) 1794
Zinc	Zn	30	65.409(4) stable	139m	heavy metal (German ’Zinke’ protuberance) antiquity, iron rust protection
Zirconium	Zr	40	91.224(2)[f] stable	160m	heavy metal (from the mineral zircon, after Arabic ‘zargum’ golden colour) 1789, chemical and surgical instruments

a. The atomic mass unit is defined as $1\,\mathrm{u} = \frac{1}{12} m(^{12}\mathrm{C}) = 1.660\,5402(10)\,\mathrm{yg}$. For elements found on earth, the *average* atomic mass for the natural occurring isotope mixture is given, with the error in the last digit in brackets. For elements not found on earth, the mass of the *longest living* isotope is given; as it is not an average, it is written in brackets, as usual in this domain. Ref. 906

b. The element is not found on earth due to its short lifetime.

c. The element contains at least one radioactive isotope.

d. The element has no stable isotopes.

e. The *atomic radius* does not exist. In fact, all atoms being clouds, they have no boundary. The size of atoms thus depends on the way it is defined. Several definitions are possible. Usually, the radius is defined in a way to be useful for the estimation of distances between atoms. This distance is different for different bond types. Radii for metallic bonds are labelled m, radii for (single) covalent bonds with carbon c, and Van der Waals radii v. Noble gas radii are denoted n. Note that values found in the literature vary by about 10 %; values in brackets lack literature references. Ref. 907 Ref. 907

The covalent radius is often up to 0.1 nm smaller than the metallic radius for elements on the (lower) left of the periodic table; on the (whole) right side it is essentially equal to the metallic radius. In between, the difference between the two decreases towards the right. Are you able to explain why? By the way, ionic radii differ considerably from atomic ones, and depend both on the ionic charge and the element itself. Challenge 1225 n

All these values are for atoms in their ground state. Excited atoms can be one hundreds of times larger than atoms in the ground state; however, excited atoms do not form solids or chemical compounds.

f. The isotopic composition and thus the average atomic mass of the element varies depending on the mining place or on subsequent human treatment, and can lie outside the values given. For example, the atomic mass of commercial lithium ranges between 6.939 and 6.996 u. The masses of isotopes

are known in atomic mass units with nine or more significant digits, and usually with one or two digits less in kilogram. The errors in the atomic mass is thus mainly to the variations in isotope composition. For a precise isotope mass list, see the http://csnwww.in2p3.fr web site.
g. The lifetime errors are asymmetric or not well known.

References

899 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 891.

900 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 891.

901 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 892.

902 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 pages 893 and 896.

903 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 893.

904 Stephen J. Gould, 1980, *The Panda's thumb*, W.W. Norton & Co., New York. This is one of several interesting and informative books on evolutionary biology by the best writer in the field. Cited on page 893.

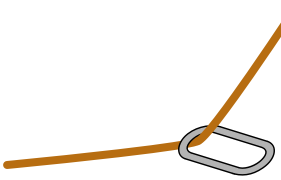
905 For a good review, see the article by P.T. Greenland, *Antimatter*, Contemporary Physics **38**, pp. 181–203, 1997. Cited on page 894.

906 The atomic masses, as given by IUPAC, can be found in Pure and Applied Chemistry **73**, pp. 667–683, 2001, or on the www.iupac.org web site. For an isotope mass list, see the http://csnwww.in2p3.fr web site. Cited on pages 896 and 902.

907 The metallic, covalent and Van-der-Waals radii are from Nathaniel W. Alcock, *Bonding and Structure*, Ellis Horwood 1999. The text also explains in detail how the radii are defined and measured. Cited on page 902.

908 Almost everything known about each element and its chemistry can be found in the encyclopedic Gmelin, *Handbuch der anorganischen Chemie*, published from 1817 onwards. It presently has over 500 volumes, now all published in English under the title *Handbook of Inorganic and Organometallic Chemistry*, and contains at least one volume dedicated to each chemical element. Cited on page 896.

Typeset in
January 2003



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. 899

In particular, we shall 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 amathematician. Challenge 1226
> 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.

Complex numbers

Complex numbers are defined by $z = a + ib$. The generators of the complex numbers, 1 and i, obey the well known algebra

$$\begin{array}{c|cc} \cdot & 1 & i \\ \hline 1 & 1 & i \\ i & i & -1 \end{array} \qquad (619)$$

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. From $|wz| = |w|\,|z|$ follows the *two-squares theorem*

$$(a_1^2 + a_2^2)(b_1^2 + b_2^2) = (a_1 b_1 - a_2 b_2)^2 + (a_1 b_2 + a_2 b_1)^2 \qquad (620)$$

valid for all real numbers a_i, b_i. It was already known, in its version for integers, to Diophantus of Alexandria.

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. Therefore, translating the definition of multiplication into geometrical language allows to
Challenge 1227 rapidly prove geometrical theorems,such as the one of Figure 259.

Complex numbers can also be represented as 2 × 2 matrices of the form

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \quad \text{with} \quad a, b \in \mathbf{R} \quad . \qquad (621)$$

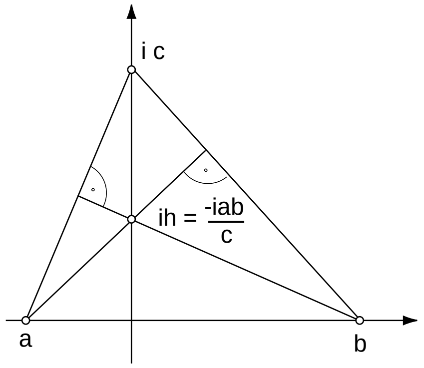


Figure 259 A property of triangles easily provable with complex numbers

Usual matrix addition and multiplication then give the same result as complex addition and multiplication. In this way, complex numbers can be represented by a special type of real
Challenge 1228 matrices. What is $|z|$ in matrix language?

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 $az^2 + bz + c = 0$
Challenge 1229 for an unknown z always have two solutions.

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 use complex numbers to describe quantities with phases, such as alternating currents or electrical fields in space.

By the way, there are as many complex numbers as there are real numbers. Are you able
Challenge 1230 to show this?

> 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 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 three-dimensional vectors. Ref. 900

The quaternions $\mathbf{H}$ form a 4-dimensional algebra over the reals with the basis $1, i, j, k$ satisfying

$$\begin{array}{c|cccc} \cdot & 1 & i & j & k \\ \hline 1 & 1 & i & j & k \\ i & i & -1 & k & -j \\ j & j & -k & -1 & i \\ k & k & j & -i & -1 \end{array} \qquad (622)$$

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. With the quaternions, the idea of a non-commutative product appeared for the first time 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; just find all solutions of the equation $X^2 + 1 = 0$ to find out. Challenge 1231

Every quaternion X can be written in the form

$$X = x_0 + x_1 i + x_2 j + x_3 k = x_0 + \mathbf{v} = (x_0, x_1, x_2, x_3) = (x_0, \mathbf{v}) \quad , \qquad (623)$$

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 multiplication of two general quaternions can be written as

$$\begin{aligned} (a_1, b_1, c_1, d_1)(a_2, b_2, c_2, d_2) = (&a_1 a_2 - b_1 b_2 - c_1 c_2 - d_1 d_2\ ,\ a_1 b_2 + b_1 a_2 + c_1 d_2 - d_1 c_2, \\ &a_1 c_2 - b_1 d_2 + c_1 a_2 + d_1 b_2, a_1 d_2 + b_1 c_2 - c_1 b_2 + d_1 a_2) \end{aligned} \qquad (624)$$

---

[*] William Rowan Hamilton (1805, Dublin–1865, Dunsink), Irish child prodigy, mathematician, named the quaternions after an expression from the vulgate (act. apost. 12, 4).

The conjugate quaternion $\overline{X}$ is defined as $\overline{X} = x_0 - \mathbf{v}$, so that $\overline{XY} = \overline{Y}\,\overline{X}$. The *norm* $|X|$ of a quaternion X is defined as $|X|^2 = X\overline{X} = \overline{X}X = x_0^2 + x_1^2 + x_2^2 + x_3^2 = x_0^2 + \mathbf{v}^2$. The norm is multiplicative, i.e. $|XY| = |X|\,|Y|$.

The relation $|XY| = |X|\,|Y|$ implies the *four-squares theorem*

$$\begin{aligned}(a_1^2+a_2^2+a_3^2+a_4^2)&(b_1^2+b_2^2+b_3^2+b_4^2)\\ &= (a_1b_1-a_2b_2-a_3b_3-a_4b_4)^2+(a_1b_2+a_2b_1+a_3b_4-a_4b_3)^2\\ &\quad +(a_1b_3+a_3b_1+a_4b_2-a_2b_4)^2+(a_1b_4+a_4b_1+a_2b_3-a_3b_2)^2 \qquad (625)\end{aligned}$$

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. The product of two pure quaternions $V = (0,\mathbf{v})$ and $W = (0,\mathbf{w})$ is given by $VW = (-\mathbf{v}\cdot\mathbf{w}, \mathbf{v}\times\mathbf{w})$, where $\cdot$ denotes the scalar product and $\times$ denotes the vector product. Note that 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 or $180\,°$ around the axis defined by the direction $\mathbf{v} = (v_x, v_y, v_z)$.

Challenge 1232

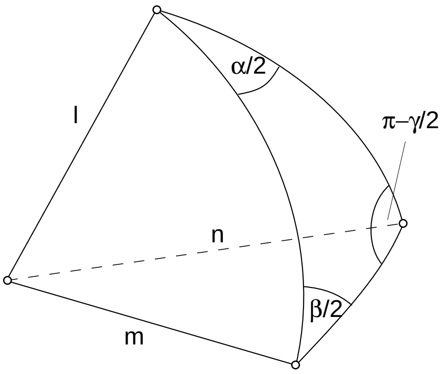


Figure 260 Combinations of rotations

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

$$W = QVQ^* \quad . \qquad (626)$$

In this case, when using pure quaternions such as V or W to describe positions, unit quaternions can be used 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

$$(\cos\gamma/2, \sin\gamma/2\mathbf{n}) = (\cos\alpha/2, \sin\alpha/2\mathbf{l})(\cos\beta/2, \sin\beta/2\mathbf{m}) \quad , \qquad (627)$$

shown graphically in Figure 260.

Quaternions can teach something about the motion of hand and arm. Keeping the left arm straight, defining its 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 1233

The reason for this behaviour is the non-commutativity of rotations. This noncommutativity can be specified more precisely using 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 570

Ref. 901

The easy description of rotations and positions with quaternions is used in robotics, in astronomy, and in flight simulators, due to the especially simple coding of coordinate transformations it provides. Inside three-dimensional graphic visualisation software, quaternions are often used 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 $\overline{X}/N(X)$. This allows to define a division of quaternions. Therefore quaternions are said to form a *division algebra*. In fact the quaternions **H**, the complex numbers **C** and the reals **R** 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 *centre* of the quaternions, i.e. the set of those quaternions commuting with all quaternions, are the reals.

Like the complex numbers, quaternions can be represented as matrices of the form

$$\begin{pmatrix} A & B \\ -B^* & A^* \end{pmatrix} \text{ with } A, B \in \mathbf{C}, \quad \text{or as} \begin{pmatrix} a & b & c & d \\ -b & a & -d & c \\ -c & d & a & -b \\ -d & -c & b & a \end{pmatrix} \text{ with } a, b, c, d \in \mathbf{R} \quad , \tag{628}$$

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 give the same result as quaternionic addition and multiplication.

The generators of the quaternions can be realised for example as

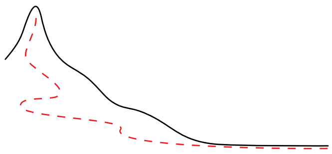
$$1 : \sigma_0 \quad , \quad i : -i\sigma_1 \quad , \quad j : -i\sigma_2 \quad , \quad k : -i\sigma_3 \tag{629}$$

where the σ_n are the Pauli spin matrices.*

---

* The *Pauli spin matrices* are the complex, hermitean matrices

$$\sigma_0 = \mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{630}$$



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net

Real 4 × 4 representations are not unique, as

$$\begin{pmatrix} a & b & -d & -c \\ -b & a & -c & d \\ d & c & a & b \\ c & -d & -b & a \end{pmatrix} \tag{631}$$

shows; however, no representation by 3 × 3 matrices is possible. Challenge 1235

These matrices contain real and complex elements, which pose 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? First of all, binary rotations are similar to positions, and thus to 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 mountain ascent.

As a remark, when Maxwell wrote down his equations of electrodynamics, he used quaternion notation. The now usual 3-vector notation was introduced later by other scientists, notably by Hertz and Heaviside. Maxwell's original equations of electrodynamics, in modern quaternion notation, read: See page 382 Challenge 1236

$$dF = -\frac{Q}{\varepsilon_0} \tag{632}$$

where the quantities are defined as following:

$$\begin{aligned} F &= E + \sqrt{-1}\,cB \\ E &= iE_x + jE_y + kE_z \\ B &= iB_x + jB_y + kB_z \\ d &= \delta + \sqrt{-1}\,\partial_t/c \\ \delta &= i\partial_x + j\partial_y + k\partial_z \\ Q &= \rho + \sqrt{-1}\,J/c \end{aligned} \tag{633}$$

and where $\sqrt{-1}$ is the complex root of -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 $1, i_n$

---

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 $i : i\sigma_3, j : i\sigma_2, k : i\sigma_1$.

with $n = 1...7$ satisfying

$$\begin{array}{c|cccccccc}
 & 1 & i_1 & i_2 & i_3 & i_4 & i_5 & i_6 & i_7 \\
\hline
1 & 1 & i_1 & i_2 & i_3 & i_4 & i_5 & i_6 & i_7 \\
i_1 & i_1 & -1 & i_3 & -i_2 & i_5 & -i_4 & i_7 & -i_6 \\
i_2 & i_2 & -i_3 & -1 & i_1 & -i_6 & i_7 & i_4 & -i_5 \\
i_3 & i_3 & i_2 & -i_1 & -1 & i_7 & i_6 & -i_5 & -i_4 \\
i_4 & i_4 & -i_5 & i_6 & -i_7 & -1 & i_1 & -i_2 & i_3 \\
i_5 & i_5 & i_4 & -i_7 & -i_6 & -i_1 & -1 & i_3 & i_2 \\
i_6 & i_6 & -i_7 & -i_4 & i_5 & i_2 & -i_3 & -1 & i_1 \\
i_7 & i_7 & i_6 & i_5 & i_4 & -i_3 & -i_2 & -i_1 & -1
\end{array} \tag{634}$$

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 $x(xy) = x^2y$ 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

$$\begin{pmatrix} A & B \\ -\bar{B} & \bar{A} \end{pmatrix} \quad \text{where} \quad A, B \in \mathbf{H} \quad , \quad \text{or as real } 8 \times 8 \text{ matrices.} \tag{635}$$

Matrix multiplication then gives the same result as octonionic multiplication.

The relation $|wz| = |w|\,|z|$ allows to deduce the impressive *eight-squares theorem*

$$\begin{aligned}
&(a_1^2+a_2^2+a_3^2+a_4^2+a_5^2+a_6^2+a_7^2+a_8^2)(b_1^2+b_2^2+b_3^2+b_4^2+b_5^2+b_6^2+b_7^2+b_8^2) \\
&\quad = (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4 - a_5b_5 - a_6b_6 - a_7b_7 - a_8b_8)^2 \\
&\quad +(a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3 + a_5b_6 - a_6b_5 - a_7b_8 + a_8b_7)^2 \\
&\quad +(a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2 + a_5b_7 + a_6b_8 - a_7b_5 - a_8b_6)^2 \\
&\quad +(a_1b_4 + a_2b_3 - a_3b_2 + a_4b_1 + a_5b_8 - a_6b_7 + a_7b_6 - a_8b_5)^2 \\
&\quad +(a_1b_5 - a_2b_6 - a_3b_7 - a_4b_8 + a_5b_1 + a_6b_2 + a_7b_3 + a_8b_4)^2 \\
&\quad +(a_1b_6 + a_2b_5 - a_3b_8 + a_4b_7 - a_5b_2 + a_6b_1 - a_7b_4 + a_8b_3)^2 \\
&\quad +(a_1b_7 + a_2b_8 + a_3b_5 - a_4b_6 - a_5b_3 + a_6b_4 + a_7b_1 - a_8b_2)^2 \\
&\quad +(a_1b_8 - a_2b_7 + a_3b_6 + a_4b_5 - a_5b_4 - a_6b_3 + a_7b_2 + a_8b_1)^2
\end{aligned} \tag{636}$$

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

$$\begin{aligned} u \times v &= -v \times u && \textit{anticommutativity} \\ (u \times v)w &= u(v \times w) && \textit{exchange rule.} \end{aligned} \qquad (637)$$

Using the definition

$$X \times Y = \frac{1}{2}(XY - YX) \quad , \qquad (638)$$

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, thus fulfilling (637). Interestingly, using definition (638) for *octonions* is possible. In that case the product of imaginary octonions, i.e. octonions of the sort $(0, \mathbf{U})$, also yields only imaginary octonions,and the $\mathbf{U}$'s also follow expression (637). In fact, this is the only other non-trivial example possible. Thus a vector product exists only in 3 and in 7 dimensions. Ref. 902 Challenge 1237 e

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-Lifshitz 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. 903 Challenge 1238 n

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 **R** converge. Can you find one example? Ref. 904 See page 469 Challenge 1239

The surreals also include infinitely small numbers. That is also the case for 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. 902

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

$$\theta^2 = 0 \quad \text{and} \quad \theta_i\theta_j + \theta_j\theta_i = 0 \quad . \tag{639}$$

Challenge 1240 n 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

$$\begin{aligned} (a+b)+c = a+(b+c) = a+b+c & \quad \textit{associativity of vector addition} \\ n+a = a & \quad \textit{existence of null vector} \\ (-a)+a = n & \quad \textit{existence of negative vector} \\ 1a = a & \quad \textit{regularity of scalar multiplication} \\ (s+r)(a+b) = sa+sb+ra+rb & \quad \textit{complete distributivity of scalar multiplication} \end{aligned} \tag{640}$$

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 first 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

$$\begin{aligned} d(a,b) \geqslant 0 & \quad \textit{positivity of metric} \\ d(a,b)+d(b,c) \geqslant d(a,c) & \quad \textit{triangle inequality} \\ d(a,a) = 0 & \quad \textit{regularity of metric} \end{aligned} \tag{641}$$

---

[*] Hermann Günther Grassmann (1809–1877) mathematician.

For example, measuring the distance between cities in France, i.e. points on a surface, by the shortest distance of travel via Paris, except in the case if they both lie on a line already going through Paris, defines a metric between the points in France. Challenge 1241 n

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

$$\begin{aligned} ||ra|| &= |r|\ ||a|| && \textit{linearity of norm} \\ ||a+b|| &\leqslant ||a|| + ||b|| && \textit{triangle inequality} \\ ||a|| = 0 \quad &\text{only if} \quad a = 0 && \textit{regularity} \end{aligned} \tag{642}$$

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 1242 n

$$d(a,b) = ||a-b|| \tag{643}$$

so that all normed spaces are also metric spaces. This norm is the standard choice. 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

$$\begin{aligned} ab &= ba && \textit{commutativity of scalar product} \\ (ra)(sb) &= rs(ab) && \textit{bilinearity of scalar product} \\ (a+b)c &= ac+bd && \textit{left distributivity of scalar product} \\ a(b+c) &= ab+ac && \textit{right distributivity of scalar product} \\ aa &\geqslant 0 && \textit{positivity of scalar product} \\ aa = 0 \quad \text{only if} \quad a &= 0 && \textit{regularity of scalar product} \end{aligned} \tag{644}$$

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*

$$\begin{aligned} ab = \overline{(ba)} &= \bar{b}\bar{a} && \textit{hermitean property} \\ (ra)(sb) &= \bar{r}s(ab) && \textit{sesquilinearity of scalar product} \\ (a+b)c &= ac+bd && \textit{left distributivity of scalar product} \\ a(b+c) &= ab+ac && \textit{right distributivity of scalar product} \\ aa &\geqslant 0 && \textit{positivity of scalar product} \\ aa = 0 \quad \text{only if} \quad a &= 0 && \textit{regularity of scalar product} \end{aligned} \tag{645}$$

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

---

* The term *sesquilinear* is Latin for 'one-and-a-half-linear'. Sometimes however, the half-linearity is assumed in the other argument.

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.

All inner product spaces are also metric spaces and thus normed spaces, if the metric is defined, as usually done, by

$$d(a,b) = \sqrt{(a-b)(a-b)} \quad . \tag{646}$$

Inner product spaces allow to speak about the length and the direction of vectors, as we are used to in physics. In addition, inner product spaces allow to define a *basis*, the mathematical concept used to define coordinate systems.

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

$$\begin{aligned} x+y &= y+x && \textit{commutativity of addition} \\ x(y+z) = xy+xz \quad , \quad (x+y)z &= xz+yz && \textit{distributivity of multiplication} \\ xx &\geqslant 0 && \textit{positivity} \\ xx = 0 \quad \text{only if} \quad x &= 0 && \textit{regularity of multiplication} \end{aligned} \tag{647}$$

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

$$x(yz) = (xy)z \quad \textit{associativity} \tag{648}$$

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

$$c(xy) = (cx)y = x(cy) \qquad \textit{linearity} \tag{649}$$

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 spaces; the difference being that also a (linear) and associative multiplication among the vectors is defined.

---

* 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

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

$$\begin{aligned}(x^*)^* &= x\\ (x+y)^* &= x^* + y^*\\ (cx)^* &= c^* x^* \quad \text{for all} \quad c \in \mathbf{C}\\ (xy)^* &= y^* x^* \end{aligned} \tag{651}$$

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, i.e. a complete normed algebra, over the complex numbers with an involution $*$ so that the norm $||x||$ of an element x can be defined as

$$||x||^2 = x^* x \quad . \tag{652}$$

Algebras are *complete* if 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 1244

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 non-negative 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*,

---

Challenge 1243 n able to give the set of all possible linear transformations of points on a plane? And in space? And in Minkowski space?

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 a transformation T has the effect

$$Te = \lambda e \tag{650}$$

the vector e is called 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. The state of the system after the measurement is given by the eigenvector of the measured eigenvalue. Therefore every expert on motion must know what an eigenvalue is.

See page **??**

See page 591

is called a real (or complex) Lie algebra if this operation fulfils the properties

$$\begin{aligned}[X,Y] &= -[Y,X] && \textit{antisymmetry}\\ [aX+bY,Z] &= a[X,Z]+b[Y,Z] && \textit{linearity}\\ [X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]] &= 0 && \textit{Jacobi identity}\end{aligned} \tag{653}$$

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 *centre* 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 every finite dimensional Lie algebra is isomorphic to a linear Lie algebra. Therefore, by picturing the elements of Lie algebras in terms of matrices all finite dimensional cases are covered.

The name 'Lie algebra' was chosen because 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 the symbol $\cdot$ stands for its multiplication, a Lie algebra appears when defining the *commutator* by

$$[X,Y] = X\cdot Y - Y\cdot X \quad ; \tag{654}$$

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:

$$[T_i,T_j] = \sum_k c_{ij}^k T_k \tag{655}$$

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 908

---

[*] 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

$$[a,a_j] = \sum_c \mathrm{ad}(a)_{cj} a_c \quad . \tag{656}$$

It implies 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, a scalar product can be defined by setting

$$(X,Y) = \mathrm{Tr}(\mathrm{ad}X\, \mathrm{ad}Y) \quad . \tag{657}$$

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.

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. Simple Lie algebras can be complex or real.

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-hermitean $n \times n$ 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. 905

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 Killing form is invariant under the action of any automorphism of the algebra L. In a given basis, one has

$$(X,Y) = \mathrm{Tr}(\,(\mathrm{ad}X)^i_k\,(\mathrm{ad}Y)^s_i) = c^i_{lk} c^k_{si} x^l y^s = g_{ls} x^l y^s \qquad (658)$$

where $g_{ls} = c^i_{lk} c^k_{si}$ is called the *Cartan metric tensor* of the Lie algebra L.

The Virasoro algebra

The *Virasoro algebra* is the infinite algebra of operators L_n satisfying

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{c}{12}(m^3 - m)\delta_{m,-n} \qquad (659)$$

where the number c, which may be zero, is called the *central charge*, and the factor $1/12$ is 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 686.* Challenge 1245 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.

Topology - what shapes exist?

> Topology is group theory.
> The Erlangen program

In a simplified view of topology sufficient for physicists, only one type of entities can possess shapes: manifolds. Manifolds are generalized examples of pullovers; manifolds are flat, have holes, boundaries, and often can be reversed. Pullovers are tricky entities. Are you able Challenge 1246 to find out how to reverse your pullover while your hands are tied to each other?(A friend may help you.) By the way, the same feat is also possible with your trousers, while your feet are tied to each other.

– CS – section on topology to be added – CS –

In other words, through homotopy and homology theory, mathematicians can classify manifolds. The properties of holes in manifolds allow to determine whether two manifolds can be deformed into each other. However, physicists are now extending these results. Deformation is a classical idea which assumes continuous space and time, as well as arbitrary small action. In nature, quantum effects cannot be neglected. Quantum effects can be used to transform physical manifold even if their topology is different, such as a torus into a sphere. Can you see how this can be achieved? Challenge 1247

Lie groups

In nature, Lagrangians of the fundamental forces are invariant under gauge transformations and under continuous space-time transformations. These symmetry groups are examples of Lie groups, which are special types of infinite continuous groups. They are named after the great Norwegian mathematician Sophus Lie (1849–1899). His name is pronounced like ‘Lee’.

See page 236 * Note that the conformal symmetry group in *four* dimensions has 15 parameters, and thus its Lie algebra is finite (fifteen) dimensional.

A (real) *Lie group* is an infinite symmetry group, i.e. a group with infinitely many elements, which at the same time is also an analytical manifold. Sloppily speaking, this definition means that the elements of the group can at the same time be seen as points on a smooth (hyper-) surface whose shape, i.e. the coordinates of the points of the surface, can be described by an analytic function, i.e. by a function so smooth that it can be expressed as a power series. In addition, the points can be multiplied in some way, as they are also elements of a group. Furthermore, the coordinates of the product have to be analytical functions of the coordinates of the factors, and the coordinates of the inverse have to be analytic functions of the coordinates of the element itself. In fact this definition is too strict; it can be proven that a Lie group is any topological group whose topological space is a finite-dimensional, locally Euclidean manifold.

A *complex Lie group* is a group where the manifold is complex and the functions holomorphic instead of analytical.

– CS – section on Lie groups to be added – CS –

References

899 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 904.

900 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 906.

901 See the fine book by 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 908.

902 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 912.

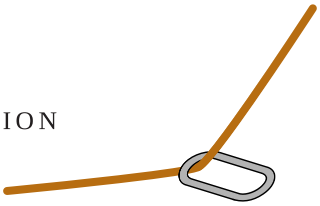
903 Gaussian numbers are presented and explained in ... Cited on page 912.

904 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 912.

905 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 918.

Typeset in
January 2003

APPENDIX E INFORMATION SOURCES ON MOTION



> No place affords a more striking conviction of the vanity of human hopes than a public library.
> Samuel Johnson (1709–1784)

> Das Internet ist die offenste Form der geschlossenen Anstalt.[*]
> Matthias Deutschmann

In the text, outstanding books introducing neighbouring domains are presented in footnotes. The reference list at the end of each chapter collects general material satisfying further curiosity about what is encountered in this mountain ascent. All citations can also be found by looking up the author's name in the index. To find additional information, 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

---

[*] 'The internet is the most open form of a closed institution.'

search for all subsequent publications which cite it. Then, using the bimonthly *Physics Abstracts*, which also exists both in paper and in electronic form, you 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 free of charge, as preprints, i.e. before official publication and check by referees. This famous service is available at the http://www.arxiv.org web site.

Table 73 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	via e-mail, add
and lattice calculations	@arxiv.org or	
phenomenological high energy physics	hep-ph	@www.arxiv.org 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@arXiv.org or
of quantum physics		gr-qc@www.arxiv.org or
experimental nuclear physics	nucl-ex	physics@babbage.sissa.it
theoretical nuclear physics	nucl-th	
astrophysics	astro-ph	
condensed matter physics	cond-mat	
mathematical physics	math-ph	
mathematics	math	
computer science	CoRR	

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.

In the last decade of the twentieth century, the internet expanded into a mix of library, media store, discussion platform, order desk and time waster. With a personal computer, a modem and free *browser* software you can look for information in millions of pages of documents. 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.*

To start using the web, ask a friend who knows, or send an electronic mail message consisting of the line ‘HELP’ to listserv@info.cern.ch, the server at the European Organisation

---

∗ 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;

for Particle Research, where the web was invented.* Searching the web for authors, organizations, books, publications, companies or simple keywords using search engines can be a rewarding experience. A few interesting servers are given below.

Table 74 Some interesting world wide web servers

Topic	web site address ('URL')
Search engines and more	
Good information search engines	http://www.altavista.com/cgi-bin/query?pg=aq http://www.metager.de
Search old usenet articles	http://groups.google.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
Physics and science	
Electronic preprints	http://www.arxiv.org and others – see above http://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
Physics problems by Yakov Kantor	http://star.tau.ac.il/QUIZ/
Physics problems by Henry Greenside	http://www.phy.duke.edu/ hsg/physics-challenges/challenges.html
Physics question of the week	http://www.physics.umd.edu/lecdem/outreach/QOTW/active/questions.htm
Ask the expert	http://www.sciam.com/askexpert_directory.cfm

- 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.

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*, Penguin, 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 can transform it into the backbone of the opportunity web mentioned by Illich; it is a social development to follow closely. It mainly 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')
Article summaries in 25 science magazines	http://www.mag.browse.com/science.html
Abstracts of papers in 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
	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.motionmountain.net
	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
Physics lecture scripts in German and English	http://kbibmp5.ub.uni-kl.de/Linksammlung/Physik/liste.html
'World' lecture hall	http://www.utexas.edu/world/lecture
Math forum internet resource collection	http://mathforum.org/library/
Purdue math problem of the week	http://www.math.purdue.edu/academics/pow/
Macalester college math problem of the week	http://mathforum.org/wagon/
Math formulas	http://dlmf.nist.gov
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/

Topic	web site address (‘URL’)
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	
Minerals	http://webmineral.com
NASA	http://oel-www.jpl.nasa.gov/basics/bsf.html
	http://gsfc.nasa.gov
Hubble space telescope	http://hubble.nasa.gov
The cosmic mirror	http://www.astro.uni-bonn.de/~dfischer/mirror
Solar system simulator	http://space.jpl.nasa.gov
Observable satellites	http://liftoff.msfc.nasa.gov/RealTime/JPass/20/
The earth from space	http://www.visibleearth.nasa.gov
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 humour	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 aeroplanes	http://pchelp.inc.net/paper_ac.htm
	http://www.ivic.qc.ca/~aleexpert/aluniversite/ klinevogelmann.html
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

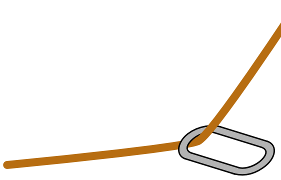
Do you want to study physics without actually going to university? Nowadays it is possible to do so via e-mail and internet, in German, at the University of Kaiserslautern.[*] In the near future, a nationwide project in Britain should allow the same for English speaking students. As introduction, use the last update of this physics text!

> Si tacuisses, philosophus mansisses.[**]
> After Boethius.

---

[*] See the http://www.fernstudium-physik.de web site.

[**] ‘If you had kept quiet, you would have remained philosopher.’ After the story Boethius tells in *De consolatione philosophiae*, 2,7, 67 ff.



Typeset in
January 2003

APPENDIX F LIST OF FIGURES

Picture credits

All drawings are copyright © 1997 – 2003 by Christoph Schiller.

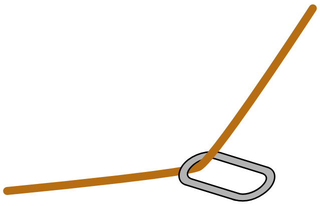
The copyrights of the figures on page 25 and on page 30 cannot be traced.

The copyright of the figure on page 727 belongs to the M.C. Escher Heirs, c/o Cordon Art, Baarn, The Netherlands. They kindly gave permission for its use.

The figure of myosotis on page 143 is copyright by Markku Savela, Markku.Savela@iki.fi, who kindly gave permission to use it.

The figures of galaxies on page 303, pages 303 to 304 and on page 326 are courtesy of NASA.

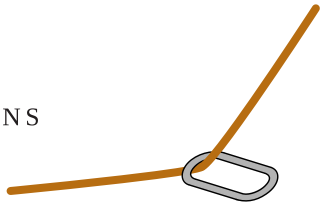
Typeset in
January 2003



APPENDIX G LIST OF TABLES

Typeset in
January 2003

APPENDIX H CHALLENGE HINTS AND SOLUTIONS



> Never make a calculation before you know the answer.
> John Wheeler's well-known motto*

There are 1247 challenges in the text. Let me know the challenge for which you want a hint or a solution to be added here; I will mail it to you and include it in the text in the future.

Challenge 1 (page 18) A honest physicist will not put his hand into fire on any statement about nature he thinks is correct, but will put it into fire to say that a statement is false.

Challenge 2 (page 20) These topics are all addressed later in the text.

Challenge 3 (page 27) There are many ways to distinguish real motion from an illusion of motion: for example, only real motion can be used to set something else into motion.

Challenge 4 (page 27) Without detailed and precise experiments, both sides can find examples to prove their point. Creation is supported by the appearance of mould or bacteria in a glass of water; creation is supported by its opposite, namely traceless disappearance, such as the disappearance of motion. Conservation is supported by all those investigations which look into assumed cases of appearance or disappearance.

Challenge 5 (page 28) Political parties, sects, helping organisations and therapists of all kinds are typical for this behaviour.

Challenge 6 (page 31) The issue is not yet completely settled for the motion of empty space, such as in the case of gravitational waves. For sure, empty space is not made of small particles of finite size, as this would contradict the transversality of gravity waves.

Challenge 7 (page 31) This methods is known work with others fears as well.

Challenge 8 (page 33) It is; we shall return to this important issue several times in our adventure. It will require a lot of patience to solve it, though.

Challenge 9 (page 34) See page 601.

Challenge 10 (page 35) A ghost can be a moving image; it cannot be a moving object, as objects cannot interpenetrate. See page 573.

Challenge 11 (page 35) How would you measure this?

Challenge 12 (page 35) Can one show at all that something has stopped moving?

---

* Wheeler wanted people to estimate, to try, and to guess; but not saying it out loud. A correct guess reinforces the physics instinct, and a wrong one leads to the pleasure of surprise.

Challenge 13 (page 35) The number of reliable digits of a measurement result is a simple quantification of precision.

Challenge 14 (page 35) No; memory is needed for observation and measurements.

Challenge 15 (page 35) Note that you never have observed zero speed.

Challenge 16 (page 35) $(2^{64}-1)$ grains of rice, given a wold harvest of 500 million tons, are about 4000 years or rice harvests.

Challenge 17 (page 35) Just try it. Can you explain the observation?

Challenge 18 (page 35) It rolls towards the centre of the table, as the centre is somewhat lower than the border.

Challenge 19 (page 35) Accelerations can be felt. Many devices measure accelerations and then deduce the position. They are used in aeroplanes when flying over the atlantic.

Challenge 20 (page 35) The necessary rope length is nh, where n is the number of wheels/pulleys.

Challenge 21 (page 35) In everyday life, this is correct; what happens when quantum effects are taken into account?

Challenge 22 (page 36) There is only one way: compare it with the speed of light.

Challenge 23 (page 38) The average distance change of two neighbouring atoms in a piece of quartz over the last million years. Do you know something still slower?

Challenge 24 (page 37) Equivalently: do points in space exist? The third part studies this issue in detail.

Challenge 25 (page 39) All electricity sources must use the same phase when they feed electric power into the net. Clocks of computers on the internet must be synchronized.

Challenge 26 (page 39) Note that the shift increases quadratically with time, not linearly.

Challenge 27 (page 40) Natural time is measured with natural motion. Natural motion is the motion of light. Natural time is this defined with the motion of light.

Challenge 28 (page 41) He measured the amount of water spilling from a water tube in the short time.

Challenge 29 (page 41) There is no way to define a local time at the poles.

Challenge 30 (page 43) The forest is full of light and thus of light rays.

Challenge 31 (page 43) This you can solve trying to think in four dimensions. Try to imagine how to switch the sequence when two pieces cross.

Challenge 32 (page 44) Light.

Challenge 36 (page 47) It is easier to work with the unit torus. Take the unit interval $[0,1]$ and equate the end points. Define a set B in which the elements are a given real number b from the interval plus all those numbers who differ from that real by a rational number. The unit circle can be thought as the union of all the sets B. (In fact, every set B is a shifted copy of the rational numbers $\mathbf{Q}$.) Now build a set A by taking one element from each set B. Then build the set family consisting of the set A and its copies A_q shifted by a rational q. The union of all these sets is the unit torus. The set family is countably infinite. Then divide it into *two* countably infinite set families. It is easy to see that each of the two families can be renumbered and its elements shifted in such a way that each of the two families forms a unit torus.

Mathematicians say There is no countably infinitely additive measure of $\mathbf{R}^n$ or that sets such as A Ref. 29

are non-measurable. As a result of their existence, the ‘multiplication’ of lengths is possible. Later on we shall explore whether bread or *gold* can be multiplied in this way.

Challenge 37 (page 48) This puzzle is left to solve to the reader.

Challenge 38 (page 48) An example is the region between the x-axis and the function which assigns 1 to every transcendental and 0 to every non-transcendental number.

Challenge 39 (page 48) We use the definition of the function of the text. The dihedral angle of a regular tetrahedron is an irrational multiple of π, so the tetrahedron has a nonvanishing Dehn invariant. The cube has a dihedral angle of $\pi/2$, so the Dehn invariant of the cube is 0. Therefore, the cube is not equidecomposable with the regular tetrahedron.

Challenge 40 (page 49) If you think you can show that empty space is continuous, you are wrong. Check your arguments. If you think you can prove the opposite, you *might* be right – but only if you already know what is explained in the third part of the text.

Challenge 41 (page 49) Obviously, we use light to check that the plumb line is straight, so the two definitions must be the same. This is the case because the field lines of gravity are also possible paths for the motion of light. However, this is not always the case; can you spot the exceptions?

Another way to check straightness is along the surface of calm water.

Challenge 42 (page 50) The hollow earth theory is correct if the distance formula is used consistently. In particular, one has to make the assumption that objects get smaller as they approach the centre of the hollow sphere. Good explanations of all events are found on http://www.geocities.com/inversedearth/ Quite some material can be found on the internet, also under the names of celestrocentric system, inner world theory or concave earth theory. There is no way to prefer one description over the other, except possibly for reasons of simplicity or intellectual lazyness.

Challenge 44 (page 50) 22 times. 2 times.

Challenge 45 (page 50) For two hands, the answer is 143 times.

Challenge 47 (page 50) Nothing, neither a proof nor a disproof.

Challenge 316 (page 162) The different usage reflects the idea that we are able to determine our position by ourselves, but not the time in which we are. The section on determinism will show how wrong this distinction is.

See page 114

Challenge 50 (page 50) For a curve, use, at each point, the curvature radius of the circle approximating the curve in that point; for a surface, define two directions in each point, and use two such circles along these directions.

Challenge 51 (page 50) It moves about 1 cm in 50 ms.

Challenge 52 (page 50) See page 202.

Challenge 54 (page 50) Because the are or were fluid.

Challenge 46 (page 50) The earth rotates with 15 minutes per minute.

Challenge 56 (page 51) Hint: draw all objects involved.

Challenge 57 (page 51) Hint: there are an infinite number of shapes.

Challenge 58 (page 51) The curve is obviously called a catenary, from latin for chain. If you approximate it by short straight segments, you can make wooden blocks that can form an arch without any need for glue.

Challenge 59 (page 51) A limit does not exist in classical physics; however, there is one in nature which appears as soon as quantum effects are taken into account.

Challenge 60 (page 51) The inverse radii, or curvatures, obey $a^2+b^2+c^2+d^2=(1/2)(a+b+c+d)^2$. This formula was discovered by René Descartes. If one continues putting circles in the remaining spaces, one gets so-called circle packings, a pretty domain fo recreational mathematics. They have many strange properties, such as intriguing relations between the coordinates of the circle centres and their curvatures.

Challenge 61 (page 51) Draw a logarithmic scale, i.e., put every number at a distance corresponding to its natural logarithm.

Challenge 62 (page 51) Two more.

Challenge 63 (page 52) From $x=gt^2/2$ you get the following rule: square the number of seconds, multiply by five, and you have the depth in metres.

Challenge 555 (page 266) Just experiment.

Challenge 66 (page 53) On horizontal ground, neglecting the height of the thrower and air resistance, the distance is $v^2 \sin 2\alpha/2g$.

Challenge 67 (page 53) Walk in the rain, measure your own speed and the angle with which the rain appears to fall.

Challenge 68 (page 53) Check your calculation with the information that the world record is juggling with nine balls.

Challenge 69 (page 53) The long jump record could surely be increased by measuring the true jumping distance with a photographic camera, and getting rid of the sand stripe.

Challenge 70 (page 53) It seems not too much. But the lead in them can poison the environment.

Challenge 71 (page 53) It is said so, as rain drops would then be ice spheres, and fall with high speed.

Challenge 72 (page 54) Stones do not fall parabolas when studied in detail, i.e. when the change of g with height is taken into account. Their path is an ellipse. This appears for long 'throws', such as throws around the earth.

Challenge 75 (page 55) Yes, it is a vector space.

Challenge 78 (page 57) One can argue that any source of light must have finite size.

Challenge 79 (page 57) What the unaided human eye perceives as a tiny black point is usually about 50 μm in diameter.

Challenge 80 (page 57) See page 398.

Challenge 81 (page 58) One has to check carefully whether the conceptual steps that lead us to extract the concept of point are physically possible. This will be discussed in the third part of the adventure.

Challenge 82 (page 58) One can rotate the hand in a way that the arm makes the motion described here. See also page 571.

Challenge 83 (page 58) Any number, without limit.

Challenge 84 (page 58) The blood supply would periodically be run over by the wheel. Could it make a propeller?

Challenge 85 (page 59) The brain in the skull, the blood factories inside bones or the growth of the eye are examples.

Challenge 86 (page 60) One can also add the sun, the sky, and the landscape to the list.

Challenge 87 (page 60) Ghosts, hallucinations, Elvis sightings or extraterrestrials must all be one or the other. There is no third option.

Challenge 88 (page 61) The issue was hotly discussed in the 17th century, and even Galileo argued for them being images. However, they are objects, as they can collide with other objects.

Challenge 89 (page 61) The minimum speed is roughly the one at which it is possible to ride without hands. If you do so, and then *gently* push on the steering wheel, you can make this experience described above. Watch out: too strong a push will make you fall badly.

Challenge 90 (page 62) If the ball is not rotating, after the collision the two balls will depart with a *right* angle between them.

Challenge 91 (page 63) Part of the energy is converted into heat; the rest is transferred as kinetic energy of the concrete block. As the block is heavy, its speed is small and easily stopped by the human body. This effect works also with anvils, it seems.

Challenge 92 (page 63) Yes, mass works also for magnetism, because the precise condition is not that the interaction be central, but that it realizes a more general condition, which includes accelerations such as those produced by magnetism. Can you deduce the condition from the definition of mass?

Challenge 93 (page 64) The weight decreased, due to the evaporated water and, to a minor degree, due to the exhaled carbon dioxide.

Challenge 94 (page 64) Rather than using the inertial effects of the earth, it is easier to deduce its mass from its gravitational effects.

Challenge 98 (page 64) Tachyons must have imaginary mass.

Challenge 99 (page 65) Legs are never perfectly vertical; they would immediately glide away. Once the cat or the person is on the floor, it is almost impossible to stand up again.

Challenge 1062 (page 623) Momentum (or centre of mass) conservation would imply that the environment would be accelerated into the opposite direction. Energy conservation would imply that a huge amount of energy would be transferred between the two locations, melting everything in between. Teleportation would thus contradict energy and momentum conservation.

Challenge 101 (page 66) The part of the tides due to the sun, the solar wind, and the interactions between both magnetic fields are examples of friction mechanisms.

Challenge 102 (page 67) With the factor $1/2$, increase of (physical) kinetic energy is equal to the (physical) work performed on a system.

Challenge 104 (page 69) It is a smart application of momentum conservation.

Challenge 105 (page 69) Neither. With brake on, the damage is higher, but still equal for both cars.

Challenge 106 (page 69) Heating systems, transport engines, engines in factories, steel plants, electricity generators covering the losses in the power grid, etc.

Challenge 109 (page 71) Use the definition of the moment of inertia and Pythagoras' theorem for every mass element of the body.

Challenge 110 (page 71) Hang up the body, attaching the rope in two different points. The crossing point of the prolonged rope lines is the centre of mass.

Challenge 111 (page 71) Spheres have an orientation, because we can always add a tiny spot on their surface. This possibility is not given for microscopic objects, and we shall study this situation in the part on quantum theory.

Challenge 114 (page 72) See Tables 102 and 56.

Challenge 117 (page 73) The points that move exactly along the radial direction of the wheel form a circle below the axis. They are the points that are sharp in Figure 24 of page 73.

Challenge 118 (page 73) Use the conservation of angular momentum around the point of contact. If all the wheel's mass is assumed in the rim, the final rotation speed is half the initial one; it is independent of the friction coefficient.

Challenge 123 (page 77) The Coriolis acceleration is behind the deviation from the straight line.

Challenge 124 (page 77) Rotation leads to a small frequency and thus colour changes of the circulating light.

Challenge 125 (page 77) The weight changes when going east or when moving west due to the Coriolis acceleration. If the rotation speed is tuned to the oscillation frequency of the balance, the effect is increased by resonance. This trick was also used by Eötvös.

Challenge 126 (page 77) The Coriolis acceleration makes the bar turn, as every moving body is deflected to the side, and the two deflections add up in this case. The direction of the deflection depends on whether the experiments is performed on the northern or the southern hemisphere.

Challenge 128 (page 80) The original result by Bessel was $0.3136''$, or 657.7 thousand orbital radii, which he thought to be 10.3 light years or 97.5 Pm.

Challenge 130 (page 82) The galaxy forms a stripe in the sky. The galaxy is thus a flattened structure. This is even clearer in the infrared, as shown more clearly in Figure 131 on page 303. From the flattening (and its circular symmetry) we can deduce that the galaxy must be rotating. Thus other matter must exist in the universe.

Challenge 131 (page 82) If the earth changed its rotation speed ever so slightly we would walk inclined, the water of the oceans would flow north, the atmosphere would be filled with storms and earthquakes would appear due to the change in earth's shape.

Challenge 132 (page 82) The first point is that moving a part never shifts the centre of gravity of a closed system. But is the universe closed? Or a system? The third part of the adventure centres on these issues.

Challenge 134 (page 83) The momentum transfer to the wall is double when the the ball rebounds perfectly.

Challenge 135 (page 83) Take the plastic cover of the cork, put the cloth around the bottle (this is for protection reasons only), and repeatedly hit the bottle on the floor or a fall in an inclined way, as shown in Figure 20 on page 66. With each hit, the cork will come out a bit.

Challenge 137 (page 83) The atomic force microscope.

Challenge 139 (page 84) Running man: $E \approx 0.5 \cdot 80\,\mathrm{kg} \cdot (5\,\mathrm{m/s})^2 = 1\,\mathrm{kJ}$; rifle bullet: $E \approx 0.5 \cdot 0.04\,\mathrm{kg} \cdot (500\,\mathrm{m/s})^2 = 5\,\mathrm{kJ}$.

Challenge 140 (page 84) The flame leans towards the inside.

Challenge 142 (page 84) It almost doubles in size.

Challenge 145 (page 84) place the tea in cups on a board, and attach the board to long ropes.

Challenge 152 (page 85) Yes; it happens twice a year. To minimize the damage, dishes should be dark in colour.

Challenge 153 (page 85) A rocket fired from the back would be a perfect defence against planes attacking from behind. However, when released, the rocket is effectively flying backwards with respect to the air, thus turns around and then becomes a danger to the plane that launched it. Engineers who did not think about this effect almost killed a pilot during the first such tests.

Challenge 155 (page 85) Whatever the ape does, whether it climbs up or down or even lets himself fall, it remains at the same height as the mass. What happens with friction at the wheel?

Challenge 159 (page 85) Yes, if he moves at a large enough angle to the direction of the boat's motion.

Challenge 161 (page 85) $\Theta = \frac{2}{5}mr^2$.

Challenge 162 (page 85) Yes. Can you imagine what happens for an observer on the equator?

Challenge 164 (page 86) The plane is described in the web sites cited; for a standing human the plane is the vertical plane containing the two eyes.

Challenge 166 (page 87) Classical or everyday nature is right-left symmetric and thus requires an even number of legs. Walking on two-dimensional surfaces naturally leads to a minimum of four legs.

Challenge 168 (page 87) Yes, the effect has been measured for skyscrapers. Can you estimate the values?

Challenge 172 (page 89) The value of the product GM for the earth is $4.0 \cdot 10^{14}\,\mathrm{m}^3/\mathrm{s}^2$.

Challenge 175 (page 90) The Atwood machine is the answer: two almost equal weights connected by a string hanging from a well-oiled wheel. The heavier one falls very slowly.

Challenge 177 (page 90) The speed is proportional to l/T, which makes it proportional to $l^{1/2}$.

Challenge 178 (page 91) He suspended a horizontal handle with a long metal wire. He then approached a large mass to the handle and noted how much the handle rotated.

Challenge 188 (page 97) This is a resonance effect, in the same way that a small vibration of a string can lead to large oscillation of the air and sound box in a guitar.

Challenge 190 (page 99) The total angular momentum of the earth and the moon must remain constant.

Challenge 195 (page 101) The centre of mass falls with the usual acceleration; the end thus falls faster.

Challenge 196 (page 102) Your weight is roughly constant; thus the earth must be round. On a flat earth, the weight would change from place to place.

Challenge 199 (page 102) That is the mass of the earth. Just turn the table on its head.

Challenge 202 (page 102) The moon will be about 1.25 times as far as it is now. The sun then will slow down the earth-moon system rotation, this time due to the much smaller tidal friction from the sun's deformation. As a result, the moon will return to smaller and smaller distances to earth. However, the sun will have become a red giant by then, and have swallowed earth and moon.

Challenge 204 (page 102) As Galileo determined, for a swing (half a period) the ratio is $\sqrt{2}/pi$. Not more than two, maybe three decimals of π can be determined this way.

Challenge 205 (page 102) Momentum conservation is not a hindrance, as any tennis racket has the same effect on the tennis ball.

Challenge 207 (page 102) This question is old (it was already asked in Newton's times) and deep. One reason is that stars are kept apart by rotation around the galaxy. The other is that galaxies are kept apart by the momentum they got in the big bang. Without the big bang, all stars would have collapsed together. In this sense, the big bang can be deduced from the attraction of gravitation and the immobile sky at night. We shall find out later that the darkness of the night sky gives a second argument for the big bang.

Challenge 208 (page 103) Due to the plateau, the effective mass of the earth is larger.

Challenge 209 (page 103) The choice is clear once you notice that there is no section of the orbit which is concave towards the sun. Can you show this?

Challenge 210 (page 103) It would be a black hole; no light could escape. Black holes are discussed in detail in the chapter on general relativity.

Challenge 224 (page 105) True.

Challenge 228 (page 105) Never. The moon points always towards the earth. The earth changes position a bit, due to the ellipticity of the moons orbit. Obviously, the earth shows phases.

Challenge 230 (page 105) What counts is local verticality; with respect to it, the river always flows down.

Challenge 1076 (page 629) There are no such bodies, as the chapter of general relativity will show.

Challenge 212 (page 103) A handle of two bodies.

Challenge 214 (page 103) Using a maximal jumping height of $h = 00.5$ m on earth and an estimated asteroid density of $\rho = 3\ \mathrm{Mg/m^3}$, we get a maximum radius of $R^2 = 3gh/4\pi G\rho \approx 703\,\mathrm{m}$.

Challenge 233 (page 106) The time is the same for all such tunnels, and thus in particular it is the same as for the pole to pole tunnel.

Challenge 235 (page 108) The electricity consumption of a rising escalator indeed increases when the person on it walks in the same direction. By how much?

Challenge 239 (page 111) True?

Challenge 244 (page 113) The light mill is an example. See page 400

Challenge 245 (page 113) Electric charge.

Challenge 246 (page 113) If you have found reasons to answer yes, you overlooked something. Just go into more details, and check whether the concepts you used apply to the universe. Also define carefully what you mean by 'universe'.

Challenge 249 (page 115) This connection shall become important in the third part of our adventure.

Challenge 248 (page 115) A system showing energy or matter motion faster than light would imply that for such systems there are observers for which the order between cause and effect are reversed. A space-time diagram (and a bit of exercise from the section on special relativity) shows this.

Challenge 253 (page 117) Of course; moral laws are summaries of what others think or will do about personal actions.

Challenge 254 (page 117) Space-time is defined using matter; matter is defined using space-time.

Challenge 255 (page 117) Fact is that physics has been based on a circular definition for hundreds of years. Thus it is possible to build even an exact science on sand. Nevertheless, the elimination of the circularity is an important aim.

Challenge 261 (page 130) The water is drawn up along the sides of the spinning egg. The fastest way to empty a bottle of water is to spin the water while emptying it.

Challenge 271 (page 136) Do not despair. Up to now, nobody has been able to imagine a universe (that is not necessarily the same as a 'world') different from the one we know. So far, such attempts have always led to logical inconsistencies.

Challenge 280 (page 142) The relation is

$$\frac{c_1}{c_2} = \frac{\sin\alpha_1}{\alpha_2} \tag{660}$$

The particular speed ratio between air (or vacuum) and a material gives the *index of refraction* n:

$$n = \frac{c_1}{c_o} = \frac{\sin\alpha_1}{\alpha_o} \tag{661}$$

Challenge 281 (page 142) Gases are mainly made of vacuum. Their index of refraction is near to one.

Challenge 282 (page 142) Diamonds also sparkle because they work as prisms; different colours have different indices of refraction. Thus their sparkle is also due to their dispersion; therefore it is a mix of all colours of the rainbow.

Challenge 285 (page 143) The universe is not a system. This topic will be discussed in detail later on.

Challenge 289 (page 146) The integers and addition form a group. Does a painter's set of oil colours with the operation of mixing form a group?

Challenge 311 (page 159) Waves can be damped to extremely low intensities. If this is not possible, the observation is not a wave.

Challenge 315 (page 162) The sound of thunder or of car traffic gets lower and lower in frequency with increasing distance.

Challenge 314 (page 162) No, as it does not move. It is an oscillation. However, it can be seen as the superposition of two waves travelling in opposite directions.

Challenge 316 (page 162) Neither; both possibilities are against the properties of water: in surface waves, the water molecules move in circles.

Challenge 318 (page 163) To reduce noise reflection and thus hall effects. They effectively diffuse the arriving wave fronts.

Challenge 319 (page 163) Waves in a river are never elliptical; they remain circular.

Challenge 339 (page 171) None.

Challenge 322 (page 163) The sun is always on a different position than the one we observe it to be. What is the difference, measured in angular diameters of the sun?

Challenge 325 (page 164) An overview of presently tested systems can be found in *Oceans of electricity - new technologies convert the motion of waves into watts*, Science News **159**, pp. 234–236, April 2001.

Challenge 329 (page 166) For jumps of an animal of mass m the necessary energy E is given as $E = mgh$, and the work available to a muscle is roughly speaking proportional to its mass $W \sim m$. Thus 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 $1.5 \pm 0.7\,\mathrm{J/kg}$.

Challenge 331 (page 167) The critical height for a column of material is given by $h_{\text{crit}}^4 = \frac{\beta}{4\pi g} m \frac{E}{\rho^2}$, where $\beta \approx 1.9$ is the constant determined by the calculation when a column buckles und its own weight.

Challenge 332 (page 170) No metal wire allows to build such a long wire. Only the idea of carbon nanotubes has raised the hope again; some dream of wire material based on them, stronger than any material known so far. However, no such material is known yet. The system faces many dangers, such as fabrication defects, lightning, storms, meteorites and space debris. All would lead to the breaking of the wires – if such wires will ever exist. But the biggest of all dangers is the lack of cash to build it.

Challenge 334 (page 170) This argument is comprehensible only when one remembers that 'twice the amount' means 'twice as many molecules'.

Challenge 335 (page 170) The third component of air is the noble gas argon, making up about 1%. The rest is made up by carbon dioxide, water vapour and other gases. Are these percentages volume or weight percentages?

Challenge 336 (page 170) Yes. The bulb will not resist two such cars though.

Challenge 317 (page 163) Swimmers are able to cover 100 m in 48 s, or slightly better than 2 m/s. With a body length of about 1.9 m, the critical speed is 1.7 m/s. That is why short distance swimming depends on training; for longer distances the technique plays a larger role, as the critical speed has not been attained yet. The formula also predicts that on the 1500 m distance, a 2 m tall swimmer has a potential advantage of over 45 s on one with body height 1.8 m. If one adds the fact that longer swimmers swim shorter distances (why?), it is predicted that successful swimmers will get taller and taller over time.

Challenge 354 (page 175) The answer depends on the size of the balloons, as the pressure is not a monotonous function of the size. If the smaller balloon is not too small, the smaller balloon wins.

Challenge 344 (page 171) The blood pressure in the feet of a standing human is about 27 kPa, double the pressure at the heart.

Challenge 345 (page 171) The soap flows down the bulb, making it thicker at the bottom and thinner at the top, until it bursts.

Challenge 347 (page 171) A stalactite contains a thin channel along its axis through which the water flows, whereas a stalagmite is massive throughout.

Challenge 348 (page 171) About 1 part in a thousand.

Challenge 349 (page 171) For this to happen, friction would have to exist on the microscopic scale, and energy would have to disappear.

Challenge 350 (page 171) The longer funnel is empty before the short one. Energy conservation yields $P/\rho + gh + v^2/2 = \text{const.}$ thus the speed v is higher for greater heights h of the funnel.

Challenge 359 (page 177) Einstein argued that dissipation and fluctuations are related, as they are both due to the effects of atoms. The kinematic viscosity ν (measuring dissipation) has thus to be proportional to the diffusion constant D (measuring fluctuations). Dissipation is usually described by the dynamic viscocity $\eta = \nu/\rho$, related through the fluid's mass density. Using the ideal gas law we thus have

$$D \sim \frac{\eta R T}{N P} \tag{662}$$

With the definition of D we get, using the square displacement λ,

$$N \sim \frac{\eta R T}{\lambda^2 N P} \quad . \tag{663}$$

Challenge 369 (page 184) Hot air is less dense, and thus wants to rise.

Challenge 372 (page 184) In general, it is impossible to draw a line through three points.

Challenge 373 (page 184) No, as a water molecule is heavier than that. However, if the water is allowed to be dirty, it is possible. What happens if the quantum of action is taken into account?

Challenge 374 (page 184) Keep the paper wet.

Challenge 375 (page 184) The danger is not due to the amount of energy, but due to the time in which it is available.

Challenge 376 (page 184) The internet is full of solutions.

Challenge 378 (page 184) Only if it is a closed system. Is the universe closed? Is it a system? This is discussed in the third part of the mountain ascent.

Challenge 381 (page 185) For such small animals the body temperature would fall too low. They could not eat fast enough to get the energy needed to keep themselves warm.

Challenge 390 (page 185) It is about 10^{-9} that of the earth.

Challenge 395 (page 186) The vortex in the tube is cold in the middle and hot at its outside; the air from the middle is sent to one end and the air from the outside to the other. The heating of the outside is due to the work that the air rotating inside has to do on the air outside to get a rotation that eats up angular momentum. For a detailed explanation, see the beautiful text by MARK P. SILVERMAN, *And yet it moves: strange systems and subtle questions in physics*, Cambridge University Press, 1993, p. 221.

Challenge 403 (page 190) There are many more butterflies than tornadoes. In addition, the belief in the butterfly effect completely neglects an aspect of nature that is essential for self-organisation: friction and dissipation.

Challenge 408 (page 193) All three statements are hogwash. A drag coefficient implies that the cross area of the car is known to the same precision. This is actually extremely difficult to measure and to keep constant. In fact, the value 0.375 for the Ford Escort was a cheat, as many other measurements showed. The fuel consumption is even more ridiculous, as it implies that fuel volumes and distances can be measured to that same precision. Opinion polls are taken by phoning at most 2000 people; due to the difficulties in selecting the right representative sample, that gives a precision of at most 3%.

Challenge 412 (page 199) A cone or a hyperboloid also look straight from all directions, provided the positioning is correct. The best method to check planarity is to use interference between an arriving and a departing coherent beam of light. If the fringes are straight, the surface is planar.

Challenge 413 (page 199) A fraction of infinity is still infinite.

Challenge 414 (page 201) Otherwise the velocity sum would be larger than c.

Challenge 415 (page 201) The drawing shows it.

Challenge 416 (page 201) There are cat-eyes on the moon; they are used to reflect laser light pulses sent there through telescopes. The timing then gives the moon distance. Of course, absolute distance is not know to high precision, but the variations are. The thickness of the atmosphere is the largest source of error.

Challenge 417 (page 201) Fizeau used a mirror about 4 km away, so that he could measure the timing by a simple shutter mechanism: the teeth of a turning wooden wheel. He only needed to measure the rotation speed and the geometry of the wheel.

Challenge 418 (page 202) The time must be shorter than $T = l/c$, in other words, shorter than 30 ps; it was a *gas* shutter, not a solid one. It was triggered by a light pulse extracted from the one to be photographed; for certain materials, such as the used gas, strong light can lead to bleaching, so that they become transparent. For more details about the shutter and its neat trigger technique, see the paper by the authors.

Challenge 419 (page 202) Just take a photograph of a lightning while moving the camera horizontally. You will see that a lightning is made of several discharges; the whole shows that lightning is much slower than light.

Challenge 423 (page 204) The motion of gravity, of gluons, and possibly of neutrinos.

Challenge 425 (page 205) $\lambda_R/\lambda_S = \gamma$.

Challenge 428 (page 207) z is determined using the spectral lines or Fraunhofer lines. The speeds are: $v(z = -0.1) = 31\,\mathrm{Mm/s} = 0.1c$ *towards* the observer and $v(z = 5) = 284\,\mathrm{Mm/s} = 0.95c$ *away* from the observer.

See page 548

Challenge 426 (page 205) To change from bright red (650 nm) to green (550 nm), $v = 0.166c$ is necessary.

Challenge 430 (page 208) Inside colour television tubes (they use higher voltages than black and white ones), electrons are described by $v/c \approx \sqrt{2\cdot 30/511}$ or $v \approx 0.3c$.

Challenge 431 (page 208) If you can imagine this, publish it immediately.

Challenge 437 (page 211) The human value is achieved in particle accelerators; the value in nature is found in cosmic rays of the highest energies.

Challenge 447 (page 216) Not with present experimental methods.

Challenge 455 (page 217) Proper velocity can only be defined for observers, i.e., for entities which can carry a clock. That is not the case for images.

Challenge 460 (page 222) Relativity makes the arguments of challenge 1062 watertight.

Challenge 465 (page 224) Annihilation of matter and antimatter.

Challenge 474 (page 229) Probably not, as all relations among physical quantities are known now. However, you might check for yourself...

Challenge 476 (page 231) Any motion with light speed.

Challenge 505 (page 243) Yes, it is true.

Challenge 507 (page 243) Yes; however, the effect is minimal and depends on the position of the sun. In fact, what is white at one height is not white at another.

Challenge 509 (page 244) Locally, light moves with speed c.

Challenge 510 (page 244) Away from earth, g decreases; it is effectively zero over most of the distance.

Challenge 511 (page 244) Light is necessary to determine distance *and* to synchronize clocks; thus there is no way to measure the speed of light from one point to another alone; the reverse motion needs to be included.

Challenge 512 (page 245) See the cited reference. The factor 2 was forgotten there; can you deduce it?

Challenge 256 (page 120) For example, speed inside materials is slowed, but between atoms, light still travels with vacuum speed.

Challenge 257 (page 125) Use the electrostatic and the nuclear repulsion to calculate the distance.

Challenge 516 (page 249) There is not enough time to send the signal to the battery that contact is made, so that the current cannot start flowing.

Challenge 1135 (page 744) They are accelerated upwards.

Challenge 534 (page 259) The energy due to the rotation can be neglected compared with all other energies in the problem.

Challenge 543 (page 264) Different nucleons, different nuclei, different atoms and different molecules have different percentages of binding energies relative to the total mass.

Challenge 550 (page 265) After about half an hour.

Challenge 554 (page 266) With $\hbar$ as smallest angular momentum one get about 100 Tm.

Challenge 555 (page 266) No. The diffraction of the beams does not allow it. Also quantum theory makes this impossible; bound states of massless particles, such as photons, are not stable.

Challenge 557 (page 267) The orbital radius is 4.2 earth radii; that makes ca. 38 μs every day.

Challenge 559 (page 267) Of course! Other spatial dimensions could exist which can be detected only with help of measurement apparatuses. For example, hidden dimensions could appear at energies not accessible in everyday life.

Challenge 570 (page 273) Since there is no negative mass, gravitoelectric fields cannot be neutralized, as is possible e.g. around a metallic conductor.

Challenge 598 (page 284) No; a line cannot have intrinsic curvature. A torus is indeed intrinsically curved; it cannot be cut open to a flat sheet of paper.

Challenge 635 (page 299) Indeed, in general relativity gravitational energy cannot be localized in space, in contrast to what one expects and requires from an interaction.

Challenge 650 (page 313) The rabbit observes that all other rabbits seem to move away from him.

Challenge 688 (page 338) The hollow earth theory is correct if usual distance are consistently changed to $r_{\mathrm{he}} = R^2_{\mathrm{earth}}/r$. This implies a quantum of action that decreases towards the centre of the hollow sphere. Then there is no way to prefer one description over the other, except for reasons of simplicity.

Challenge 714 (page 348) This happens in the same way that the static electric field comes out of a charge. In both cases, the transverse fields do not get out, but the longitudinal fields do. Quantum theory provides the deeper reason. Real radiation particles, which are responsible for free, transverse fields, cannot leave a black hole because of the escape velocity. However, virtual particles can, as their speed is not bound by the speed of light. All static, longitudinal fields are produced by virtual particles. In addition, there is a second reason. Classical field can come out of a black hole because for an outside observer everything making it up is continuously falling, and nothing has actually crossed the horizon. The field sources thus are not yet out of reach.

Challenge 717 (page 348) The description says it all. A visual impression can be found in the room on black holes in the 'Deutsches Museum' in München.

Challenge 729 (page 368) The liquid drops have to detach from the flow exactly inside the metal counterelectrodes. Opel simply earthed the metal piece they had built into the cars without any contact to the rest of the car.

Challenge 730 (page 369) A lot of noise while banging up and down.

Challenge 733 (page 371) A simple geometrical effect: anything flowing out homogeneously from a sphere diminishes with the square of the distance.

Challenge 735 (page 372) No; they only separate charges and pump them around.

Challenge 736 (page 372) Uncharged bodies can attract each other if they are made of charged constituents neutralizing each other, and if the charges are constrained in their mobility. The charge fluctuations then lead to attraction. Most molecules interact among each other in this way; such forces are also at the basis of surface tension in liquids and thus of droplet formation.

Challenge 738 (page 374) See challenge 430.

Challenge 739 (page 374) Th electrons move slowly, but the speed of electrical signals is given by the time at which the electrons move. Imagine long queue of cars (representing electrons) waiting in front of a red traffic light. All drivers look at the light. As soon as it turns green, everybody starts driving. Even though the driving speed might be only $10\,\mathrm{m/s}$, the speed of traffic flow onset was that of light. It is this latter speed which gives the speed of electrical signals.

Challenge 771 (page 385) Field lines and equipotential surfaces are always orthogonal to each other. Thus a field line cannot cross an equipotential surface twice.

Challenge 770 (page 384) Some momentum is carried away by the electromagnetic field.

Challenge 781 (page 389) Other asymmetries in nature include the helicity of the DNA molecules making up the chromosomes and many other molecules in living systems, the right hand preference of most humans, the asymmetry of fish species which usually stay flat on the bottom of the seas.

Challenge 782 (page 390) This is not possible at all using gravitational or electromagnetic systems or effects. The only way is to use the weak nuclear interaction, as shown in the chapter on the nucleus.

Challenge 784 (page 391) Imagine E and B as the unite vectors of two axes in complex space. Then any rotation of these axes is also a generalized duality symmetry.

Challenge 787 (page 393) In every case of interference, the energy is redistributed into other directions. This is the general rule; sometimes it is quite tricky to discover this other direction.

Challenge 790 (page 393) He noted that when a prism produces a rainbow, a thermometer placed in the region after the colour red shows a temperature rise.

Challenge 796 (page 398) Such an observer would experience a wavy but static field, which cannot exist, as the equations for the electromagnetic field show.

Challenge 797 (page 398) Syrup shows an even more beautiful effect in the following setting. Take a long transparent tube closed at one end, and fill it with syrup. shine a red Helium-Neon laser into the tube from the bottom. Then introduce a linear polarizer into the beam: the light seen in the tube will form a spiral. By rotating the polarizer you can make the spiral advance or retract. This effect, called the *optical activity* of sugar,, is due to the ability of sugar to rotate light polarization and to the fact that plants make only one of the two forms of handed sugar.

Challenge 798 (page 398) The 1 mm beam would return 1000 times wider than the 1 m beam.

Challenge 803 (page 400) A surface of $1\,\mathrm{m}^2$ perpendicular to the light receives about $1\,\mathrm{kW}$ of radiation. That makes the same pressure as produced by the weight of about $0.3\,\mathrm{mg}$ of matter.

Challenge 804 (page 400) The shine side gets twice the momentum transfer as the black side, and thus should be pushed backwards.

Challenge 808 (page 401) A polarizer can do this.

Challenge 811 (page 402) The interference patterns change when colours are changed. Rainbows also appear because different colours are due to different frequencies.

Challenge 815 (page 404) Film a distant supernova explosion and check whether it happens at the same time for each colour separately.

Challenge 817 (page 405) It would imply an energy speed higher than that of light. See page 406.

Challenge 819 (page 406) The light is pulsed; thus it is the energy velocity.

Challenge 820 (page 406) Inside matter, the energy is transferred to atoms, than back to light, to the next atoms, etc. That takes time and slows down the propagation.

Challenge 822 (page 408) This is true even in general relativity, when the bending of the vacuum is studied.

Challenge 752 (page 380) Move them to form a T shape.

Challenge 753 (page 380) For four and more switches, on uses inverters; an inverter is a switch with two inputs and two outputs which in one position, connects first and second input to first and second output respectively, and in the other position connects the first input to the second output and vice versa. (There are other possibilities, though; wires can be saved using electromagnetic relay switches.) For three switches, there is a simpler solution than with inverters.

Challenge 754 (page 381) Glue two mirrors together at a right angle. Or watch yourself on TV using a video camera.

Challenge 755 (page 381) The image flips up: a 90 degree rotation turns the image by 180 degrees.

Challenge 756 (page 381) This is again an example of combined triboluminescence and triboelectricity. See also http://scienceworld.wolfram.com/physics/Triboluminescence.com or http://www.geocities.com/RainForest/9911/tribo.htm

Challenge 758 (page 381) The angular size of the sun is too large; diffraction plays no role here.

Challenge 763 (page 382) Light makes seven turns of the earth in one second.

Challenge 830 (page 409) The charges in a metal rearrange in a way that the field inside remains vanishing. This makes cars and aeroplanes safe against lightning. Of course, if the outside field varies so quickly that the rearrangement cannot follow, fields can enter the Faraday cage.

For gravity and solid cages, mass rearrangement is not possible, so that there is no gravity shield.

However, one should wait a bit before stepping out of a car after lightning has hit, as the car is on rubber wheels; the idea is to let the charge flow into the ground.

Challenge 839 (page 411) The number of photons times the quantum of action $\hbar$.

Challenge 842 (page 411) The charging stops because a negatively charged satellite repels electrons, and thus stops any electron collecting mechanism. Electrons are captured more than ions because it is easier for them to have an inelastic collision with the satellite.

Challenge 843 (page 411) Any loss mechanism will explain the loss of energy, such as electrical resistance or electromagnetic radiation. After a fraction of a second, the energy will be lost. This little problem is often discussed on the internet.

Challenge 844 (page 411) Show that even though the radial magnetic field of a spherical wave is vanishing by definition, Maxwell's equations would require it to be different from zero. Since electromagnetic waves are transversal, it is also sufficient to show that it is impossible to comb a hairy sphere without having a (double) vortex or two simple vortices. Despite these statements, quantum theory changes the picture somewhat: the emission probability of a photon from an excited atom in a degenerate state is spherically symmetric exactly.

Challenge 851 (page 417) The eye and vision system subtract constant patterns.

Challenge 845 (page 412) The human body is slightly conducting and changes the shape of the field and thus effectively short circuits it. Usually, the field cannot be used to generate energy, as the currents involved are much too small. (Lightning bolts are a different story, of course. They are due - very indirectly - to the field of the earth, but they are too irregular to be used consistently. Franklin's lightning rod is such an example.)

Challenge 853 (page 418) A hologram is always transparent; one can always see the background through the hologram. A hologram thus always gives an impression similar to what movies show as ghosts.

Challenge 866 (page 429) Indeed, the sun emits about $4 \cdot 10^{26}$ W from its mass of $2 \cdot 10^{30}$ kg, about $0.2\,\mathrm{mW/kg}$. The adult human body (at rest) emits about 100 W (you can check this in bed at night), thus about $1.2\,\mathrm{W/kg}$ per ton. This is about 6000 times more than the sun.

Challenge 879 (page 448) The issue is: is the 'universe' a concept? More about this issue in the third part of the text.

Challenge 881 (page 451) When thinking, physical energy, momentum and angular momentum are conserved, and thermodynamic entropy is not destroyed. Any experiment that this would not be so would point to unknown processes. However, there is no evidence for this.

Challenge 884 (page 452) Is it possible to use the term 'complete' when describing nature?

Challenge 889 (page 453) Yes.

Challenge 893 (page 459) Neither has a defined content, clearly stated limits, nor a domain of application.

Challenge 894 (page 459) Impossible! That would not be a concept, as it has no content. The solution to the issue must be and will be different.

Challenge 896 (page 462) The most famous is the class of all sets that do not contain themselves. This is not a set, but a class.

Challenge 898 (page 462) $(x,y) := \{x, \{x,y\}\}$.

Challenge 899 (page 463) Hint: show that any countable list of reals misses at least one number. This was proven for the first time by Cantor. His way was to write the list in decimal expansion, and then find a number that is surely not in the list. Second hint: his world-famous trick is called the 'diagonal argument'.

Challenge 900 (page 464) Hint: all reals are limits of series of rationals.

Challenge 903 (page 466) Yes.

Challenge 904 (page 466) There are infinitely many of them. But the smallest is already quite large.

Challenge 905 (page 467) $0 := \emptyset \quad , \quad 1 := \{\emptyset\} \quad , \quad 2 := \{\{\emptyset\}\}$ etc.

Challenge 906 (page 470) Subtraction is easy. Addition is not commutative only for cases when infinite numbers are involved: $\omega + 2 \neq 2 + \omega$.

Challenge 907 (page 470) Examples are $1 - \varepsilon$ or $1 - 4\varepsilon^2 - 3\varepsilon^3$.

Challenge 908 (page 471) The answer is 57; the cited reference gives the details.

Challenge 909 (page 473) $2^{2^{22}}$ and $4^{4^{4^4}}$.

Challenge 912 (page 482) 'All Cretans lie' is *false*, since the opposite, namely 'some Cretans say the truth' is true in the case given. The trap is that the opposite of the original sentence is usually, but *falsely*, assumed to be 'all Cretans say the truth'.

Challenge 913 (page 482) The statement cannot be false, due to the first half and the 'or' construction. Since it is true, the second half must be true, and you are an angel.

Challenge 914 (page 487) Only induction allows to make use of similarities, and thus to define concepts.

See page 772

Challenge 915 (page 489) Yes, as we shall find out.

Challenge 916 (page 489) Yes, as observation implies interaction.

Challenge 917 (page 490) Lack of internal contradictions means that a concept is valid as a thinking tool; as we use our thoughts to describe nature, mathematical existence is a very specialized version of physical existence, as thinking is itself a natural process. Indeed, mathematical concepts are also useful for the description of the working of computers and the like.

Another way to make the point is to stress that all mathematical concepts are built from sets and relations, or some suitable generalizations of them. These basic building blocks are taken from our physical environment. Sometimes the idea is expressed differently; many mathematicians have acknowledged that certain mathematical concepts, such as natural numbers, are taken directly from experience.

Challenge 918 (page 490) Examples are Achilles, Odysseus, mickey mouse, the gods of polytheism and spirits.

Challenge 920 (page 492) Torricelli made vacuum in a U-shaped glass tube, using mercury, the same liquid metal used in thermometers. Can you imagine how? A more difficult question: where did he get mercury from?

Challenge 921 (page 493) Stating that something is infinite can be allowed, if the statement is falsifiable. An example is the statement "There are infinitely many mosquitoes."

Other statements are not falsifiable, such as "The universe continue without limit behind the horizon." Such a statement is a belief, not a fact.

Challenge 922 (page 494) They are not sets either, and thus not collections of points.

Challenge 923 (page 495) There is still no possibility to interact with all matter and energy, as this includes oneself.

Challenge 924 (page 500) No. There is only a generalization encompassing the two.

Challenge 926 (page 500) An explanation of the universe is not possible, as the term explanation require the possibility to talk about systems outside the one under consideration. The universe is not part of a larger set.

Challenge 927 (page 502) Both can in fact be seen as two sides of the same argument: there is no other choice;there is only one possibility. The rest of nature shows that it has to be that way, as everything depends on everything.

See page 166

Challenge 934 (page 518) Due to the quantum of action, atoms in all people, be they giants or dwarfs, have the same size. That giants do not exist was shown already by Galilei. The argument is based on the given strength of materials, which thus implies that atoms are the same everywhere. That dwarfs cannot exist is due to the same reason; nature is not able to make people smaller than usual (except in the womb) as this would require smaller atoms.

Challenge 967 (page 543) The quantum of action implies that two subsequent observations always differ. Thus the surface of a liquid cannot be at rest.

Challenge 977 (page 553) The difficulties to see hydrogen atoms are due to their small size and their small number of electrons. As a result, hydrogen atoms produce only weak contrasts in X-ray images. For the same reasons it is difficult to image them with electrons; the Bohr radius of hydrogen is only slightly larger than the electron Compton wavelength.

Challenge 982 (page 554) $r = 86\,\mathrm{pm}$, thus $T = 12\,\mathrm{eV}$. That compares to the actual value of $13.6\,\mathrm{eV}$. The trick for the derivation of the formula is to use $< \psi|r_x^2|\psi > = \frac{1}{3} < \psi|\mathbf{rr}|\psi >$, a relation valid for states with no orbital angular momentum. It is valid for all coordinates and also for the three momentum observables, as long as the system is nonrelativistic.

Challenge 984 (page 554) The fields are crated by neutrons or protons, which have a smaller Compton wavelength.

Challenge 985 (page 559) Point particles cannot be marked; nearby point particles cannot be distinguished, due to the quantum of action.

Challenge 992 (page 561) For a large number of particles, the interaction energy will introduce errors. For very large numbers, the gravitational binding energy will do so as well.

Challenge 999 (page 568) In the way their intestines are folded, in the lines of their hands and other skin lines; often features like black points on the skin are mirror inverted on the two twins.

Challenge 1003 (page 573) Angels can be distinguished by name, can talk and sing; thus they are made of a large number of fermions. In fact, many angels are human sized, so that they do not even fit on the tip of a pin.

Challenge 1011 (page 577) Ghosts, like angels, can be distinguished by name, can talk and can be seen; thus they contain fermions. However, they can pass through walls and they are transparent; thus they cannot be made of fermions, but must be images, made of bosons. That is a contradiction.

Challenge 1023 (page 598) The moon is on contact with baths like the solar wind, the infalling meteorites, the electromagnetic background radiation of the universe, the neutrino flux from the sun, etc.

Challenge 1019 (page 591) Such a computer requires clear phase relations between components; such phase relations are extremely sensitive to outside disturbances.

Challenge 1025 (page 602) A virus is an example. It has no own metabolism. (By the way, the fact that some viruses can form crystals is not a proof that they are not living beings, in contrast to what is often said.)

Challenge 1026 (page 603) The navigation systems used by flies are an example.

Challenge 1027 (page 604) The thermal energy kT is about $4\,\mathrm{zJ}$ and a typical relaxation time is $0.1\,\mathrm{ps}$.

Challenge 1028 (page 607) This is not possible at present. If you know a way, publish it. It would help a sad single mother who has to live without financial help from the father, despite a lawsuit, as it was yet impossible to decide which of the two candidates is the right one.

Challenge 1031 (page 607) Well, men are more similar to chimpanzees than to women. More seriously, the above data, even though often quoted, is wrong. Newer measurements by Roy Britten in 2002 have shown that the difference in genome between humans and chimpanzees is about 5%. (PNAS Online) In addition, the difference between man and woman is smaller than one whole chromosome. Newest measurements suggest that all humans have at least 99.9% of common genes.

Challenge 1035 (page 608) Chemical processes, including diffusion and reaction rates, are strongly temperature dependent. They affect the speed of motion of the individual and thus its chance of survival. Keeping temperature in the correct range is thus important for evolved life forms.

Challenge 1036 (page 608) Haven't you tried yet? Physics is an experimental science.

Challenge 1038 (page 610) Radioactive dating methods can be said to be based on the nuclear interactions, even though the detection is again electromagnetic.

Challenge 1049 (page 616) With a combination of the methods of Table 49 it is possible; but whether there will ever be an organisation willing to pay for this to happen is another question.

Challenge 1051 (page 617) For example, a heavy mountain will push down the earth's crust into the mantle, making it melt again on the bottom side, and thus lowering the position of the top.

Challenge 1054 (page 617) Since the height of the potential is always finite, walls can always be overcome by tunnelling.

Challenge 1055 (page 617) The lid of a box can never be at rest, as is required for a tight closure, but is always in motion, due to the quantum of action.

Challenge 1066 (page 626) For example, you could change gravity between two mirrors.

Challenge 1068 (page 626) Echoes do not work one the speed of sound is reached, and do not work well when it is approached. Both the speed of light and that of sound have a finite value. Moving with a mirror still gives a mirror image. This means that the speed of light cannot be reached. If it cannot be reached, it must be the same for all observers.

Challenge 1069 (page 627) Mirrors do not usually work for matter; in addition, if they did, matter would require much higher acceleration values.

Challenge 1071 (page 628) The overhang can have any value whatsoever. There is no limit. Taking the indeterminacy principle into account introduces a limit as the last brick or card must not allow the centre of gravity, through its indeterminacy, to be over the edge of the table.

Challenge 1072 (page 628) A larger charge would lead to field that spontaneously generate electron positron pairs, the electron would fall into the nucleus and reduce its charge by one unit.

Challenge 1075 (page 629) The Hall effect results from the deviation of electrons in a metal due to an applied magnetic field. Therefore it depends on their speed. One gets values around 1 mm. Inside atoms, one can use Bohr's atomic model as approximation.

Challenge 1076 (page 629) The usual way to pack oranges on a table is the densest way to pack spheres.

Challenge 1077 (page 630) Just use a paper drawing. Draw a polygon, and draw it again at latter times, taking into account how the sides grow over time. You will see by yourself how the faster growing sides disappear over time.

Challenge 1079 (page 630) Mud is a suspension of sand; sand is not transparent, even if made of clear quartz, because of the scattering of light at the irregular surface of its grains. A suspension cannot be transparent if the index of refraction of the liquid and the suspended particles is different. It is never transparent if the particles, as in most sand types, are themselves not transparent.

Challenge 1080 (page 631) No. Bound states of massless particles are always unstable.

Challenge 1082 (page 631) Methods to move on perfect ice from mechanics:

- do nothing, just wait that the higher centrifugal acceleration at body height pulls you away;
- to rotate yourself, just rotate your arm above your head;
- throw a shoe or any other object away;
- breathe in vertically, breathing out (or talking) horizontally (or vice versa);
- wait to be moved by the centrifugal acceleration due to the rotation of the earth (and its oblateness);

- jump vertically repeatedly: the Coriolis acceleration will lead to horizontal motion;
- wait to be moved by the sun or the moon, like the tides are;
- ‘swim’ in the air using hands and feet;
- wait to be hit by a bird, a flying wasp, inclined rain, wind, lava, earthquake, plate tectonics, or any other macroscopic object (all objects pushing count only as one solution);
- wait to be moved by the change in gravity due to convection in earth’s mantle;
- wait to be moved by the gravitation of some comet passing by;
- counts only for kids: spit, sneeze, cough, fart, pee; or move your ears and use them as wings. Note that gluing your tongue is not possible on perfect ice.

Challenge 1083 (page 631) Methods to move on perfect ice using thermodynamics and electrodynamics

- use the radio/tv stations to push you around;
- use your portable phone and a mirror;
- switch on a pocket lam, letting the light push you;
- wait to be pushed around by Brownian motion in air;
- heat up one side of your body: black body radiation will push you;
- heat up one side of your body, e.g. by muscle work: the changing airflow or the evaporation will push you;
- wait for one part of the body to be cooler than the other, and the corresponding black body radiation effects;
- wait for the magnetic field of the earth to pull on some ferromagnetic or paramagnetic metal piece in your clothing or in your body;
- wait to be pushed by the light pressure, i.e. by the photons, from the sun or from the stars, maybe using a pocket mirror to increase the efficiency;
- rub some polymer object to charge it electrically, and then move it in circles, thus creating a magnetic field that interacts with the one of the earth.

Note that perfect frictionless surfaces do not melt.

Challenge 1084 (page 631) Methods to move on perfect ice using quantum effects:

- wait for your wavefunction to spread out and collapse at the end of the ice surface;
- wait for the pieces of metal in the clothing to attract to the metal in the surrounding through the Casimir effect;
- wait to be pushed around by radioactive decays in your body.

Challenge 1085 (page 631) Methods to move on perfect ice using general relativity:

- move an arm to emit gravitational radiation;
- deviate the cosmic background radiation with a pocket mirror;
- wait to be pushed by gravitational radiation from star collapses;
- wait to the universe to contract.

Challenge 1086 (page 631) Methods to move on perfect ice using material science, geophysics, astrophysics:

- be pushed by the radio waves emitted by thunderstorms and absorbed in painful human joints;
- wait to be pushed around by cosmic rays;
- wait to be pushed around by the solar wind;
- wait to be pushed around by solar neutrinos;
- wait to be pushed by the transformation of the sun into a red giant;
- wait to be hit by a meteorite.

Challenge 1087 (page 631) A method to move on perfect ice using selforganisation, chaos theory, and biophysics:

- wait that the currents in the brain interact with the magnetic field of the earth by controlling your thoughts.

Challenge 1088 (page 631) Methods to move on perfect ice using quantum gravity, supersymmetry and string theory:

- accelerate your pocket mirror with your hand;
- deviate the Unruh radiation of the earth with a pocket mirror;
- wait for proton decay to push you through the recoil.

Challenge 1098 (page 641) The universe has about 10^{22} stars; the sun has a luminosity of about 10^{26} W; the total luminosity of the visible matter in the universe is thus about 10^{48} W. A gamma ray burster emits up to $3 \cdot 10^{47}$ W.

Challenge 1103 (page 642) They are carried away by the gravitational radiation.

Challenge 1113 (page 646) No system is known in nature which emits or absorbs only one graviton at a time. This is another point speaking against the existence of gravitons.

Challenge 1119 (page 694) Most macroscopic matter properties fall in this class, such as the change of water density with temperature.

Challenge 1122 (page 706) Never expect a correct solution for personal choices. Do what you yourself think and feel is correct.

Challenge 1127 (page 716) The infinite sum is not defined for numbers; however, it is defined for a knotted string.

Challenge 1135 (page 744) Accelerators scale in size with the ...power of energy. Thus one would get ...However, this assumes that gravity plays no effect. Taking gravity into account is not simple. In any case the highest energy would be limited to that energy at which matter and radiation can be distinguished, as this distinction is necessary to build accelerators. Therefore a Planck energy accelerator is impossible; the best possible is a unification energy accelerator.

Nature shows no accelerator of this power; cosmic rays always have much smaller energy. The maximum known is about one thousandth of this value. Black holes are no sources for unification energy particles, due to the gravitational potential. Only the cosmic horizon could be such a source. For some yet unclear reasons, this is not the case.

Challenge 1136 (page 745) The Planck energy is $E_{\mathrm{Pl}} = \sqrt{\hbar c^5/G} = 2.0\,\mathrm{GJ}$. Car fuel delivers about 43 MJ/kg. Thus the Planck energy corresponds to the energy of 47 kg of car fuel, about a tankful.

Challenge 1137 (page 745) Not really, as the mass error is equal to the mass only in the Planck case.

Challenge 1140 (page 745) There is no gravitation at those energies and there are no particles. There is thus no paradox.

Challenge 1142 (page 746) The Planck acceleration is given by $a_{\mathrm{Pl}} = \sqrt{c^7/\hbar G} = 5.6 \cdot 10^{51}\,\mathrm{m/s^2}$.

Challenge 1144 (page 746) The energy is the unification energy, about 800 times smaller than the Planck energy.

Challenge 1145 (page 751) Good! Publish it.

Challenge 1180 (page 779) Any change in rotation speed of the earth would change the sea level.

Challenge 1181 (page 780) Just measure the maximum water surface the oil drop can cover, by looking at the surface under a small angle.

Challenge 1182 (page 781) Keep the fingers less than 1 cm from your eye.

Challenge 1184 (page 788) As vacuum and matter cannot be distinguished, both share the same properties. In particular, both scatter strongly at high energies.

Challenge 1187 (page 799) The number of spatial dimensions must be given first, in order to talk about spheres.

Challenge 1194 (page 815) The lid of a box must obey the indeterminacy relation. It cannot be at perfect rest with respect to the rest of the box.

Challenge 1195 (page 815) Of course not, as there are no infinite quantities in nature. The question is whether the detector would be as large as the universe or smaller. What is the answer?

Challenge 1200 (page 815) Yes, as nature's inherent measurement errors cannot clearly distinguish among them.

Challenge 1201 (page 815) Of course.

Challenge 1202 (page 815) We still have the chance to find the best approximate concepts possible. There is no reason to give up.

Challenge 1208 (page 864) That would be a beautiful example of a logical error. (Just imagine a foreign writing system explaining itself in this way.) The debates about the pronunciation of ancient greek also shows the problems. The definitions of pronunciations found in dictionaries avoid this issue by referring to the memory of pronounced words or to sounds found in nature.

Challenge 1214 (page 880) Do not forget the relativistic time dilation.

Challenge 1216 (page 880) Since the temperature of the triple point of water is fixed, the temperature of the boiling point is fixed as well. Historically, the value of the triple point has not been well chosen.

Challenge 1217 (page 881) Probably the quantity with the biggest variation is mass, where a prefix for $1\,\mathrm{eV}/c^2$ would be useful and one for the total mass in the universe, which is about 10^{90} times larger.

Challenge 1218 (page 881) The formula with $n-1$ is a better fit. Why?

Challenge 1220 (page 886) The double of that number, the number made of the sequence of all even numbers, etc.

Challenge 1215 (page 880) About $10\,\mu\mathrm{g}$.

Challenge 1222 (page 888) There are still many discoveries to be made in modern mathematics, especially in topology, number theory and algebraic geometry. Mathematics has a good future.

Challenge 1224 (page 892) The gauge coupling constants determine the size of atoms, the strength of chemical bonds, and thus the size of all things.

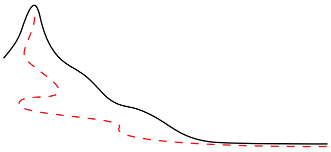
Challenge 1225 (page 902) Covalent bonds tend to produce full shells; this is a smaller change on the right side of the periodic table.

Challenge 1238 (page 912) 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?

Challenge 1241 (page 914) The metric is regular, positive definite and obeys the triangle inequality.

Challenge 1243 (page 916) The solution is the set of all two by two matrices, as each two by two matrix specifies a linear transformation, if one defines a transformed point as the product of the point and this matrix. (Only multiplication with a fixed matrix can give a linear transformation.) Can you recognize from a matrix wether it is a rotation, a reflection, a dilation, a shear, or a stretch along two axes? What are the remaining possibilities?

So far, out of 1247 challenges, 391 solutions are given; in addition, 130 challenges require no entry in this list.



MOTION MOUNTAIN

Hiking beyond space and time
along the concepts of modern physics

To the kind reader

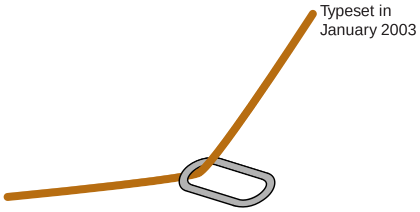
In exchange for getting this section for free, I ask you to send a short email that comments on one or more the following:

- What was hard to understand?
- Did you find any mistakes?
- What figure or explanation were you expecting?
- Which page was boring?

Of course, any other suggestions are 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. If you send a particularly useful contribution (send it in English, Italian, Dutch, German, Spanish, Portuguese or French) you will be mentioned in the foreword of the text, or receive a small reward, or both.

Enjoy!

Christoph Schiller
mm@motionmountain.net



Appendix I Subject and Name Index

Typeset in
January 2003

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. The index thus also works as glossary.

...
ma la religione di voi è qui
e passa
di generazione in generazione
ammonendo
che SCIENZA È LIBERTÀ.

Giosuè 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 inscribed the entry hall of the University of Bologna, the oldest university of the world.