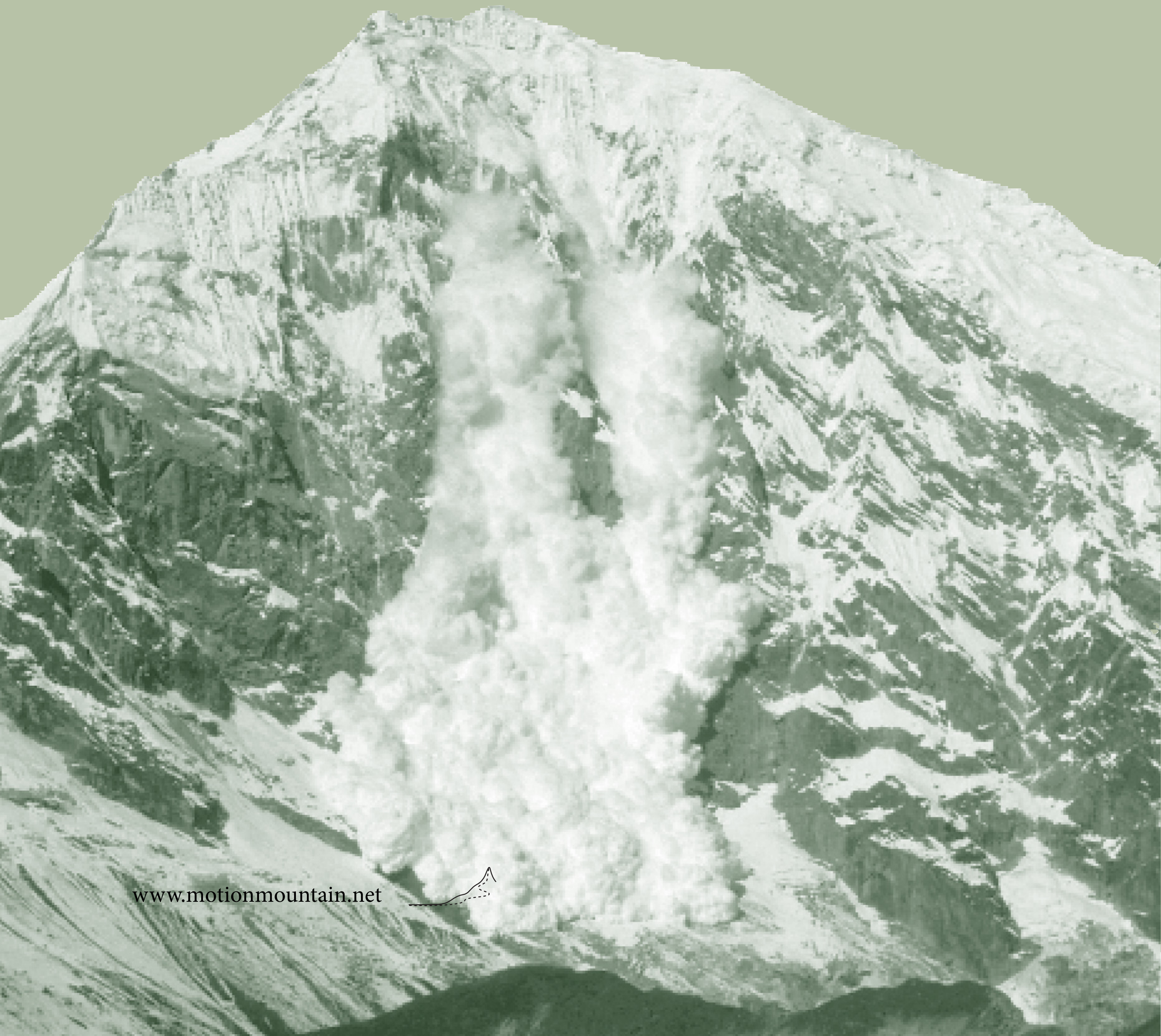
Christoph Schiller

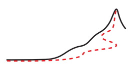
MOTION MOUNTAIN

THE ADVENTURE OF PHYSICS – VOL. V

PLEASURE, TECHNOLOGY AND STARS



www.motionmountain.net



Christoph Schiller

Motion Mountain

The Adventure of Physics
Volume V

Pleasure, Technology and Stars

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

Editio vicesima quarta.

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

To Britta, Esther and Justus Aaron

τῷ ἐμοὶ δαὶμονι

Die Menschen stärken, die Sachen klären.

PREFACE

> Primum movere, deinde docere.*
>
> Antiquity

This book is written for anybody who is curious about nature and motion. Curiosity about how bodies, images and empty space move leads to many adventures. This volume presents the best adventures about the motion inside people and animals, as well as about the motion inside matter – from the largest stars to the smallest nuclei.

Motion inside bodies – dead or alive – is described by quantum theory. Quantum theory is the description of motion based on a smallest action, or better, a smallest change. With this basic idea, the text shows how to describe life, death and pleasure. The smallest change also explains the observations of chemistry, geology, material science and astrophysics. In the structure of physics, these topics correspond to the three 'quantum' points in Figure 1. The topics form applied quantum physics; they are introduced in this text. The text arose from a threefold aim that I have pursued since 1990: to present the basics of motion in a way that is simple, up to date and captivating.

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

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

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

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

Munich, 1 January 2011.

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

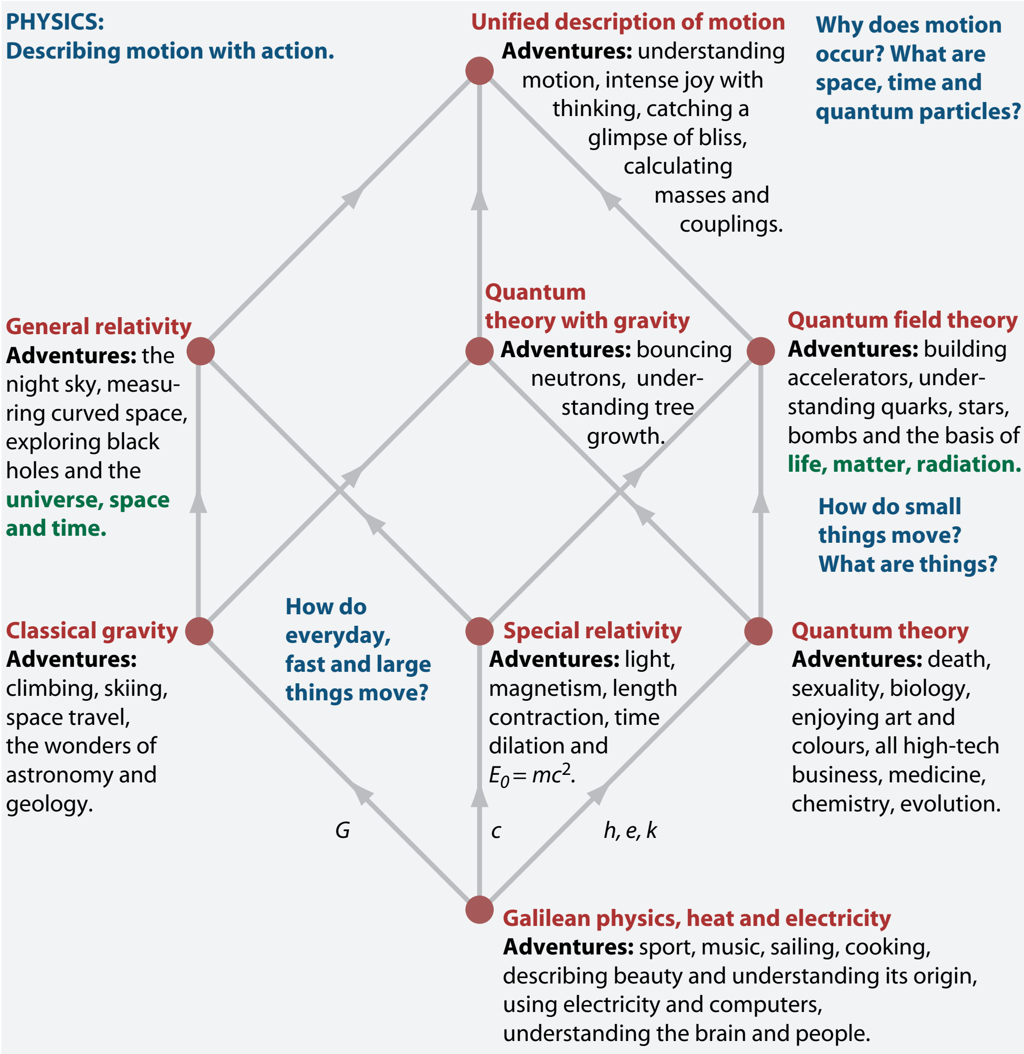


FIGURE 1 A complete map of physics: the connections are defined by the speed of light c, the gravitational constant G, the Planck constant h, the Boltzmann constant k and the elementary charge e.

Advice for learners

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

Using this book

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

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

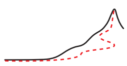
A request

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

Challenge 1 s

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

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



Contents

Pleasure, Technology and Stars

In our quest to understand how things move
as a result of a minimal change in nature, we discover
why a smallest change is necessary to make pleasure possible,
why the floor does not fall but keeps on carrying us,
that interactions are exchanges of radiation particles,
that matter is not permanent,
why empty space pulls mirrors together,
why the stars shine,
how the atoms formed that make us up,
and why swimming and flying is not so easy.

Chapter 1

MOTION FOR ENJOYING LIFE

> Homo sum, humani nil a me alienum puto.*
>
> Terence

Since we are able to explore quantum effects without ideological baggage, let us have some serious fun in the world of quantum physics. The quantum of action $\hbar$ has significant consequences for medicine, biology, chemistry, material science, engineering and the light emitted by stars. Also art, the colours and materials it uses, and the creative process in the artist, are based on the quantum of action.** From a physics standpoint, all these domains study small motions of quantum particles; thus the understanding and the precise description requires quantum physics. We will only explore a cross-section of these topics, but it will be worth it.

We start with three special forms of motion of charged particles that are of special importance to humans: life, reproduction and death. We mentioned at the start of quantum Vol. IV, page 14 physics that none of them can be described by classical physics. Indeed, life, sexuality and death are quantum effects. And in the domain of life, every perception and every sense, and thus every kind of pleasure, are due to quantum effects. The same is true for all our actions. Let us find out why.

FROM BIOLOGICAL MACHINES TO MINIATURIZATION

Living beings are physical systems that show 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 self-reproduce.*

From your biology lessons in secondary school you might remember the main properties of reproduction*** and heredity. Reproduction is characterized by random changes from

* 'I am a man and nothing human is alien to me.' Terence is Publius Terentius Afer (*c.* 190–159 BCE), the important roman poet. He writes this in his play *Heauton Timorumenos*, verse 77.

** The photograph on page 14 shows a soap bubble, the motion of the fluid in it, and the interference colours; it was taken and is copyright by Jason Tozer for Creative Review/Sony.

*** However, there are examples of objects which reproduce and which nobody would call living. Can you Challenge 2 s find some examples, together with a sharper definition? To avoid misunderstandings, whenever we say 'reproduction' in the following, we mean 'self-reproduction'.

one generation to the next. The statistics of mutations, for example Mendel's 'laws' of heredity, and the lack of intermediate states, are direct consequences of quantum theory. In other words, reproduction and growth are quantum effects.

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.

Reproduction and growth are simpler the smaller the adult system is. Therefore, most living beings are extremely small machines for the tasks they perform. This is especially clear when they are compared to human-made machines. This smallness of living beings is often surprising, because the design of human-made machines has considerably fewer requirements: human-made 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 strong restrictions life is subjected to, living beings hold many miniaturization world records for machines:

— 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. Or look at the size of any computer that attempts to speak.
— The brain has the densest and fastest memory of any device so far. The set of compact discs (CDs) or digital versatile discs (DVDs) that compare with the brain is many Page 186 thousand times larger.
— 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 in the future. Nevertheless, the sensor *sizes* developed by evolution – think also about the ears or eyes of a common fly – are still unbeaten.
— Living beings that fly, swim or crawl – such as fruit flies, plankton or amoebas – are still thousands of times smaller than anything comparable that is built by humans. In particular, already the navigation systems built by nature are far smaller than anything built by human technology.
— Challenge 3 s Can you spot more examples?

The superior miniaturization of living beings – compared to human-built machines – 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, in humans, breathing helps digestion; head movements pump liquid through the spine; a single hormone influences many chemical processes. In addition, 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. Living systems use many such optimizations.

When is a machine *well* miniaturized? When it makes efficient use of quantum effects. In short, miniaturization, reproduction, growth and 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 told us that only a *similar* object is possible, as an *exact* copy would contradict the quantum of action, Vol. IV, page 103 as we found out earlier on. But this is not a disadvantage: a similar, thus imperfect copy is all that is required for life and is essential for evolution.

Since reproduction requires mass increase, reproducing objects show both metabolism and growth. In order that growth can lead 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. The observation of reproduction thus implies the existence of atoms and the necessity of 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 from generation to generation requires that living beings have ways to store information. Living beings must have some built-in memory storage. 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, composed 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. Page 48 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 move. 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: we need 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 1 gives an overview of motion processes in living beings. The table shows that all motion

* 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. Radioactivity can also terminate life or be of use in medicine. Ref. 1

The nuclear interactions are also implicitly involved in life in several other ways. The nuclear interactions were necessary to form the atoms – 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).

TABLE 1 Motion and motors found in living beings

Motion type	Examples	Involved motors
Growth	collective molecular processes in cell growth	ion pumps
	gene turn-on and turn-off	linear molecular motors
	ageing	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	muscle working	linear molecular motors, ion pumps
	metabolism (respiration, digestion)	muscles, ion pumps
	thermodynamics of whole living system and of its parts	muscles
	nerve signalling	ion motion, ion pumps
	brain working, thinking	ion motion, ion pumps
	memory: long-term potentiation	chemical pumps
	hormone production	chemical 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
	blood clotting	chemical pumps
	bronchial cleaning	hair motors
Sensing	eye	chemical pumps, ion pumps
	ear	hair motion sensors, ion pumps
	smell	ion pumps
	touch	ion pumps
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

FIGURE 2 A quantum machine (© Elmar Bartel)

in living beings can be summarized in a few classes that are defined by the underlying motor.

Nature only needs a few small but powerful devices to realize all the motion types used by humans and by all other living beings: *pumps* and *motors*. 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 machines. 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 machines*. 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.

1. *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 Challenge 4 s the power due to the 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.

Vol. I, page 82 2. We encountered *rotational motors* already above; nature uses them to rotate the cilia

Summing up, the nuclear interactions play a role in the appearance and in the destruction of life; but they usually play no role for the actions or functioning of particular living beings.

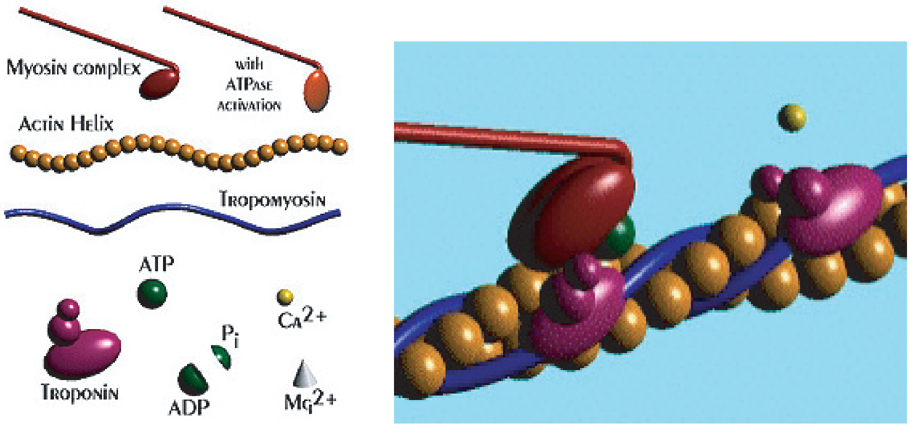


FIGURE 3 Myosin and actin: the building bricks and the working of a linear molecular motor (image and QuickTime film © San Diego State University, Jeff Sale and Roger Sabbadini)

Ref. 2 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 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 Ref. 3 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.

3. *Molecular pumps* are essential to life. They pump chemicals, such as ions or specific molecules, into every cell or out of it, using energy, even if the concentration gradient tries to do the opposite. Molecular pumps are thus essential in ensuring that life is a process *far* from equilibrium. Malfunctioning molecular pumps are responsible for many problems, for example the water loss in cholera.

In the following, we concentrate on linear motors. The ways molecules produce movement in linear motors was uncovered during the 1990s. Ref. 4 The results started a wave of research on all other molecular motors found in nature. All molecular motors share several characteristic properties: molecular motors do *not* involve temperature gradients involved, as car engines do, they do *not* involve electrical currents, as electrical motors do, and they do *not* rely on concentration gradients, as chemically induced motion, such as the rising of a cake, does.

Linear molecular motors

The central element of the most important linear molecular motor 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, as shown in Figure 3. The motion step size has been measured with beautiful experiments to always be an integer Ref. 4 multiple of 5.5 nm. A step, usually forward, but sometimes backwards, 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

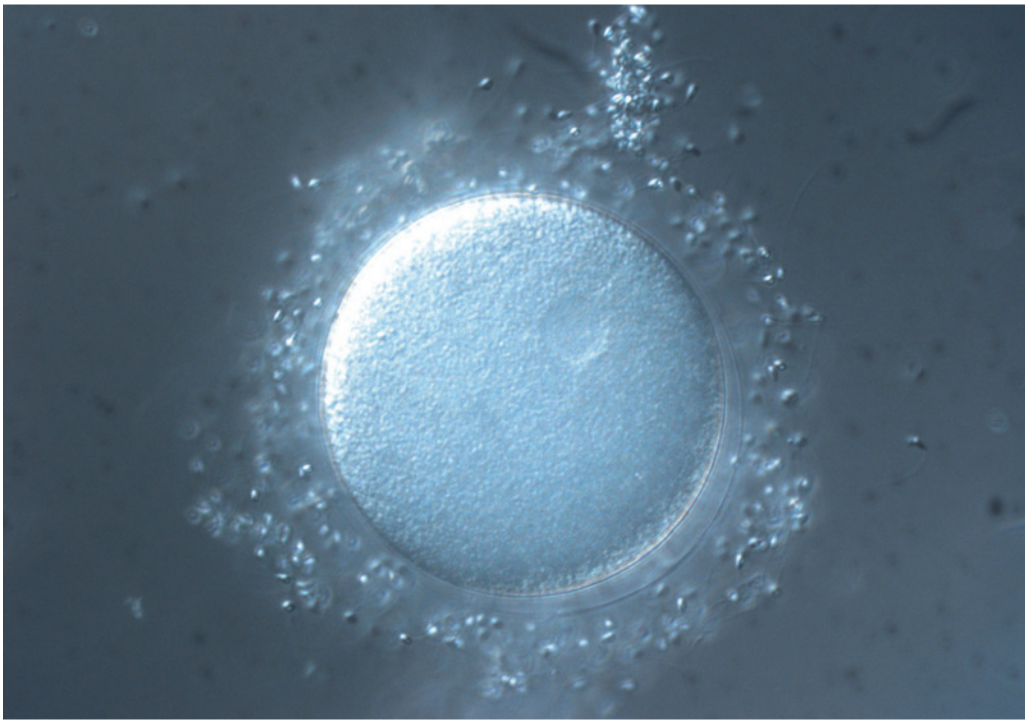


FIGURE 4 A sea urchin egg surrounded by sperm, or molecular motors in action: molecular motors make sperm move, make fecundation happen, and make cell division occur (photo by Kristina Yu, © Exploratorium www.exploratorium.edu)

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? Molecular motors are so small that the noise due to the Brownian motion of the molecules of the liquid around them is not negligible. But evolution 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 evolution is the use of an asymmetric, but periodic potential, a so-called *ratchet*.* The second trick of evolution is a temporal variation of the potential, together with an energy input to make
Ref. 5 it happen. The most important realizations are shown in Figure 5.

The periodic potential variation in a molecular motor ensures that for a short, recurring time interval the free Brownian motion of the moving molecule – typically 1 μm/s – affects its position. Subsequently, the molecule is fixed again. In most of the short time intervals of free Brownian motion, the position will not change. But if the position does change, the intrinsic asymmetry of the ratchet shape ensures that with high probability the molecule advances in the preferred direction. (The animation of Figure 3 lacks this irregularity.) 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

* It was named by Walt Disney after Ratchet Gearloose, the famous inventor from Duckburg.

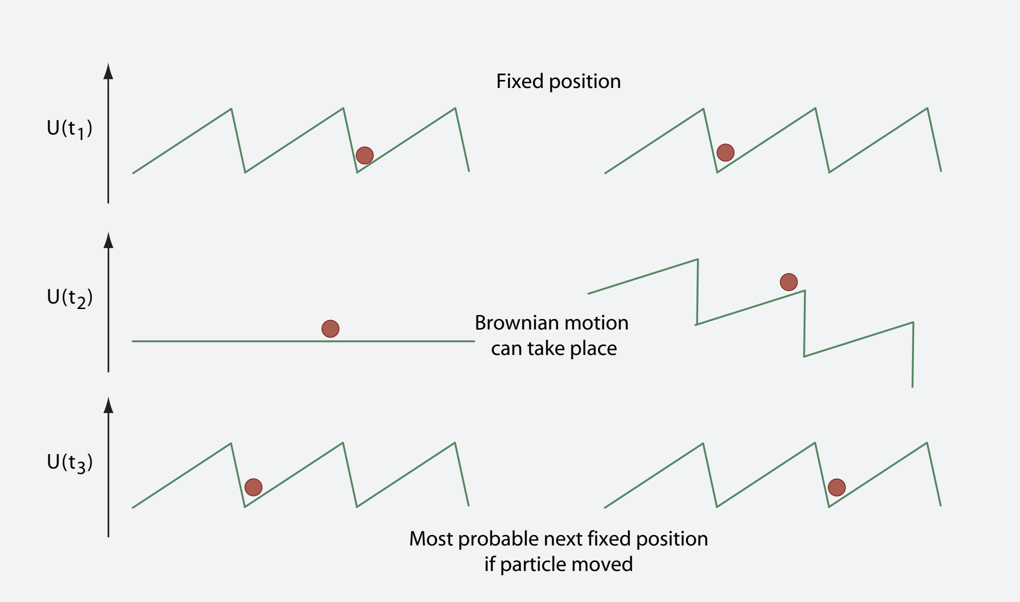


FIGURE 5 Two types of Brownian motors: switching potential (left) and tilting potential (right)

of single molecules can be followed in special experimental set-ups. These experiments confirm that muscles use such a ratchet mechanism. The ATP molecule 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 that 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 ratchet motors do not exist only in living systems; they also exist as human-built systems. Examples are electrical ratchets that move single electrons and optical ratchets that drive small particles. Extensive experimental research is going on in these fields.

Classical ratchets exist in many forms. For example, many piezoelectric actuators work as ratchets. All atomic force microscopes and scanning electron microscopes use such actuators.

Curiosities and fun challenges about biology

> Una pelliccia è una pelle che ha cambiato bestia.*
>
> Girolamo Borgogelli Avveduti

* 'A *fur* is a skin that has changed beast.'

Discuss the following argument: If nature were classical instead of quantum, there would not be just *two* sexes – or any other discrete number of them, as in some lower animals – but there would be a *continuous range* of them. In a sense, there would be an infinite number of sexes. True? Challenge 5 s

* *

Biological *evolution* can be summarized in three statements:

1. All living beings are different – also in a species.
2. All living beings have a tough life – due to competition.
3. Living beings with an advantage will survive and reproduce.

As a result of these three points, with each generation, *species and living beings can change*. The result of accumulated generational change is called biological evolution. The last point is often called the 'survival of the fittest'.

These three points explain, among others, the change from unicellular to multicellular life, from fish to land animals, and from animals to people.

We note that quantum physics enters in every point that makes up evolution. For example, the differences mentioned in the first point are due to quantum physics: perfect copies of macroscopic systems are impossible. And of course, life and metabolism are quantum effects. The second point mentions competition; that is a type of measurement, which, as we saw, is only possible due to the existence of a quantum of action. The third point mentions reproduction: that is again a quantum effect, based on the copying of genes, which are quantum structures. In short, both life and its evolution are quantum effects.

* *

Challenge 6 s 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 than the original? Challenge 7 s

It is well known that the first cloned cat, *copycat*, born in 2002, looked completely different from the 'original' (in fact, its mother). The fur colour and its patch pattern were completely different from that of the mother. Analogously, identical human twins have different finger prints, iris scans, blood vessel networks and intrauterine experiences, among others.

* *

A famous unanswered question on evolution: how did the first *kefir grains* form? Kefir grains produce the kefir drink when covered with milk for about 8 to 12 hours. The grains consist of a balanced mixture of about 40 types of bacteria and yeasts. All kefir grains in the world are related. But how did the first ones form, about 1000 years ago? Challenge 8 r

* *

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. Challenge 9 s How can this be?

* *

How is it possible that the genetic difference between man and chimpanzee is regularly given as about 1 %, whereas the difference between man and woman is one chromosome Challenge 10 s in 46, in other words, about 2.2 %?

* *

What is the longest time a single bacterium has survived? It is more than the 5000 years of the bacteria found in Egyptian mummies. For many years, the time was estimated to lie Ref. 6 at over 25 million years, a value claimed for the bacteria spores resurrected from the intestines in insects enclosed in amber. Then it was claimed to lie at over 250 million years, the time estimated that certain bacteria discovered in the 1960s by Heinz Dombrowski in (low-radioactivity) salt deposits in Fulda, in Germany, have hibernated there before being brought back to life in thelaboratory. A similar result has been recently claimed by the discovery of another bacterium in a North-American salt deposit in the Salado Ref. 7 formation.

However, these values are now disputed, as DNA sequencing has shown that these bacteria were probably due to sample contamination in the laboratory, and were not part Ref. 8 of the original sample. So the question is still open.

* *

Molecular motors are quite capable. The molecular motors in the sooty shearwater (*Puffinus griseus*), a 45 cm long bird, allow it to fly 74 000 km in a year, with a measured record of 1094 km a day.

* *

In 1967, a TV camera was deposited on the Moon. Unknown to everybody, it contained a small patch of *Streptococcus mitis*. Three years later, the camera was brought back to Ref. 9 Earth. The bacteria were still alive. They had survived for three years without food, water or air. Life can be resilient indeed.

* *

In biology, classifications are extremely useful. (This is similar to the situation in astrophysics, but in full contrast to the situation in physics.) Table 2 gives an overview of the Ref. 10 magnitude of the task. This wealth of material can be summarized in one graph, shown in Figure 6. Newer research seems to suggest some slight changes to the picture. So far however, there still is only a single root to the tree.

* *

Muscles produce motion through electrical stimulation. Can technical systems do the same? There is a candidate. So-called electroactive polymers change shape when they are activated with electrical current or with chemicals. They are lightweight, quiet and simple to manufacture. However, the first arm wrestling contest between human and artificial muscles held in 2005 was won by a teenage girl. The race to do better is ongoing.

TABLE 2 Approximate number of living species

Life group	Described species	Estimated species min.	Estimated species max.
Viruses	4 000	$50 \cdot 10^3$	$1 \cdot 10^6$
Prokaryotes (‘bacteria’)	4 000	$50 \cdot 10^3$	$3 \cdot 10^6$
Fungi	72 000	$200 \cdot 10^3$	$2.7 \cdot 10^6$
Protozoa	40 000	$60 \cdot 10^3$	$200 \cdot 10^3$
Algae	40 000	$150 \cdot 10^3$	$1 \cdot 10^6$
Plants	270 000	$300 \cdot 10^3$	$500 \cdot 10^3$
Nematodes	25 000	$100 \cdot 10^3$	$1 \cdot 10^6$
Crustaceans	40 000	$75 \cdot 10^3$	$200 \cdot 10^3$
Arachnids	75 000	$300 \cdot 10^3$	$1 \cdot 10^6$
Insects	950 000	$2 \cdot 10^6$	$100 \cdot 10^6$
Molluscs	70 000	$100 \cdot 10^3$	$200 \cdot 10^3$
Vertebrates	45 000	$50 \cdot 10^3$	$55 \cdot 10^3$
Others	115 000	$200 \cdot 10^3$	$800 \cdot 10^3$
Total	$1.75 \cdot 10^6$	$3.6 \cdot 10^6$	$112 \cdot 10^6$

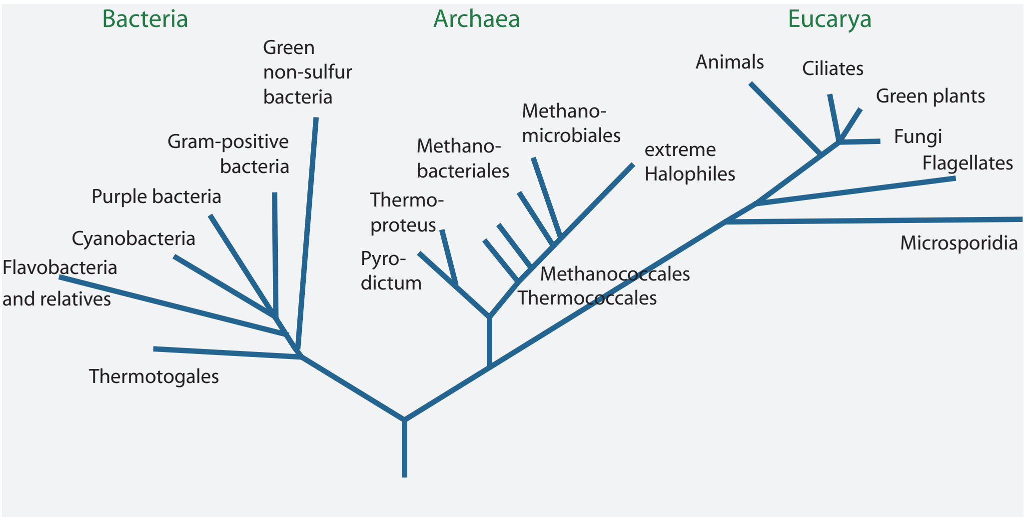


FIGURE 6 A modern version of the evolutionary tree

* *

Challenge 11 ny

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? Or is the definition of living beings as ‘what is made of cells’ more precise?

* *

Also growth is a type of motion. Some is extremely complex. Take the growth of acne. It requires a lack of zinc, a weak immune system, several bacteria, as well as the help of *Demodex brevis*, a mite (a small insect) that lives in skin pores. With a size of 0.3 mm, somewhat smaller than the full stop at the end of this sentence, this and other animals living on the human face can be observed with the help of a strong magnifying glass.

* *

Humans have many living beings on board. For example, humans need bacteria to live. It is estimated that 90% of the bacteria in the human mouth alone are not known yet; only about 500 species have been isolated so far. These useful bacteria help us as a defence against the more dangerous species.

Bacteria are essential for human life: they help us to digest and they defend us against Ref. 11 illnesses. In fact, the number of bacteria in a human body is estimated to be 10^{14}, whereas the number of cells in a human body is estimated to lie between 10^{13} and $5 \cdot 10^{13}$. In short, a human body contains more bacteria than own cells! Nevertheless, the combined mass of all bacteria is estimated to be only about 1 kg, because bacteria are much smaller than human cells, on average.

* *

How do trees grow? When a tree – a *monopodal phanerophyte* – grows and produces leaves, between 40% and 60% of the mass it consists of, namely the water and the minerals, has to be lifted upwards from the ground. (The rest of the mass comes from the CO_2 in the air.) How does this happen? The materials are pulled upwards by the water columns inside the tree; the pull is due to the negative pressure that is created when the top of the column evaporates. This is called the *transpiration-cohesion-tension model*. (This summary is the result of many experiments.) In other words, no energy is needed for the tree to pump its materials upwards.

Trees do not need energy to transport water. As a consequence, a tree grows purely by adding material to its surface. This implies that when a tree grows, a branch that is formed at a given height is also found at that *same* height during the rest of the life of Challenge 12 e that tree. Just check this observation with the trees in your garden.

* *

Mammals have a narrow operating temperature. In contrast to machines, humans func-Challenge 13 d tion only if the internal temperature is within a narrow range. Why? And does this requirement also apply to extraterrestrials – provided they exist?

* *

How did the first cell arise? This question is still open. However, researchers have found several substances that spontaneously form closed membranes in water. Such substances also form foams. It might well be that life formed in foam. Other options discussed are that life formed underwater, at the places where magma rises into the ocean. Elucidating Challenge 14 r the question is one of the great open riddles of biology.

* *

Challenge 15 s Could life have arrived to Earth from outer space?

* *

Is there life elsewhere in the universe? The answer is clear. First of all, there *might* be life elsewhere, though the probability is extremely small, due to the long times involved and the requirements for a stable stellar system, a stable planetary system, and a stable geological system. In addition, all statements that claim to have detected an example were lies.

* *

What could *holistic medicine* mean to a scientist, i.e., avoiding bullshit and beliefs? Holistic medicine means treating illness with view on the whole person. That translates to four domains:
— *physical* support, to aid mechanical or thermal healing processes in the body;
— *chemical* support, with nutrients or vitamins;
— *signalling* support, with electrical or chemical means, to support the signalling system of the body;
— *psychologic* support, to help all above processes.
When all theses aspects are taken care of, healing is as rapid and complete as possible. However, one main rule remains: *medicus curat, natura sanat*.*

* *

Life is, above all, beautiful. For example, go to www.thedeepbook.org to enjoy the beauty of life deep in the ocean.

* *

What are the effects of environmental pollution on life? Answering this question is an intense topic of modern research. Here are some famous stories.

— Herbicides and many genetically altered organisms kill bees. For this reason, bees are dying (since 2007) in the United States; as a result, many crops – such as almonds and oranges – are endangered there. In countries where the worst herbicides and genetically modified crops have been banned, bees have no problems.
— Chemical pollution leads to malformed babies. In mainland China, one out of 16 children is malformed for this reason (in 2007). In Japan, malformations have been much reduced – though not completely – since strict anti-pollution laws have been passed.
— Radioactive pollution kills. In Russia, the famous Lake Karachay had to be covered by concrete because its high radioactivity killed anybody that walked along it for half an hour.
— Smoking kills – though slowly. Countries that have lower smoking rates or that have curbed smoking have reduced rates for cancer and several other illnesses.
— Eating tuna is dangerous for your health, because of the heavy metals it contains.
— Cork trees are disappearing. The wine industry has started large research programs to cope with this problem.

* The physician helps, but nature heals

— Even arctic and antarctic animals have livers full of human-produced chemical poisons.
— Burning fuels rises the CO_2 level of the atmosphere. This leads to many effects for the Earth's climate, including a slow rise of average temperature and sea level.

Ecological research is uncovering many additional connections. Let us hope that the awareness for these issues increases across the world.

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 measurements rely on quantum theory. 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.

Of course, no researcher will admit that he studies pleasure. Therefore the researcher will say that he or she studies the senses, and that he or she is doing *perception research*. But pleasure, and with it, all human sensors, exist to let life continue. Pleasure is highest when life is made to continue. In the distant past, the appearance of new sensors in living systems has always had important effects of evolution, for example during the Cambrian explosion. Researching pleasure and sensors is indeed a fascinating field that is still evolving; here we can only have a quick tour of the present knowledge.

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 sound sensor can cover an energy range of 10^{13}; indeed, the detected sound intensities range from 1 pW/m^2 (some say 50 pW/m^2) to 10 W/m^2, the corresponding air pressure variations from 20 μPa to 60 Pa. The lowest intensity that can be heard is that of a 20 W sound source heard at a distance of 10 000 km, if no sound is lost in between.

Audible sound wavelengths span from 17 m (20 Hz) to 17 mm (20 kHz). In this range, the ear, with its 16 000 to 20 000 hair cells, is able to distinguish at least 1500 pitches. 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. They give the eye a resolving power of $1'$ and the capacity to consciously detect down to 60 *incident* photons in 0.15 s, or 4 *absorbed* photons in the same time interval.

The eye contains 120 million highly sensitive general light intensity detectors, the *rods*. They are responsible for the mentioned high sensitivity. Rods cannot distinguish colours. Before the late twentieth century, 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 that were as sensitive as the human eye.

FIGURE 7 The different speed of the eye's colour sensors, the cones, lead to a strange effect when this picture (in colour version) is shaken right to left in *weak* light

Page 141 The human eye contains about 6 million not so sensitive colour detectors, the *cones*, whose distribution we have seen earlier on. The different chemicals in the three cone types (red, green, blue) lead to different sensor speeds; this can be checked with the simple test shown in Figure 7. The sensitivity difference between the colour-detecting cones and the colour-blind rods is the reason that at night all cats are grey. Ref. 12

The images of the eye are only sharp if the eye constantly moves in small random motions. If this motion is stopped, for example with chemicals, the images produced by the eye become unsharp.

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; there are separate sensors for heat, for coldness, and for pain. Some react proportionally to the stimulus intensity, some differentially, giving signals only when the stimulus changes. Many of these sensors are also found inside the body – for example on the tongue. The sensors are triggered when external pressure deforms them; this leads to release of Na^+ and K^+ ions through their membranes, which then leads to an electric signal that is sent via nerves to the brain.

The taste mechanisms of *tongue* are only partially known. The tongue is known to produce six taste signals* – sweet, salty, bitter, sour, proteic and fatty – and the mechanisms are just being unravelled. The sense for proteic, also called *umami*, has been discovered in 1907, by Ikeda Kikunae; the sense for 'fat' has been discovered only in 2005. Democritus imagined that taste depends on the shape of atoms. Today it is known that sweet taste is connected with certain shape of molecules. Modern research is still unravelling the various taste receptors in the tongue. At least three different sweetness receptors, dozens of bitterness receptors, and one proteic and one fattiness receptor are known. In contrast, the sour and salty taste sensation are known to be due to ion channels. Despite all this knowledge, no sensor with a distinguishing ability of the same degree as the tongue has yet been built by humans. A good taste sensor would have great commercial value for the Ref. 13

* Taste sensitivity is *not* separated on the tongue into 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 16 s

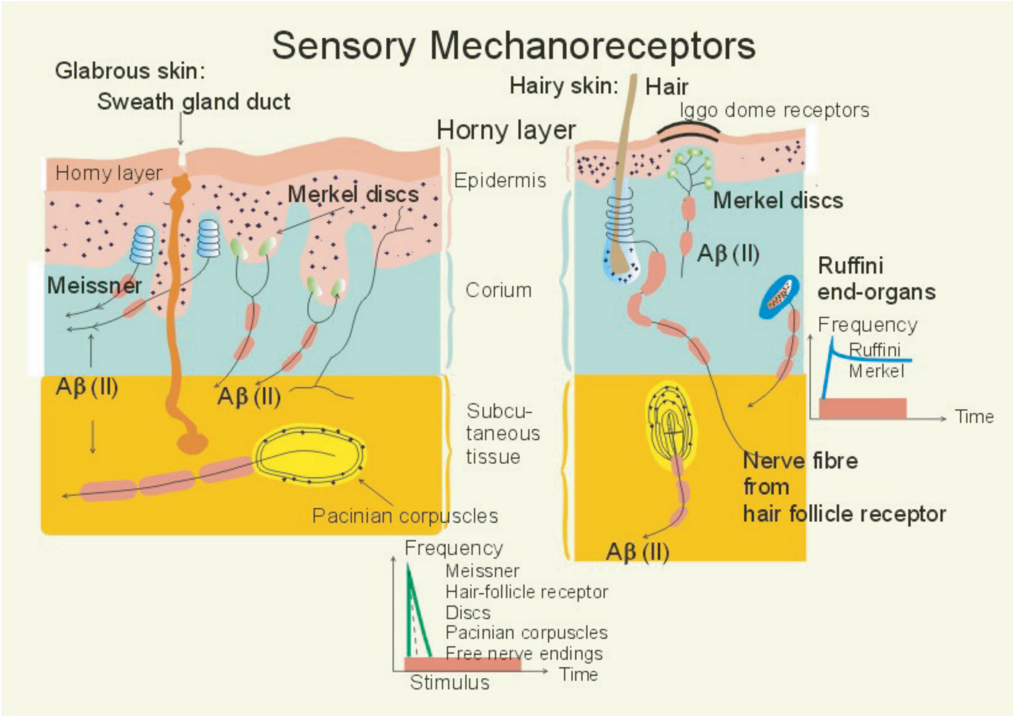


Figure 8 The five sensors of touch in humans: hair receptors, Meissner's corpuscules, Merkel cells, Ruffini corpuscules, and Pacinian corpuscules

food industry. Research is also ongoing to find substances to block taste receptors, with the aim to reduce the bitterness of medicines or of food.

The *nose* has about 350 different smell receptors; through combinations it is estimated that the nose can detect about 10 000 different smells.* Together with the five signals that the sense of taste can produce, the nose also produces a vast range of taste sensations. It protects against chemical poisons, such as smoke, and against biological poisons, such as faecal matter. In contrast, artificial gas sensors exist only for a small range of gases. Good artificial taste and smell sensors would allow to check wine or cheese during their production, thus making its inventor extremely rich. At the moment, humans are not even capable of producing sensors as good as those of a bacterium; it is known that *Escherichia coli* can sense at least 30 substances in its environment.

Challenge 17 ny

The human body also contains *orientation sensors* in the ear, *extension sensors* in each muscle, and *pain sensors* distributed with varying density over the skin and inside the body.

Page 30

Other animals feature additional types of sensors. Sharks can feel electrical fields, many snakes have sensors for infrared light, such as the pit viper. These sensors are used to locate prey. Pigeons, trout and sharks can feel magnetic fields, and use this sense for navigation. Many birds and certain insects can see UV light. Bats and dolphins are able to

* Linda Buck and Richard Axel received the 2004 Nobel Prize for medicine and physiology for their unravelling of the working of the sense of smell.

hear ultrasound up to 100 kHz and more. Whales and elephants can detect and localize infrasound signals.

In summary, the sensors with which nature provides us 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 Ref. 14 possible. Studying physics is one way to do this.

> 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 the signals arriving from the sensors, the brain, is essential for all feelings of pleasure. The human brain has the highest complexity of all brains known.** In addition, the processing power and speed of the human brain is still larger than any device build by man.

We saw already earlier on how electrical signals from the sensors are transported into Vol. I, page 249 the brain. In the brain itself, the arriving signals are classified and stored, sometimes for a short time, sometimes for a long time. The various storage mechanisms, essentially taking place in the structure and the connection strength between brain cells, were elucidated by modern neuroscience. The remaining issue is the process of classification. For certain low level classifications, such as 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 yet known how to describe the processes of reading, understanding and talking in terms of signal motions. Research is still in full swing and will probably remain so for a large part of the twenty-first century.

In the following we look at a few abilities of our brain, of our body and of other bodies that are important for the types of pleasure that we experience when we study of motion.

* Victor Friedrich Weisskopf (b. 1908 Vienna, d. 2002 Cambridge), acclaimed theoretical physicist who worked with Einstein, Born, Bohr, Schrödinger and Pauli. He catalysed the development of quantum electrodynamics and nuclear physics. He worked on the Manhattan project but later in life intensely campaigned against the use of nuclear weapons. During the cold war he accepted the membership in the Soviet Academy of Sciences. He was professor at MIT and for many years director of CERN, in Geneva. He wrote several successful physics textbooks. I heard him making the above statement at CERN, in 1981, 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.

Living clocks

> L'horologe fait de la réclame pour le temps.*
>
> Georges Perros

Vol. I, page 41 We have given an overview of living clocks already at the beginning of our adventure. They are common in bacteria, plants and animals. As Table 3 shows, without biological clocks, life and pleasure would not exist.

When we sing a musical note that we just heard we are able to reproduce the original frequency with high accuracy. We also know from everyday experience that humans are able to keep the beat to within a few per cent for a long time. When doing sport or when dancing, we are able to keep the timing to high accuracy. (For shorter or longer times, the internal clocks are not so precise.) All these clocks are located in the brain. Ref. 15

Brains process information. Also computers do this, and like computers, all brains need a clock to work well. Every clock is made up of the same components. It needs an *oscillator* determining the rhythm and a mechanism to feed the oscillator with energy. In addition, every clock needs an oscillation *counter*, i.e., a mechanism that reads out the clock signal, and a means of *signal distribution* throughout the system is required, synchronizing the processes attached to it. Finally, a clock needs a *reset mechanism*. If the clock has to cover many time scales, it needs several oscillators with different oscillation frequencies and a way to reset their relative phases.

Even though physicists know fairly well how to build good clocks, we still do not know many aspects of biological clocks. Most biological oscillators are chemical systems; some, like the heart muscle or the timers in the brain, are electrical systems. The general elucidation of chemical oscillators is due to Ilya Prigogine; it has earned him a Nobel Prize for chemistry in 1977. But not all the chemical oscillators in the human body are known yet, not to speak of the counter mechanisms. For example, a 24-minute cycle inside each human cell has been discovered only in 2003, and the oscillation mechanism is not yet fully clear. (It is known that a cell fed with heavy water ticks with 27–minute instead of 24–minute rhythm.) It might be that the daily rhythm, the circadian clock, is made up of or reset by 60 of these 24–minute cycles, triggered by some master cells in the human body. The clock reset mechanism for the circadian clock is also known to be triggered by daylight; the cells in the eye who perform this resetting action have been pinpointed only in 2002. The light signal from these cells is processed by the superchiasmatic nuclei, two dedicated structures in the brain's hypothalamus. The various cells in the human body act differently depending on the phase of this clock. Ref. 20 Ref. 19

The clocks with the longest cycle in the human body control *ageing*. One of the more famous ageing clock limits the number of division that a cell can undergo. The number of cell divisions, typically between 50 and 200, is finite for most cell types of the human body. (An exception are reproductory cells – we would not exist if they would not be able to divide endlessly.) The cell division counter has been identified; it is embodied in the *telomeres*, special structures of DNA and proteins found at both ends of each chromosome. (This work won the Nobel Prize in Medicine in 2009.) These structures are reduced by a small amount during each cell division. When the structures are too short, cell division stops. The purely theoretical prediction of this mechanism by Alexei Olovnikov

* 'Clocks are ads for time.'

TABLE 3 Examples of biological rhythms and clocks

Living being	Oscillating system	Period
Sand hopper (*Talitrus saltator*)	knows in which direction to flee from the position of the Sun or Moon	circadian
Human (*Homo sapiens*)	gamma waves in the brain	0.023 to 0.03 s
	alpha waves in the brain	0.08 to 0.13 s
	heart beat	0.3 to 1.5 s
	delta waves in the brain	0.3 to 10 s
	blood circulation	30 s
	cellular circahoral rhythms	1 to 2 ks
	rapid-eye-movement sleep period	5.4 ks
	nasal cycle	4 to 14 ks
	growth hormone cycle	11 ks
	suprachiasmatic nuclei (SCN), circadian hormone concentration, temperature, etc.; leads to jet lag	90 ks
	monthly period	2.4(4) Ms
	built-in aging	3.2(3) Gs
Common fly (*Musca domestica*)	wing beat	30 ms
Fruit fly (*Drosophila melanogaster*)	wing beat for courting	34 ms
Most insects (e.g. wasps, fruit flies)	winter approach detection (diapause) by length of day measurement; triggers metabolism changes	yearly
Algae (*Acetabularia*)	Adenosinetriphosphate (ATP) concentration	
Moulds (e.g. *Neurospora crassa*)	conidia formation	circadian
Many flowering plants	flower opening and closing	circadian
Tobacco plant	flower opening clock; triggered by length of days, discovered in 1920 by Garner and Allard	annual
Arabidopsis	circumnutation	circadian
	growth	a few hours
Telegraph plant (*Desmodium gyrans*)	side leaf rotation	200 s
Forsythia europaea, *F. suspensa*, *F. viridissima*, *F. spectabilis*	Flower petal oscillation, discovered by Van Gooch in 2002	5.1 ks

in 1971 was later proven by a number of researchers. (Only the latter received the Nobel Prize in medicine, in 2009, for this confirmation.) Research into the mechanisms and the exceptions to this process, such as cancer and sexual cells, is ongoing.

Not all clocks in human bodies have been identified, and not all mechanisms are known. For example, basis of the monthly period in women is both interesting and complex.

Other fascinating clocks are those at the basis of *conscious* time. Of these, the brain's stopwatch or *interval timer*, has been most intensely studied. Only recently was its mechanism uncovered by combining data on human illnesses, human lesions, magnetic resonance studies, and effects of specific drugs. The basic interval timing mechanism takes place in the striatum in the basal ganglia of the brain. The striatum contains thousands of timer cells with different periods. They can be triggered by a 'start' signal. Due to their large number, for small times of the order of one second, every time interval has a different pattern across these cells. The brain can read these patterns and learn them. In this way we can time music or specific tasks to be performed, for example, one second after a signal.

Ref. 21

Even though not all the clock mechanisms in humans are known, natural clocks share a property with human-built clocks: they are limited by quantum mechanics. Even the simple pendulum is limited by quantum theory. Let us explore the topic.

When do clocks exist?

> Die Zukunft war früher auch besser.*
>
> Karl Valentin.

In general relativity, we found out that purely gravitational 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.

Page 254

First of all, in quantum theory, the time is *not* an observable. Indeed, the time operator is not Hermitean. In other words, quantum theory states that 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 most people's satisfaction. Observations 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 energy – lacks a Hermitean operator whose expectation value increases monotonically with time. This result can be proven rigorously. In other words, quantum theory states that time cannot be measured.

Ref. 22

Challenge 18 ny

That time cannot be measured is not really a surprise. 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 when the chain end is reached or when the battery is empty. In other words, in all real clocks the Hamiltonian is bounded from below.

In short, quantum theory shows that exact *clocks do not exist in nature*. Quantum theory states that any clock can only be *approximate*. Obviously, this result is of importance for high precision clocks. What happens if we try to increase the precision of a clock as much as possible?

* 'Also the future used to be better in the past.' Karl Valentin (b. 1882 Munich, d. 1948 Planegg), German author and comedian.

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 strongly suppressed; however, it is still possible, though with low probability. This is true even for a sundial.
Challenge 19 ny

In other words, clocks necessarily have to be macroscopic in order to work properly. A clock must be as large as possible, in order to average out its fluctuations. Astronomical systems are good 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.

The precision of clocks

Given the limitations due to quantum theory, what is the ultimate precision of a clock? To start with, the indeterminacy relation provides the limit on the mass of a clock. The mass M must be larger than
Challenge 20 ny

$$M > \frac{\hbar}{c^2 \tau} \tag{1}$$

Challenge 21 e
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
Ref. 16

$$M > \frac{\hbar}{c^2 \tau} \frac{T}{\tau} \tag{2}$$

where T is the time to be measured. You might check that this condition directly requires that any clock must be *macroscopic*.
Challenge 22 e

Let us play with the formula by Salecker and Wigner. It can be rephrased in the following way. For a clock that can measure a time t, the size l is connected to the mass m by

$$l > \sqrt{\frac{\hbar t}{m}} \; . \tag{3}$$

Ref. 18
How close can this limit be achieved? It turns out that the 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 (3) 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.

Note that the requirement by Salecker and Wigner is not in contrast with the possibility to make the *oscillator* of the clock very small; researchers have built oscillators made
Ref. 17 of a single atom. In fact, such oscillations promise to be the most precise human built
Page 32 clocks. But the oscillator is only one part of any clock, as explained above.

In the real world, the clock limit can be tightened even more. 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*; it limits the accuracy of their frequency ν by

$$\frac{\delta\nu}{\nu} = \sqrt{\frac{\Delta E}{E_{tot}}} \tag{4}$$

where $\Delta E = \hbar/T$ is the energy indeterminacy stemming from the finite measuring time T and $E_{tot} = NE_{bind}$ is the total binding energy of the atoms in the metre bar. So far, the quantum limit has not yet been achieved for any clock, even though experiments are getting close to it.

In summary, clocks exist only in the limit of $\hbar$ being negligible. In practice, the errors made by using clocks and metre bars can be made as small as required; it suffices to make the clocks large enough. Clock built into human brains comply with this requirement. We can thus continue our investigation into the details of matter without much worry, at least for a while. Only in the last part of our mountain ascent, where the requirements for precision will be even higher and where general relativity will limit the size of physical systems, trouble will appear again: the impossibility to build precise clocks will then
Vol. VI, page 58 become a central issue.

Why are predictions so difficult, especially of the future?

> “Future: that period of time in which our affairs prosper, our friends are true, and our happiness is assured.”
>
> Ambrose Bierce

We have found in our adventure that predictions of the future are made difficult by nonlinearities and by the divergence from similar conditions; we have seen that many particles make it difficult to predict the future due to the statistical nature of their initial conditions; we have seen that quantum theory makes it hard to fully determine initial states; we have seen that a non-trivial space-time topology can limit predictability; finally, we will discover that black hole and similar horizons can limit predictability due to their one-way transmission of energy, mass and signals.

Predictability and time measurements are thus limited. The main reason for this limit is the quantum of action. If due to the quantum of action perfect clocks do not exist, is determinism still the correct description of nature? Yes and no. We learned that all the mentioned limitations of clocks can be overcome for limited time intervals; in practice, these time intervals can be made so large that the limitations do not play a role in everyday life. As a result, in quantum systems both determinism and time remain applicable, as long as we do not extend it to infinite space and time. However, when extremely large dimensions and intervals need to be taken into account, quantum theory cannot be applied alone; in those cases, general relativity needs to be taken into account. Page 52

Decay and the golden rule

> I prefer most of all to remember the future.
>
> Salvador Dalì

The decoherence of superposition of macroscopically distinct states plays an important role in another common process: the decay of unstable systems or particles. Decay is any spontaneous change. Like the wave aspect of matter, decay is a process with no classical counterpart. It is true that decay, including the ageing of humans, can be observed classically; however, its *origin* is a pure quantum effect.

Experiments show that the prediction of decay, like that of scattering of particles, is only possible on average, for a large number of particles or systems, never for a single one. These results confirm the quantum origin of the process. In every decay process, the superposition of macroscopically distinct states – in this case those of a decayed and an undecayed particle – is made to decohere rapidly by the interaction with the environment. Usually the 'environment' vacuum, with its fluctuations of the electromagnetic, weak and strong fields, is sufficient to induce decoherence. As usual, the details of the involved environment states are unknown for a single system and make any prediction for a specific system impossible.

Decay, including that of radioactive nuclei, is influenced by the environment, even in the case that it is 'only' the vacuum. The statement can be confirmed by experiment. By enclosing a part of space between two conducting plates, one can change the degrees of freedom of the vacuum electromagnetic field contained between them. Putting an electromagnetically unstable particle, such as an excited atom, between the plates, indeed changes the lifetime of the particle. Can you explain why this method is not useful to lengthen the lifespan of humans? Ref. 23 Challenge 23 ny

What is the *origin* of decay? Decay is always due to tunnelling. With the language of quantum electrodynamics, we can say that decay is motion induced by the vacuum fluctuations. Vacuum fluctuations are random. The experiment between the plates confirms the importance of the environment fluctuations for the decay process.

Quantum theory gives a simple description of decay. For a system consisting of a large number N of decaying identical particles, any *decay* is described by

$$\dot{N} = -\frac{N}{\tau} \quad \text{where} \quad \frac{1}{\tau} = \frac{2\pi}{\hbar} \left|\langle \psi_{\text{initial}} | H_{\text{int}} | \psi_{\text{final}} \rangle\right|^2 . \tag{5}$$

This result was named the *golden rule* by Fermi,* because it works so well despite being an approximation whose domain of applicability is not easy to specify.

The golden rule leads to

$$N(t) = N_0 \,\mathrm{e}^{-t/\tau} . \tag{6}$$

* Originally, the golden rule is a statement from the christian bible, namely the sentence 'Do to others what you want them to do to you'.

Decay is thus predicted to follow an exponential law, independently of the details of the physical process. In addition, the decay time τ depends on the interaction and on the square modulus of the transition matrix element. For over half a century, all experiments confirmed that decay is exponential.

On the other hand, when quantum theory is used to derive the golden rule, it is found that decay is exponential only in certain special systems. A calculation that takes into account higher order terms predicts two deviations from exponential decay for completely isolated systems: for short times, the decay rate should *vanish*; for long times, the decay rate should follow an *algebraic* – not an exponential – dependence on time, in some cases even with superimposed oscillations. After an intense experimental search, deviations for short times have been observed. The observation of deviations at long times are rendered impossible by the ubiquity of thermal noise. In summary, it turns out that decay is exponential only when the environment is noisy, the system made of many weakly interacting particles, or both. Since this is usually the case, the mathematically exceptional exponential decrease becomes the (golden) rule in the description of decay.

Ref. 24

Ref. 25

Can you explain why human life, despite being a quantum effect, is not observed to follow an exponential decay?

Challenge 24 ny

The present in quantum theory

> Utere tempore.*
>
> Ovidius

Many thinkers advise to enjoy the present. As shown by perception research, what humans call 'present' has a duration of between 20 and 70 milliseconds. This leads us to ask whether the *physical* present might have a duration as well.

Ref. 26

In everyday life, we are used to imagine that shortening the time taken to measure the position of a point object as much as possible will approach the ideal of a particle fixed at a given point in space. When Zeno discussed flight of an arrow, he assumed that this is possible. However, quantum theory changes the situation.

We know that the quantum of action makes rest an impossibility. However, the issue here is different: we are asking whether we can say that a moving system is at a given spot at a given time. In order to determine this, we could use a photographic camera whose shutter time can be reduced at will. What would we find? When the shutter time approaches the oscillation period of light, the sharpness of the image would decrease; in addition, the colour of the light would be influenced by the shutter motion. We can increase the energy of the light used, but the smaller wavelengths only shift the problem, they do not solve it. Worse, at extremely small wavelengths, matter becomes transparent, and shutters cannot be realized any more. Whenever we reduce shutter times as much as possible, observations become unsharp. Quantum theory thus does not confirm the naive expectation that shorter shutter times lead to sharper images. In contrast, the quantum aspects of nature show us that there is no way in principle to approach the limit that Zeno was discussing.

This counter-intuitive result is due to the quantum of action: through the indeterminacy relation, the smallest action prevents that moving objects are at a fixed position at a

* 'Use your time.' *Tristia* 4, 3, 83

given time. Zeno's discussion was based on an extrapolation of classical physics into domains where it is not valid any more. Every observation, like every photograph, implies a *time average*: observations average interactions over a given time.* For a photograph, the duration is given by the shutter time; for a measurement, the average is defined by the details of the set-up. Whatever this set-up might be, the averaging time is never zero. There is no 'point-like' instant of time that describes the present. The observed *present* is always an average over a non-vanishing interval of time. In nature, the present has a duration.

Why can we observe motion?

Zeno of Elea was thus wrong in assuming that motion is a sequence of specific positions in space. Quantum theory implies that motion is *not* the change of position with time. The investigation of the issue showed that this statement is only an approximation for low energies or for long observation times.

Why then can we describe motion in quantum theory? Quantum theory shows that motion is the low energy approximation of quantum evolution. Quantum evolution assumes that space and time measurements of sufficient precision can be performed. We know that for any given observation energy, we can build clocks and metre bars with much higher accuracy than required, so that quantum evolution is applicable in all cases. This is the case in everyday life.

Obviously, this pragmatic description of motion rests on the assumption that for any observation energy we can find a still higher energy used by the measurement instruments to define space and time. We deduce that if a highest energy would exist in nature, we would get into big trouble, as quantum theory would then break down. As long as energy has no limits, all problems are avoided, and motion remains a sequence of quantum observables or quantum states, whichever you prefer.

The assumption of energy without limit works extremely well; it lies at the basis of Page 42 quantum theory, even though it is rather hidden. In the final part of our ascent, we will discover that there indeed is a maximum energy in nature, so that we will need to change our approach. However, this energy value is so huge that it does not bother us at all at this point of our exploration. But it will do so later on.

Rest and the quantum Zeno effect

The quantum of action implies that there is no rest in nature. Rest is always either an approximation or a time average. For example, if an electron is bound in an atom, not freely moving, the probability cloud, or density distribution, is stationary in time.

There is another apparent case of rest in quantum theory, the quantum Zeno effect. Usually, observation changes the system. However, for certain systems, observation can have the opposite effect.

Quantum mechanics predicts that an unstable particle can prevented from decaying, if it is continuously observed. The reason is that an observation, i.e., the interaction with the observing device, yields a non-zero probability that the system does not evolve. If the frequency of observations is increased, the probability that the system does not decay at

Page 39 * The discussion of the quantum Zeno effect does not change the conclusions of this section.

all approaches 1. Three research groups – Alan Turing by himself in 1954, the group of A. Degasperis, L. Fonda and G.C. Ghirardi in 1974, and George Sudarshan and Baidyanath Misra in 1977 – have independently predicted this effect, today called the *quantum Zeno effect*. In sloppy words, the quantum Zeno effect states: if you look at a system all the time, nothing happens.

The quantum Zeno effect is a natural consequence of quantum theory; nevertheless, its strange circumstances make it especially fascinating. After the prediction, the race for the first observation began. The effect was partially observed by David Wineland and his group in 1990, and definitively observed by Mark Raizen and his group in 2001. Ref. 27 Quantum theory has been confirmed also in this aspect. The effect is also connected to the deviations from exponential decay – due to the golden rule – that are predicted by quantum theory. These issues are research topics to this day.

Ref. 28 In a fascinating twist, Saverio Pascazio and his team have predicted that the quantum Zeno effect can be used to realize X-ray tomography of objects with the lowest radiation levels imaginable.

Consciousness – a result of the quantum of action

In the pleasures of life, consciousness plays an essential role. Consciousness is our ability to observe what is going on in our mind. Page 246 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. Simply said, 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 these processes with measurement apparatus. The best systems allowing this are magnetic resonance imaging machines, described below. Page 119 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 25 ny

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 of consciousness are not clear yet. Can you add a few arguments to the ones given here? Challenge 26 ny

Why can we observe motion? – Again

Studying nature can be one of the most intense pleasures of life. All pleasures are 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 animals, motion would not exist. In many ways motion is thus an illusion, as Zeno of Elea had claimed a long time ago. To say the least, the observation of motion is a result of 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. Vol. I, page 15

Curiosities and fun challenges about quantum experience

Most clocks used in everyday life, those built inside the human body and those made by humans, are electromagnetic. Any clock on the wall, be it mechanical, quartz controlled, radio or solar controlled, is based on electromagnetic effects. Do you know an exception? Challenge 27 s

* *

The sense of smell is quite complex. For example, the substance that smells most badly to humans is *skatole*, also called *3-methylindole*. This is the molecule to which the human nose is most sensitive. Skatole makes faeces smell bad; it is a result of haemoglobin entering the digestive tract through the bile. (In contrast to humans, skatole attracts flies; it is also used by some plants for the same reason.)

On the other hand, *small* levels of skatole do not smell bad to humans. It is also used by the food industry in small quantities to give smell and taste to vanilla ice cream.

* *

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

Challenge 28 s energies as large as about 10 μJ. Is one of the two systems relativistic?

* *

Compared to all primates, the human eye is special: it is white, thus allowing others to see the direction in which one looks. Comparison with primates shows that the white colour has evolved to allow more communication between individuals. Ref. 29

* *

Even at perfect darkness, the eye does not yield a black impression, but a slightly brighter one, called *eigengrau*. This is a result of noise created inside the eye, probably triggered by spontaneous decay of rhodopsin, or alternatively, by spontaneous release of neurotransmitters.

* *

The high sensitivity of the ear can be used to *hear* light. To do this, take an empty 750 ml jam glass. Keeping its axis horizontal, blacken the upper half of the inside with a candle. The lower half should remain transparent. After doing this, close the jam glass with its lid, and drill a 2 to 3 mm hole into it. If you now hold the closed jam glass with the hole to your ear, keeping the black side up, and shining into it from below with a 50 W light bulb, something strange happens: you hear a 100 Hz sound. Why? Challenge 29 s

* *

Most senses work already before birth. It is well-known since many centuries that playing the violin to a pregnant mother every day during the pregnancy has an interesting effect. Even if nothing is told about it to the child, it will become a violin player later on. In fact, most musicians are 'made' in this way.

* *

There is ample evidence that not using the senses is damaging. People have studied what happens when in the first years of life the vestibular sense – the one used for motion detection and balance restoration – is not used enough. Lack of rocking is extremely hard to compensate later in life. Equally dangerous is the lack of use of the sense of touch. Babies, like all small mammals, that are generally and systematically deprived of these experiences tend to violent behaviour during the rest of their life. Ref. 30

* *

It is still unknown why people *yawn*. This is still a topic of research.

* *

Nature has invented the senses to increase pleasure and avoid pain. But neurologists have found out that nature has gone even further; there is a dedicated *pleasure system* in the brain, whose function is to decide which experiences are pleasurable and which not. The main parts of the pleasure system are the *ventral tegmental area* in the midbrain and the *nucleus accumbens* in the forebrain. The two parts regulate each other mainly through *dopamine* and *GABA*, two important neurotransmitters. Research has shown that dopamine is produced whenever pleasure exceeds expectations. Nature has thus de-

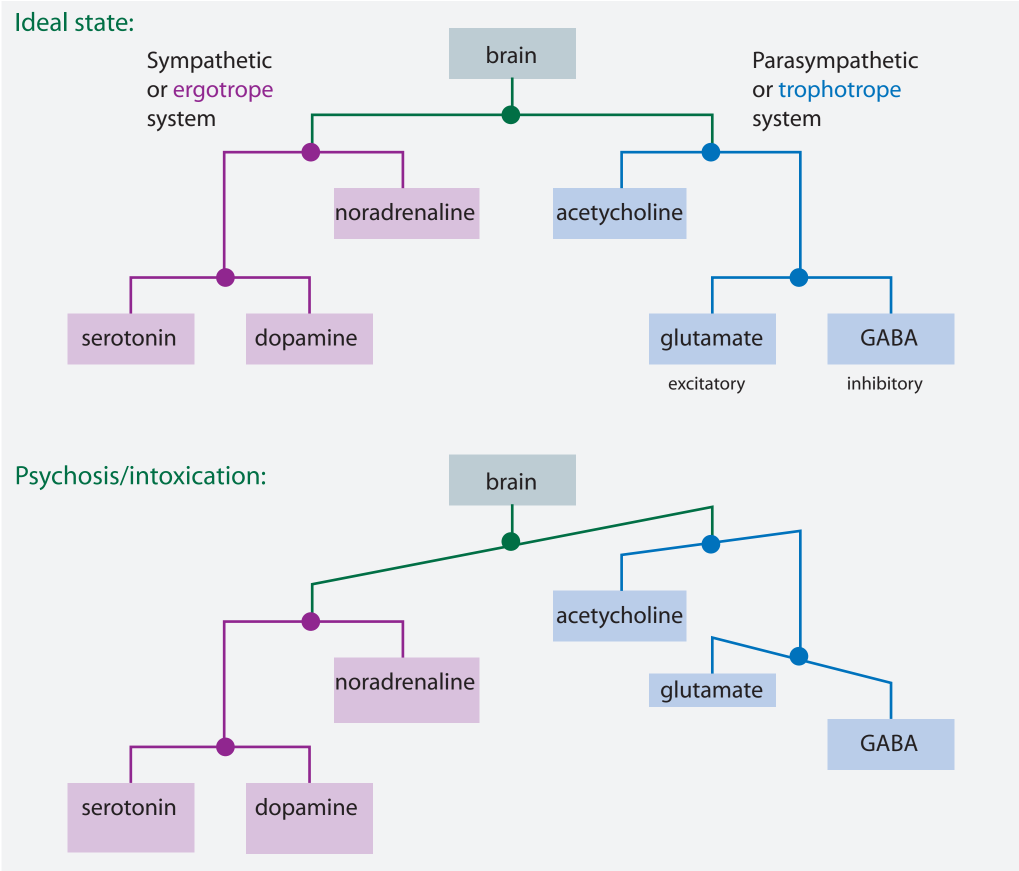


FIGURE 9 The 'neurochemical mobile' model of well-being, with one of the way it can get out of balance

veloped a special signal for this situation.

In fact, well-being and pleasure are controlled by a large number of neurotransmitters and by many additional regulation circuits. Researchers are trying to model the pleasure system with hundreds of coupled differential equations, with the distant aim being to understand addiction and depression, for example. On the other side, also simple models are possible. One, shown in Figure 9, is the 'neurochemical mobile' model of the brain. In this model, well-being is achieved whenever the six most important neurotransmitters are in relative equilibrium. The different possible departures from equilibrium, at each joint of the mobile, can be used to describe depression, schizophrenia, psychosis, the effect of nicotine or alcohol intake, alcohol dependency, delirium, drug addiction, detoxication, epilepsy and more.

Ref. 31

* *

The pleasure system is not only responsible for addiction. It is also responsible, as Helen Fisher showed through MRI brain scans, for romantic love. Romantic love, directed to one single other person, is a state that is created in the *ventral tegmental area* and in the *nucleus accumbens*. Romantic love is thus a part of the reptilian brain; indeed, romantic

Ref. 32

love is found in many animal species. Romantic love is an addiction, and works like cocaine. In short, in life, we can all chose between addiction and love.

Summary on pleasure

To increase pleasure and avoid pain, evolution has supplied the human body with numerous sensors, sensor mechanisms, and a pleasure system deep inside the brain.

In short, nature has invented pleasure as a guide for human behaviour. Neurologists have thus proven what Epicurus* said 25 centuries ago and Sigmund Freud said one century ago: pleasure controls human life. Now, all biological pleasure sensors and systems are based on chemistry and materials science. We therefore explore both fields.

* Though it is often forgotten that Epicurus also said: “It is impossible to live a pleasant life without living wisely and honourably and justly, and it is impossible to live wisely and honorably and justly without living pleasantly.”

CHAPTER 2

CHANGING THE WORLD WITH QUANTUM THEORY

THE discovery of quantum theory has changed everyday life. It has allowed the distribution of speech, music and films. The numerous possibilities of telecommunications and of the internet, the progress in chemistry, material science, medicine and electronics would not have been possible without quantum theory. Many improvements of our everyday life are due to quantum theory, and many are still expected. In the following, we give a short overview.

CHEMISTRY – FROM ATOMS TO DNA

> “Bier macht dumm.*
>
> Albert Einstein”

It is an old truth that Schrödinger's equation contains all of chemistry.** With quantum theory, for the first time people were able to calculate the strengths of chemical bonds, and what is more important, the angle between them. Quantum theory thus explains the *shape* of molecules and thus indirectly, the shape of all matter.

To understand molecules, the first step is to understand atoms. The early quantum theorists, lead by Niels Bohr, dedicated their life to understanding their structure. The main result of their efforts is what you learn in high school: in atoms with more than one electron, the various electron clouds form spherical layers around the nucleus. The layers can be grouped into groups of related clouds, called *shells*. For electrons outside a given shell, the nucleus and the inner shells, the atomic *core*, can often be approximated as a single charged entity.

Shells are numbered from the inside out. This *principal quantum number*, usually written n, is deduced and related to the quantum number that identifies the states in the hydrogen atom.

Quantum theory shows that the first shell has room for two electrons, the second for 8, the third for 18, and the general n-th shell for $2n^2$ electrons. A way to picture this connection is shown in Figure 11. It is called the *periodic table of the elements*. The standard way to show the table is shown in Appendix B.

Ref. 34
Page 264

* 'Beer makes stupid.'

** The correct statement is: the *Dirac* equation contains all of chemistry. The relativistic effects that distinguish the two equations are necessary, for example, to understand why gold is yellow and does not rust or why mercury is liquid.

Ref. 33

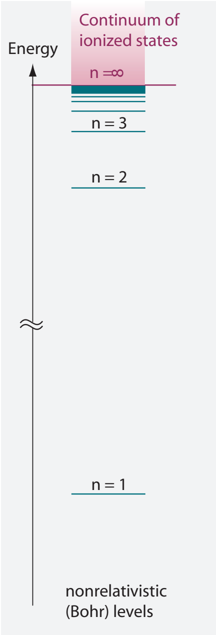


FIGURE 10 The principal quantum numbers in hydrogen

Experiments show that different atoms that share the same number of electrons in their outermost shell show similar chemical behaviour. Chemical behaviour is decided by the ability of atoms to bond. For example, the elements with one electron in their out s shell, are the *alkali metals* lithium, sodium, potassium, rubidium, caesium and francium; hydrogen, the exception, is conjectured to be metallic at high pressures. The elements with filled outermost shells are the *noble gases* helium, neon, argon, krypton, xenon, radon and ununoctium.

Bonds

When two atoms approach each other, their electron clouds are deformed and mixed. The reason for these changes is the combined influence of the two nuclei. These cloud changes are highest for the outermost electrons: they form chemical *bonds*.

Bonds can be pictured, in the simplest approximation, as cloud overlaps that *fill* the outermost shell of both atoms. These overlaps lead to a gain in energy. The energy gain is the reason that fire is hot. In wood fire, chemical reactions between carbon and oxygen atoms lead to a large release of energy. After the energy has been released, the atomic bond produces a fixed distance between the atoms, as shown in Figure 12. This distance is due to an energy minimum: a lower distance would lead to electrostatic repulsion between the atomic cores, a higher distance would increase the electron cloud energy.

Many atoms can bind to several neighbours. In this case, energy minimization also leads to specific bond angles. Do you remember those funny pictures of school chemistry about orbitals and dangling bonds? Well, dangling bonds can now be measured and seen. Several groups were able to image them using scanning force or scanning tunnelling microscopes.

Ref. 35

Ref. 36

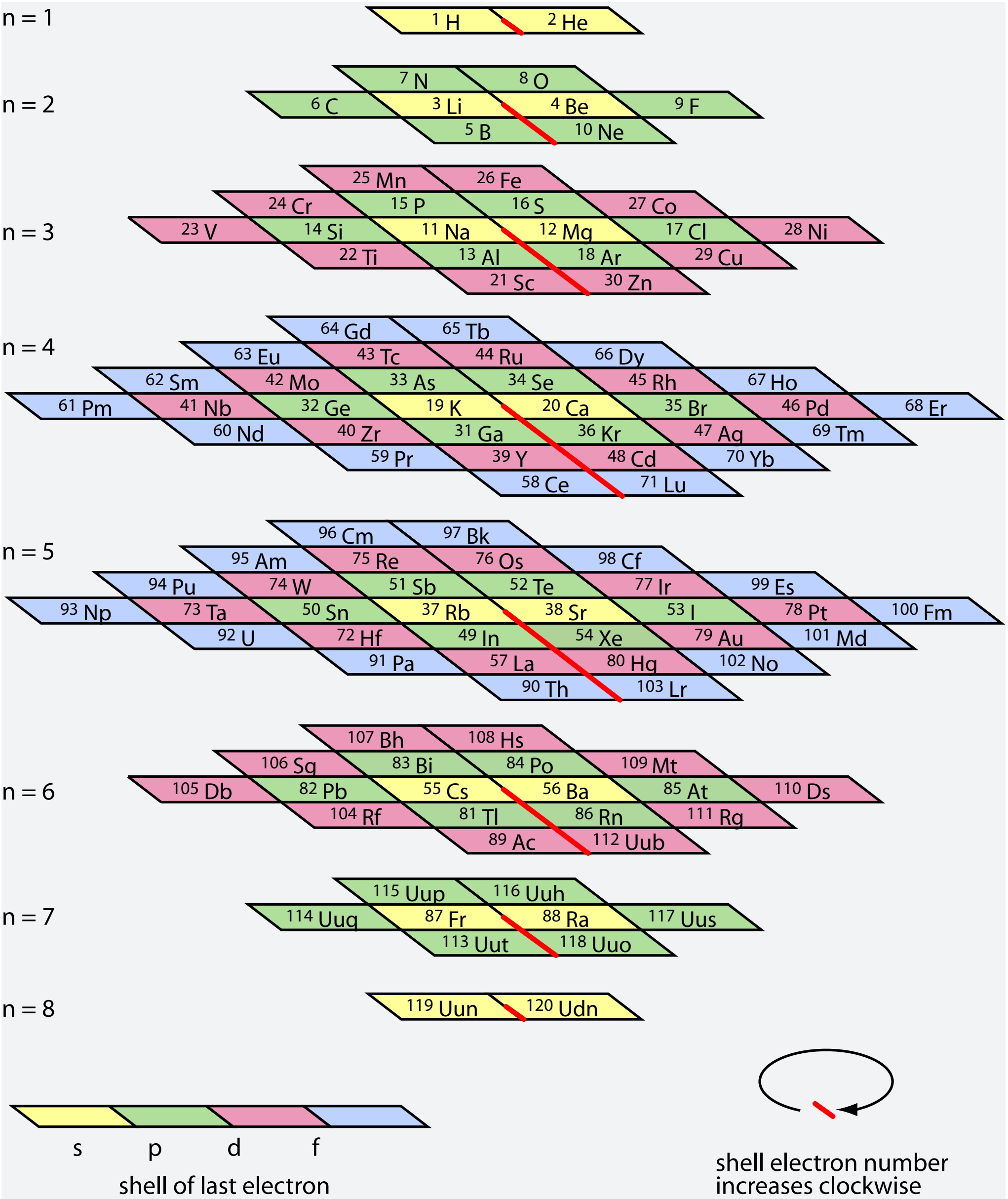


FIGURE 11 An unusual form of the periodic table of the elements

The repulsion between the clouds of each bond explains why angle values near that of tetrahedral skeletons ($2 \arctan \sqrt{2} = 109.47°$) are so common in molecules. For example, the H-O-H angle in water molecules is 107°.

Challenge 30 e

By the way, it is now known that the uranium U_2 molecule has a quintuple bond, and that the tungsten W_2 molecule has a hextuple bond.

Ref. 37

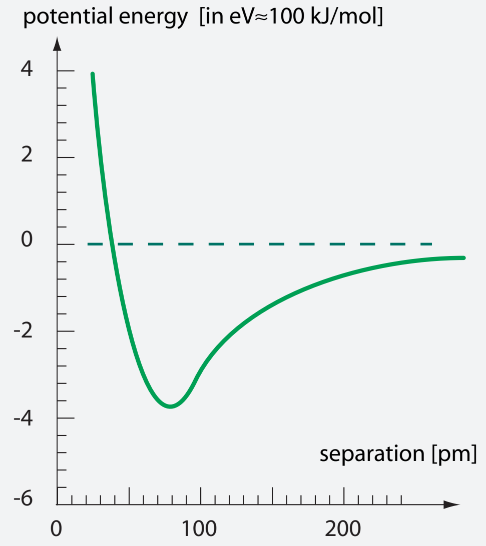


FIGURE 12 The forming of a chemical bond between two atoms, and the related energy minimum

Ribonucleic acid and deoxyribonucleic acid

Probably the most fascinating molecule of all is human deoxyribonucleic acid, better known with its abbreviation DNA. The nucleic acids where discovered in 1869 by the Swiss physician Friedrich Miescher (1844–1895) in white blood cells. In 1874 he published an important study showing that the molecule is contained in spermatozoa, and discussed the question if this substance could be related to heredity. With his work, Miescher paved the way to a research field that earned many colleagues Nobel Prizes (though not for himself).

DNA is, as shown in Figure 13, a polymer. A polymer is a molecule built of many similar units. In fact, DNA is among the longest molecules known. Human DNA molecules, for example, can be up to 5 cm in length. Inside each human cell there are 46 chromosomes. In other words, inside each human cell there are molecules with a total length of 2 m. The way nature keeps them without tangling up and knotting is a fascinating topic in itself. All DNA molecules consist of a double helix of sugar derivates, to which four nuclei acids are attached in irregular order. Nowadays, it is possible to make images of single DNA Ref. 38 molecules; an example is shown in Figure 14.

At the start of the twentieth century it became clear that Desoxyribonukleinsäure (DNS) – translated as deoxyribonucleic acid (DNA) into English – was precisely what Erwin Schrödinger had predicted to exist in his book *What Is Life?* As central part of the chromosomes contained the cell nuclei, DNA is responsible for the storage and reproduction of the information on the construction and functioning of Eukaryotes. The information is coded in the ordering of the four nucleic acids. DNA is the carrier of hereditary information. DNA determines in great part how the single cell we all once have been grows into the complex human machine we are as adults. For example, DNA determines the hair colour, predisposes for certain illnesses, determines the maximum size one can

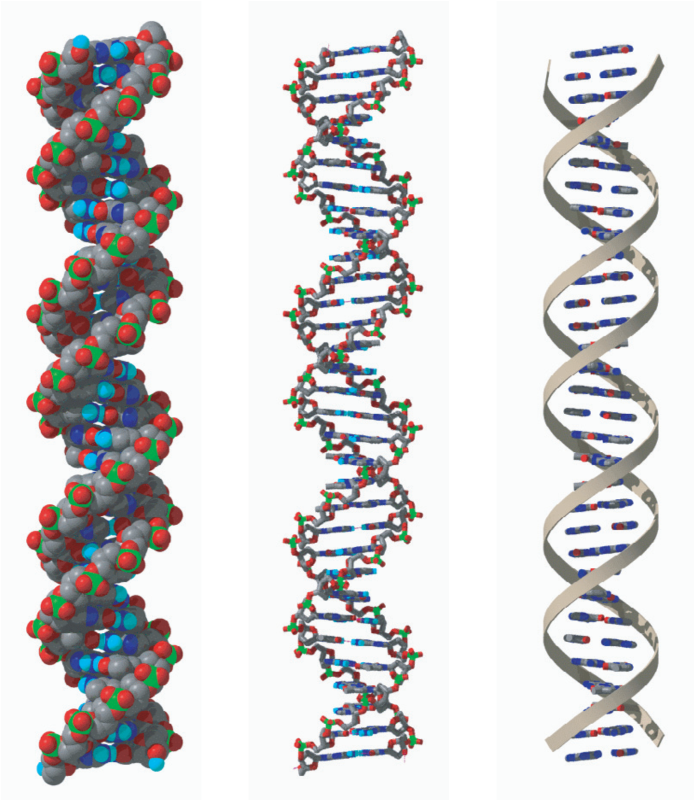


FIGURE 13 Several ways to picture B-DNA, all in false colours (© David Deerfield)

grow to, and much more. Of all known molecules, human DNA is thus most intimately related to human existence. The large size of the molecules is the reason that understanding its full structure and its full contents is a task that will occupy scientists for several generations to come.

Curiosities and fun challenges about chemistry

One of the most fascinating topics of chemistry is that of poisons. Over 50 000 poisons are known, starting with water (usually kills when drunk in amounts larger than about 10 l) and table salt (can kill when 100 g are ingested) up to polonium 210 (kills in doses as low as 5 ng, much less than a spec of dust). Most countries have publicly accessible poison databases; see for example www.gsbl.de.

Can you imagine why 'toxicology', the science of poisons, actually means 'bow science' in Greek? Challenge 31 e

However, not all poisons are chemical. Paraffin and oil for lamps, for example, regularly kill children because the oil enters the lung and forms a thin film over the alveoles, preventing oxygen intake. This so-called *lipoid pneumonia* can be deadly even when only a single drop of oil is in the mouth and then inhaled by a child.

* *

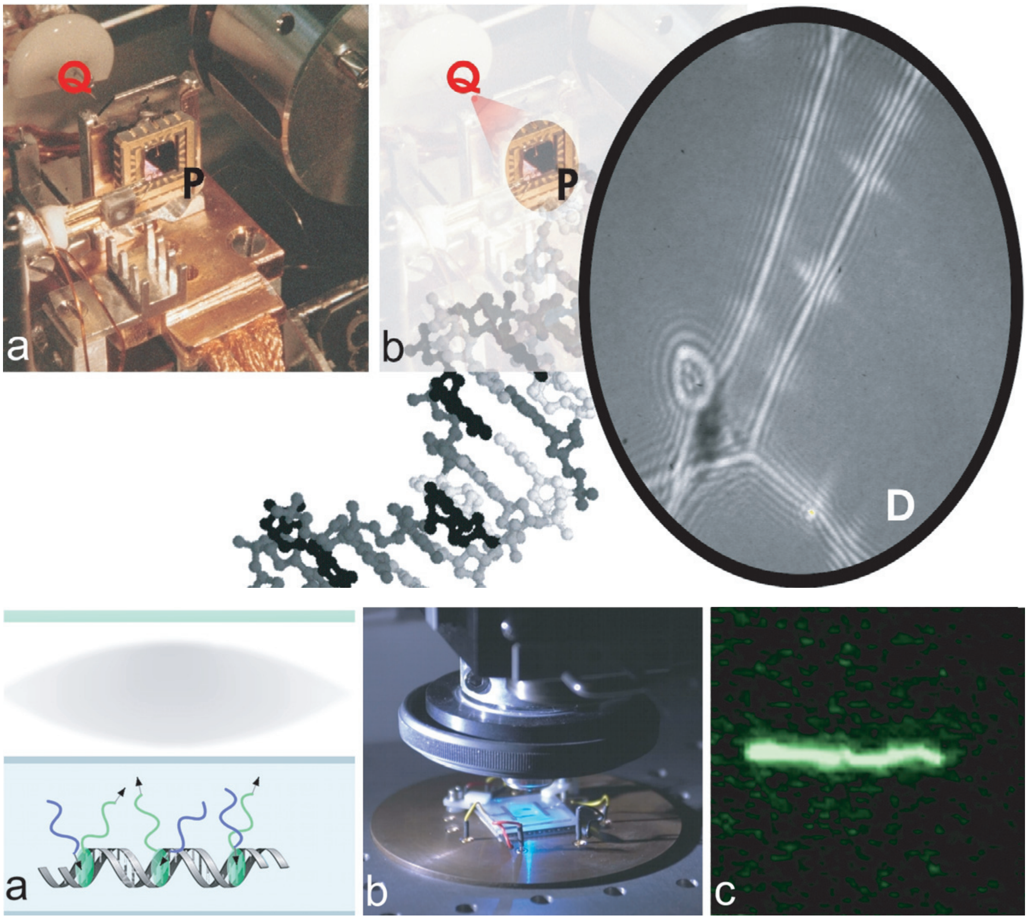


FIGURE 14 Two ways to image single DNA molecules: by holography with electrons emitted from atomically sharp tips (top) and by fluorescence microscopy, with a commercial optical microscope (bottom) (© Hans-Werner Fink/Wiley VCH)

A cube of sugar does not burn. However, if you put some cigarette ash on top of it, it burns. Why? Challenge 32 ny

* *

When one mixes 50 ml of distilled water and 50 ml of ethanol (alcohol), the volume of the mixture has less than 100 ml. Why? Challenge 33 e

* *

Why do organic materials burn at much lower temperature than inorganic materials? Challenge 34 ny

* *

An important aspect of life is death. When we die, conserved quantities like our energy, momentum, angular momentum and several other quantum numbers are redistributed. They are redistributed because conservation means that nothing is lost. What does all this imply for what happens after death? Challenge 35 s

* *

Chemical reactions can be slow but still dangerous. Spilling mercury on aluminium will lead to an amalgam that reduces the strength of the aluminium part after some time. That is the reason that bringing mercury thermometers on aeroplanes is strictly forbidden.

* *

Challenge 36 s
What happens if you take the white power potassium iodide () and the white power lead nitrate ($Pb(NO_3)_2$) and mix them with a masher?(This needs to be done with proper protection and supervision.)

* *

Writing on paper with a pen filled with lemon juice instead of ink produces invisible writing. Later on, the secret writing can be made visible by carefully heating the paper on top of a candle flame.

* *

In 2008, it was shown that perispinal infusion of a single substance, *etanercept*, reduced Alzheimer's symptoms in a patient with late-onset Alzheimer's disease, within a few minutes. Curing Alzheimer's disease is one of the great open challenges for modern medicine.

MATERIALS SCIENCE

> “Did you know that one cannot use a boiled egg as a toothpick?”
>
> Karl Valentin

We mentioned several times that the quantum of action explains all properties of matter. Many researchers in physics, chemistry, metallurgy, engineering, mathematics and biology have cooperated in the proof of this statement. In our mountain ascent we have only a little time to explore this vast but fascinating topic. Let us walk through 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 in two ideas: First, atoms do not penetrate each other: despite being mostly empty clouds, atoms keep a distance. Secondly, atoms cannot be compressed in size. Both properties are due to Pauli's exclusion principle between electrons. The fermion character of electrons avoids that atoms shrink or interpenetrate – on Earth.
Vol. IV, page 113

In fact, not all floors keep up due to the fermion character of electrons. Atoms are not impenetrable at all pressures. At sufficiently large pressures, atoms can collapse, and form new types of floors. Such floors do not exist on Earth. These 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 usual *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.

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.

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 $2.3 \cdot 10^{17}\,\mathrm{kg/m^3}$. A drop weighs about 200 000 tons. In these so-called *neutron stars*, the floor – or better, the size – is also determined by Pauli pressure; however, it is the Pauli Page 138 pressure between neutrons, triggered by the nuclear interactions. These neutron stars are all around 10 km in radius.

If the pressure increases still further, the star becomes a *black hole*, and never stops Page 235 collapsing. Black holes have no floor at all; they still have a constant size though, determined by the horizon curvature.

The question whether other star types exist in nature, with other floor forming mechanisms – such as *quark stars* – is still a topic of research.

Rocks and stones

If a geologist takes a stone in his hands, he is usually able to give, within an error of a few per cent, the age of the stone simply by looking at it. The full story behind this ability forms a large part of geology, but the general lines should be known to every physicist.

Every stone arrives in your hand through the *rock cycle*. The rock cycle is a process that transforms magma from the interior of the Earth into *igneous (or magmatic) rocks* through cooling and crystallization. Igneous rocks, such as basalt, can transform through erosion, transport and deposition into *sedimentary rocks*. Either of these two rock types can be transformed through high pressures or temperatures into *metamorphic rocks*, such as marble. Finally, most rocks are generally – but not always – transformed back into magma.

The full rock cycle takes around 110 to 170 million years. For this reason, rocks that are older than this age are much less common on Earth. Any stone is the product of erosion of one of the rock types. A geologist can usually tell, simply by looking at it, the type of rock it belongs to; if he sees the original environment, he can also give the age, without any laboratory.

For a physicist, most rocks are mixtures of crystals. *Crystals* are solids with a regular arrangement of atoms. They form a fascinating topic by themselves.

Some interesting crystals

Every crystal, like every structure in nature, is the result of growth. Every crystal is thus the result of motion. To form a crystal whose regularity is as high as possible and whose shape is as symmetric as possible, the required motion is a *slow* growth of facets from the liquid (or gaseous) basic ingredients. The growth requires a certain pressure, temperature and temperature gradient for a certain time. For the most impressive crystals, the

TABLE 4 The types of rocks and stones

Type	Properties	Subtype	Example
Igneous rocks (magmatites)	formed from magma, 95% of all rocks	volcanic or extrusive	basalt (ocean floors, Giant's Causeway), andesite, obsidian
		plutonic or intrusive	granite, gabbro
Sedimentary rocks (sedimentites)	often with fossils	clastic	shale, siltstone, sandstone
		biogenic	limestone, chalk, dolostone
		precipitate	halite, gypsum
Metamorphic rocks (metamorphites)	transformed by heat and pressure	foliated	slate, schist, gneiss (Himalayas)
		non-foliated (grandoblastic or hornfelsic)	marble, skarn, quartzite
Meteorites	from the solar system	rock meteorites	
		iron meteorites	

gemstones, the conditions are usually quite extreme; this is the reason for their durability. The conditions are realized in specific rocks deep inside the Earth, where the growth process can take thousands of years. Mineral crystals can form in all three types of rocks: Page 52 igneous (magmatic), metamorphic, and sedimentary. Other crystals can be made in the laboratory in minutes, hours or days and have led to a dedicated industry. Only a few crystals grow from liquids at standard conditions; examples are gypsum and several other sulfates, which can be crystallized at home, potassium bitartrate, which appears in the making of wine, and the crystals grown inside plants or animals, such as teeth, bones or magnetosensitive crystallites.

Growing, cutting, treating and polishing crystals is an important industry. Especially the growth of crystals is a science in itself. Can you show with pencil and paper that Challenge 37 e only the *slowest* growing facets are found in crystals? In the following, a few important crystals are presented.

* *

Quartz, amethyst (whose colour is due to radiation and iron Fe^{4+} impurities), citrine (whose colour is due to Fe^{3+} impurities), smoky quartz (with colour centres induced by radioactivity), agate and onyx are all forms of crystalline silicon dioxide or SiO_2. Quartz forms in igneous and in magmatic rocks; crystals are also found in many sedimentary rocks. Quartz crystals can be larger than human. By the way, most amethysts lose their colour with time, so do not waste money buying them.

Quartz is the most common crystal on Earth's crust and is also grown synthetically for many high-purity applications. The structure is rombohedral, and the ideal shape is a six-sided prism with six-sided pyramids at its ends. Quartz melts at 1986 K and is piezo- and pyroelectric. Its piezoelectricity makes it useful as electric oscillator and filter. Vol. I, page 235 A film of an oscillating clock quartz is part of the first volume. Quartz is also used for glass production, in communication fibres, for coating of polymers, in gas lighters, as source of silicon and for many other applications.

FIGURE 15 Quartz found at St. Gotthard, Switzerland, picture size 12 cm (© Rob Lavinsky)

FIGURE 16 Citrine found on Magaliesberg, South Africa, crystal height 9 cm (© Rob Lavinsky)

FIGURE 17 Amethystine and orange quartz found in the Orange River, Namibia, picture size 6 cm (© Rob Lavinsky)

∗ ∗

Corundum, ruby and sapphire are crystalline variations of alumina. Corundum is pure and colourless Al_2O_3, ruby is Cr doped and blue sapphire is Ti or Fe doped. They have trigonal crystal structure and melt at 2320 K. Natural gems are formed in metamorphic rocks. Yellow, green, purple, pink, brown, grey and salmon-coloured sapphires also exist, when doped with other impurities. The colours of natural sapphires, like that of many other gemstones, are often changed by baking and other treatments.

Corundum, ruby and sapphire are used in jewellery, as heat sink and growth substrate, and for lasers. Corundum is also used as scratch-resistant 'glass' in watches. Ruby was the first gemstone that was grown synthetically in gem quality, in 1892 by Auguste Verneuil (1856-1913), who made his fortune in this way.

∗ ∗

Tourmaline is a frequently found mineral and can be red, blue, green, orange, yellow, pink or black, depending on its composition. The chemical formula is astonishingly complex and varies from type to type. Tourmaline has trigonal structure and usually forms columnar crystals that have triangular cross-section. It is only used in jewellery. Paraiba tourmalines, a very rare type of green or blue tourmaline, are among the most beautiful gemstones and can be, if untreated, more expensive than diamonds.

FIGURE 18 Corundum found in Laacher See, Germany, picture size 4 mm (© Stephan Wolfsried)

FIGURE 19 Ruby found in Jagdalak, Afghanistan, picture height 2 cm (© Rob Lavinsky)

FIGURE 20 Sapphire found in Ratnapura, Sri Lanka, crystal size 1.6 cm (© Rob Lavinsky)

FIGURE 21 Natural bicoloured tourmaline found in Paprok, Afghanistan, picture size 9 cm (© Rob Lavinsky)

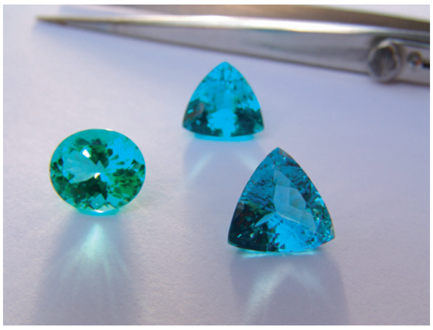


FIGURE 22 Cut Paraiba tourmaline from Brazil, picture size 3 cm (© Manfred Fuchs)

* *

Garnets are a family of compounds of the type $X_2Y_3(SiO_4)_3$. They have cubic crystal structure. They can have any colour, depending on composition. They show no cleavage and their common shape is a rhombic dodecahedron. Some rare garnets differ in colour when looked at in daylight or in incandescent light. Natural garnets form in metamorphic rocks and are used in jewellery, as abrasive and for water filtration. Synthetic garnets are used in many important laser types.

FIGURE 23 Red garnet with smoky quartz found in Lechang, China, picture size 9 cm (© Rob Lavinsky)

FIGURE 24 Green demantoid, a garnet owing its colour to chromium doping, found in Tubussis, Namibia, picture size 5 cm (© Rob Lavinsky)

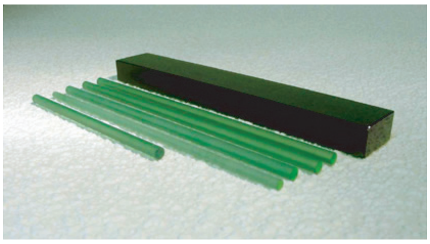


FIGURE 25 Synthetic Cr,Tm,Ho:YAG, a doped yttrium aluminium garnet, picture size 25 cm (© Northrop Grumman)

* *

Alexandrite, a chromium-doped variety of chrysoberyl, is used in jewellery and in lasers. Its composition is $BeAl_2O_4$; the crystal structure is orthorhombic. Chrysoberyl melts at 2140 K. Alexandrite is famous for its colour-changing property: it is green in daylight or fluorescent light but amethystine in incandescent light, as shown in Figure 23. The effect is due to its chromium content: the ligand field is just between that of chromium in red ruby and that in green emerald. A few other gems also show this effect, in particular the rare blue garnet and some Paraiba tourmalines.

FIGURE 26 Alexandrite found in the Setubal river, Brazil, crystal height 1.4 cm, illuminated with daylight (left) and with incandescent light (right) (© Trinity Mineral)

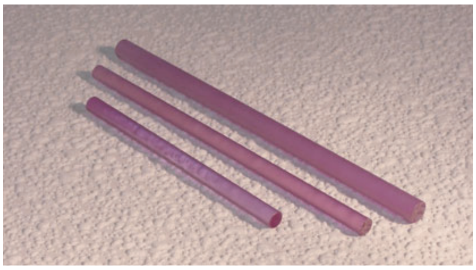


FIGURE 27 Synthetic alexandrite, picture size 20 cm (© Northrop Grumman)

* *

Perovskites are a large class of cubic crystals used in jewellery and in tunable lasers. Their general composition is XYO_3, XYF_3 or $XYCl_3$.

* *

Diamond is a metastable variety of graphite, thus pure carbon. Theory says that graphite

FIGURE 28 Perovskite found in Hillesheim, Germany, picture width 3 mm (© Stephan Wolfsried)

to be added

FIGURE 29 Synthetic perovskite, picture size c... cm (© Aa Bb)

is the stable form; practice says that diamond is still more expensive. In contrast to graphite, diamond has face-centred cubic structure, is a large band gap semiconductor and typically has octahedral shape. Diamond burns at 1070 K; in the absence of oxygen it converts to graphite at around 1950 K. Diamond can be formed in magmatic and in metamorphic rocks. Diamonds can be synthesized in reasonable quality, though gemstones of large size and highest quality are not yet possible. Diamond can be coloured and be doped to achieve electrical conductivity in a variety of ways. Diamond is mainly used in jewellery, for hardness measurements and as abrasive.

FIGURE 30 Natural diamond from Saha republic, Russia, picture size 4 cm (© Rob Lavinsky)

to be added

FIGURE 31 Synthetic diamond, picture size c... cm (© Aa Bb)

to be added

FIGURE 32 Ophtalmic diamond knife, picture size c... cm (© Aa Bb)

* *

Silicon, Si, is not found in nature in pure form; all crystals are synthetic. The structure is face-centred cubic, thus diamond-like. It is moderately brittle, and can be cut in thin wafers which can be further thinned by grinding or chemical etching, even down to a thickness of 10 μm. Being a semiconductor, the band structure determines its black colour, its metallic shine and its brittleness. Silicon is widely used for silicon chips and electronic semiconductors. Today, human-sized silicon crystals can be grown free of dislocations and other line defects. (They will still contain some point defects.)

* *

Teeth are the structures that allowed animals to be so successful in populating the Earth.

FIGURE 33 Silicon crystal growing machine, and two resulting crystals, with a length of c. 2 m (© www.PVATePla.com)

They are composed of several materials; the outer layer, the *enamel*, is 97% hydroxylapatite, mixed with a small percentage of two proteins groups, the amelogenins and the enamelins. The growth of teeth is still not fully understood; neither the molecular level nor the shape-forming mechanisms are completely clarified. Hydroxylapatite is soluble in acids; addition of fluorine ions changes the hydroxylapatite to fluorapatite and greatly reduces the solubility. This is the reason for the use of fluorine in tooth paste.

Hydroxylapatite (or hydroxyapatite) has the chemical formula $Ca_{10}(PO_4)_6(OH)_2$, possesses hexagonal crystal structure, is hard (more than steel) but relatively brittle. It occurs as mineral in sedimentary rocks (see Figure 34), in bones, renal stones, bladders tones, bile stones, atheromatic plaque, cartilage arthritis and teeth. Hydroxylapatite is mined as a phosphorus ore for the chemical industry, is used in genetics to separate single and double-stranded DNA, and is used to coat implants in bones.

FIGURE 34 Hydroxylapatite found in Snarum, Norway, picture size 5 cm (© Rob Lavinsky)

FIGURE 35 The main and the reserve teeth on the jaw bone of a shark, all covered in hydroxylapatite, picture size 15 cm (© Peter Doe)

* *

Pure metals, such as gold, silver and even copper, are found in nature, usually in magmatic rocks. But only a few metallic compounds form crystals, such as pyrite. Monocrystalline pure metal crystals are all synthetic. Monocrystalline metals, for example iron, aluminium, gold or copper, are extremely soft and ductile. Either bending them repeatedly – a process called cold working – or adding impurities, or forming alloys makes them hard and strong. Stainless steel, a carbon-rich iron alloy, is an example that uses all three processes.

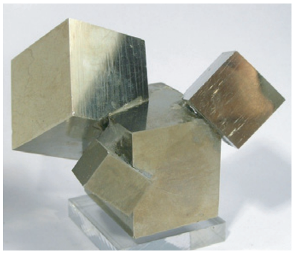


FIGURE 36 Pyrite, found in Navajún, Spain, picture width 5.7 cm (© Rob Lavinsky)

FIGURE 37 Silver from Colquechaca, Bolivia, picture width 2.5 cm (© Rob Lavinsky)

FIGURE 38 Synthetic copper single crystal, picture width 30 cm (© Lachlan Cranswick)

* *

In 2009, Luca Bindi of the Museum of Natural History in Florence, Italy, made headlines Ref. 39 across the world with his discovery of the first natural *quasicrysal*. Quasicrystals are materials that show non-crstallographic symmetries. Until 2009, only synthetic materials

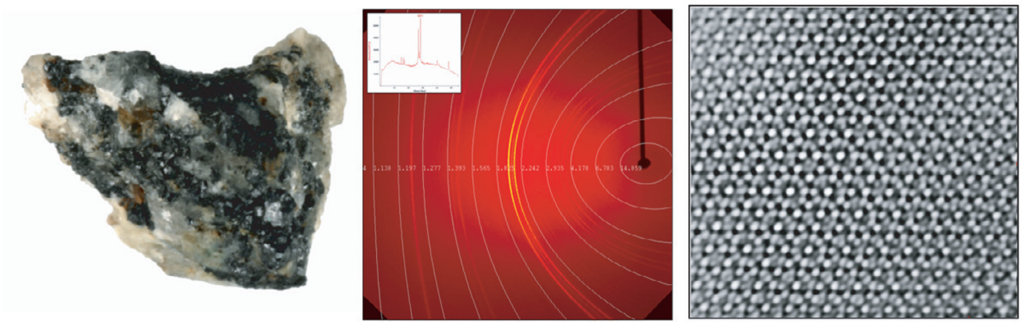


FIGURE 39 The specimen, found in the Koryak Mountains in Russia, is part of a triassic mineral, about 220 million years old; the black material is mostly khatyrkite ($CuAl_2$) and cupalite ($CuAl_2$) but also contains quasicrystal grains with composition $Al_{63}Cu_{24}Fe_{13}$ that have fivefold symmetry, as clearly shown in the X-ray diffraction pattern and in the transmission electron image. (© Luca Bindi)

were known. Then, in 2009, after years of searching, Bindi discovered a specimen in his collection whose grains clearly show fivefold symmetry.

* *

There are about 4000 known mineral types. On the other hand, there are ten times as many obsolete mineral names, namely around 40 000. An official list can be found in various places on the internet, including www.mindat.org or www.minieralienatlas.de. To explore the world of crystal shapes, see the www.smorf.nl website. Around new 40 minerals are discovered each year. Searching for minerals and collecting them is a fascinating pastime.

How can we 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 5 gives an overview of the possibilities.

TABLE 5 Signals penetrating mountains and other matter

Signal	Penetration depth in stone	Achieved resolution	Material dependence	Use
Matter				
Diffusion of water or liquid chemicals	*c.* 5 km	*c.* 100 m	medium	mapping hydrosystems
Diffusion of gases	*c.* 5 km	*c.* 100 m	medium	studying vacuum systems
Electromagnetism				
Infrasound and earthquakes	100 000 km	100 km	high	mapping of Earth crust and mantle

TABLE 5 (Continued) Signals penetrating mountains and other matter

Signal	Penetration depth in stone	Achieved resolution	Material dependence	Use
Sound, explosions, seismic waves	0.1 – 10 m	*c.* $\lambda/100$	high	oil and ore search, structure mapping in rocks, searching for underwater treasures in sunken ship with sub-bottom-profilers
Ultrasound		1 mm	high	medical imaging, acoustic microscopy
Static magnetic field variation			medium	cable search, cable fault localization, search for structures and metal inside soil, rocks and seabed
Electrical currents				soil and rock investigations, search for tooth decay
Electromagnetic sounding, 0.2 – 5 Hz				soil and rock investigations in deep water and on land
Radio waves	10 m	30 m to 1 mm	small	soil radar (up to 10 MW), magnetic imaging, research into solar interior
Ultra-wide band radio	10 cm	1 mm	sufficient	searching for wires and tubes in walls, breast cancer detection
Mm and THz waves	below 1 mm	1 mm		see through clothes, envelopes and teeth Ref. 40
Infrared	*c.* 1 m	0.1 m	medium	mapping of soil over 100 m
Visible light	*c.* 1 cm	0.1 µm	medium	imaging of many sorts
X-rays	a few metres	5 µm	high	medicine, material analysis, airports, food production check
γ-rays	a few metres	1 mm	high	medicine
Neutrons from a reactor	up to *c.* 1 m	1 mm	medium	tomography of metal structures, e.g., archeological statues or engines
Muons created by cosmic radiation	up to *c.* 300 m	0.1 m	small	finding caves in pyramids, imaging interior of trucks
Positrons	up to *c.* 1 m	2 mm	high	used in medicine for tomography
Electrons	up to *c.* 1 µm	10 nm	small	used in transmission electron microscopes
Weak interactions				
Neutrino beams	light years	zero	very weak	studies of Sun
Strong interactions				
Radioactivity	1 mm to 1 m			airport security checks

TABLE 5 (Continued) Signals penetrating mountains and other matter

Signal	Penetration depth in stone	Achieved resolution	Material dependence	Use
Gravitation				
Variation of g		50 m	low	oil and 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? Challenge 38 s A motion expert like the reader should be able to give a definite answer.

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 metre or less: magnetic resonance imaging. We will discuss it later on. Page 119

The other modern imaging technique, ultrasound imaging, is useful in medicine. However, its use for prenatal diagnostics of embryos is not recommended. Page 246 Studies have found that ultrasound produces extremely high levels of audible sound to the baby, especially when the ultrasound is repeatedly switched on and 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. In 1996, a team of Dutch scientists found a material that can be switched from mirror mode to transparent mode using an electrical signal. Ref. 41 This seems a first step to realize the dream to become invisible and visible at will.

In 2006, and repeatedly since then, researchers made the headlines in the popular press by claiming that they could build a *cloak of invisibility*. This is a blatant lie. This lie is frequently used to get funding from gullible people, such as buyers of bad science fiction books or the military. It is claimed that objects can be made invisible by covering them with metamaterials. The impossibility of this aim has been already shown earlier on. Vol. III, page 131 But we now can say more.

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

TABLE 6 Matter at lowest temperatures

Phase	Type	Low temperature behaviour	Example
Solid	conductor	superconductivity	lead, MgB_2 (40 K)
		antiferromagnet	chromium, MnO
		ferromagnet	iron
	insulator	diamagnet	
Liquid	bosonic	Bose–Einstein condensation, i.e., superfluidity	^{4}He
	fermionic	pairing, then BEC, i.e., superfluidity	^{3}He
Gas	bosonic	Bose–Einstein condensation	^{87}Rb, ^{7}Li, ^{23}Na, H, ^{4}He, ^{41}K
	fermionic	pairing, then Bose–Einstein condensation	^{40}K, ^{6}Li

the eye, due to their refractive index. Quartz can be invisible only when submerged in liquids with the same refractive index.

In short, anything that has a shape cannot be invisible. If we want to become invisible, we must transform ourselves into a diffuse gas cloud of non-absorbing atoms. On the way to become invisible, we would lose all memory and all genes, in short, we would lose all our individuality. But an individual cannot be made of gas. An individual is defined through its boundary. There is no way that we can be invisible and alive at the same time; a way to switch back to visibility is even less impossible. In summary, quantum theory shows that only the dead can be invisible. Quantum theory is thus reassuring. We saw Vol. IV, page 114 already that quantum theory forbids ghosts; we now find that it also forbids any invisible beings.

How does matter behave at the lowest temperatures?

The low-temperature behaviour of matter has numerous experimental and theoretical aspects. The first issue is whether matter is always solid at low temperatures. The answer is no: all phases exist at low temperatures, as shown in Table 6.

Concerning the electric properties of matter at lowest temperatures, the present status is that matter is either insulating or superconducting. Finally, one can ask about the magnetic properties of matter at low temperatures. We know already that matter can not be paramagnetic at lowest temperatures. It seems that matter is either ferromagnetic, diamagnetic or antiferromagnetic at lowest temperatures.

Page 70 More about superfluidity and superconductivity will be told below.

Curiosities and fun challenges about materials science

What is the maximum height of a mountain? This question is of course of interest to all Ref. 42 climbers. Many effects limit the height. The most important is the fact that under heavy Challenge 39 ny 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
Challenge 40 s 80 km 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 factory, is cheap and can substitute duck feathers in bed covers, sleeping bags or in badminton Challenge 41 r shuttlecocks. Another industrial challenge is to find an artificial substitute for latex, and a third one is to find a substitute for a material that is rapidly disappearing due to pollution: cork.

* *

How much does the Eiffel tower change in height over a year due to thermal expansion Challenge 42 s and contraction?

* *

Challenge 43 ny What is the difference between solids, liquids and gases?

* *

What is the difference between the makers of bronze age knifes and the builders of the Eiffel tower? Only their control of defect distributions. The main defects in metals are *disclinations* and *dislocations*. Disclinations are crystal defects in form of surfaces; they are the microscopic aspect of grain boundaries. Dislocations are crystal defects in form of curved lines; above all, their distribution and their motion in a metal determines the Page 221 stiffness. For a picture of dislocations, see below.

* *

Challenge 44 s 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 a box made of such Challenge 45 s walls can never be tightly shut. Can you provide the argument?

* *

Quantum theory predicts that heat transport at a given temperature is quantized. Can Challenge 46 ny you guess the unit of thermal conductance?

* *

Robert Full has shown that van der Waals forces are responsible for the way that geckos Ref. 43 walk on walls and ceilings. The gecko, 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 (or alternatively, capillary forces) to stick to the surface. As a result, the gecko can walk on vertical glass walls or even on glass ceilings; the sticking force can be as high as 100 N per foot.

The same mechanism is used by jumping spiders (*Salticidae*). For example, *Evarcha arcuata* have hairs at their feet which are covered by hundred of thousands of setules. Ref. 44 Again, the van der Waals force in each setule helps the spider to stick on surfaces.

Researchers have copied these mechanisms for the first time in 2003, using microlithography on polyimide, and hope to make durable sticky materials in the near future.

* *

Ref. 45 One of the most fascinating material in nature are bones. Bones are light, stiff, and can heal after fractures. If you are interested in composite materials, read more about bones: their structure is incredibly complex.

* *

Millimetre waves or terahertz waves are emitted by all bodies at room temperature. Modern camera systems allow to image them. In this way, it is possible to see through clothes. This ability could 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 under way. The waves can also be used to see through paper, thus making it unnecessary to open letters in order to read them. Secret services are exploiting this technique. A third application of terahertz waves might be in medical diagnostic, for example for the search of tooth decay. Terahertz waves are almost without side effects, and thus superior to X-rays. The lack of low-priced quality sources is still an obstacle to their application.

* *

Ref. 46 Does the melting point of water depend on the magnetic field? This surprising claim was made in 2004 by Inaba Hideaki and colleagues. They found a change of 0.9 mK/T. It is known that the refractive index and the near infrared spectrum of water is affected by magnetic fields. Indeed, not everything about water might be known yet.

* *

Plasmas, or ionized gases, are useful for many applications. Not only can they be used for heating or cooking and generated by chemical means (such plasmas are variously called fire or flames) but they can also be generated electrically and used for lighting or deposition of materials. Electrically generated plasmas are even being studied for the disinfection of dental cavities.

* *

Ref. 47 Challenge 47 s It is known that the concentration of CO_2 in the atmosphere between 1800 and 2005 has increased from 280 to 380 parts per million, as shown in Figure 40. (How would you measure this?) It is known without doubt that this increase is due to human burning of fossil fuels, and not to natural sources such as the oceans or volcanoes. There are three arguments. First of all, there was a parallel decline of the $^{14}C/^{12}C$ ratio. Second, there was a parallel decline of the $^{13}C/^{12}C$ ratio. Finally, there was a parallel decline of the oxygen concentration. All three measurements independently imply that the CO_2 increase is due to the burning of fuels, which are low in ^{14}C and in ^{13}C, and at the same time decrease the oxygen ratio. Natural sources do not have these three effects. Since CO_2 is a major greenhouse gas, the data implies that humans are also responsible for a large part of the temperature increase during the same period. Global warming exists and is mainly due to humans. The size of the effect, however, is still a matter of heated dispute.

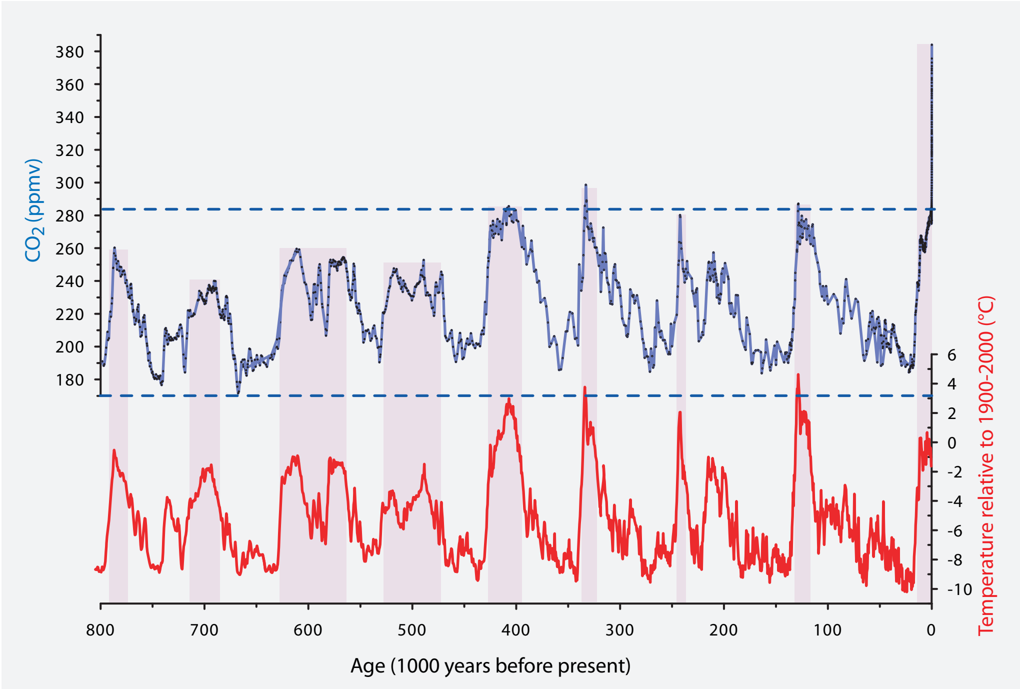


FIGURE 40 The concentration of CO_2 and the change of average atmospheric temperature in the past 0.8 million years (© Dieter Lüthi)

* *

Making crystals can make one rich. The first man who did so, the Frenchman Auguste Verneuil (1856–1913), sold rubies grown in his laboratory for many years without telling anybody. Many companies now produce synthetic gems.

Synthetic diamonds have now displaced natural diamonds in almost all applications. In the last years, methods to produce large, white, jewel-quality diamonds of ten carats and more are being developed. These advances will lead to a big change in all the domains that depend on these stones, such as the production of the special surgical knives used in eye lens operation. Ref. 48

* *

The technologies to produce *perfect* crystals, without grain boundaries or dislocations, are an important part of modern industry. Perfectly regular crystals are at the basis of the integrated circuits used in electronic appliances, are central to many laser and telecommunication systems and are used to produce synthetic jewels.

* *

Challenge 48 ny How can a small plant pierce through tarmac?

* *

If you like abstract colour images, do not miss looking at liquid crystals through a microscope. You will discover a wonderful world. The best introduction is the text by Ingo

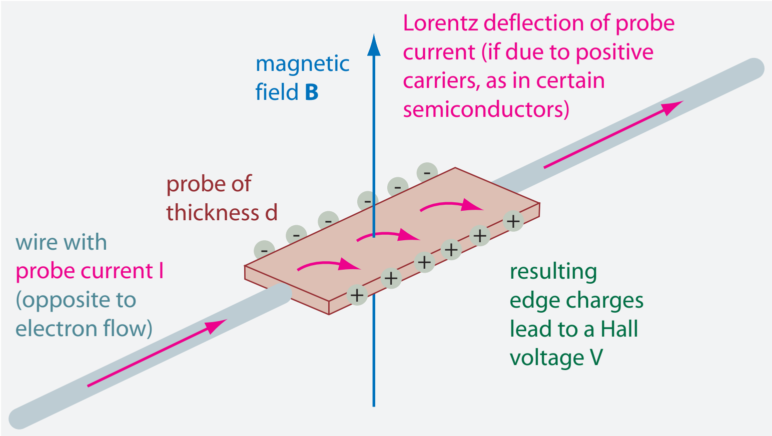


FIGURE 41 The Hall effect

Ref. 49 Dierking.

* *

The Lorentz force leads to an interesting effect inside materials. If a current flows along a conducting strip that is in a (non-parallel) magnetic field, a voltage builds up between two edges of the conductor, because the charge carriers are deflected in their flow. This effect is called the (classical) *Hall effect* after the US-American physicist Edwin Hall (1855–1938), who discovered it in 1879, during his PhD. The effect, shown in Figure 41, is regularly used, in so-called *Hall probes*, to measure magnetic fields; the effect is also used to read data from magnetic storage media or to measure electric currents (of the order of 1 A or more) in a wire without cutting it. Typical Hall probes have sizes of around 1 cm down to 1 μm and less. The Hall voltage V turns out to be given by

$$V = \frac{IB}{ned} , \qquad (7)$$

where n is the electron number density, e the electron charge, and d is the thickness of the probe, as shown in Figure 41. Deducing the equation is a secondary school exercise. (Challenge 49 e) The Hall effect is a material effect, and the material parameter n determines the Hall voltage. The sign of the voltage also tells whether the material has positive or negative charge carriers; indeed, for metal strips the voltage polarity is opposite to the one shown in the figure. (Challenge 50 e)

Many variations of the Hall effect have been studied. For example, the *quantum* Hall effect will be explored below. Page 71

In 1998, Geert Rikken and his coworkers found that in certain materials photons can also be deflected by a magnetic field; this is the *photonic Hall effect*. Ref. 50

In 2005, again Geert Rikken and his coworkers found a material, a terbium gallium garnet, in which a flow of phonons in a magnetic field leads to temperature difference on the two sides. They called this the *phonon Hall effect*. Ref. 51

* *

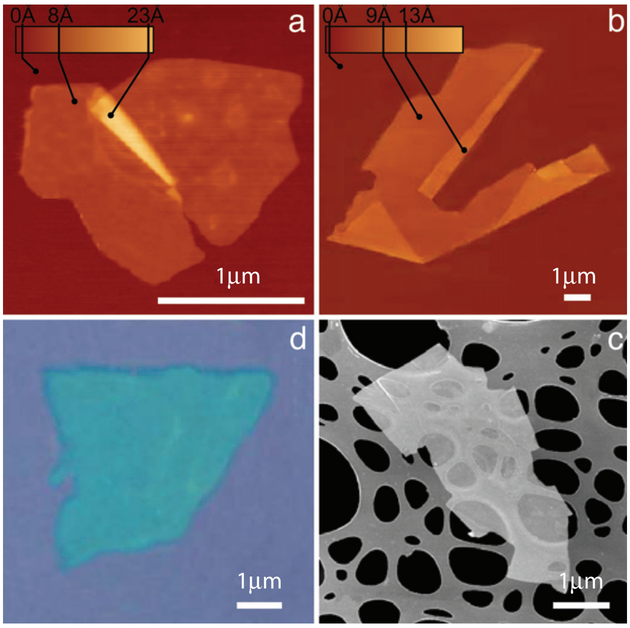


FIGURE 42 Single atom sheets, mapped by atomic force microscopy, of $NbSe_2$ (a) and of graphite or graphene (b), a single atom sheet of MoS_2 imaged by optical microscopy (d), and a single atom sheet of $Bi_2Sr_2CaCu_2O_x$ on a holey carbon film imaged by scanning electron microscopy (c) (from Ref. 53, © 2005 National Academy of Sciences)

Do magnetic fields influence the crystallization of calcium carbonate in water? This issue is topic of intense debates. It might be, or it might not be, that magnetic fields change the crystallization seeds from calcite to aragonite, thus influencing whether water tubes are covered on the inside with carbonates or not. The industrial consequences of reduction in *scaling*, as this process is called, would be enormous. But the issue is still open, as are convincing data sets. Ref. 52

* *

Ref. 53 It has recently become possible to make the thinnest possible sheets of graphite and other materials (such as BN, MoS_2, $NbSe_2$, $Bi_2Sr_2CaCu_2O_x$): these crystal sheets are precisely one atom thick! The production of *graphene* – that is the name of a monoatomic graphite layer – is extremely complicated: you need graphite from a pencil and a roll of adhesive tape. That is probably why it was necessary to wait until 2004 for the development of the technique. (In fact, the stability of monoatomic sheets was questioned for many years before that. Some issues in physics cannot be decided with paper and pencil; sometimes you need adhesive tape as well.) Graphene and the other so-called *two-dimensional crystals* (this is, of course, a tabloid-style exaggeration) are studied for their electronic and mechanical properties; in future, they might even have applications in high-performance batteries.

* *

Gold absorbs light. Therefore it is used, in expensive books, to colour the edges of pages. Apart from protecting the book from dust, it also prevents that sunlight lets the pages turn yellow near the edges.

FIGURE 43 The beauty of materials science: the surface of a lotus leaf leads to almost spherical water droplets; plasma-deposited PTFE, or teflon, on cotton leads to the same effect for the coloured water droplets on it (© tapperboy, Diener Electronics)

FIGURE 44 A piece of aerogel, a solid that is so porous that it is translucent (courtesy NASA)

∗ ∗

Like trees, crystals can have growth rings. Smoke quartz is known for these so-called *phantoms*, but also fluorite and calcite.

∗ ∗

The science and art of surface treatment is still in full swing, as Figure 43 shows. Making hydrophobic surfaces is an important part of modern material science, that copies what the lotus, *Nelumbo nucifera*, does in nature. Hydrophobic surfaces allow that water droplets bounce on them, like table tennis balls on a table. The lotus surface uses this property to clean itself, hence the name *lotus effect*. This is also the reason that lotus plants have become a symbol of purity.

Vol. I, page 38

∗ ∗

Sometimes research produces bizarre materials. An example are the so-called *aerogels*, highly porous solids, shown in Figure 44. Aerogels have a density of a few g/l, thus a few hundred times lower than water and only a few times that of air. Like any porous material, aerogels are good insulators; however, they are easily destroyed and therefore have not found important applications up to now.

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 have shaped modern life in many ways.

Motion without friction – superconductivity and superfluidity

We are used to thinking that friction is inevitable. We even learned that friction was an inevitable result of the particle structure of matter. It should come to the surprise of every physicist that motion without friction is possible.

In 1911 Gilles Holst and Heike Kamerlingh Onnes discovered that at low temperatures, electric currents can flow with no resistance, i.e., with no friction, through lead. The observation is called *superconductivity*. In the century after that, many metals, alloys and ceramics have been found to show the same behaviour.

The condition for the observation of motion without friction is that quantum effects play an essential role. That is the reason for the requirement of low temperature in such experiments. Nevertheless, it took over 50 years to reach a full understanding of the effect. This happened in 1957, when Bardeen, Cooper and Schrieffer published their results. Ref. 54 At low temperatures, electron behaviour is dominated by an attractive interaction that makes them form pairs – today called *Cooper pairs* – that are effective bosons. And bosons can all be in the same state, thus effectively moving without friction.

For superconductivity, the attractive interaction between electrons is due to the deformation of the lattice. Two electrons attract each other in the same way as two masses attract each other due to deformation of the space-time mattress. However, in the case of solids, the deformations are quantized. With this approach, Bardeen, Cooper and Schrieffer explained the lack of electric resistance of superconducting materials, their complete diamagnetism ($\mu_r = 0$), the existence of an energy gap, the second-order transition to normal conductivity at a specific temperature, and the dependence of this temperature on the mass of the isotopes. Last but not least, they received the Nobel Prize in 1972.*

Another type of motion without friction is *superfluidity*. Already in 1937, Pyotr Kapitsa had predicted that normal helium (^{4}He), below a transition observed at the temperature of 2.17 K, would be a superfluid. In this domain, the fluid moves without friction through

* For John Bardeen (1908–1991), this was his second, after he had got the first Nobel Prize in Physics in 1956, shared with William Shockley and Walter Brattain, for the discovery of the transistor. The first Nobel Prize was a problem for Bardeen, as he needed time to work on superconductivity. In an example to many, he reduced the tam-tam around himself to a minimum, so that he could work as much as possible on the problem of superconductivity. By the way, Bardeen is topped by Frederick Sanger and by Marie Curie. Sanger first won a Nobel Prize in Chemistry in 1958 by himself and then won a second one shared with Walter Gilbert in 1980; Marie Curie first won one with her husband and a second one by herself, though in two different fields.

tubes. (In fact, the fluid remains a mixture of a superfluid component and a normal component.) Helium is even able, after an initial kick, to flow over obstacles, such as glass walls, or to flow out of bottles. ^{4}He is a boson, so no pairing is necessary for it to flow without friction. This research earned Kapitsa a Nobel Prize in 1978.

The explanation of superconductivity also helped for fermionic superfluidity. In 1972, Richardson, Lee, and Osheroff found that even ^{3}He is superfluid, if temperatures are lowered below 2.7 mK. ^{3}He is a fermion, and requires pairing to become superfluid. In fact, below 2.2 mK, ^{3}He is even superfluid in two different ways; one speaks of phase A and phase B.*

In the case of ^{3}He, the theoreticians had been faster than the experimentalists. The theory for superconductivity through pairing had been adapted to superfluids already in 1958 – before any data were available – by Bohr, Mottelson and Pines. This theory was then adapted and expanded by Anthony Leggett.** The attractive interaction between ^{3}He atoms, the basic mechanism that leads to superfluidity, turns out to be the spin-spin interaction.

In superfluids, like in ordinary fluids, one can distinguish between laminar and turbulent flow. The transition between the two regimes is mediated by the behaviour of vortices. But in superfluids, vortices have properties that do not appear in normal fluids. In the superfluid ^{3}He-B phase, vortices are *quantized*: vortices only exist in integer multiples of the elementary circulation $h/2m_{^3\text{He}}$. Present research is studying how these vortices behave and how they induce the transition form laminar to turbulent flows. Ref. 55

In recent years, studying the behaviour of *gases* at lowest temperatures has become very popular. When the temperature is so low that the de Broglie wavelength is comparable to the atom-atom distance, bosonic gases form a Bose–Einstein condensate. The first one were realized in 1995 by several groups; the group around Eric Cornell and Carl Wieman used ^{87}Rb, Rand Hulet and his group used ^{7}Li and Wolfgang Ketterle and his group used ^{23}Na. For fermionic gases, the first degenerate gas, ^{40}K, was observed in 1999 by the group around Deborah Jin. In 2004, the same group observed the first gaseous Fermi condensate, after the potassium atoms paired up.

The fractional quantum Hall effect

The fractional quantum Hall effect is one of the most intriguing discoveries of materials science. The effect concerns the flow of electrons in a two-dimensional surface. In 1982, Robert Laughlin predicted that in this system one should be able to observe objects with electrical charge $e/3$. This strange and fascinating prediction was indeed verified in 1997. Ref. 56 Ref. 57

The story begins with the discovery by Klaus von Klitzing of the quantum Hall effect. In 1980, Klitzing and his collaborators found that in two-dimensional systems at low temperatures – about 1 K – the electrical conductance S is quantized in multiples of the quantum of conductance Ref. 58

$$S = n \frac{e^2}{\hbar} . \tag{8}$$

* They received the Nobel Prize in 1996 for this discovery.

** Aage Bohr, son of Niels Bohr, and Ben Mottelson received the Nobel Prize in 1975, Anthony Leggett in 2003.

The explanation is straightforward: it is the quantum analogue of the classical Hall effect, which describes how conductance varies with applied magnetic field. The corresponding resistance values are

$$R = \frac{1}{n}\frac{\hbar}{e^2} = \frac{1}{n}\, 25\,812,807557(18)\,\Omega\ . \qquad (9)$$

The values are independent of material, temperature, or magnetic field. They are constants of nature. Von Klitzing received the Nobel Prize for physics for the discovery, because the effect was unexpected, allows a highly precise measurement of the fine structure constant, and also allows one to build detectors for the smallest voltage variations measurable so far. His discovery started a large wave of subsequent research.

Only two years later, it was found that in extremely strong magnetic fields and at extremely low temperatures, the conductance could vary in steps *one third* that size. Shortly afterwards, even stranger numerical fractions were also found. In fact, all fractions of the form $m/(2m+1)$ or of the form $(m+1)/(2m+1)$, m being an integer, are possible. This is the *fractional quantum Hall effect*. In a landmark paper, Robert Laughlin explained all these results by assuming that the electron gas could form collective states showing quasiparticle excitations with a charge $e/3$. This was confirmed experimentally 15 years later and earned him a Nobel Prize 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.

Ref. 59

Ref. 56

Subsequent experiments confirmed Laughlin's deduction. He had predicted the appearance of a new form of a *composite* quasi-particle, built of electrons and of one or several magnetic flux quanta. If an electron bonds with an *even* number of quanta, the composite is a fermion, and leads to Klitzing's *integral* quantum Hall effect. If the electron bonds with an *odd* number of quanta, the composite is a boson, and the *fractional* quantum Hall effect appears. The experimental and theoretical details of these quasi-particles might well be the most complex and fascinating aspects of solid state physics, but exploring them would lead us too far from the aim of our adventure.

In 2007, a new chapter in the story was opened by Andre Geim and his own and a second team, when they discovered a new type of quantum Hall effect at room temperature. They used graphene, i.e., single-atom layers of graphite, and found a relativistic analogue of the quantum Hall effect. This effect was even more unexpected than the previous ones, is equally interesting, and can be performed on a table top. The groups are good candidates for a trip to Stockholm.*

Ref. 60
Page 68

What do we learn from these results? Systems in two dimensions have states which follow different rules than systems in three dimensions. The fractional charges in superconductors have no relation to quarks. Quarks, the constituents of protons and neutrons, have charges $e/3$ and $2e/3$. Might the quarks have something to do with superconductivity? At this point we need to stand the suspense, as no answer is possible; we come back to this issue in the last part of this adventure.

* This prediction from December 2008 became reality in December 2010.

Lasers and other spin-one vector boson launchers

Ref. 61

Photons are vector bosons; a lamp is thus a vector boson launcher. All lamps fall into one of three classes. Incandescent lamps use emission from a hot solid, gas discharge lamps use excitation of atoms, ions or molecules through collision, and solid state lamps generate (cold) light through recombination of charges in semiconductors.

TABLE 7 A selection of lamps and lasers

Lamp type, application	Wave-length	Bright-ness or power	Cost	Life-time
Incandescent lamps				
Tungsten wire light bulbs, halogen lamps, for illumination	300 to 800 nm	5 to 25 lm/W	0.1 cent/lm	700 h
Stars, for production of heavy elements	full spectrum			thou-sands of millions of years
Gas discharge lamps				
Oil lamps, candles, for illumination	white	up to 500 lm	1 cent/lm	5 h
Neon lamps, for advertising	red			up to 30 kh
Mercury lamps, for illumination	UV plus spectrum	45 to 110 lm/W	0.05 cent/lm	3000 to 24 000 h
Metal halogenide lamps (ScI_3 or 'xenon light', NaI, DyI_3, HoI_3, TmI_5) for car headlights and illumination	white	110 lm/W	1 cent/lm	up to 20 kh
Sodium low pressure lamps for street illumination	589 nm yellow	200 lm/W	0.2 cent/lm	up to 18 kh
Sodium high pressure lamps for street illumination	broad yellow	120 lm/W	0.2 cent/lm	up to 24 kh
Xenon arc lamps, for cinemas	white	30 to 150 lm/W, up to 15 kW		100 to 2500 h
Stars, for production of heavy elements	many lines			thou-sands of millions of years
Recombination lamps				
Foxfire in forests, e.g. due to *Armillaria mellea* and other bioluminescent fungi in rotting wood	green	just visible		years
Firefly, to attract mates	green-yellow			*c.* 10 h

TABLE 7 A selection of lamps and lasers (continued)

Lamp type, application	Wave-length	Bright-ness or power	Cost	Life-time
Large deep sea squid, *Taningia danae*, producing light flashes, to confuse prey	red			years
Deep-sea fish, such as *angler fish*, to attract prey or find mates	white			years
Deep-sea medusae, to produce attention so that predators of predators are attracted	blue and all other colours			years
Light emitting diodes, for measurement, illumination and communication	red, green, blue, UV	up to 100 lm/W	10 cent/lm	15k to 100 kh
Synchroton radiation sources				
Electron synchroton source	X-rays to radio waves	pulsed	many MEuro	
Possibly some stars	broad spectrum			thou-sands of years
Ideal white lamp or laser	visible	*c.* 300 lm/W	n.a.	n.a.
Ideal coloured lamp or laser	green	683 lm/W	n.a.	n.a.
Gas lasers				
He-Ne laser (obsolete), for school experiments	632.8 nm	550 lm/W	2000 cent/lm	300 h
Argon laser, for pumping and laser shows	several blue and green lines	up to 100 W	10 kEuro	
Krypton laser, for pumping and laser shows	several blue, green, red lines	50 W		
Xenon laser	many lines in the IR, visible and near UV	20 W		
Nitrogen (or 'air') laser, for pumping of other lasers, for hobby	337.1 nm	pulsed up to 1 MW	down to a few hundred Euro	limited by metal elec-trode lifetime
Water vapour laser, for research	many lines between 7 and 220 μm, often 118 μm	CW 0.5 W, pulsed much higher	a few kEuro	

TABLE 7 A selection of lamps and lasers (continued)

Lamp type, application	Wave-length	Bright-ness or power	Cost	Life-time
CO_2 laser, for cutting, welding, glass welding and surgery	10.6 µm	CW up to 100 kW, pulsed up to 10 TW		1500 h
Excimer laser, for lithography in silicon chip manufacturing, eye surgery, pumping, psoriasis treatment, laser deposition	193 nm (ArF), 248 nm (KrF), 308 nm (XeCl), 353 nm (XeF)	100 W		
Metal vapour lasers (Cu, Cd, Se, Ca, Ag, Au, Mn, Tl, In, Hg)				
Copper vapour laser, for pumping, photography, dermatology, laser cutting, hobby constructions and explorative research	248 nm, 511 nm and 578 nm	pulses up to 5 MW		
Cadmium vapour laser, for printing, typesetting and recognition of forged US dollar notes	325 nm and 442 nm	up to 200 mW	12 kEuro	10 kh
Gold vapour laser, for explorative research, dermatology	627 nm	pulses up to 1 MW	from a few hundred Euro upwards	
Chemical gas lasers				
HF, DF and oxygen-iodine laser, used as weapons, pumped by chemical reactions	1.3 to 4.2 µm	up to MW in CW mode	over 10 MEuro	un-known
Liquid Dye lasers				
Rhodamine, stilbene, coumarin etc. lasers, for spectroscopy and medical uses	tunable, range depends on dye in 300 to 1100 nm range	up to 10 W	10 kEuro	dye-dependent
Beer, vodka, whiskey and many other liquids work as laser material	IR, visible			
Solid state lasers				
Ruby laser (obsolete), for holography and tattoo removal	694 nm		1 kEuro	
Nd:YAG (neodym-YAG) laser, for material processing, surgery, pumping, rangefinding, velocimetry	1064 nm	CW 10 kW, pulsed 300 MW	50 to 500 kEuro	1000 h
Er:YAG laser, for dermatology	2940 nm			

TABLE 7 A selection of lamps and lasers (continued)

Lamp type, application	Wave-length	Bright-ness or power	Cost	Life-time
Ti:sapphire laser, for ultrashort pulses for spectroscopy, LIDAR, and research	650 to 1200 nm	CW 1 W, pulsed 300 TW	from 5 kEuro upwards	
Alexandrite laser, for laser machining, dermatology, LIDAR	700 to 840 nm			
Cr:LiSAF laser		pulsed 10 TW, down to 30 fs		
Cr:YAG laser	1.35 to 1.6 µm	pulsed, down to 100 fs		
Cr:Forsterite laser, optical tomography	1200 to 1300 nm	pulsed, below 100 fs		
Erbium doped glass laser, used in optical communications (undersea cables) and optical amplifiers	1.53 to 1.56 µm			years
Perovskite laser, such as $Co:KZnF_3$, for research	NIR tunable, 1650 to 2070 nm	100 mW	2 kEuro	
F-centre laser, for spectroscopy (NaCl:OH-, KI:Li, LiF)	tuning ranges between 1.2 and 6 µm	100 mW	20 kEuro	
Semiconductor lasers				
GaN laser diode, for optical recording	400 to 500 nm	10 mW	a few Euro	a few 100 h
AlGaAs laser diode, for optical recording, pointers, data communication, laser fences, bar code readers (normal or vertical cavity)	620 to 900 nm	up to 1 W	below 1 Euro to 100 Euro	*c.* 10 000 h
InGaAsP laser diode, for fiberoptic communication, laser pumping, material processing, medial uses (normal and vertical cavity or VCSEL)	1 to 2.5 µm	up to 100 W	below 1 Euro up to a few k Euro	up to 20 000 h
Lead salt (PbS/PbSe) laser diode, for spectroscopy and gas detection	3 to 25 µm	0.1 W	a few 100 Euro	
Quantum cascade laser, for research and spectroscopy	2.7 to 350 µm	CW achieved		
Hybrid silicon lasers, for research	IR	nW	0.1 MEuro	
Free electron lasers				

TABLE 7 A selection of lamps and lasers (continued)

Lamp type, application	Wave-length	Bright-ness or power	Cost	Life-time
Used for material science	5 nm to 1 mm	CW 20 kW, pulsed in GW range	10 MEuro	years
Nuclear-reaction pumped lasers				
Have uses only in science fiction and for getting money from gullible military				

Most solid state lamps are light emitting diodes. The large progress in brightness of light emitting diodes could lead to a drastic reduction in future energy consumption, if their cost is lowered sufficiently. Many engineers are working on this task. Since the cost is a good estimate for the energy needed for production, can you estimate which lamp is
Challenge 51 s the most friendly to the environment?

Nobody thought much about lamps, until Albert Einstein and a few other great physicists came along, such as Theodore Maiman and Hermann Haken. Many other researchers later received Nobel Prizes by building on their work. In 1916, Einstein showed that there are two types of sources of light – or of electromagnetic radiation in general – both of which actually 'create' light. He showed that every lamp whose brightness is turned up high enough will change 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'.) After a passionate worldwide research race, in 1960 Maiman was the first to build a laser emitting visible light. (So-called *masers* emitting microwaves were already known for several decades.) In summary, Einstein and the other physicists 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 strongly pumped LEDs or light emitting diodes), He–Ne lasers (strongly pumped neon lamps), liquid lasers (essentially strongly pumped fire flies) and ruby lasers (strongly 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. Most materials can be used to make lasers, including water, beer and vodka.

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, Ref. 62 only eleven photons on average were moving between the two mirrors. Quite a small lamp. Also lasers emitting light in two dimensions have been built. They produce a light plane instead of a light beam.

Lasers have endless applications. Lasers read out data from compact discs (CDs), are used in the production of silicon integrated circuits, and transport telephone signals; we Vol. I, page 312 already encountered lasers that work as loudspeakers. The biggest advances in recent years came from the applications of femtosecond laser pulses. Femtosecond pulses generate high-temperature plasmas in the materials they propagate, even in air. Such short pulses can be used to cut material without heating it, for example to cut bones in heart operations. Femtosecond lasers have been used to make high resolution hologram of human heads within a single flash. Recently such lasers have been used to guide lightning Ref. 63 along a predetermined path and seem promising candidates for laser ligtning rods. A curious application is to store information in fingernails (up to 5 Mbit for a few months) Ref. 64 using such lasers, in a way not unlike that used in recordable compact discs (CD-R).

Can two photons interfere?

Page 51 In 1930, Dirac made the famous statement already mentioned above:

> Each photon interferes only with itself. Interference between two different photons never occurs. Ref. 65

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 Ref. 66 the 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. 67 have shown light interference from the fifties onwards. In 1963, Magyar and Mandel 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

Figure yet to be included

FIGURE 45 An electron hologram

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.

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

Ref. 68 * In 2002, the first holograms have been produced that made use of neutron beams.

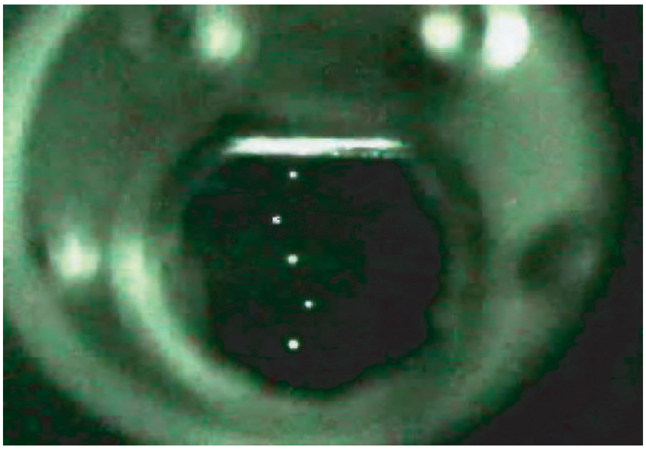


FIGURE 46 Imaging isolated electrons: single electrons surrounded by bubbles that explode in liquid helium under negative pressure produce white spots (mpg film © Humphrey Maris)

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.

Challenges, dreams and curiosities about quantum technology

Nowadays, we carry many electronic devices in our jacket or trousers. Almost all use batteries. In the future, there is a high chance that some of these devices will extract energy from the human body. There are several options. One can extract thermal energy with thermoelements, or one can extract vibrational energy with piezoelectric, electrostatic or electromagnetic transducers. The challenge is to make these elements small and cheap. It will be interesting to find out which technology will arrive to the market first.

* *

Ref. 69 In 2007, Humphrey Maris and his student Wei Guo performed an astonishing experiment: they filmed single electrons with a video camera. Well, the truth is a bit more complicated, but it is not a lie to summarize it in this way.

Maris is an expert on superfluid helium. For many years he knew that free electrons in superfluid helium repel helium atoms, and can move, surrounded by a small vacuum bubble, about 2 nm across, through the fluid. He also discovered that under negative pressure, these bubbles can grow and finally explode. When they explode, they are able to scatter light. With his student Wei Guo, he then injected electrons into superfluid helium through a tungsten needle under negative voltage, produced negative pressure by focussing waves from two piezoelectric transducers in the bulk of the helium, and shone light through the helium. When the pressure became negative enough they saw the explosions of the bubbles. Figure 46 shows the video. If the experiment is confirmed, it is one of the highlights of experimental physics in the last decade.

* *

Challenge 52 d Is it possible to make A4-size, thin and flexible colour displays for an affordable price?

* *

Challenge 53 r Will there ever be desktop laser engravers for 1000 euro?

* *

Challenge 54 r Will there ever be room-temperature superconductivity?

* *

Challenge 55 s Will there ever be teleportation of everyday objects?

* *

One process that quantum physics does not allow is telephathy. An unnamed space agency found this out during the Apollo 14 mission, when, during the flight to the moon, cosmonaut Edgar Mitchell tested telepathy as communication means. Unsurprisingly, he found that it was useless. (Ref. 70) It is unclear why the unnamed space agency spent so much money for a useless experiment – an experiment that could have been performed, at a cost of a phone call, also down here on earth.

* *

Challenge 56 d Will there ever be applied quantum cryptology?

* *

Will there ever be printable polymer electronic circuits, instead of lithographically patterned silicon electronics as is common now? (Challenge 57 d)

* *

Challenge 58 r Will there ever be radio-controlled flying toys in the size of insects?

Chapter 3

QUANTUM ELECTRODYNAMICS – THE ORIGIN OF VIRTUAL REALITY

The central concept that 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.

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. It might be that this effect was known before the nineteenth century, when many places still lacked harbours.*

Waves 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

$$E = mgh\,\alpha^2/2 \tag{10}$$

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 $h = 1.5\,\mathrm{m}$.

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

* Sipko Boersma published a paper in which he gave his reading of shipping manuals, advising captains to
Ref. 71 let the ships be pulled apart using a well-manned rowing boat. This reading has been put into question by
Ref. 72 subsequent research, however.

given, the radiated power W is

$$W = 2\pi \frac{E}{QT} . \quad (11)$$

We saw above that radiation force (radiation pressure times area) is W/c, where c is the wave propagation velocity. For water waves, we have the famous relation

$$c = \frac{gT}{2\pi} . \quad (12)$$

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

$$ma = m2\pi^2 \frac{h\alpha^2}{QT^2} . \quad (13)$$

Inserting typical values such as $Q = 2.5$, $T = 10$ s, $\alpha = 0.14$ rad and a ship mass of 700 tons, we get about 1.9 kN. Long swells thus make ships attract each other. The strength of the attraction is comparatively small and could 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 hide or overshadow this attraction.

Sound waves or noise in air show 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. 73

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. The light-emitting surface of a television tube is made by deposing small neutral, but conductive particles on glass. 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 metal plates (or metal mirrors) 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 neutral bodies attract each other in vacuum, with a force depending on their geometry. Ref. 74, Ref. 75

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 conducting bodies. Casimir understood that the space between two parallel conducting mirrors, due to the geometrical constraints, had different zero-point fluctuations than the free vacuum. Like in the case of two ships, the result would be the attraction of the two mirrors.

Casimir predicted that the attraction for two mirrors of mass m and surface A is given by

$$\frac{ma}{A} = \frac{\pi^3}{120} \frac{\hbar c}{d^4} . \quad (14)$$

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 confirmation was by Derjaguin, Abrikosova and Lifshitz in 1956; the second experimental confirmation was by Marcus Sparnaay, 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. (Note that at very small distances, the dependence is not $1/d^4$, but $1/d^3$.) In summary, uncharged bodies attract through electromagnetic field fluctuations.

Ref. 76 Ref. 77 Ref. 78 Ref. 79

In a cavity, spontaneous emission is suppressed, if it is smaller than the wavelength of the emitted light! This effect has also been observed. It 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 59 s

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. 80 Ref. 81

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?

Challenge 60 s

In 2006, the Casimir effect provided another surprise. The ship story just presented is beautiful, interesting and helps understanding the effect; but it seems that the story is based on a misunderstanding. Alas, the interpretation of the old naval text given by Sipko Boersma seems to be wishful thinking. There might be such an effect for ships, but it has never been observed nor put into writing by seamen, as Fabrizio Pinto has pointed out after carefully researching naval sources. As analogy, it remains valid.

Ref. 72

The Lamb shift

It is not frequent that a person receives the Nobel Prize in Physics for observing the colour of a lamp. But it happens. In 1947, Willis Lamb performed a careful measurement of the spectrum of hydrogen. He found the first effect due to virtual particles. More precisely, he found that the $2S_{1/2}$ energy level in atomic hydrogen lies slightly above the $2P_{1/2}$ level. This is in contrast to the calculation performed above, where the two levels are predicted to have the same energy. In reality, they have an energy difference of 1057.864 MHz, or 4.3 μeV. This discovery had important consequences for the description of quantum theory and yielded Lamb a share of the 1955 Nobel Prize in Physics. Why?

Page 159

The reason for the level difference is an unnoticed approximation performed in the simple relativistic calculation of the hydrogen levels. There are two equivalent ways to explain it. One is to say that the calculation neglects the coupling terms between the Dirac equation and the Maxwell equations. This explanation lead to the first calculations of the Lamb shift, around the year 1950. The other, equivalent explanation is to say that the calculation *neglects virtual particles*. In particular, the calculation neglects the virtual photons emitted and absorbed during the motion of the electron around the nucleus. This is the explanation in line with the modern vocabulary of quantum electrodynamics. QED is the perturbative approach to solve the coupled Dirac and Maxwell equations.

The QED Lagrangian and its symmetries

In simple terms, *quantum electrodynamics* is the description of electron motion. This implies that the description is fixed by the effects of mass and charge, and by the quantum of action. The QED Lagrangian density is given by:

$$\begin{aligned} \mathcal{L}_{\rm EW} = \ & \overline{\psi}(i\partial\!\!\!/ - m_k)\psi && \Big\} \ \text{the matter term} \\ & -\frac{1}{4\mu_0}F_{\mu\nu}F^{\mu\nu} && \Big\} \ \text{the electromagnetic field term} \\ & -q\overline{\psi}\gamma^{\mu}\psi A_{\mu} && \Big\} \ \text{the eelectromagnetic interaction term} \end{aligned} \qquad (15)$$

We know the matter term from the Dirac equation for free particles; it describes the kinetic energy of free electrons. We know the term of the electromagnetic field from the Maxwell's equations; it describes the kinetic energy of photons. The interaction term is the term that encodes the gauge symmetry of electromagnetism, also called 'minimal coupling'; it encodes the potential energy. In other words, the Lagrangian describes the motion of electrons and photons.

All experiments ever performed agree with the prediction by this Lagrangian. In other words, this Lagrangian is the final and correct description of the motion of electrons and photons. In particular, the Lagrangian describes the size, shape and colour of atoms, the size, shape and colour of molecules, as well as all interactions of molecules. In short, the Lagrangian describes all of materials science, all of chemistry, and all of biology. Exaggerating a bit, this is the Lagrangian that describes life. (In fact, the description of atomic Page 119 nuclei must be added; we will explore it below.)

All electromagnetic effects, including the growth of the coloured spots on butterfly wings, the functioning of the transistor or the cutting of paper with scissors, are completely described by the QED Lagrangian. Since the Lagrangian is part of the final description of motion, it is worth thinking about it in more detail.

Which requirements are necessary to deduce the QED Lagrangian? This issue has been explored in great detail. The answer is given by the following list:

— compliance with the observer-invariant quantum of action for the motion of electrons and photons,
— symmetry under the permutation group among many electrons, i.e., fermion behaviour of electrons,
— compliance with the invariance of the speed of light, i.e., symmetry under transformations of special relativity,

— symmetry under U(1) gauge transformations for the motion of photons and of charged electrons,
— symmetry under renormalization group,
— low-energy interaction strength described by the fine structure constant $\alpha \approx 1/137.036$.

The last two points require some comments.

As in all cases of motion, the action is the time-volume integral of the Lagrangian density. All fields, be they matter and radiation, move in such a way that this action remains minimal. In fact there are no known differences between the prediction of the least action principle based on the QED Lagrangian density and observations. Even though the Lagrangian density is known since 1926, it took another twenty years to learn how to calculate with it. Only in the years around 1947 it became clear, through the method of *renormalization*, that the Lagrangian density of QED is the *final* description of all motion of matter due to electromagnetic interaction in flat space-time. The details were shown independently by Julian Schwinger, Freeman Dyson, Richard Feynman and Tomonaga Shin-Itiro.*

The QED Lagrangian density contains the strength of the electromagnetic interaction in the form of the fine structure constant α. This number is part of the Lagrangian; no explanation for its value is given. The explanation was still unknown in the year 2000. Also the U(1) gauge group is specific to electromagnetism. All others requirements are valid for every type of interaction. Indeed, the search for the Lagrangians of the two nuclear interactions became really focused and finally successful only when the necessary Page 119 requirements were clearly spelled out, as we will discover in the chapters on the nucleus and on the interactions that describe its behaviour.

The Lagrangian density retains all symmetries that we know from classical physics. Vol. I, page 331 Motion is continuous, it conserves energy–momentum and angular momentum, it is relative, it is right–left symmetric, it is reversible, i.e., symmetric under change of velocity sign, and it is lazy, i.e., it minimizes action. In short, within the limits given by the quantum of action, also motion due to QED remains predictable.

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.

But virtual particles have also other surprising properties: virtual photons cannot be counted.

* Tomonaga Shin-Itiro (1906–1979) Japanese developer of quantum electrodynamics, winner of the 1965 Nobel Prize for physics together with Feynman and Schwinger. Later he became an important figure of science politics; together with his class mate from secondary school and fellow physics Nobel Prize winner, Yukawa Hidei, he was an example to many scientists in Japan.

** 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.

Vacuum energy: infinite or zero?

The strangest result of quantum field theory is the energy density of the vacuum. On one side, the vacuum has, to an excellent approximation, no mass and no energy content. The vacuum energy of vacuum is thus measured and expected to be *zero* (or at least extremely small).*

On the other side, the energy density of the zero-point fluctuations of the electromagnetic field is given by

$$\frac{E}{V} = \frac{4\pi h}{c^3} \int_0^\infty \nu^3 \mathrm{d}\nu \ . \tag{16}$$

The result of this integration is infinite. Quantum field theory thus predicts an *infinite* energy density of the vacuum.

Vol. VI, page 35
We can try to moderate the problem in the following way. As we will discover in the last part of our adventure, there are good arguments that a smallest measurable distance exists in nature; this smallest length appears when gravity is taken into account. The minimal distance is of the order of the *Planck length*

$$l_{\mathrm{Pl}} = \sqrt{\hbar G / c^3} \approx 1.6 \cdot 10^{-35}\,\mathrm{m} \ . \tag{17}$$

Vol. VI, page 35
A minimal distance leads to a maximum cut-off frequency. But even in this case the vacuum density that follows is still a huge number, and is much larger than observed by over 100 orders of magnitude. In other words, QED seems to predict an infinite, or, when gravity is taken into account, a huge vacuum energy. But measurements show a tiny vale. What exactly is wrong in this simple calculation? The answer cannot be given at this point; it will become clear in the last volume of our adventure.

Moving mirrors

Mirrors also work when they or the light source is in motion. In contrast, walls, i.e., sound mirrors, do not produce echoes for all sound sources or for all wall speeds. For example, experiments show that walls do not produce echoes if the wall or the sound source moves faster than sound. On the other hand, light mirrors always produce an image, whatever the involved speeds. These observations confirm that the speed of light is the same for all observers: it is *invariant*. (Can you detail the argument?) In contrast, the speed of sound in air depends on the observer; it is *not* invariant.

Challenge 62 s

Vol. II, page 21
Light mirrors also differ from tennis rackets. (Rackets are tennis ball mirrors, to continue the previous analogy.) We have seen that light 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 a *limit* speed. In short, the simple existence and observation 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 they undergo *accelerated* motion. This issue yields a

* In 1998, this side of the issue was confused even further. Astrophysical measurements, confirmed in the subsequent years, have found that the vacuum energy has a small, but non-zero value, of the order of $0.5\,\mathrm{nJ/m^3}$. The reason for this value is not yet understood, and is one of the open issues of modern physics.

surprising result.

In the 1970s, quite a number of researchers independently found that there is no vacuum for accelerated observers. This effect is called *Fulling–Davies–Unruh effect*. (The incorrect and rarely used term *dynamical Casimir effect* has been abandoned.) For an accelerated observer, the vacuum is full of heat radiation. We will discuss this below. This fact has an interesting consequence for accelerated mirrors: a mirror in accelerated motion reflects the heat radiation it encounters. In short, *an accelerated mirror emits light*! Unfortunately, the intensity of this so-called *Unruh radiation* is so weak that it has not Page 103 been measured directly, up to now. We will explore the issue in more detail below. (Can Challenge 63 s you explain why accelerated mirrors emit light, but not matter?)

PHOTONS HITTING PHOTONS

When virtual particles are taken into account, light beams can 'bang' onto each other. This result is in full contrast to classical electrodynamics. Indeed, QED shows that the appearance of virtual electron-positron pairs allow photons to hit each other. And such pairs are found in any light beam.

However, the cross-section for photons banging onto each other is small. In fact, the bang is extremely weak. When two light beams cross, most photons will pass undisturbed. The cross-section A is approximately

$$A \approx \frac{973}{10\,125\pi}\,\alpha^4 \left(\frac{\hbar}{m_e c}\right)^2 \left(\frac{\hbar\omega}{m_e c^2}\right)^6 \qquad (18)$$

for the everyday case that the energy $\hbar\omega$ of the photon is much smaller than the rest energy $m_e c^2$ of the electron. This low-energy value is about 18 orders of magnitude smaller than what was measurable in 1999; the 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 settings one observes not only interaction through virtual electron–antielectron pairs, but also through virtual muon–antimuon pairs, virtual quark–antiquark pairs, and much more.

Everybody who consumes science fiction knows that matter and antimatter annihilate and transform into pure light. More precisely, a matter particle and an antimatter particle annihilate into two or more photons. Interestingly, quantum theory predicts that the opposite process is also possible: photons hitting photons can produce matter! In 1997, Ref. 82 this prediction was also 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 acquired a large energy, when seen in the inertial frame of the experimenter. The green laser pulse, of 527 nm wavelength or 2.4 eV photon energy, had a peak power density of 10^{22} W/m^2, about the highest achievable so far. That is a photon density of 10^{34} /m^3 and an electric field of 10^{12} V/m, both of which were record values at the time. When this green laser pulse was reflected off a 46.6 GeV electron beam, the returning photons had an energy of 29.2 GeV Challenge 64 e and thus had become high-energy gamma rays. These gamma rays then collided with other, still incoming green photons and produced electron–positron pairs through the reaction

$$\gamma_{29.2} + n\,\gamma_{\text{green}} \rightarrow e^{+} + e^{-} \tag{19}$$

for which both final particles were detected by special apparatuses. The experiment thus showed that light can hit light in nature, and above all, that doing so can produce matter. This is the nearest we can get to the science fiction fantasy of light swords or of laser swords banging onto each other.

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, Ref. 83 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 Fulling–Davies–Unruh effect, also called the *dynamical Casimir effect*, as described above. The damping depends on the shape and thus also on the direction of the body's motion.

Vol. I, page 302 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 Ref. 84 of the vacuum should produce a mean square displacement

$$\langle d^2 \rangle = \hbar / m t \tag{20}$$

that increases linearly with time; however, the extremely small displacement produced in this way is out of experimental reach so far. But the result is not a surprise. Are you Challenge 65 ny able to give another, less complicated explanation for it?

Renormalization – why is an electron so light?

In classical physics, the field energy of a point-like charged particle, and hence its mass, Vol. III, page 167 was predicted to be infinite. QED effectively *smears out* the charge of the electron over its Compton wavelength; as a result, the field energy contributes only a small correction to Challenge 66 s its total mass. Can you confirm this?

However, in QED, many intermediate results in the perturbation expansion are divergent integrals, i.e., integrals with infinite value. The divergence is due to the assumption that infinitely small distances are possible in nature. The divergences thus are artefacts that can be eliminated; the elimination procedure is called renormalization.

Sometimes it is claimed that the infinities appearing in quantum electrodynamics in the intermediate steps of the calculation show that the theory is incomplete or wrong. However, this type of statement would imply that classical physics is also incomplete or wrong, on the ground that in the definition of the velocity v with space x and time t, namely

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} = \lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t} = \lim_{\Delta t \rightarrow 0} \Delta x \frac{1}{\Delta t} , \tag{21}$$

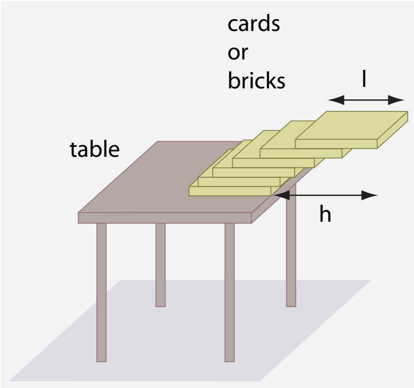


FIGURE 47 What is the maximum possible value of h/l?

one gets an infinity as intermediate step. Indeed, dt being vanishingly small, one could argue that one is dividing by zero. Both arguments show the difficulty to accept that the result of a limit process can be a finite quantity even if infinite quantities appear in the calculation. The parallel between electron mass and velocity is closer than it seems; both intermediate 'infinities' stem from the assumption that space-time is continuous, i.e., infinitely divisible. The infinities necessary in limit processes for the definition of differentiation, integration or for renormalization appear only when space-time is approximated, as physicists say, as a 'continuous' set, or as mathematicians say, as a 'complete' set.

On the other hand, the conviction that the appearance of an infinity might be a sign of incompleteness of a theory was an interesting development in physics. It shows how uncomfortable many physicists had become with the use of infinity in our description of nature. Notably, this was the case for Paul Dirac himself, who, after having laid in his youth the basis of quantum electrodynamics, has tried for the rest of his life to find a way, without success, to change the theory so that intermediate infinities are avoided. Ref. 85

Renormalization is a procedure that follows from the requirement that continuous space-time and gauge theories must work together. In particular, renormalization follows form the requirement that the particle concept is consistent, i.e., that perturbation expansions are possible. Intermediate infinities are not an issue. In a bizarre twist, a few decades after Dirac's death, his wish has been fulfilled after all, although in a different manner than he envisaged. The final part of this mountain ascent will show the way out of the issue. Vol. VI, page 33

Curiosities and fun 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.

* *

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 67 s

* *

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 polarization of the vacuum, also plays a role in much larger systems, such as charged black holes, as we will see shortly.

Ref. 86 Challenge 68 s Page 109

* *

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. 87

* *

Is there a *critical magnetic field* in nature, like there is a critical electric field, limited by spontaneous pair production?

Challenge 69 ny

* *

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.

In radioactive decay, the slowness record is held by ^{209}Bi, with over 10^{19} years of half-life.

Ref. 88 Page 261

* *

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 70 s

* *

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

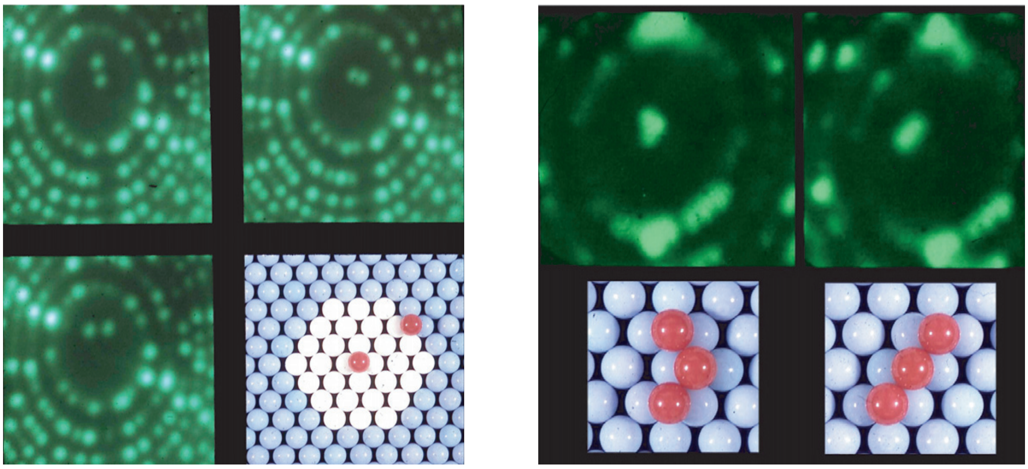


FIGURE 48 On tungsten tips, rhenium atoms, visible at the centre of the images, do not form dimers (left) but do form trimers (right) (© Hans-Werner Fink, APS, AIP from Ref. 91)

Challenge 71 s regular dense packing 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. (Ref. 89)

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 many similar results 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. 90)

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 – two atoms moving together – 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. 91)

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

FIGURE 49 Some snow flakes (© Furukawa Yoshinori)

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 72 s

Ref. 92

However, not always the slowest growing planes win out. Figure 49 shows some well-known exceptions. Explaining such shapes is possible today, and Furukawa Yoshinori is one of the experts in the field, heading a dedicated research team. Indeed, 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 non-linear processes. The details are still a topic of research.

Ref. 93

Vol. I, page 318

* *

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. 94

* *

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. 95

* *

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 73 s

Ref. 96

* *

In the past, the description of motion with formulae was taken rather seriously. Before computers appeared, only those examples of motion were studied that could be described with simple formulae. But this narrow-minded approach turns out to be too restrictive. Indeed, mathematicians showed 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 formulae, but on the description by clear equations based on space and time.

* *

Challenge 74 d The Casimir effect, as well as other experiments, imply that there is a specific and *finite* energy density that can be ascribed to the vacuum. Does this mean that we can apply the Banach–Tarski effect to pieces of vacuum?

* *

Challenge 75 s Can you explain why mud is not clear?

* *

Challenge 76 s 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 77 s Can the universe ever have been smaller than its own Compton wavelength?

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.

Challenge 78 s *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? 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 79 s

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 80 s

Quantum theory offers a wealth of methods. Of course, quantum mechanics shows that we actually are always moving, since the indeterminacy 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 81 s

Materials 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 82 s

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 83 s

Assuming that you read already the section following the present one, on the effects of semiclassical *quantum gravity*, here is an additional puzzle: is it possible to move by accelerating a pocket mirror, using the emitted Unruh radiation? Can you find at least two other methods to move yourself using quantum gravity effects? Can you find one from string theory?
Challenge 84 s

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.

A summary of quantum electrodynamics

The shortest possible summary of quantum electrodynamics is the following:

▷ *Everyday matter is made of charged particles that interact through photon exchange in the way described by Figure 50.*

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 Feynman diagram of Figure 50. It contains the Lagrangian and the Hamiltonian of QED, as well as a method to calculate results for actual systems. As QED is a perturbative theory, the diagram directly describes the first order effects; its composite diagrams 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 50.

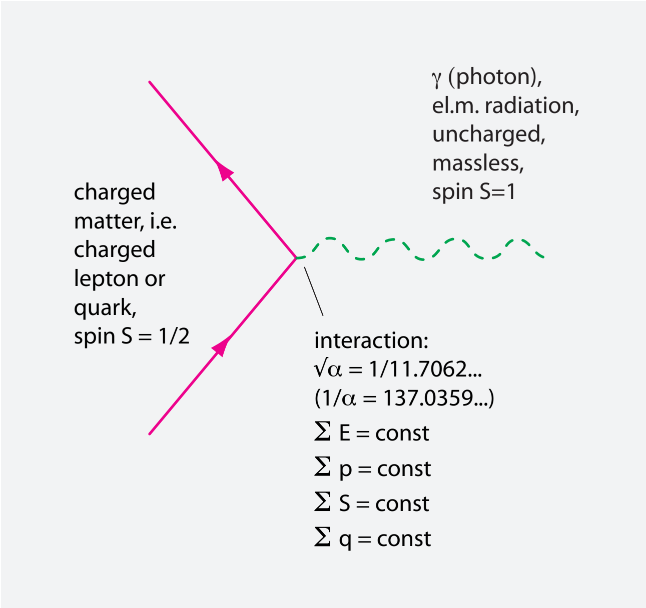


FIGURE 50 QED as perturbation theory in space-time

Matter is divisible because the interactions are of finite strength, matter is also divisible 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

$$d \approx \lambda_C = \frac{h}{m\,c} \,. \tag{22}$$

In situations where the dimensions are of the order of the de Broglie wavelength, or equivalently, where the action is of the order of the Planck value, simple quantum mechanics is sufficient:

$$d \approx \lambda_{dB} = \frac{h}{m\,v} \,. \tag{23}$$

For even 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 processes. 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 aurorae 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. 98
Ref. 99

Materials 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 twenty-first 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 explosion.

Ref. 100

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. 101

In the theory of quantum electrodynamics, discoveries are expected by all those who study it in sufficient detail. For example, Dirk Kreimer has discovered that higher order interaction diagrams built using the fundamental diagram of Figure 50 contain relations to the 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 non-perturbative 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.

Ref. 102

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.

* On the other hand, outside QED, there is beautiful work going on how humans move their limbs; it seems that any general human movement is constructed in the brain by combining a small set of fundamental movements.

Ref. 97

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.

Chapter 4

QUANTUM MECHANICS WITH GRAVITATION – A FIRST APPROACH

Gravitation is a weak effect. Every seaman knows that storms, not gravity, cause the worst accidents. Despite its weakness, the inclusion of gravity into quantum theory raises a number of issues. We must solve them in order to complete our ascent.

First of all, does gravity act on quantum systems? Yes, it does. In the chapter on general relativity we already mentioned that light frequency changes with height. Thus gravity has a simple and measurable effect on photons. But gravity also acts on all other quantum systems, such as atoms.

Falling atoms

In 2004 it became possible to repeat Galileo's leaning tower experiment with single atoms instead of steel balls. The result is as expected: single atoms do fall like stones. In particular, atoms of different mass fall with the same acceleration, within the experimental precision of one part in 6 million.

Ref. 103

This result is not surprising, as all falling everyday objects are made of atoms. But what is the effect of gravity on wave functions?

Playing table tennis with neutrons

The gravitational potential also has directly measurable effects on quantum particles. Classically, a table tennis ball follows a parabolic path when bouncing over a table tennis table, when friction is neglected. How does a quantum particle behave in the same setting? The experiment was first performed in 2002, with neutrons.

Ref. 104

In the gravitational field, a bouncing quantum particle is still described by a wave function. In contrast to the classical case, the possible energy values are discrete. Indeed, the quantization of the action implies that for a bounce of energy E_n and duration t_n,

Challenge 85 e

$$n\hbar \sim E_n\, t_n \sim \frac{E_n^{3/2}}{g m^{1/2}} . \qquad (24)$$

In other words, only discrete bounce heights are possible in the quantum case. This leads to a probability density that changes with height in discrete steps, as shown in Figure 51.

The best experimental procedure is to produce an intense beam of neutral particles, because neutral particles are not affected by stray electromagnetic fields. Neutrons are

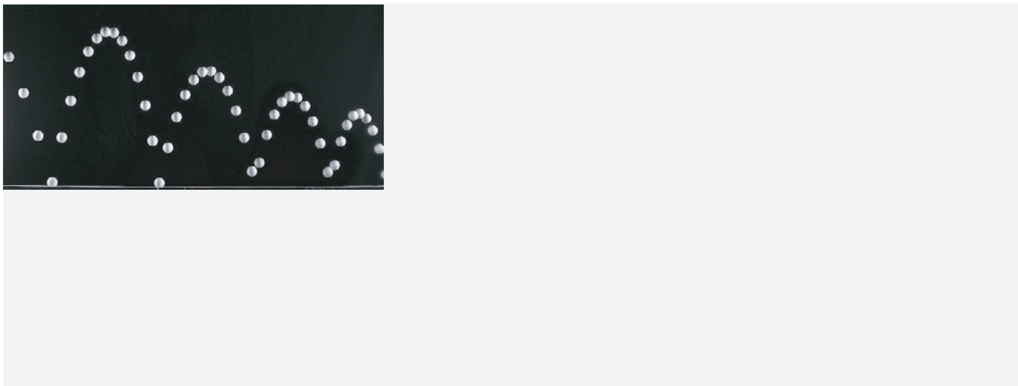


FIGURE 51 Table tennis and neutrons

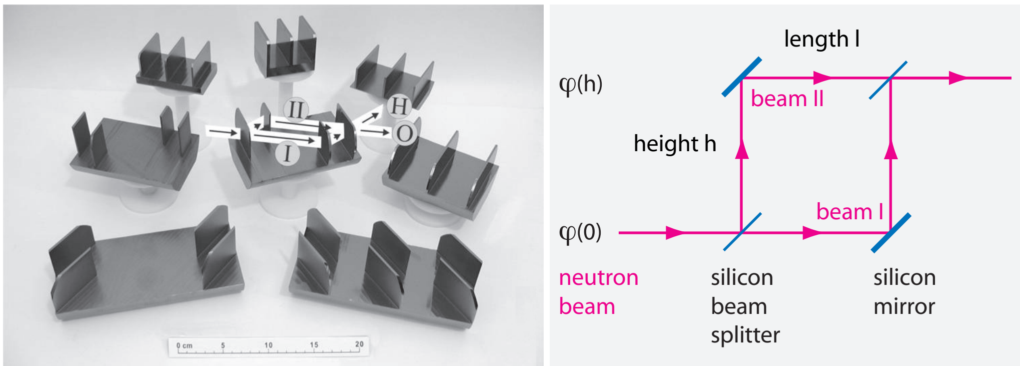


FIGURE 52 The weakness of gravitation. A neutron interferometer made of a silicon single crystal (with the two neutron beams I and II) can be used to detect the effects of gravitation on the phase of wave functions (photo © Helmut Rauch and Erwin Seidl).

ideal, as they are produced in large quantities by nuclear reactors. Using a few clever tricks, the experimenters managed to slow down neutrons to the incredibly small value of 8 m/s, comparable to a table tennis ball. They then directed the neutrons onto a neutron mirror, the analogue of the table tennis table, and observed the neutrons jumping back up.

Why is the experiment so difficult? The lowest energy levels for neutrons due to gravity are $2.3 \cdot 10^{-31}$ J, or 1.4 peV, followed by 2.5 peV,3.3 peV, 4.1 peV, and so forth. To get an impression of the smallness of these values, 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. Nevertheless, the team managed to measure the first few discrete energy levels. The results confirmed the prediction of the Schrödinger equation to the achievable measurement precision. In short, gravity influences wave functions, and does so as expected.

The gravitational phase of wave functions

Not only does gravity change the shape of wave functions; it also changes their *phase*. Can you imagine why? The prediction was first confirmed in 1975, using a device invented

Challenge 86 s

Ref. 105 by Helmut Rauch and his team. Rauch had developed neutron interferometers based on single silicon crystals, shown in Figure 52, in which a neutron beam from a nuclear reactor is split into two beams and the two beams are then recombined and brought to interference.

By rotating the interferometer mainly around the horizontal axis, Samuel Werner and his group let the two neutron beams interfere after having climbed a small height h at Ref. 106 two different locations. The experiment is shown schematically on the right of Figure 52. The neutron beam is split; the two beams are deflected upwards, one directly, one a few centimetres further on, and then recombined.

For such a experiment in gravity, quantum theory predicts a phase difference $\Delta\varphi$ between the two beams given by

$$\Delta\varphi = \frac{mghl}{\hbar v} , \tag{25}$$

where l is the horizontal distance between the two climbs and v and m are the speed and mass of the neutrons. This and other beautifully simple experiments have confirmed the prediction by quantum theory within experimental errors.

Ref. 107 In the 1990s, similar experiments have even been performed with complete atoms. These atom interferometers are so sensitive that local gravity g can be measured with a precision of more than eight significant digits.

In short, neutrons, atoms and photons show no surprises in gravitational fields. Gravity can be included into all quantum systems of everyday life. By including gravity, the Dirac equation can thus be used, for example, to describe the growth and the processes inside trees.

The gravitational Bohr atom

Can gravity lead to *bound* quantum systems? A short calculation shows that an electron circling a proton due to gravity alone, without electrostatic attraction, would do so at a Challenge 87 ny gravitational Bohr radius of

$$r_{\text{gr.B.}} = \frac{\hbar^2}{G\, m_e^2\, m_p} = 1.1 \cdot 10^{29}\ \text{m} \tag{26}$$

which is about a thousand times the distance to the cosmic horizon. A gravitational Bohr atom would be larger than the universe. This enormous size is the reason that in a normal hydrogen atom there is not a *single way* to measure gravitational effects between its Challenge 88 ny components. (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.

We might conclude from all this that gravity does not require a quantum description. Indeed, we stumbled onto quantum effects because classical electrodynamics implies, in stark contrast with reality, that atoms decay in about 0.1 ns. Classically, an orbiting electron would emit radiation until it falls into the nucleus. Quantum theory is thus necessary to explain the existence of matter.

Challenge 89 ny

When the same stability calculation is performed for the emission of gravitational radiation by orbiting electrons, one finds a decay time of around 10^{37} s. (True?) This extremely large value, trillions of times longer than the age of the universe, is a result of the low emission of gravitational radiation by rotating masses. Therefore, the existence of normal atoms does not require a quantum theory of gravity.

Page 156

Indeed, quantum gravity is unnecessary in every single domain of everyday life. However, we will see now that quantum gravity is necessary in domains which are more remote, but also more fascinating.

Gravitation and 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 'entropy'. We have seen that the Boltzmann constant k is the smallest entropy found in nature.

Vol. I, page 229

What is the influence of gravitation on entropy, and thermodynamics in general? For a long time, nobody was interested in this question. In parallel, for many decades nobody asked whether there also exists a theoretical *maximum* for entropy. The situations changed dramatically in 1973, when Jacob Bekenstein discovered that the two issues are related.

Ref. 108

Bekenstein was investigating the consequences gravity has for quantum physics. He found that the entropy S of an object of energy E and size L is bound by

$$S \leqslant EL\frac{k\pi}{\hbar c} \tag{27}$$

for all physical systems, where k is the Boltzmann constant. In particular, he deduced that (nonrotating) black holes saturate the bound, with an entropy given by

Challenge 90 s

$$S = A\,\frac{kc^3}{4G\hbar} = M^2\,\frac{4\pi kG}{\hbar c} \tag{28}$$

where A is now the area of the *horizon* of the black hole. It is 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.***

---

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

** The precise discussion that black holes are the most disordered systems in nature is quite subtle. The issue is summarized by Bousso. Bousso claims that the area appearing in the maximum entropy formula cannot be taken naively as the area at a given time, and gives four arguments why this should be not allowed. However, all four arguments are wrong in some way, in particular because they assume that lengths smaller than the Planck length or larger than the universe's size can be measured. Ironically, he brushes aside some

Ref. 109

As an interesting note, the maximum entropy also implies an upper memory limit for memory chips. Can you find out how? Challenge 91 s

Black hole entropy is somewhat mysterious. What are the different microstates leading to this macroscopic entropy? It took many years to convince physicists that the microstates are due to the various possible states of the black hole horizon itself, and that they are due to the diffeomorphism invariance at this boundary. Ref. 110 As Gerard 't Hooft explains, the entropy expression implies that the number of degrees of freedom of a black hole is about (but not exactly) one per Planck area of the horizon. Challenge 92 s

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 short time. All these results about black holes were waiting to be discovered since the 1930s; incredibly, nobody had thought about them for the subsequent 40 years.

Measuring acceleration with a thermometer: Fulling–Davies–Unruh radiation

Independently, Stephen Fulling in 1973, Paul Davies in 1975 and William Unruh in 1976 made the same theoretical discovery while studying quantum theory: Ref. 111 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. The appearance of radiation for an accelerated observer in vacuum is called the *Fulling–Davies–Unruh effect*.

The radiation has a spectrum corresponding to the temperature

$$T = a \frac{\hbar}{2\pi k c} , \qquad (29)$$

where a is the value of the acceleration. The result means that there is no vacuum on Earth, because any observer on its surface can maintain that he is accelerated with $9.8\,\mathrm{m/s^2}$, thus leading to $T = 40\,\mathrm{zK}$! 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 in this way. But if it will, it will be a beautiful experimental result. Ref. 112

When this effect was predicted, people explored all possible aspects of the argument. For example, also an observer in rotational motion detects radiation following expression (29). But that was not all. It was found that the simple 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 emitted and 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 accelerations. This result will play an important role in the search for unification, as we will discover later on. Vol. VI, page 58

of the arguments himself later in the paper, and then deduces an improved formula, which is exactly the same as the one he criticizes first, just with a different interpretation of the area A. Later in his career, Bousso revised his conclusions; he now supports the maximum entropy bound. In short, the expression of black hole entropy is indeed the maximum entropy for a physical system with surface A.

Surprisingly, at the end of the twentieth century it became clear that the Fulling–Davies–Unruh effect had already been observed *before* it was predicted! The Fulling–Davies–Unruh effect turned out be equivalent to a well-established observation: the so-called *Sokolov–Ternov effect*. The Russian physicist Igor Ternov, together with Arsenji Sokolov, had used the *Dirac equation* to predicted in 1963 that electrons in circular accelerators and in storage rings that circulate at high energy would automatically polarize. The prediction was first confirmed by experiments at the Russian Budker Institute of Nuclean Physics in 1971, and then confirmed by experiments in Orsay, in Stanford and in Hamburg. Nowadays, the effect is used in many accelerator experiments. In the 1980s, Bell and Leinaas realized that the Sokolov–Ternov effect is the *same* effect as the Fulling–Davies–Unruh effect, but seen from a different reference frame! The equivalence is somewhat surprising, but is now well-established. In charges moving in a storage ring, the emitted radiation is not thermal, so that the analogy is not obvious or simple. But the effect that polarizes the beam – namely the difference in photon emission for spins that are parallel and antiparallel to the magnetic field – is the same as the Fulling–Davies–Unruh effect. We thus have again a case of a theoretical discovery that was made much later than necessary.

Ref. 113

Ref. 113

Black holes aren't black

In 1973 and 1974, Jacob Bekenstein, and independently, 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. They 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 virtual antiparticle is captured by the black hole. The virtual antiparticle is thus of negative energy, and reduces the mass of the black hole. The mechanism applies both to fermions and bosons. From far away this effect looks like the emission of a particle. A detailed investigation showed that the effect is most pronounced for photon emission. In short, Bekenstein and Hawking showed that *black holes radiate as black bodies*.

Black hole radiation confirms both the result on black hole entropy by Bekenstein and the effect for observers accelerated in vacuum found by Fulling, Davies and Unruh. When all this became clear, a beautiful Gedanken experiment was published by William Unruh and Robert Wald, showing that the whole result could have been deduced already 50 years earlier!

Ref. 114

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 53. Imagine a mirror box full of heat radiation, thus full of 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 con-

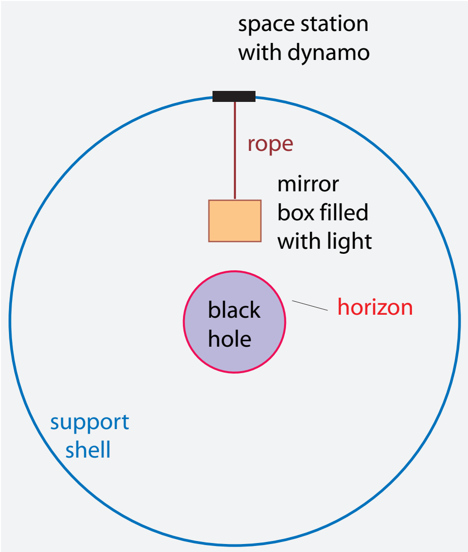


FIGURE 53 A *Gedanken experiment* allowing to deduce the existence of black hole radiation

verted 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, William Unruh and Robert Wald showed that black hole radiation is precisely the forgotten effect that puts everything right. Because of black hole radiation, the box feels buoyancy, so that it cannot be lowered down to the horizon completely. The box floats somewhat above the horizon, 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 when the box is opened. 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.

Well, the second principle of thermodynamics 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. At the horizon, the acceleration due to gravity is $g_{surf} = c^4/4GM$. The energy E gained by lowering the box is

$$E = mc^2 - mg_{surf}\frac{d}{2} = mc^2\left(1 - \frac{dc^2}{8GM}\right) \quad (30)$$

The efficiency of the process is $\eta = E/mc^2$. To be consistent with the second law of thermodynamics, this efficiency must obey

$$\eta = \frac{E}{mc^2} = 1 - \frac{T_{BH}}{T} , \tag{31}$$

where T is the temperature of the radiation inside the box. We thus find a black hole temperature T_{BH} that is determined by the hovering distance d. The hovering distance 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 $d \approx \hbar c/kT$. A precise calculation introduces a factor π, giving the result

$$T_{BH} = \frac{\hbar c^3}{8\pi kGM} = \frac{\hbar c}{4\pi k}\frac{1}{R} = \frac{\hbar}{2\pi kc}\, g_{surf} \quad \text{with} \quad g_{surf} = \frac{c^4}{4GM} \tag{32}$$

where R and M are the radius and the mass of the black hole. The quantity T_{BH} is either called the *black-hole temperature* or the *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 for the weakness of black hole radiation is 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 93 ny

Ref. 115

Black hole radiation is thus so weak that we must speak of an academic effect! It leads to a luminosity that increases with decreasing mass or size as

Challenge 94 ny

$$L \sim \frac{1}{M^2} \sim \frac{1}{R^2} \quad \text{or} \quad L = nA\sigma T^4 = n\frac{c^6\hbar}{G^2M^2}\frac{\pi^2}{15\cdot 2^7} \tag{33}$$

Page 164

where σ is the Stefan–Boltzmann or *black body radiation constant*, n is the number of particle degrees of freedom that can be radiated; as long as only photons are radiated – the only case of practical importance – we have $n = 2$.

Ref. 116

Black holes thus shine, and the more the smaller they are. This is a genuine *quantum* effect, since classically, black holes, as the name says, cannot emit any light. Even though the effect is academically weak, it will be of importance later on. In actual systems, many other effects around black holes increase the luminosity far above the Bekenstein–Hawking value; indeed, black holes are usually brighter than normal stars, due to the radiation emitted by the matter falling into them. But that is another story. Here we are only treating isolated black holes, surrounded only by pure vacuum.

TABLE 8 The principles of thermodynamics and those of horizon mechanics

Principle	Thermodynamics	Horizons
Zeroth principle	the temperature T is constant in a body at equilibrium	the surface gravity a is constant on the horizon
First principle	energy is conserved: $dE = TdS - pdV + \mu dN$	energy is conserved: $d(mc^2) = \frac{ac^2}{8\pi G} dA + \Omega dJ + \Phi dq$
Second principle	entropy never decreases: $dS \geqslant 0$	surface area never decreases: $dA \geqslant 0$
Third principle	$T = 0$ cannot be achieved	$a = 0$ cannot be achieved

The lifetime of black holes

Challenge 95 ny

Due to the emitted radiation, black holes gradually lose mass. Therefore their theoretical lifetime is *finite*. A calculation shows that the lifetime is given by

$$t = M^3 \frac{20\,480\,\pi\,G^2}{\hbar c^4} \approx M^3\ 3.4 \cdot 10^{-16}\ \mathrm{s/kg^3} \qquad (34)$$

as function of their initial mass M. For example, a black hole with mass of 1 g 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. Again, these numbers are purely academic. The important point is that *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. Bekenstein–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 William 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. 117

Ref. 118

Black holes are all over the place

Around the year 2000, astronomers amassed a large body of evidence that showed something surprising. The seems to be a supermassive black hole at the centre of almost all galaxies. The most famous of all is of course the black hole at the centre of our own galaxy. Quasars, active galactic nuclei and gamma ray bursters also seem to be due to supermassive black holes at the centre of galaxies. The masses of these black holes are typically a million solar masses.

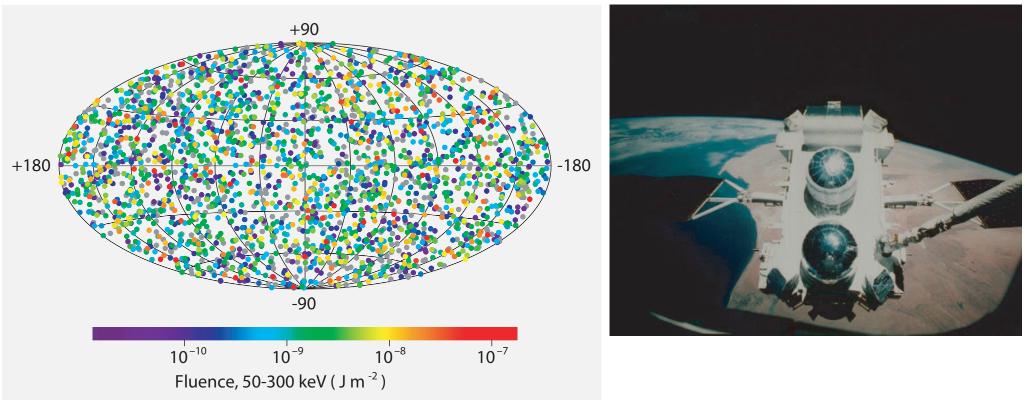


FIGURE 54 The location and energy of the 2704 gamma ray bursts observed in the sky between 1991 and 2000 by the BATSE experiment on board of the Compton Gamma Ray Observatory, a large satellite deployed by the space shuttle after over 20 years of planning and construction; the Milky Way is located around the horizontal line running from +180 to −180. (NASA)

Astronomers also think that many other, smaller astrophysical objects contain black holes: ultraluminous X-ray sources and x-ray binary stars are candidates for black holes of intermediate mass.

Finally, one candidate explanation for dark matter on the outskirts of galaxies is the possibility of small black holes.

In short, black holes seem to be quite common across the universe. Whenever astronomers observe a new class of objects, two questions arise directly: how do the objects form? And how do they disappear?

We have seen that quantum mechanics puts an upper limit to the life time of a black hole. The upper limit is academic, but that is not important. The main point is that it exists. Indeed, astronomers think that most black holes die in other ways, and much before the Bekenstein–Hawking limit, for example through mergers. All this is still a topic of research. The detectors of gravitational waves might clarify these processes in Vol. II, page 151 the future.

How are black holes born? It turns out that the birth of black holes can actually be observed.

Gamma ray bursts

Nuclear explosions produce flashes of gamma rays. In the 1960, several countries thought that detecting gamma ray flashes, or better, their absence, using satellites, would be the best way to ensure that nobody was detonating nuclear bombs above ground. But when the military sent such satellites into the sky, they found something surprising. They found about two gamma flashes *every* day. For fear of being laughed at, the military kept this result secret for many years.

It took the military six years to understand what an astronomer could have told them in five minutes: the flashes, today called *gamma ray bursts*, were coming from outer Ref. 120 space. Finally, the results were published; this is probably the only discovery about nature that was made by the military. Another satellite, this time built by normal scientists, the Compton Gamma Ray Observatory, confirmed that the bursts were *extragalactic* in origin, as shown by the map of Figure 54.

Measurements of gamma ray burst measurements are done by satellites because most gamma rays do not penetrate the atmosphere. 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 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.* Ref. 121

All this data together showed that gamma ray bursts have durations between milliseconds and about an hour. Gamma ray bursts seem to fall into (at least) two classes: the *short bursts*, usually below 3 s in duration and emitted from closer sources, and the *long bursts*, emitted from distant galaxies, typically with a duration of 30 s and more, and with a softer energy spectrum. The long bursts produce luminosities estimated to be up to 10^{45} W. This is about one hundredth of the brightness all stars of the 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. Ref. 123 Challenge 96 s

In fact, the measured luminosity of long bursts is near the theoretical maximum luminosity a body can have. This limit is given by Vol. II, page 96

$$L < L_{\text{Pl}} = \frac{c^5}{4G} = 0.9 \cdot 10^{52}\ \text{W}\ , \qquad (35)$$

Challenge 97 e as you might want to check yourself. In short, the sources of gamma ray bursts are the biggest bombs found in the universe, explosions of almost unimaginable proportions. Recent research seems to suggest that long gamma ray bursts are not isotropic, but that they are *beamed*, so that the huge luminosity values just mentioned might need to be divided by a factor of 1000.

However, the mechanism that leads to the emission of gamma rays is still unclear. It is often speculated that short bursts are due to merging neutron stars, whereas long bursts are emitted when a black hole is formed in a supernova or hypernova explosion. In this case, long gamma ray bursts would be 'primal screams' of black holes *in formation*. However, a competing explanation states that long gamma ray bursts are due to the *death* of black holes. Ref. 121

Indeed, already 1975, a powerful radiation emission mechanism was predicted for dying *charged* black holes by Damour and Ruffini. Charged black holes have a much shorter lifetime than neutral black holes, 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 Ref. 119

* For more about this fascinating topic, see the www.aip.de/~jcg/grb.html website by Jochen Greiner.

of seconds. That is quite shorter than the 10^{68} years predicted by Bekenstein–Hawking radiation! This process thus produces an extremely intense gamma ray burst.

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 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. 122

Understanding long gamma ray bursts is one of the most fascinating open questions in astrophysics. The relation to black holes is generally accepted. Many processes leading to emission of radiation from black holes are possible. Examples are matter falling into the black hole and heating up, or matter being ejected from rotating black holes through the Penrose process, or charged particles falling into a black hole. These mechanisms are known; they 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. But even 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.

Page 243

Material properties of black holes

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 $4\pi R^3/3$, where R is their radius. This density is then given by

$$\rho = \frac{1}{M^2} \frac{3c^6}{32\pi G^3} \qquad (36)$$

and can be quite low for large black holes. For the largest black holes known, with 1000 million solar masses or more, the density is of the order of the density of air. Nevertheless, even in this case, the density is the highest possible in nature for that mass.

Challenge 98 e

By the way, the gravitational acceleration at the horizon is still appreciable, as it is given by

$$g_{\text{surf}} = \frac{1}{M} \frac{c^4}{4G} = \frac{c^2}{2R} \qquad (37)$$

which is still 15 km/s^2 for an air density black hole.

Challenge 99 ny

Obviously, the black hole temperature is related to the entropy S by its usual definition

$$\frac{1}{T} = \left.\frac{\partial S}{\partial E}\right|_{\rho} = \left.\frac{\partial S}{\partial (Mc^2)}\right|_{\rho} \tag{38}$$

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 100 ny

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 101 ny Challenge 102 s Ref. 124

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. 125

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 electrical 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, we can study and describe macroscopic black holes like any other macroscopic material body. The topic is not closed. Ref. 116 Ref. 126

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 (34) 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), or 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 103 e 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.

Ref. 127 A clear exposition of the topic was given by Don Page. He calculated what information would be measured in the radiation if 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 Challenge 104 ny

$$S_m = \frac{1-m}{2n} + \sum_{k=n+1}^{mn} \frac{1}{k} \qquad (39)$$

which is approximately given by

$$S_m = \ln m - \frac{m}{2n} \quad \text{for} \quad m \gg 1 \, . \qquad (40)$$

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 (40) 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 (40) 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 one half of a bit of the radiated information. It is thus necessary to measure almost the *complete* radiation to obtain a sizeable chunk of the radiated information. In other words, it is extremely hard to determine the information contained in black hole radiation. In summary, at any given instant, the amount of information radiated by a black hole is negligible when compared with the total black hole radiation; it is practically impossible to obtain valuable information through measurements or even through calculations that use usual approximations.

Ref. 128

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 we explore a few additional ones.

Page 248

* *

Apart from the questions of entropy, we can look for methods to cheat around conservation of energy, angular momentum, or charge. Every Gedanken experiment 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 105 ny

* *

A black hole transforms matter into antimatter with a certain efficiency. Indeed, a black hole formed by collapsing matter also radiated antimatter. Thus one might look for departures from particle number conservation. Are you able to find an example?

Challenge 106 ny

* *

Black holes deflect light. Is the effect polarization dependent? Gravity itself makes no difference of polarization; however, if virtual particle effects of QED are included, the story might change. First calculations seem to show that such an effect exists, so that gravitation might produce rainbows. Stay tuned.

Ref. 129

* *

If lightweight boxes made of mirrors can float in radiation, one gets a strange consequence: such a box might self-accelerate in free space. In a sense, an accelerated box could float on the Fulling–Davies–Unruh radiation it creates by its own acceleration.

Are you able to show the following: one reason why this is impossible is a small but significant difference between gravity and acceleration, namely the absence of tidal effects. (Other reasons, such as the lack of perfect mirrors, also make the effect impossible.)

Challenge 107 ny

* *

In 2003, Michael Kuchiev has made the spectacular prediction that matter and radiation with a wavelength *larger* than the diameter of a black hole is partly reflected when it hits a black hole. The longer the wavelength, the more efficient the reflection would be. For stellar or even larger black holes, he predicts that only photons or gravitons are reflected. Black holes would thus not be complete trash cans. Is the effect real? The discussion is still ongoing.

Ref. 130

Curiosities about quantum theory and gravity

Due to the influence of gravity on phases of wave functions, 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 calculated results do not reproduce experiments.

Ref. 131

* *

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 R_0, and the number of baryons N_b:

Ref. 132

$$(N_b)^3 \approx \left(\frac{R_0}{l_{Pl}}\right)^4 = \left(\frac{t_0}{t_{Pl}}\right)^4 \approx 10^{244} \qquad (41)$$

in which $N_b = 10^{81}$ and $t_0 = 1.2 \cdot 10^{10}$ a were used. There is no known reason why the number of baryons and the horizon size R_0 should be related in this way. This coincidence is equivalent to the one originally stated by Dirac,* namely

$$m_p^3 \approx \frac{\hbar^2}{Gct_0} \, . \qquad (43)$$

* The equivalence can be deduced using $Gn_b m_p = 1/t_0^2$, which, as Weinberg explains, is required by several cosmological models. Indeed, this can be rewritten simply as

Ref. 133
Page 96

$$m_0^2/R_0^2 \approx m_{Pl}^2/R_{Pl}^2 = c^4/G^2 \, . \qquad (42)$$

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 last part of the mountain ascent.

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 (41) or (43) express some long-sought relation between local and global topological properties 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, and that they should not be taken too seriously. Ref. 134

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 bring to our mind the similarity between the electromagnetic field and the gravitational 'field,' our next step should be to find the quantum description of the gravitational field. However, despite attempts by many brilliant minds for almost a century, this approach was not successful. Indeed, modern approaches take another direction, as will be explained later on. Let us see what was achieved and why the results are not sufficient.

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 quanta of the gravitational field have *vanishing* mass and move at light speed. The independence of gravity from electromagnetic effects implies a *vanishing* electric charge.

We observe that gravity is always attractive and never repulsive. This 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', spin $S = 2$ 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. 135

The coupling strength of gravity, corresponding to the fine structure constant of electromagnetism, is given either by

$$\alpha_{G1} = \frac{G}{\hbar c} = 2.2 \cdot 10^{-15}\,\mathrm{kg}^{-2} \quad \text{or by} \quad \alpha_{G2} = \frac{Gmm}{\hbar c} = \left(\frac{m}{m_{Pl}}\right)^2 = \left(\frac{E}{E_{Pl}}\right)^2 \qquad (44)$$

However, the first expression is not a pure number; the second expression is, but depends on the mass we insert. 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 $\alpha_{em} = 1/137.04$.

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

Challenge 108 s

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. It might thus be that relations such as $E = \hbar\omega$ or $p = \hbar/2\pi\lambda$ are not applicable to gravitational waves. In short, 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.

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

$$\Delta g \approx 2 \frac{l_{\text{Pl}}{}^2}{L^2} , \tag{45}$$

Challenge 109 ny

where $l_{\text{Pl}} = \sqrt{\hbar G/c^3}$ 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 $\hbar$ cannot be observed. This implies that the observable values for the metric g in a region of size L are bound by

$$g \geqslant \frac{2\hbar G}{c^3} \frac{1}{L^2} . \tag{46}$$

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 be essential in the last volume of our mountain ascent.

Vol. VI, page 54

Decoherence of space-time

General relativity taught us that the gravitational field and space-time are the same. If the gravitational field evolves like a quantum system, we may ask why no superpositions of different macroscopic space-times are observed.

Ref. 136

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

Page 123 value g'. As in the case of a superposition of macroscopic distinct wave functions, such a superposition *decays*. In particular, it decays when particles cross the volume. A short Challenge 110 ny calculation yields a decay time given by

$$t_d = (\frac{2kT}{\pi m})^{3/2} \frac{nl^4}{(g - g')^2} , \qquad (47)$$

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 Challenge 111 ny *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.

In short, matter not only tells space-time how to curve, it also tells it to behave with class.

Quantum theory as the enemy of science fiction

How does quantum theory change our ideas of space-time? The end of the twentieth century has brought several unexpected but strong results in semiclassical quantum gravity.

Ref. 137 In 1995 Ford and Roman found that worm holes, which are imaginable in general relativity, cannot exist if quantum effects are taken into account. They showed that macroscopic worm holes require unrealistically large negative energies. (For microscopic worm holes the issue is still unclear.)

Ref. 138 In 1996 Kay, Radzikowski and Wald showed that closed time-like curves do not exist in semiclassical quantum gravity; there are thus no time machines in nature.

Ref. 139 In 1997 Pfenning and Ford showed that warp drive situations, which are also imaginable in general relativity, cannot exist if quantum effects are taken into account. Such situations require unrealistically large negative energies.

In short, the inclusion of quantum effects destroys all those fantasies which were started by general relativity.

No vacuum means no particles

Gravity has an important consequence for quantum theory. To count and define particles, quantum theory needs a defined vacuum state. However, the vacuum state cannot be Page 97 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 final part of our mountain ascent.

Summary on quantum theory and gravity

Everyday gravitational fields can be included in quantum theory. *Weak* gravitational fields have predictable and measurable effects on wave functions. The inclusion of *strong* gravitational fields into quantum theory leads to problems with the particle concept.

Conversely, the inclusion of quantum effects into general relativity leads to space-time foam, space-time superpositions, and probably of gravitons. The inclusion also forbids the existence of wormholes, time-like curves and negative energy regions. For high curvatures, problems with the concept of space-time arise.

In summary, the combination of quantum theory and gravitation leads to problems with both the particle concept and the space-time concept. We are thus forced putting into question the foundations of our description of nature so far.

In fact, up to now we hid a simple fact: quantum theory and general relativity *contradict* each other. This contradiction was the real reason that we stepped back to special relativity before we started exploring quantum theory. By stepping back we avoided many problems, because quantum theory does not contradict *special* relativity. However, quantum theory does contradict *general* relativity. The issues are dramatic, changing everything from the basis of classical physics to the results of quantum theory. There will be surprising consequences for the nature of space-time, for the nature of particles, and for motion itself. Before we study these issues, however, we complete the theme of the present, quantum part of the mountain ascent, namely exploring the essence of matter and interactions.

Vol. VI, page 17

Chapter 5

THE STRUCTURE OF THE NUCLEUS – THE DENSEST CLOUDS

Nuclear physics was born in 1896 in France, but is now a small activity. Not many researchers are working on the topic now. The field produced not more than one daughter, experimental high energy physics, which was born around 1930. These activities have been in strong decline since 1985, with the exception of the latest CERN experiment, the *Large Hadron Collider*. Given the short time nuclear physics has been in existence, the history of the field is impressive: it discovered why stars shine, how powerful bombs work, how cosmic evolution produced the atoms we are made of and how medical doctors can dramatically improve their healing rate.

Ref. 140

> “Nuclear physics is just low-density astrophysics.”
> Anonymous

A physical wonder – magnetic resonance imaging

Arguably, the most spectacular tool that physical research produced in the twentieth century was magnetic resonance imaging, or MRI for short. This technique allows us to image human bodies with a high resolution and with (almost) no damage, in strong contrast to X-ray imaging. Though the machines are still expensive – costing up to several million euro – there is hope that they will become cheaper in the future. Such a machine, shown in Figure 55, consists essentially of a large magnetic coil, a radio transmitter and a computer. Some results of putting part of a person into the coil are shown in Figure 56.

In MRI machines, a radio transmitter emits radio waves that are absorbed because hydrogen nuclei are small spinning magnets. The magnets can be parallel or antiparallel to the magnetic field produced by the coil. The transition energy E is absorbed from a radio wave whose frequency ω is tuned to the magnetic field B. The energy absorbed by a single hydrogen nucleus is given by

$$E = \hbar\omega = \hbar\gamma B \qquad (48)$$

The material constant $\gamma/2\pi$ has a value of 42.6 MHz/T for hydrogen nuclei; it results from the non-vanishing spin of the proton. This is a quantum effect, as shown by the appearance of the quantum of action $\hbar$. Using some cleverly applied magnetic fields, typically with a strength between 0.3 and 7 T for commercial and up to 21 T for experimental machines, the machines are able to measure the absorption for each volume element separately. Interestingly, the precise absorption *level* depends on the chemical compound the nucleus is built into. Thus the absorption value will depend on the chemical substance.

FIGURE 55 A commercial MRI machine (© Royal Philips Electronics)

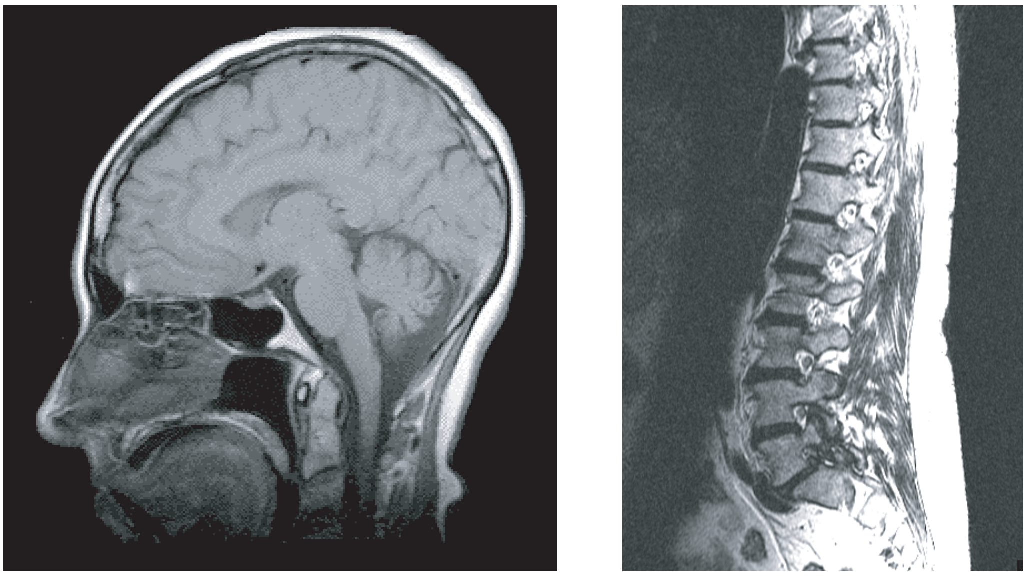


FIGURE 56 Sagittal images of the head and the spine (used with permission from Joseph P. Hornak, *The Basics of MRI*, www.cis.rit.edu/htbooks/mri, Copyright 2003)

When the intensity of the absorption is plotted as grey scale, an image is formed that retraces the different chemical compositions. Two examples are shown in Figure 56. Using additional tricks, modern machines can picture blood flow in the heart or air flow in
Ref. 141 lungs; they now routinely make films of the heart beat. Other techniques show how the location of sugar metabolism in the brain depends on what you are thinking about.*

* The website www.cis.rit.edu/htbooks/mri by Joseph P. Hornak gives an excellent introduction to magnetic

Also what many are thinking about all the time has been imaged: the first scan of Ref. 142 people making love has been taken by Willibrord Weijmar Schultz and his group in 1999. It is shown in Figure 57.

Each magnetic resonance image thus proves that (many) *atoms have spinning nuclei*. Like any other object, nuclei have size, shape, colour, composition and interactions. Let us explore them.

The size of nuclei

The magnetic resonance signal shows that hydrogen nuclei are quite sensitive to magnetic fields. The g-factor of protons, defined using the magnetic moment μ, their mass and charge as $g = \mu 4m/e\hbar$, is about 5.6. Using expression (47) that relates the g-factor Page 91 and the radius of a composite object, we deduce that the radius of the proton is about 0.9 fm; this value is confirmed by many experiments. Protons are thus much smaller, about 30 000 times smaller, than hydrogen atoms, the smallest of atoms, whose radius is about 30 pm. In turn, the proton is the smallest of all nuclei; the largest known nuclei have radii 7 times the proton value.

The small size of nuclei is no news. It is known since the beginning of the twentieth century. The story starts on the first of March in 1896, when Henri Becquerel* discovered a puzzling phenomenon: minerals of uranium potassium sulphate blacken photographic plates. Becquerel had heard that the material is strongly fluorescent; he conjectured that fluorescence might have some connection to the X-rays discovered by Conrad Röngten the year before. His conjecture was wrong; nevertheless it led him to an important new discovery. Investigating the reason for the effect of uranium on photographic plates, Becquerel found that these minerals emit an undiscovered type of radiation, different from anything known at that time; in addition, the radiation is emitted by any substance containing uranium. In 1898, Bémont named the property of these minerals *radioactivity*.

Radioactive rays are also emitted from many elements other than uranium. The radiation can be 'seen': it can be detected by the tiny flashes of light that are emitted when the rays hit a scintillation screen. The light flashes are tiny even at a distance of several metre from the source; thus the rays must be emitted from point-like sources. Radioactivity has to be emitted from single atoms. Thus radioactivity confirmed unambiguously that atoms do exist. In fact, radioactivity even allows to *count* atoms, as we will find out shortly.

The intensity of radioactivity cannot be influenced by magnetic or electric fields; it does not depend on temperature or light irradiation. In short, radioactivity does not depend on electromagnetism and is not related to it. Also the high energy of the emitted radiation cannot be explained by electromagnetic effects. Radioactivity must thus be due to another, new type of force. In fact, it took 30 years and a dozen of Nobel Prizes to fully

resonance imaging, both in English and Russian, including the physical basis, the working of the machines, and numerous beautiful pictures. The method of studying nuclei by putting them at the same time into magnetic and radio fields is also called *nuclear magnetic resonance*.

* Henri Becquerel (b. 1852 Paris, d. 1908 Le Croisic), important French physicist; his primary topic was the study of radioactivity. He was the thesis adviser of Marie Curie, the wife of Pierre Curie, and was central to bringing her to fame. The SI unit for radioactivity is named after him. For his discovery of radioactivity he received the 1903 Nobel Prize for physics; he shared it with the Curies.

FIGURE 57 The origin of human life (© Willibrord Weijmar Schultz)

FIGURE 58 Henri Becquerel (1852–1908)

understand the details. It turns out that several types of radioactivity exist; the types of emitted radiation behave differently when they fly through a magnetic field or when they encounter matter. They are listed in Table 9. In the meantime, all these rays have been studied in great detail, with the aim to understand the nature of the emitted entity and its interaction with matter.

In 1909, radioactivity inspired the 37 year old physicist Ernest Rutherford,* who had won the Nobel Prize just the year before, to another of his brilliant experiments. He asked his collaborator Hans Geiger to take an emitter of alpha radiation – a type of radioactivity

* Ernest Rutherford (1871–1937), important New Zealand physicist. He emigrated to Britain and became professor at the University of Manchester. He coined the terms alpha particle, beta particle, proton and neutron. A gifted experimentalist, he discovered that radioactivity transmutes the elements, explained the nature of alpha rays, discovered the nucleus, measured its size and performed the first nuclear reactions. Ironically, in 1908 he received the Nobel Prize for chemistry, much to the amusement of himself and of the world-wide physics community; this was necessary as it was impossible to give enough physics prizes to the numerous discoverers of the time. He founded a successful research school of nuclear physics and many famous physicists spent some time at his institute. Ever an experimentalist, Rutherford deeply disliked quantum theory, even though it was and is the only possible explanation for his discoveries.

TABLE 9 The main types of radioactivity and rays emitted by matter

Type	Particle	Example	Range	Danger	Shield	Use
α rays 3 to 10 MeV	helium nuclei	^{235}U, ^{238}U, ^{238}Pu, ^{238}Pu, ^{241}Am	a few cm in air	when eaten, inhaled, touched	any material, e.g. paper	thickness measurement
β rays 0 to 5 MeV	electrons and	^{14}C, ^{40}K, ^{3}H, ^{101}Tc	< 1 mm in metal	serious	metals	cancer treatment
	antineutrinos		light years	none	none	research
β^{+} rays	positrons and	^{40}K, ^{11}C, ^{11}C, ^{13}N, ^{15}O	less than β	medium	any material	tomography
	neutrinos		light years	none	none	research
γ rays	high energy photons	^{110}Ag	several m in air	high	thick lead	preservation of herbs, disinfection
n reactions *c.* 1 MeV	neutrons	^{252}Cf, Po-Li (α,n), ^{38}Cl-Be (γ,n)	many m in air	high	0.3 m of paraffin	nuclear power, quantum gravity experiments
n emission typ. 40 MeV	neutrons	^{9}He, ^{24}N, ^{254}Cf	many m in air	high	0.3 m of paraffin	research experiments
p emission typ. 20 MeV	protons	^{5}Be, ^{161}Re	like α rays	small	solids	
spontaneous fission typ. 100 MeV	nuclei	^{232}Cm, ^{263}Rf	like α rays	small	solids	detection of new elements

which Rutherford had identified and named 10 years earlier – and to point the radiation at a thin metal foil. The quest was to find out where the alpha rays would end up. The research group followed the path of the particles by using scintillation screens; later on they used an invention by Charles Wilson: the cloud chamber. A *cloud chamber*, like its successor, the bubble chamber, produces white traces along the path of charged particles; the mechanism is the same as the one than leads to the white lines in the sky when an aeroplane flies by.

The radiation detectors gave a strange result: most alpha particles pass through the metal foil undisturbed, whereas a few are scattered and a few are reflected. In addition, those few which are reflected are not reflected by the surface, but in the inside of the foil. (Can you imagine how they showed this?) Rutherford deduced from this scattering experiment that first of all, the atoms in the metal foil are mainly transparent. Only

Challenge 112 s

FIGURE 59 Marie Curie (1867–1934)

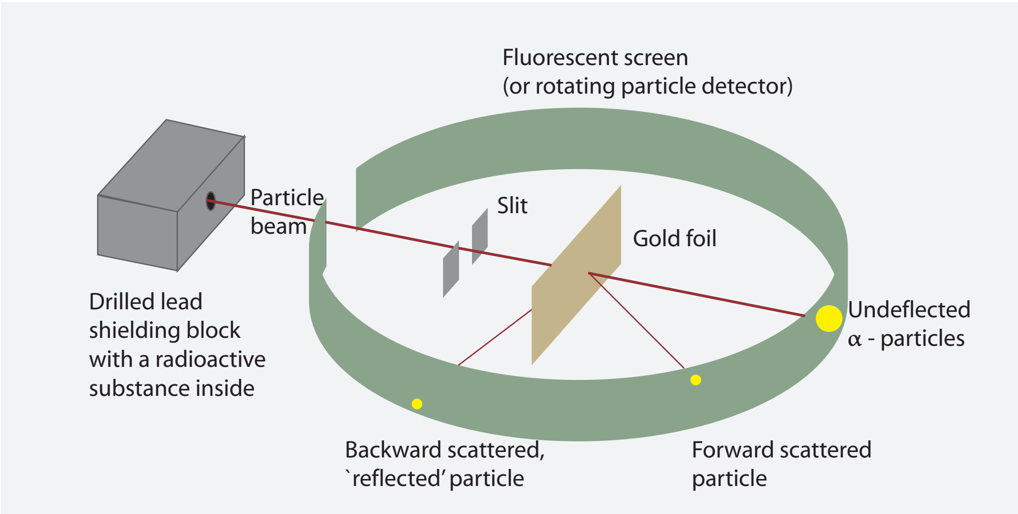


FIGURE 60 The schematics of the Rutherford–Geiger scattering experiment.

transparency explains why most alpha particles pass the foil without disturbance, even though it was over 2000 atoms thick. But some particles were scattered by large angles or even reflected. Rutherford showed that the reflections must be due to a single scattering point. By counting the particles that were reflected (about 1 in 20000 for his 0.4 µm gold foil), Rutherford was also able to deduce the *size* of the reflecting entity and to estimate its *mass*. He found that the reflecting entity contains practically all of the mass of the atom in a diameter of around 1 fm. Rutherford thus named it the atomic *nucleus*. Using the knowledge that atoms contain electrons, Rutherford then deduced from this experiment that atoms consist of an electron cloud that determines the size of atoms – of the order of 0.1 nm – and of a tiny but heavy nucleus at the centre. If an atom had the size of a basketball, its nucleus would have the size of a dust particle, yet contain 99.9 % of the basketball's mass. Atoms resemble thus candy floss around a heavy dust particle. Even though the candy floss – the electron cloud – around the nucleus is extremely thin and light, it is strong enough to avoid that two atoms interpenetrate. The candy floss, i.e., the electron cloud, keeps the neighbouring nuclei at constant distance. For the tiny and massive alpha particles however, the candy floss is essentially empty space, so that they

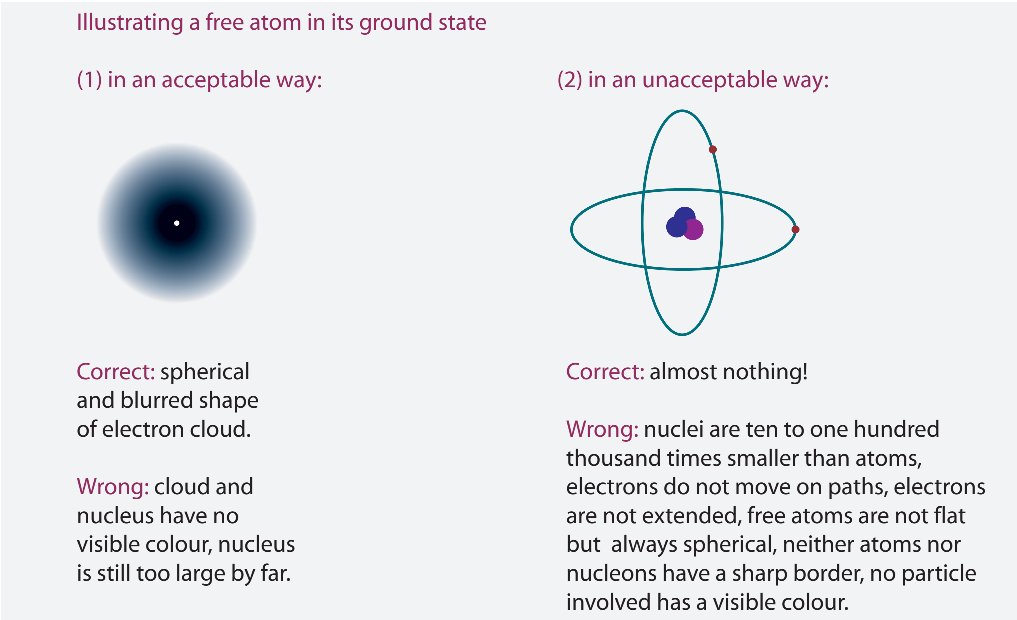


FIGURE 61 A reasonably realistic (left) and a misleading illustration of an atom (right) as is regularly found in school books. *Atoms in the ground state are spherical electron clouds with a tiny nucleus, itself a cloud, at its centre.* Interacting atoms, chemically bound atoms and some, but not all excited atoms have electron clouds of different shapes.

simply fly through the electron clouds until they either exit on the other side of the foil or hit a nucleus.

The density of the nucleus is impressive: about $5.5 \cdot 10^{17}\ \mathrm{kg/m^3}$. At that density, the mass of the Earth would fit in a sphere of 137 m radius and a grain of sand would have a mass larger than the largest existing oil tanker. (True?) Challenge 113 e

> I now know how an atom looks like!
> Ernest Rutherford

Nuclei are composed

The magnetic resonance images also show that nuclei are composed. Images can be taken also using heavier nuclei instead of hydrogen, such as certain fluorine or oxygen nuclei. The g-factors of these nuclei also depart from the value 2 characteristic of point particles; the more massive they are, the bigger the departure. Such objects have a finite size; indeed, the size of nuclei can be measured directly and confirm the values predicted by the g-factor. Both the values of the g-factor and the non-vanishing sizes show that nuclei are composed. Vol. IV, page 90

Interestingly, the idea that nuclei are composed is older than the concept of nucleus itself. Already in 1815, after the first mass measurements of atoms by John Dalton and others, researchers noted that the mass of the various chemical elements seem to be almost perfect multiples of the weight of the hydrogen atom. William Prout then formulated the hypothesis that all elements are composed of hydrogen. When the nucleus was discov-

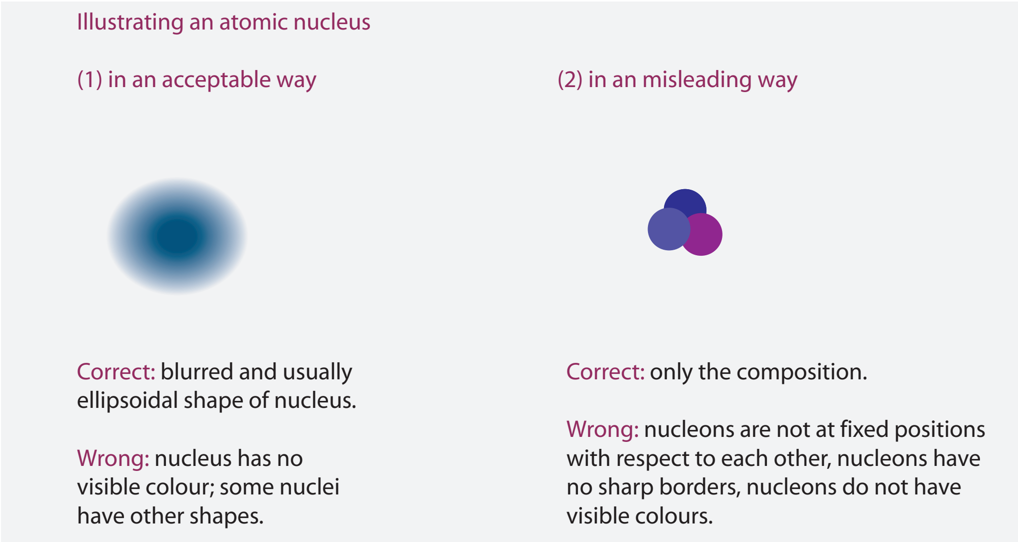


FIGURE 62 A reasonably realistic (left) and a misleading illustration of a nucleus (right) as is regularly found in school books. *Nuclei are spherical nucleon clouds.*

ered, knowing that it contains almost all mass of the atom, it was therefore first thought that all nuclei are made of hydrogen nuclei. Being at the origin of the list of constituents, the hydrogen nucleus was named *proton*, from the greek term for 'first' and reminding the name of Prout at the same time. Protons carry a positive unit of electric charge, just the opposite of that of electrons, but are almost 2000 times as heavy.

However, the charge and the mass numbers of the other nuclei do not match. On average, a nucleus that has n times the charge of a proton, has a mass that is about 2.6 n times than of the proton. Additional experiments then confirmed an idea formulated by Werner Heisenberg: all nuclei heavier than hydrogen nuclei are made of positively charged *protons* and neutral *neutrons*. Neutrons are particles a tiny bit more massive than protons (the difference is less than a part in 700), but without any electrical charge. Since the mass is almost the same, the mass of nuclei – and thus that of atoms – is still an (almost perfect) integer multiple of the proton mass. But since neutrons are neutral, the mass and the charge number of nuclei differ. Being neutral, neutrons do not leave tracks in clouds chambers and are more difficult to detect. For this reason, they were discovered much later than other subatomic particles.

Today it is possible to keep single neutrons suspended between suitably shaped coils, with the aid of teflon 'windows'. Such traps were proposed in 1951 by Wolfgang Paul. They work because neutrons, though they have no charge, do have a small magnetic moment. (By the way, this implies that neutrons are themselves composed of charged particles.) With a suitable arrangement of magnetic fields, neutrons can be kept in place, in other words, they can be levitated. Obviously, a trap only makes sense if the trapped particle can be observed. In case of neutrons, this is achieved by the radio waves absorbed when the magnetic moment switches direction with respect to an applied magnetic field. The result of these experiments is simple: the lifetime of free neutrons is around 888(1) s. Nevertheless, inside most nuclei we are made of, neutrons do not decay, as the result

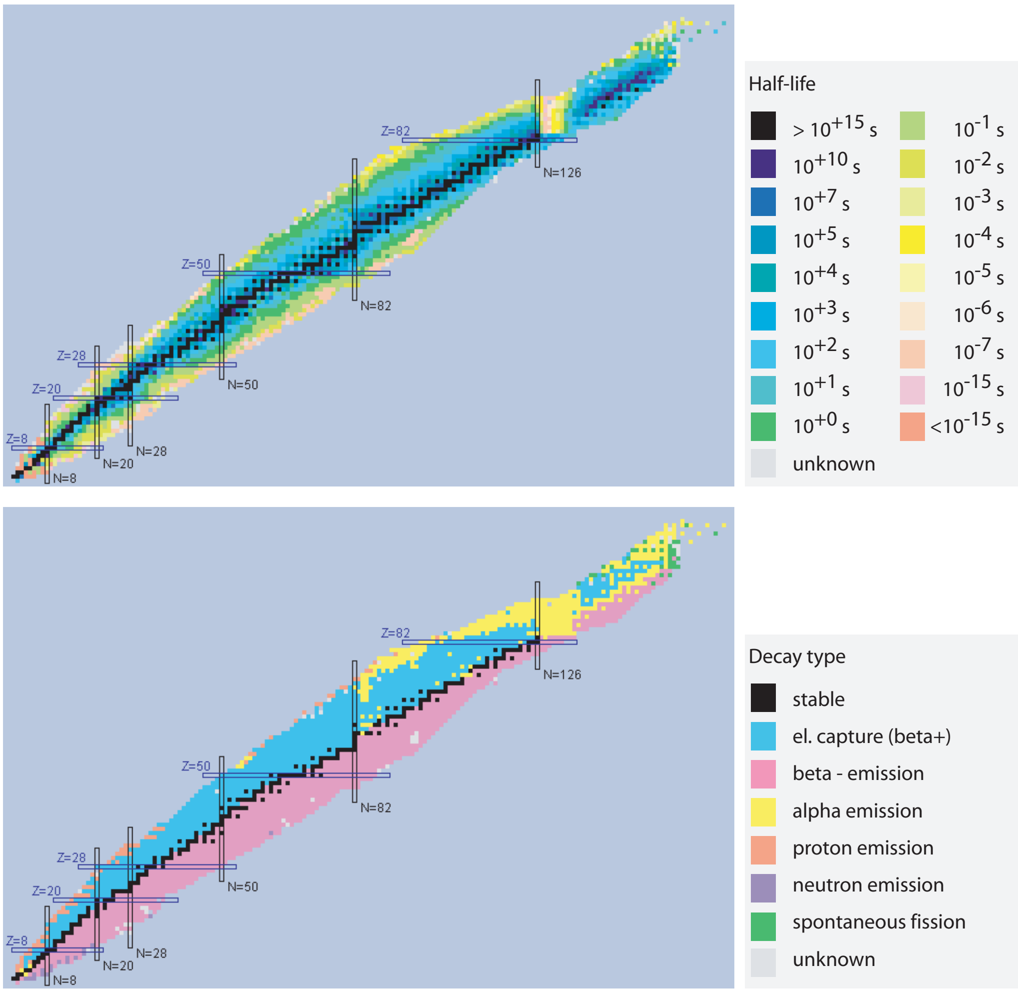


FIGURE 63 All known nuclides with their lifetimes (above) and main decay modes (below), data from www.nndc.bnl.gov/nudat2)

Challenge 114 s

does not lead to a state of lower energy. (Why not?)

Magnetic resonance images also show that some elements have different types of atoms. These elements have atoms that with the same number of protons, but with different numbers of neutrons. One says that these elements have several isotopes.* This also explains why some elements radiate with a mixture of different decay times. Though chemically they are (almost) indistinguishable, isotopes can differ strongly in their nuclear properties. Some elements, such as tin, caesium, or polonium, have over thirty isotopes each. Together, the about 100 known elements have over 2000 *nuclides*.**

The motion of protons and neutrons inside nuclei allows to understand the spin and the magnetic moment of nuclei. Since nuclei are so extremely dense despite containing

* The name is derived from the Greek words for 'same' and 'spot', as the atoms are on the same spot in the periodic table of the elements.

** Nuclides is the standard expression for a nucleus with a given number of neutrons and protons.

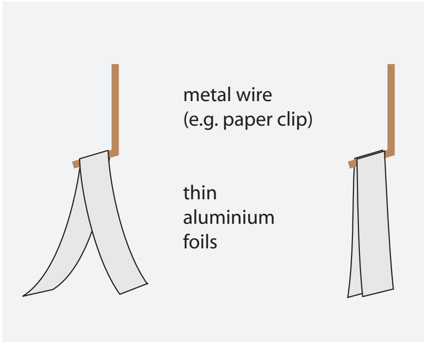


FIGURE 64 An electroscope (or electrometer) (© Harald Chmela) and its charged (middle) and uncharged state (right)

FIGURE 65 Viktor Heß (1883–1964)

numerous positively charged protons, there must be a force that keeps everything together against the electrostatic repulsion. We saw that the force is not influenced by electromagnetic or gravitational fields; it must be something different. The force must be short range; otherwise nuclei would not decay by emitting high energy alpha rays. The new force is called the strong nuclear interaction. We shall study it in detail shortly.

Nuclei can move alone – cosmic rays

In everyday life, nuclei are mostly found inside atoms. But in some situations, they move all by themselves, without surrounding electron clouds. The first to discover an example was Rutherford; with a clever experiment he showed that alpha particles are helium nuclei. Like all nuclei, alpha particles are small, so that they are quite useful as projectiles.

Then, in 1912, Viktor Heß* made a completely unexpected discovery. Heß was intrigued by electroscopes (also called electrometers). Page 22 These are the simplest possible detectors of electric charge. They mainly consist of two hanging, thin metal foils, such as two strips of aluminium foil taken from a chocolate bar. When the electroscope is charged,

* Viktor Franz Heß, (1883–1964), Austrian nuclear physicist, received the Nobel Prize for physics in 1936 for his discovery of cosmic radiation. Heß was one of the pioneers of research into radioactivity. Heß' discovery also explained why the atmosphere is always somewhat charged, a result important for the formation and behaviour of clouds. Twenty years after the discovery of cosmic radiation, in 1932 Carl Anderson discovered the first antiparticle, the positron, in cosmic radiation; in 1937 Seth Neddermeyer and Carl Anderson discovered the muon; in 1947 a team led by Cecil Powell discovered the pion; in 1951, the Λ^0 and the kaon K^0 are discovered. All discoveries used cosmic rays and most of these discoveries led to Nobel Prizes.

the strips repel each other and move apart, as shown in Figure 64. (You can build one easily yourself by covering an empty glass with some transparent cellophane foil and suspending a paper clip and the aluminium strips from the foil. You can charge the electroscope with the help of a rubber balloon and a woollen pullover.) An electroscope thus measures electrical charge. Like many before him, Heß noted that even for a completely isolated electroscope, the charge disappears after a while. He asked: why? By careful study he eliminated one explanation after the other. Heß (and others) were left with only one possibility: that the discharge could be due to charged rays, such as those of the recently discovered radioactivity, emitted from the environment. To increase the distance to the environment, Heß prepared a sensitive electrometer and took it with him on a balloon flight.

Challenge 115 e

As expected, the balloon flight showed that the discharge effect diminished with height, due to the larger distance from the radioactive substances on the Earth's surface. But above about 1000 m of height, the discharge effect increased again, and the higher he flew, the stronger it became. Risking his health and life, he continued upwards to more than 5000 m; there the discharge effect was several times stronger than on the surface of the Earth. This result is exactly what is expected from a radiation coming from outer space and absorbed by the atmosphere. In one of his most important flights, performed during an (almost total) solar eclipse, Heß showed that most of the 'height radiation' did not come from the Sun, but from further away. He thus called the radiation *cosmic rays*. During the last few centuries, many people have drunk from a glass and eaten chocolate; but only Heß combined these activities with such careful observation and deduction that he earned a Nobel Prize.*

Today, the most impressive detectors for cosmic rays are Geiger–Müller counters and spark chambers. Both share the same idea; a high voltage is applied between two metal parts kept in a thin and suitably chosen gas (a wire and a cylindrical mesh for the Geiger-Müller counter, two plates or wire meshes in the spark chambers). When a high energy ionizing particle crosses the counter, a spark is generated, which can either be observed through the generated spark (as you can do yourself in the entrance hall of the CERN main building), or detected by the sudden current flow. Historically, the current was first amplified and sent to a loudspeaker, so that the particles can be heard by a 'click' noise. With a Geiger counter, one cannot see atoms or particles, but one can hear them. Finally, ionized atoms could be counted. Finding the right gas mixture is tricky; it is the reason that the counter has a double name. One needs a gas that extinguishes the spark after a while, to make the detector ready for the next particle. Müller was Geiger's assistant; he made the best counters by adding the right mixture of alcohol to the gas in the chamber. Nasty rumours maintained that this was discovered when another assistant tried, without success, to build counters while Müller was absent. When Müller, supposedly a heavy drinker, came back, everything worked again. However, the story is apocryphal. Today, Geiger–Müller counters are used around the world to detect radioactivity; the smallest fit in mobile phones and inside wrist watches.

The particle energy in cosmic rays spans a large range between 10^3 eV and at least 10^{20} eV; the latter is the same energy as a tennis ball after serve. Understanding the origin of cosmic rays is a science by its own. Some are galactic in origin, some are extragalactic.

* In fact, Hess used gold foils in his electrometer, not aluminium foils.

TABLE 10 The main types of cosmic radiation

Particle	Energy	Origin	Detector	Shield
At high altitude, the primary particles:				
Protons (90 %)	10^9 to 10^{22} eV	stars, supernovae, extragalactic, unknown	scintillator	in mines
Alpha rays (9 %)	typ. $5 \cdot 10^6$ eV	stars, galaxy	ZnS, counters	1 mm of any material
Other nuclei, such as Le, Be, B, Fe (1 %)	10^9 to 10^{19} eV	stars, novae	1 mm of any material	counters, films
Neutrinos	MeV, GeV	Sun, stars	chlorine, gallium, water	none
Electrons (0.1 %)	10^6 to $> 10^{12}$ eV	supernova remnants		
Gammas (10^{-6})	1 eV to 50 TeV	stars, pulsars, galactic, extragalactic	semiconductor detectors	in mines
At sea level, secondary particles are produced in the atmosphere:				
Muons	3 GeV, $150/\,m^2s$	protons hit atmosphere, produce pions which decay into muons	drift chamber, bubble chamber, scintillation detector	15 m of water or 2.5 m of soil
Oxygen, radiocarbon and other nuclei	varies	e.g., $n + {}^{16}O \rightarrow p + {}^{14}C$	soil	
Positrons	varies		counters	soil
Neutrons	varies	reaction product when proton hits ^{16}O nucleus	counters	soil
Pions	varies	reaction product when proton hits ^{16}O nucleus	counters	soil
In addition, there are slowed down primary beam particles.				

For most energies, supernova remnants – pulsars and the like – seem the best candidates. However, the source of the highest energy particles is still unknown.

In other words, cosmic rays are probably the only type of radiation discovered without the help of shadows. But shadows have been found later on. In a beautiful experiment performed in 1994, the shadow thrown by the Moon on high energy cosmic rays (about 10 TeV) was studied. When the position of the shadow is compared with the actual position of the Moon, a shift is found. Indeed, due to the magnetic field of the Earth, the cosmic ray Moon shadow is expected to be shifted westwards for protons and eastwards for antiprotons. The data are consistent with a ratio of antiprotons between 0 % and 30 %. Ref. 143

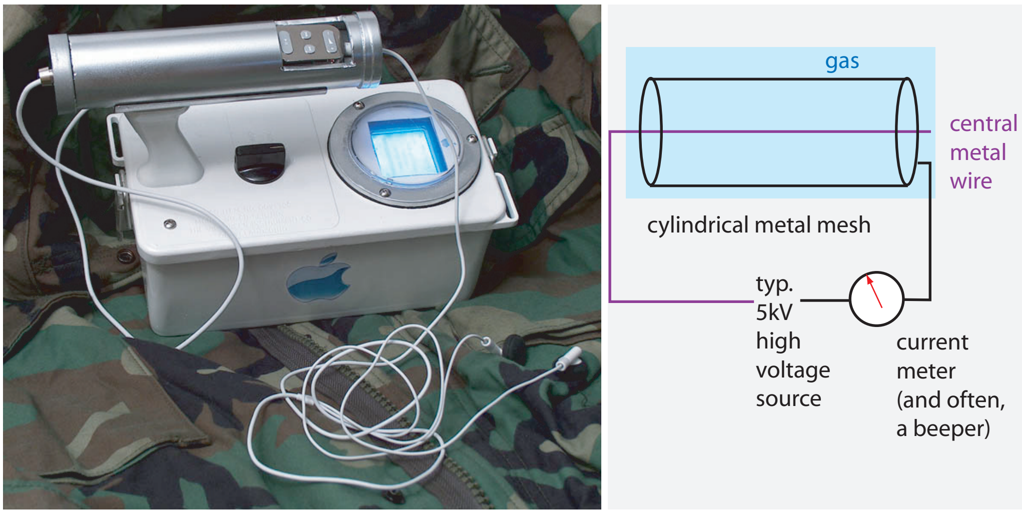


FIGURE 66 A Geiger–Müller counter with the detachable detection tube, the connection cable to the counter electronics, and, for this model, the built-in music player (© Joseph Reinhardt)

By studying the shadow's position, the experiment thus showed that high energy cosmic rays are mainly positively charged and thus consist mainly of matter, and only in small part, if at all, of antimatter.

Detailed observations showed that cosmic rays arrive on the surface of the Earth as a mixture of many types of particles, as shown in Table 10. They arrive from outside the atmosphere as a mixture of which the largest fraction are protons, followed by alpha particles, iron and other nuclei. Nuclei can thus travel alone over large distances.

The flux of charged cosmic rays arriving at the surface of the Earth depends on their energy. At the lowest energies, charged cosmic rays hit the human body many times a second. Measurements also show that the rays arrive in irregular groups, called *showers*. Page 187 In contrast, the neutrino flux is many orders of magnitude higher, but does not have any effect on human bodies.

The distribution of the incoming direction of cosmic rays shows that many rays must be extragalactic in origin. Indeed, the typical nuclei of cosmic radiation are ejected from stars and accelerated by supernova explosions. When they arrive on Earth, they interact with the atmosphere before they reach the surface of the Earth. The detailed acceleration mechanisms are still a topic of research.

Cosmic rays have several effects on everyday life. Through the charges they produce in the atmosphere, they are probably responsible for the jagged, non-straight propagation of lightning. (Lightning advances in pulses, alternating fast propagation for about 30 m with slow propagation, until they hit connect. The direction they take at the slow spots depends on the wind and the charge distribution in the atmosphere.) Cosmic rays are also important in the creation of rain drops and ice particles inside clouds, and thus indirectly in the charging of the clouds. Cosmic rays, together with ambient radioactivity, Page 17 also start the Kelvin generator.

If the magnetic field of the Earth would not exist, we would die from cosmic rays. The magnetic field diverts most rays towards the magnetic poles. Also the upper atmosphere

FIGURE 67 An aurora borealis, produced by charged particles in the night sky (© Jan Curtis)

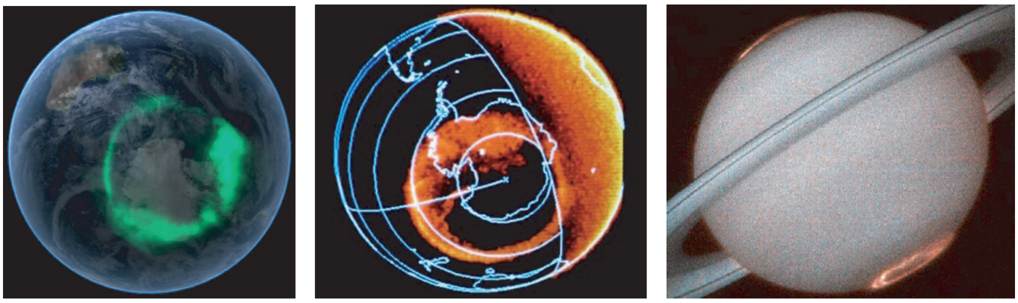


FIGURE 68 Two aurorae australes on Earth, seen from space (a composed image with superimposed UV intensity, and a view in the X-ray domain) and a double aurora on Saturn (all NASA)

helps animal life to survive, by shielding life from the harmful effects of cosmic rays. Indeed, aeroplane pilots and airline employees have a strong radiation exposure that is not favourable to their health. Cosmic rays are also one of several reasons that long space travel, such as a trip to Mars, is not an option for humans. When cosmonauts get too much radiation exposure, the body weakens and eventually they die. Space heroes, including those of science fiction, would not survive much longer than two or three years.

Cosmic rays also produce beautifully coloured flashes inside the eyes of cosmonauts; they regularly enjoy these events in their trips. But cosmic rays are not only dangerous and beautiful. They are also useful. If cosmic rays would not exist at all, we would not exist either. Cosmic rays are responsible for mutations of life forms and thus are one of the causes of biological evolution. Today, this effect is even used artificially; putting cells into a radioactive environment yields new strains. Breeders regularly derive new mutants in this way.

Cosmic rays cannot be seen directly, but their cousins, the 'solar' rays, can. This is most spectacular when they arrive in high numbers. In such cases, the particles are inevitably deviated to the poles by the magnetic field of the Earth and form a so-called *aurora borealis* (at the North Pole) or an *aurora australis* (at the South pole). These slowly moving and variously coloured curtains of light belong to the most spectacular effects in the night sky. (Have a look at www.nasa.gov/mov/105423main_FUV_2005-01_v01.mov.) Visible light and X-rays are emitted at altitudes between 60 and 1000 km. Seen from space, the aurora curtains typically form a circle with a few thousand kilometres diameter around the magnetic poles.*

Cosmic rays are mainly free nuclei. With time, researchers found that nuclei appear without electron clouds also in other situations. In fact, the vast majority of nuclei in the universe have no electron clouds at all: in the inside of stars, no nucleus is surrounded by bound electrons; similarly, a large part of intergalactic matter is made of protons. It is known today that most of the matter in the universe is found as protons or alpha particles inside stars and as thin gas between the galaxies. In other words, in contrast to what the Greeks said, matter is not usually made of atoms; it is mostly made of bare nuclei. Our everyday environment is an exception when seen on cosmic scales. In nature, atoms are rare, bare nuclei are common.

Incidentally, nuclei are in no way forced to move; nuclei can also be stored with almost no motion. There are methods – now commonly used in research groups – to superpose electric and magnetic fields in such a way that a single nucleus can be kept floating in mid-air; we discussed this possibility in the section on levitation earlier on. Page 153

Nuclei decay

Not all nuclei are stable over time. The first measurement that provided a hint was the way radioactivity changes with time. The number N of emitted rays decreases with time. More precisely, radioactivity follows an exponential decay:

$$N(t) = N(0)\, e^{-t/\tau} \qquad (49)$$

The parameter τ, the so-called *life time*, depends on the type of nucleus emitting the rays. Life times can vary from much less than a microsecond to millions of millions of years. The expression has been checked for as long as 34 multiples of the duration τ; its validity and precision is well-established by experiments. Formula (49) is an approximation for large numbers of atoms, as it assumes that $N(t)$ is a continuous variable. Despite this approximation, deriving this expression from quantum theory is not a simple exercise, as we saw above. Page 37 Though the quantum Zeno effect can appear for small times t, for the case of radioactivity it has not been observed so far.

Radioactivity is the decay of unstable nuclei. Most of all, radioactivity allows to count the number of atoms in a given mass of material. Imagine to have measured the mass of radioactive material at the beginning of your experiment; you have chosen an element that has a lifetime of about a day. Then you put the material inside a scintillation box. After a few weeks the number of flashes has become so low that you can count them; using expression (49) you can then determine how many atoms have been in the mass

* In the solar system, aurorae due to core magnetic fields have been observed on Jupiter, Saturn, Uranus, Neptune, Earth, Io and Ganymede. Aurorae due to other mechanisms have been seen on Venus and Mars.

to begin with. Radioactivity thus allows us to determine the number of atoms, and thus their size, in addition to the size of nuclei.

The decay (49) and the release of energy is typical of metastable systems. In 1903, Rutherford and Soddy discovered what the state of lower energy is for alpha and beta emitters. In these cases, radioactivity changes the emitting atom; it is a spontaneous transmutation of the atom. An atom emitting alpha or beta rays changes its chemical nature.
Page 310 Radioactivity confirms for nuclei what statistical mechanics of gases had concluded long time before for atoms: they have a structure that can change.

In *alpha decay*, the radiating nucleus emits a (doubly charged) helium nucleus. The kinetic energy is typically a handful of MeV. After the emission, the nucleus has changed to one situated two places earlier in the periodic system of the elements.

In *beta decay*, a neutron transforms itself into a proton, emitting an electron and an antineutrino. Also beta decay changes the chemical nature of the atom, but to the place following the original atom in the periodic table of the elements. A variation is the beta+ decay, in which a proton changes into a neutron and emits a neutrino and a positron. We
Page 176 will study these important decay processes below.

In *gamma decay*, the nucleus changes from an excited to a lower energy state by emitting a high energy photon. In this case, the chemical nature is not changed. Typical energies are in the MeV range. Due to the high energy, such rays ionize the material they encounter; since they are not charged, they are not well absorbed by matter and penetrate deep into materials. Gamma radiation is thus by far the most dangerous type of (outside) radioactivity.

By the way, in every human body about nine thousand radioactive decays take place every second, mainly 4.5 kBq (0.2 mSv/a) from ^{40}K and 4 kBq from ^{14}C (0.01 mSv/a).
Challenge 116 s Why is this not dangerous?

All radioactivity is accompanied by emission of energy. The energy emitted by an atom through radioactive decay or reactions is regularly a million time large than that emitted by a chemical process. That is the reason for the danger of nuclear weapons. More than a decay, a radioactive process is thus an explosion.

What distinguishes those atoms that decay from those which do not? An exponential decay law implies that the probability of decay is independent of the age of the atom.
Page 95 Age or time plays no role. We also know from thermodynamics, that all atoms have exactly identical properties. So how is the decaying atom singled out? It took around 40 years to discover that decays are triggered by the statistical fluctuations of the vacuum, as described by quantum theory. Indeed, radioactivity is one of the clearest observations that classical physics is not sufficient to describe nature. Radioactivity, like all decays, is a pure quantum effect. Only a finite quantum of action makes it possible that a system remains unchanged until it suddenly decays. Indeed, in 1928 George Gamow explained alpha decay with the tunnelling effect. The tunnelling effect explains the relation between the lifetime and the range of the rays, as well as the measured variation of lifetimes – between 10 ns and 10^{17} years – as the consequence of the varying potentials to be overcome.

By the way, massless particles cannot decay. There is a simple reason for it: massless particles do not experience time, as their paths are 'null'. A particle that does not experi-
Challenge 117 s ence time cannot have a half-life. (Can you find another argument?)

Radiometric dating

As a result of the chemical effects of radioactivity, the composition ratio of certain elements in minerals allows to determine the age of the mineral. Using radioactive decay to deduce the age of a sample is called *radiometric dating*. With this technique, geologists determined the age of mountains, the age of sediments and the age of the continents. They determined the time that continents moved apart, the time that mountains formed when the continents collided and the time when igneous rocks were formed. The times found in this way are consistent with the relative time scale that geologists had defined independently for centuries before the technique appeared. With the appearance of radiometric dating, all fell into place. Equally successful was the *radiocarbon dating method*; with it, historians determined the age of civilizations and the age of human artefacts.* Many false beliefs were shattered. In some communities the shock is still not over, even though over hundred years have passed since these results became known. An overview of the isotopes used, together with the possible applications of radiometric dating, is given in Table 11.

Page 135
Ref. 144

The technique of radiometric dating has deeply impacted astronomy, geology, evolutionary biology, archaeology and history. (And it has reduced the number of violent believers.) Life times can usually be measured to within one or two per cent of accuracy, and they are known both experimentally and theoretically not to change over geological time scales. As a result, radiometric dating methods can be surprisingly precise. Can you imagine how one measure half-lives of thousands of millions of years to high precision?

Ref. 144
Challenge 118 s

A famous technique is the mentioned radiocarbon dating method. The beta decay of the radioactive carbon isotope ^{14}C has a decay time of 5730 a. This isotope is continually created in the atmosphere through the influence of cosmic rays. This happens through the reaction $^{14}\mathrm{N} + \mathrm{n} \rightarrow \mathrm{p} + {}^{14}\mathrm{C}$. As a result, the concentration of radiocarbon in air is relatively constant over time. Inside living plants, the metabolism thus (unknowingly) maintains the same concentration. In dead plants, the decay sets in. The decay time of a few thousand years is particularly useful to date historic material. The method, called radiocarbon dating, has been used to determine the age of mummies, the age of prehistoric tools and the age of religious relics. The original version of the technique measured the radiocarbon content through its radioactive decay and the scintillations it produced. A quality jump was achieved when accelerator mass spectroscopy became commonplace. It was not necessary any more to wait for decays: it is now possible to determine the ^{14}C content directly. As a result, only a tiny amount of carbon, as low as 0.2 mg, is necessary for a precise dating. This technique showed that numerous religious relics are forgeries, such as a cloth in Turin, and several of their wardens turned out to be crooks.

Researchers have even developed an additional method to date stones using radioactivity. Whenever an alpha ray is emitted, the emitting atom gets a recoil. If the atom is part of a crystal, the crystal is damaged by the recoil. The damage can be seen under the microscope. By counting the damaged regions it is possible to date the time at which rocks have been crystallized. In this way it has been possible to determine when material from volcanic eruptions has become rock.

* In 1960, the developer of the radiocarbon dating technique, Willard Libby, received the Nobel Prize for chemistry.

TABLE 11 Natural isotopes used in radiometric dating

Isotope	Decay product	Half-life	Method using it	Examples
^{147}Sm	^{143}Nd	106 Ga	samarium–neodymium method	rocks, lunar soil, meteorites
^{87}Rb	^{87}Sr	48.8 Ga	rubidium–strontium method	rocks, lunar soil, meteorites
^{187}Rh	^{187}Os	42 Ga	rhenium–osmium method	rocks, lunar soil, meteorites
^{176}Lu	^{176}Hf	38 Ga	lutetium–hafnium method	rocks, lunar soil, meteorites
^{232}Th	^{208}Pb	14 Ga	thorium–lead method, lead–lead method	rocks, lunar soil, meteorites
^{238}U	^{206}Pb	4.5 Ga	uranium–lead method, lead–lead method	rocks, lunar soil, meteorites
^{40}K	^{40}Ar	1.26 Ga	potassium–argon method, argon–argon method	rocks, lunar soil, meteorites
^{235}U	^{207}Pb	0.7 Ga	uranium–lead method, lead–lead method	rocks, lunar soil, meteorites
^{10}Be	^{10}B	1.52 Ma	cosmogenic radiometric dating	ice cores
^{26}Al	^{26}Mg	0.72 Ma	supernova debris dating	checking that nucleosynthesis still takes place in the galaxy
^{60}Fe	^{60}Ni	2.6 Ma	supernova debris dating	deep sea crust; lifetime updated in 2009 from the previously accepted 1.5 Ma
^{36}Cl	^{36}Ar	0.3 Ma	cosmogenic radiometric dating	ice cores
^{234}U	^{230}Th	248 ka	uranium–thorium method	corals, stalactites, bones, teeth
^{230}Th	^{226}Ra	75, 4 ka	thorium-radon method	plant dating
^{14}C	^{14}N	5730 a	radiocarbon method	wood, clothing, bones, organic material, wine
^{137}Cs	^{137}B	30 a	gamma-ray counting	dating food and wine after Chernobyl nuclear accident
^{210}Pb		22 a	gamma-ray counting	dating wine
^{3}H	^{3}He	12.3 a	gamma-ray counting	dating wine

FIGURE 69 The lava sea in the volcano Erta Ale in Ethiopia (© Marco Fulle)

With the advent of radiometric dating, for the first time it became possible to reliably date the age of rocks, to compare it with the age of meteorites and, when space travel became fashionable, with the age of the Moon. The result of the field of *radiometric dating* was beyond all estimates and expectations: the oldest rocks and the oldest meteorites studied independently using different dating methods, are 4570(10) million years old. But if the Earth is that old, why did the Earth not cool down in its core in the meantime?

Why is hell hot?

The lava seas and streams found in and around volcanoes are the origin of the images that many cultures ascribe to hell: fire and suffering. Because of the high temperature of lava, hell is inevitably depicted as a hot place. A striking example is the volcano Erta Ale, shown in Figure 69. But why is lava still hot, after 4570 million years?

A straightforward calculation shows that if the Earth had been a hot sphere in the beginning, it should have cooled down and solidified already long time ago. The Earth should be a solid object, like the moon: the Earth should not contain any lava and hell would not be hot.

Challenge 119 ny

The solution to the riddle is provided by radioactivity: the centre of the Earth contains an oven fuelled by radioactive potassium ^{40}K, radioactive uranium ^{235}U and ^{238}U and radioactive thorium ^{232}Th. The radioactivity of these elements, and to minor degree a few others, keeps the centre of the Earth glowing. More precise investigations, taking into account the decay times and material concentrations, show that this mechanism indeed explains the internal heat of the Earth. (By the way, the decay of radioactive potassium is the origin for the 1 % of argon found in the Earth's atmosphere.)

Ref. 145

Page 136

This brings up a challenge: why is the radioactivity of lava and of the Earth in general not dangerous to humans?

Challenge 120 s

Nuclei can form composites

Nuclei are highly unstable when they contain more than about 280 nucleons. Higher mass values inevitably decay into smaller fragments. But when the mass is above 10^{57} nucleons, nuclear composites are stable again: such systems are then called neutron stars. This is the most extreme example of pure nuclear matter found in nature. Neutron stars are left overs of (type II) supernova explosions. They do not run any fusion reactions any more, as other stars do; in first approximation neutron stars are simply large nuclei.

Neutron stars are made of degenerate matter. Their density of $10^{18}\,\mathrm{kg/m^3}$ is a few times that of a nucleus, as gravity compresses the star. This density value means that a tea spoon of such a star has a mass of several hundred million tons. Neutron stars are about 10 km in diameter. They are never much smaller, as such smaller stars are unstable. They are never much larger, because much larger neutron stars turn into black holes.

Nuclei have colours and shapes

In everyday life, the colour of objects is determined by the wavelength of light that is least absorbed, or if they shine, by the wavelength that is emitted. Also nuclei can absorb photons of suitably tuned energies and get into an excited state. In this case, the photon energy is converted into a higher energy of one or several of the nucleons whirling around inside the nucleus. Many radioactive nuclei also emit high energy photons, which then are called gamma rays, in the range of 1 keV (or 0.2 fJ) to about 20 MeV (or 3.3 pJ). The process is similar to the emission of light by electrons in atoms. From the energy, the number and the lifetime of the excited states – they range from 1 ps to 300 d – researchers can deduce how the nucleons move inside the nucleus.

The photon energies define the 'colour' of the nucleus. It can be used, like all colours, to distinguish nuclei from each other and to study their motion. In particular, the colour of the γ-rays emitted by excited nuclei can be used to determine the chemical composition of a piece of matter. Some of these transition lines are so narrow that they can been used to study the change due to the chemical environment of the nucleus, to measure their motion or to detect the gravitational Doppler effect.

The study of γ-rays also allows to determine the shape of nuclei. Many nuclei are spherical; but many are prolate or oblate ellipsoids. Ellipsoids are favoured if the reduction in average electrostatic repulsion is larger than the increase in surface energy. All nuclei – except the lightest ones such as helium, lithium and beryllium – have a constant mass density at their centre, given by about 0.17 fermions per fm^3, and a skin thickness of about 2.4 fm, where their density decreases. Nuclei are thus small clouds, as shown in Figure 70.

We know that molecules can be of extremely involved shape. In contrast, nuclei are mostly spheres, ellipsoids or small variations of these. The reason is the short range, or better, the fast spatial decay of nuclear interactions. To get interesting shapes like in molecules, one needs, apart from nearest neighbour interactions, also next neighbour interactions and next next neighbour interactions. The strong nuclear interaction is too short ranged to make this possible. Or does it? It might be that future studies will discover that some nuclei are of more unusual shape, such as smoothed pyramids. Some predictions have been made in this direction; however, the experiments have not been performed yet. Ref. 146

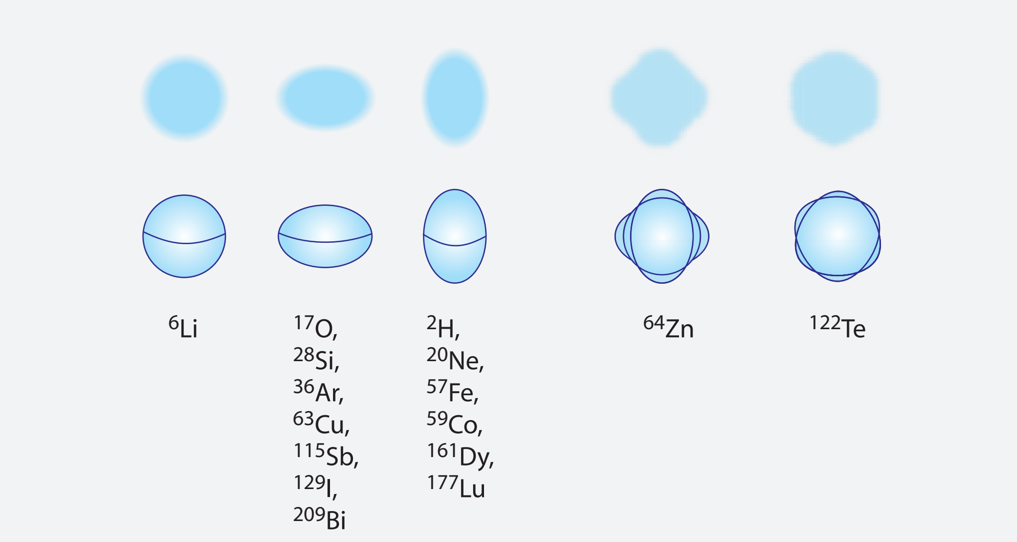


FIGURE 70 Various nuclear shapes – fixed: spherical, oblate, prolate (left) and oscillating (right), shown realistically as clouds (above) and simplified as geometric shapes (below)

The shape of nuclei does not have to be fixed; nuclei can also oscillate in shape. Such oscillations have been studied in great detail. The two simplest cases, the quadrupole and octupole oscillations, are shown in Figure 70. Obviously, nuclei can also rotate. Rapidly spinning nuclei, with a spin of up to $60\hbar$ and more, exist. They usually slow down step by step, emitting a photon and reducing their angular momentum at each step. Recently it was discovered that nuclei can also have bulges that rotate around a fixed core, a bit like tides rotate around the Earth. Ref. 147

The four types of motion in the nuclear domain

Nuclei are small because the nuclear interactions are short-ranged. Due to this short range, nuclear interactions play a role only in four types of motion: scattering, bound motion, decay and a combination of these three called *nuclear reactions*. The history of nuclear physics showed that the whole range of observed phenomena can be reduced to these four fundamental processes. In each motion type, the main interest is what happens at the start and at the end; the intermediate situations are less interesting. Nuclear interactions thus lack the complex types of motion which characterize everyday life. That is also the reason for the shortness of this chapter.

Scattering is performed in all accelerator experiments. Such experiments repeat for nuclei what we do when we look at an object. Seeing is a scattering process, as seeing is the detection of scattered light. Scattering of X-rays was used to see atoms for the first time; scattering of high energy alpha particles was used to discover and study the nucleus, and later the scattering of electrons with even higher energy was used to discover and study the components of the proton.

Bound motion is the motion of protons and neutrons inside nuclei or the motion of quarks inside mesons and baryons. Bound motion determines shape and shape changes of compounds.

Decay is obviously the basis of radioactivity. Decay can be due to the electromagnetic, the strong or the weak nuclear interaction. Decay allows to study the conserved quantities of nuclear interactions.

Nuclear reactions are combinations of scattering, decay and possibly bound motion. Nuclear reactions are for nuclei what the touching of objects is in everyday life. Touching an object we can take it apart, break it, solder two objects together, throw it away, and much more. The same can be done with nuclei. In particular, nuclear reactions are responsible for the burning of the Sun and the other stars; they also tell the history of the nuclei inside our bodies.

Quantum theory showed that all four types of nuclear motion can be described in the same way. Each type of motion is due to the exchange of virtual particles. For example, scattering due to charge repulsion is due to exchange of virtual photons, the bound motion inside nuclei due to the strong nuclear interaction is due to exchange of virtual gluons, beta decay is due to the exchange of virtual W bosons, and neutrino reactions are due to the exchange of virtual Z bosons. The rest of this chapter explains these mechanisms in more details.

Nuclei react

Page 99 The first man thought to have made transuranic elements, the Italian genius Enrico Fermi, received the Nobel Prize for the discovery. Shortly afterwards, Otto Hahn and his collaborators Lise Meitner and Fritz Strassmann showed that Fermi was wrong, and that his prize was based on a mistake. Fermi was allowed to keep his prize, the Nobel committee gave Hahn and Strassmann the Nobel Prize as well, and to make the matter unclear to everybody and to women physicists in particular, the prize was not given to Lise Meitner. (After her death, a new element was named after her.)

When protons or neutrons were shot into nuclei, they usually remained stuck inside them, and usually lead to the transformation of an element into a heavier one. After having done this with all elements, Fermi used uranium; he found that bombarding it with neutrons, a new element appeared, and concluded that he had created a transuranic element. Alas, Hahn and his collaborators found that the element formed was well-known: it was barium, a nucleus with less than half the mass of uranium. Instead of remaining stuck as in the previous 91 elements, the neutrons had split the uranium nucleus. Hahn, Meitner and Strassmann had observed reactions such as:

$$^{235}\mathrm{U} + \mathrm{n} \rightarrow {}^{143}\mathrm{Ba} + {}^{90}\mathrm{Kr} + 3n + 170\,\mathrm{MeV}\ . \qquad (50)$$

Meitner called the splitting process *nuclear fission*. The amount of energy liberated in fission is unusually large, millions of times larger than in chemical interactions. In addition, several neutrons are emitted; they can thus start a chain reaction. Later, and (of course) against the will of the team, the discovery would be used to make nuclear bombs.

Reactions and decays are transformations. In each transformation, already the Greek taught us to search, first of all, for conserved quantities. Besides the well-known cases of energy, momentum, electric charge and angular momentum conservation, the results of nuclear physics lead to several new conserved quantities. The behaviour is quite constrained. Quantum field theory implies that particles and antiparticles (commonly denoted by a bar) must behave in compatible ways. Both experiment and quantum field theory show for example that every reaction of the type

$$A + B \rightarrow C + D \quad (51)$$

implies that the reactions

$$A + \overline{C} \rightarrow \overline{B} + D \quad (52)$$

or

$$\overline{C} + \overline{D} \rightarrow \overline{A} + \overline{B} \quad (53)$$

or, if energy is sufficient,

$$A \rightarrow C + D + \overline{B} \,, \quad (54)$$

are also possible. Particles thus behave like conserved mathematical entities.

Experiments show that antineutrinos differ from neutrinos. In fact, all reactions confirm that the so-called *lepton number* is conserved in nature. The lepton number L is zero for nucleons or quarks, is 1 for the electron and the neutrino, and is -1 for the positron and the antineutrino.

In addition, all reactions conserve the so-called *baryon number*. The baryon number B is 1 for protons and neutrons (and 1/3 for quarks), and -1 for antiprotons and antineutrons (and thus $-1/3$ for antiquarks). So far, no process with baryon number violation has ever been observed. Baryon conservation is one reason for the danger of radioactivity, fission and fusion.

Bombs and nuclear reactors

Uranium fission is triggered by a neutron, liberates energy and produces several additional neutrons. Therefore, uranium fission can trigger a chain reaction which can lead to an explosion or a controlled generation of heat. Once upon a time, in the middle of the twentieth century, these processes were studied by quite a number of researchers. Most of them were interested in making weapons or in using nuclear energy, despite the high toll these activities place on the economy, on human health and on the environment.

Most stories around the development of nuclear weapons are absurd. The first such weapons were built during the second world war with the smartest physicists that could be found. Everything was ready, including the most complex physical models, several huge factories and an organization of incredible size. There was just one little problem: there was no uranium of sufficient quality. The mighty United States thus had to go around the world to shop for good uranium. They found it in the Belgian colony of Congo, in central Africa. In short, without the support of Belgium, which sold the Congolese uranium to the USA, there would have been no nuclear bomb, no early war end and no superpower status.

Congo paid a high price for this important status. It was ruled by a long chain of military dictators up to this day. But the highest price was paid by the countries that actually built nuclear weapons. Some went bankrupt, others remained underdeveloped; even the richest countries have amassed huge debts and a large underprivileged population. There is *no* exception. The price of nuclear weapons has also been that some regions of our planet became uninhabitable, such as numerous islands, deserts and marine environments. But it could have been worse. When the most violent physicist ever, Edward Teller, made his first calculations about the hydrogen bomb, he predicted that the bomb would set the atmosphere into fire. Nobel Prize winner Hans Bethe* corrected the mistake and showed that nothing of this sort would happen. Nevertheless, the military preferred to explode the hydrogen bomb in the Bikini atoll, the most distant place from their homeland they could find. Today it is even dangerous simply to fly over that island!

Ref. 148

It was then noticed that nuclear test explosions increased ambient radioactivity in the atmosphere all over the world. Of the produced radioactive elements, ^{3}H is absorbed by humans in drinking water, ^{14}C and ^{90}Sr through food, and ^{137}Cs in both ways. In the meantime, all countries have agreed to perform their nuclear tests underground.

But even peaceful nuclear reactors are dangerous. The reason was discovered in 1934 by Frédéric Joliot and his wife Irène, the daughter of Pierre and Marie Curie: *artificial radioactivity*. The Joliot–Curies discovered that materials irradiated by alpha rays become radioactive in turn. They found that alpha rays transformed aluminium into radioactive phosphorus:

$$ {}^{27}_{13}\mathrm{Al} + {}^{4}_{2}\alpha \rightarrow {}^{30}_{15}\mathrm{P} + {}^{30}_{15}\mathrm{n}\ . \qquad (55) $$

In fact, almost all materials become radioactive when irradiated with alpha particles, neutrons or gamma rays. As a result, radioactivity itself can only be contained with difficulty. After a time which depends on the material and the radiation, the box that contains radioactive material has itself become radioactive.

The dangers of natural and artificial radioactivity are the reason for the high costs of nuclear reactors. After about thirty years of operation, reactors have to be dismantled. The radioactive pieces have to be stored in specially chosen, inaccessible places, and at the same time the workers' health must not be put in danger. The world over, many dismantlings are now imminent. The companies performing the job sell the service at high price. All operate in a region not far from the border to criminal activity, and since radioactivity cannot be detected by the human senses, many crossed it. In fact, an important nuclear reactor is (usually) not dangerous to humans: the Sun. We explore it shortly.

Curiosities and challenges on radioactivity and nuclei

The SI units for radioactivity are now common around the world; in the old days, 1 Sievert, or 1 Sv, was called 100 rem or 'Röntgen equivalent man'; the SI unit for dose, 1 Gray, defined as 1 Gy = 1 J/kg, replaces what used to be called 100 rd or Rad. The SI unit for exposition, 1 C/kg, replaces the older unit 'Röntgen', with the relation 1 R =

* Hans Bethe (b. 1906 Strasbourg, d. 2005) was one of the great physicists of the twentieth century, even though he was head of the theory department that lead to the construction of the first atomic bombs. He worked on nuclear physics and astrophysics, helped Richard Feynman in developing quantum electrodynamics, and worked on solid state physics. When he got older and wiser, he became a strong advocate of arms control; he also was essential in persuading the world to stop atmospheric nuclear test explosions and saved many humans from cancer in doing so.

TABLE 12 Some radioactivity measurements

Material	Activity in Bq/kg
Air	*c.* 10^{-2}
Sea water	10^{1}
Human body	*c.* 10^{2}
Cow milk	max. 10^{3}
Pure ^{238}U metal	*c.* 10^{7}
Highly radioactive α emitters	$> 10^{7}$
Radiocarbon: ^{14}C (β emitter)	10^{8}
Highly radioactive β and γ emitters	$> 10^{9}$
Main nuclear fallout: ^{137}Cs, ^{90}Sr (α emitter)	$2 \cdot 10^{9}$
Polonium, one of the most radioactive materials (α)	10^{24}

$2.58 \cdot 10^{-4}$ C/kg.

* *

Not all γ-rays are due to radioactivity. In the year 2000, an Italian group discovered that thunderstorms also emit γ-rays, of energies up to 10 MeV. The mechanisms are still being investigated. Ref. 149

* *

Chain reactions are quite common in nature. *Fire* is a chemical chain reaction, as are exploding fireworks. In both cases, material needs heat to burn; this heat is supplied by a neighbouring region that is already burning.

* *

Radioactivity can be extremely dangerous to humans. The best example is plutonium. Only 1 μg of this alpha emitter inside the human body are sufficient to cause lung cancer. Polonium 210 is also present in tobacco leaves that were grown with artificial fertilizers. In addition, tobacco leaves filter radioactive substances from the air. These radioactive substances in tobacco are one of the reasons that smoking produces cancer.

* *

Why is nuclear power a dangerous endeavour? The best argument is Lake Karachay near Mayak, in the Urals in Russia. In less than a decade, the nuclear plants of the region have transformed it into the most radioactive place on Earth. Walking on the shore of the lake for an hour leads to death on the shore. The radioactive material in the lake was distributed over large areas in several catastrophic explosions in the 1950s and 1960s, leading to widespread death and illness. Several of these accidents were comparable to the Chernobyl accident of 1986. The lake is now covered in concrete.

* *

TABLE 13 Human exposure to radioactivity and the corresponding doses

Exposure	Dose
Daily human exposure:	
Average exposure to cosmic radiation in Europe	
at sea level	0.3 mSv/a
at a height of 3 km	1.2 mSv/a
Average (and maximum) exposure from the soil, not counting radon effects	0.4 mSv/a (2 mSv/a)
Average (and maximum) inhalation of radon	1 mSv/a (100 mSv/a)
Average exposure due to internal radionuclides	0.3 mSv/a
natural content of ^{40}K in human muscles	10^{-4} Gy and 4500 Bq
natural content of Ra in human bones	$2 \cdot 10^{-5}$ Gy and 4000 Bq
natural content of ^{14}C in humans	10^{-5} Gy
Total average (and maximum) human exposure	2 mSv/a (100 mSv/a)
Common situations:	
Dental X-ray	*c.* 10 mSv equivalent dose
Lung X-ray	*c.* 0.5 mSv equivalent dose
Short one hour flight (see www.gsf.de/epcard)	*c.* 1 μSv
Transatlantic flight	*c.* 0.04 mSv
Maximum allowed dose at work	30 mSv/a
Deadly exposures:	
Ionization	0.05 C/kg can be deadly
Dose	100 Gy=100 J/kg is deadly in 1 to 3 days
Equivalent dose	more than 3 Sv/a leads to death

All lead is slightly radioactive, because it contains the ^{210}Pb isotope, a beta emitter. This lead isotope is produced by the uranium and thorium contained in the rock from where the lead is extracted. For sensitive experiments, such as for neutrino experiments, one needs radioactivity shields. The best shield material is lead, but obviously it has to be low radioactivity lead. Since the isotope ^{210}Pb has a half-life of 22 years, one way to do it is to use *old* lead. In a precision neutrino experiment in the Gran Sasso in Italy, the research team uses lead mined during Roman times, thus 2000 years old, in order to reduce spurious signals.

* *

Not all nuclear reactors are human made. Natural reactors have been predicted in 1956 by Paul Kuroda. In 1972 the first example was found. In Oklo, in the African country of Gabon, there is a now famous geological formation where uranium is so common that two thousand million years ago a natural nuclear reactor has formed spontaneously – albeit a small one, with an estimated power generation of 100 kW. It has been burning for over 150 000 years, during the time when the uranium 235 percentage was 3% or more, as required for chain reaction. (Nowadays, the uranium 235 content on Earth is 0.7%.) The water of a nearby river was periodically heated to steam during an estimated 30 minutes; then the reactor cooled down again for an estimated 2.5 hours, since water is necessary to moderate the neutrons and sustain the chain reaction. The system has been studied in great detail, from its geological history up to the statements it makes about the constancy of the 'laws' of nature. The studies showed that 2000 million years ago the mechanisms were the same as those used today.

* *

Nuclear reactors exist in many sizes. The largest are used in power plants and can produce over 1000 MW in electrical power; the smallest are used in satellites, and usually produce around 10 kW for many years.

* *

High energy radiation is dangerous to humans. In the 1950s, when nuclear tests were still made above ground by the large armies in the world, the generals overruled the orders of the medical doctors. They positioned many soldiers nearby to watch the explosion, and worse, even ordered them to walk to the explosion site as soon as possible after the explosion. One does not need to comment on the orders of these generals. Several of these unlucky soldiers made a strange observation: during the flash of the explosion, Challenge 121 ny they were able to see the bones in their own hand and arms. How can this be?

* *

Nuclear bombs are terrible weapons. To experience their violence but also the criminal Ref. 151 actions of many military people during the tests, have a look at the pictures of explosions. In the 1950 and 60s, nuclear tests were performed by generals who refused to listen to doctors and scientists. Generals ordered to explode these weapons in the air, making the complete atmosphere of the world radioactive, hurting all mankind in doing so; worse, they even obliged soldiers to visit the radioactive explosion site a few minutes after the explosion, thus doing their best to let their own soldiers die from cancer and leukaemia. Generals are people to avoid.

* *

Several methods to date wine are used, and more are in development. A few are given in Table 11.

* *

In 1958, six nuclear bombs were made to explode in the stratosphere by a vast group of criminals. A competing criminal group performed similar experiments in 1961, followed by even more explosions by both groups in 1962. (For reports and films, see en.wikipedia.org/wiki/High_altitude_nuclear_explosion.) As a result of most of these explosions, an *artificial aurora* was triggered the night following each of them. In addition, the electromagnetic pulse from the blasts destroyed satellites, destroyed electronics on Earth, disturbed radio communications, injured people on the surface of the Earth, caused problems with power plants, and distributed large amounts of radioactive material over the Earth – during 14 years following the blasts. The van Allen radiation belts around the Earth were strongly affected; it is expected that the lower van Allen belt will recover from the blasts only in a few hundred years. Fortunately for the human race, after 1962, this activity was stopped by international treaties.

* *

Selected radioactive decay times can be changed by external influence. Electron capture, as observed in beryllium-7, is one of the rare examples were the decay time can change, by up to 1.5%, depending on the chemical environment. The decay time for the same isotope has also been found to change by a fraction of a per cent under pressures of 27 GPa. On the other hand, these effects are predicted (and measured) to be negligible for nuclei of larger mass.

* *

The non-radioactive isotopes ^2H and ^{18}O can be used for measuring energy production in humans in an easy way. Give a person a glass of doubly labelled water to drink and collect his urine samples for a few weeks. Using a mass spectrometer one can determine his energy consumption. Why? Doubly labelled water $^2H_2{}^{18}O$ is processed by the body in three main ways. The oxygen isotope is expired as $C^{18}O_2$ or eliminated as $H_2{}^{18}O$; the hydrogen isotope is eliminated as ^2H_2O. Measurements on the urine allow to determine carbon dioxide production, therefore to determine how much has food been metabolized, and thus to determine energy production.

Human energy consumption is usually given in joule per day. Measurements showed that high altitude climbers with 20 000 kJ/d and bicycle riders with up to 30 000 kJ/d are the most extreme sportsmen. Average humans produce 6 000 kJ/d.

* *

Many nuclei oscillate in shape. The calculation of these shape oscillations is a research subject in itself. For example, when a spherical nucleus oscillates, it can do so in three mutually orthogonal axes. A spherical nucleus, when oscillating at small amplitudes, thus behaves like a three-dimensional harmonic oscillator. Interestingly, the symmetry of the three-dimensional harmonic oscillator is SU(3), the same symmetry that characterizes the strong nuclear interaction. However, the two symmetries are unrelated – at least following present knowledge. A relation might appear in the future, though.

Vol. VI, page 242

* *

Magnetic resonance machines pose no danger; but they do have some effects, as Peter Mansfield, one of the inventors of the technique, explains. The first effect is due to the conductivity of blood. When blood in the aorta passes through a magnetic field, a voltage is induced. This effect has been measured and it might interfere with cardiac functioning at 7 T; usual machines have 1.5 T and pose no risk. The second effect is due to the switching of the magnetic field. Some people sense the switching in the thorax and in the shoulders. Not much is known about the details of peripheral nerve stimulation yet.

Ref. 150

Summary on nuclei

Nuclei are composed of protons and neutrons. Their radius is between one and a few femtometres and they rotate. Their rotation allows to produce magnetic resonance images. Nuclei can be spherical or ellipsoidal, they can be excited to higher energy states, and they can oscillate in shape. Nuclei have a colour that is determined by their spectra. Nuclei can decay, can break up and can react among each other. The last property is the reason that we exist, as we will show now.

CHAPTER 6

THE SUN, THE STARS AND THE BIRTH OF MATTER

> “Lernen ist Vorfreude auf sich selbst.*”
> Peter Sloterdijk

NUCLEAR physics is the most violent part of physics. But despite this bad image, nuclear physics has also something to offer that is deeply fascinating: exploring nuclei, we can understand the Sun, the stars and the early universe. Ref. 152

Nuclei consist of protons and neutrons. Since protons are positively charged, they repel each other. Inside nuclei, protons must be bound by a force strong enough to keep them together against their electromagnetic repulsion. This is the *strong nuclear interaction*; it is needed to avoid that nuclei explode. The strong nuclear interaction is the strongest of the four interactions; nevertheless, we do not experience it in everyday life, because its range is limited to distances of a few femtometres, or a few diameters of a proton. Despite this limitation, the strong interaction tells a good story about the flesh and blood we are made of.

THE SUN

The Sun emits 385 YW of light. Where does this energy come from? If it came by burning coal, the Sun would stop burning after a few thousands of years. When radioactivity was discovered, researchers tested the possibility that this process was at the heart of the Sun's shining. However, even though radioactivity – or the process of fission that was discovered later – can produce more energy than chemical burning, the composition of the Sun – mostly hydrogen and helium – makes this impossible. The origin of the energy in the Sun was settled in 1939 by Hans Bethe: the Sun burns by hydrogen *fusion*. Fusion is the composition of a large nucleus from smaller ones. In the Sun, the fusion reaction Ref. 153

$$4\,{}^{1}\mathrm{H} \rightarrow {}^{4}\mathrm{He} + 2\,e^{+} + 2\,\nu + 4.4\,\mathrm{pJ} \qquad (56)$$

is the result of a continuous cycle of three separate nuclear reactions:

$$\begin{aligned} {}^{1}\mathrm{H} + {}^{1}\mathrm{H} &\rightarrow {}^{2}\mathrm{H} + e^{+} + \nu \quad \text{(a weak nuclear reaction)} \\ {}^{2}\mathrm{H} + {}^{1}\mathrm{H} &\rightarrow {}^{3}\mathrm{He} + \gamma \quad \text{(a strong nuclear reaction)} \\ {}^{3}\mathrm{He} + {}^{3}\mathrm{He} &\rightarrow {}^{4}\mathrm{He} + 2\,{}^{1}\mathrm{H} + \gamma\,. \end{aligned} \qquad (57)$$

* ‘Learning is anticipated joy about yourself.’

In total, four protons are thus fused to one helium nucleus; if we include the electrons, four hydrogen atoms are fused to one helium atom with the emission of neutrinos and light with a total energy of 4.4 pJ (26.7 MeV). Most of the energy is emitted as light; around 10 % is carried away by neutrinos. The first of the three reaction of equation 57 is due to the weak nuclear interaction; this avoids that it happens too rapidly and ensures that the Sun will shine still for some time. Indeed, in the Sun, with a luminosity of
Ref. 154 385 YW, there are thus about 10^{38} fusions per second. This allows to deduce that the Sun will last another handful of Ga (Gigayears) before it runs out of fuel.

The fusion reaction (57) takes place in the centre of the Sun. The energy carried away by the photons arrives at the Sun's surface about two hundred thousand years later; this delay is due to the repeated scattering of the photon by the constituents inside the Sun. After two-hundred thousand years, the photons take another 8.3 minutes to reach the Earth and to sustain the life of all plants and animals.

Why do the stars shine?

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

All stars shine because of fusion. When two light nuclei are fused to a heavier one, some energy is set free, as the average nucleon is bound more strongly. This energy gain is possible until the nuclei of iron ^{56}Fe are made. For nuclei beyond this nucleus, the binding energies per nucleon then decrease again; thus fusion is not energetically possible.* The heavier nuclei found on Earth and across the universe were formed through neutron capture.

The different stars observed in the sky** can be distinguished by the type of fusion nuclear reaction that dominates. Most stars, in particular young or light stars, run hydrogen fusion. In fact, there are at least two main types of hydrogen fusion: the direct hydrogen–hydrogen (p-p) cycle and the CNO cycle(s).

Page 148 The hydrogen cycle was described above and can be summarized as

$$4\,{}^{1}\mathrm{H} \rightarrow {}^{4}\mathrm{He} + 2\,e^{+} + 2\,\nu + 4.4\,\mathrm{pJ}\ . \qquad (58)$$

This simple description does not fully purvey the fascination of the process. On average, protons in the Sun's centre move with 600 km/s. Only if they hit each other precisely head-on can a nuclear reaction occur; in all other cases, the electrostatic repulsion between the protons keeps them apart. For an average proton, a head-on collision happens once every 7 thousand million years. Nevertheless, there are so many proton collisions in the Sun that every second four million tons of hydrogen are burned to helium.

Fortunately for us, the photons generated in the Sun's centre are 'slowed' down by the outer parts of the Sun. In this process, gamma photons are progressively converted to visible photons. As a result, the sunlight of today was in fact generated at the time of the

* Thus fission becomes interesting as energy source for heavy nuclei.
** To find out which stars are in the sky above you at present, see the www.surveyor.in-berlin.de/himmel website.

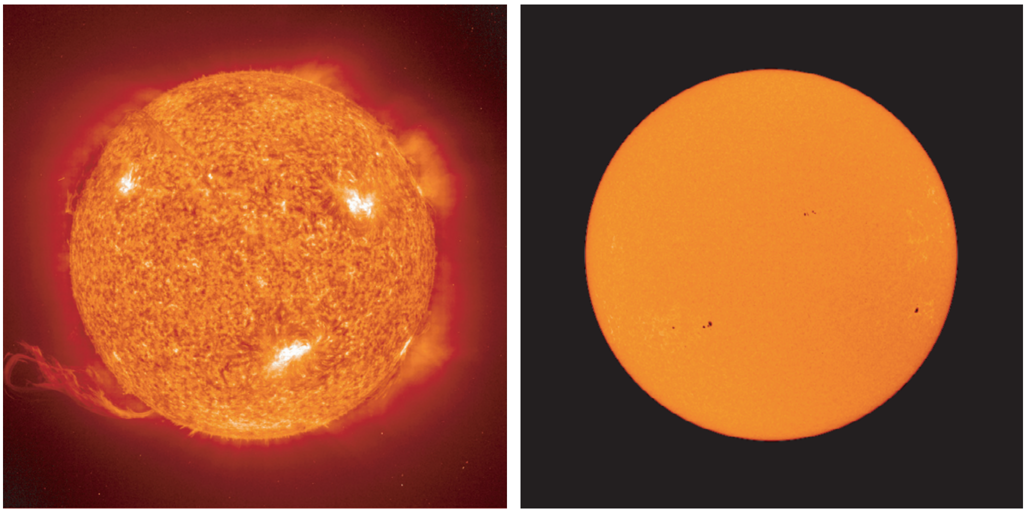


FIGURE 71 Photographs of the Sun at wavelengths of 30.4 nm (in the extreme ultraviolet, left) and around 677 nm (visible light, right, at a different date), by the SOHO mission (ESA and NASA)

Ref. 155 Neandertalers: a typical estimate is about 200 000 years ago. In other words, the effective speed of light right at the centre of the Sun is estimated to be around 10 km/year.

If a star has heavier elements inside it, the hydrogen fusion uses these elements as catalysts. This happens through the so-called CNO cycle, which runs as

$$\begin{aligned}
{}^{12}\mathrm{C} + {}^{1}\mathrm{H} &\to {}^{13}\mathrm{N} + \gamma \\
{}^{13}\mathrm{N} &\to {}^{13}\mathrm{C} + \mathrm{e}^{+} + \nu \\
{}^{13}\mathrm{C} + {}^{1}\mathrm{H} &\to {}^{14}\mathrm{N} + \gamma \\
{}^{14}\mathrm{N} + {}^{1}\mathrm{H} &\to {}^{15}\mathrm{O} + \gamma \\
{}^{15}\mathrm{O} &\to {}^{15}\mathrm{N} + \mathrm{e}^{+} + \nu \\
{}^{15}\mathrm{N} + {}^{1}\mathrm{H} &\to {}^{12}\mathrm{C} + {}^{4}\mathrm{He}
\end{aligned} \qquad (59)$$

The end result of the cycle is the same as that of the hydrogen cycle, both in nuclei and in energy. The CNO cycle is faster than hydrogen fusion, but requires higher temperatures, as the protons must overcome a higher energy barrier before reacting with carbon or nitrogen than when they react with another proton. (Why?) Challenge 122 s Due to the comparatively low temperature of a few tens of million kelvin inside the Sun, the CNO cycle is less important than the hydrogen cycle. (This is also the case for the other CNO cycles that exist.)

These studies also explain why the Sun does not collapse. The Sun is a ball of hot gas, and the high temperature of its constituents prevents their concentration into a small volume. For some stars, the radiation pressure of the emitted photons prevents collapse; for others it is the Pauli pressure; for the Sun, like for the majority of stars, it is the usual thermal motion of the gas.

The nuclear reaction rates at the interior of a star are extremely sensitive to temperature T. The carbon cycle reaction rate is proportional to between T^{13} for hot massive O stars and T^{20} for stars like the Sun. In red giants and supergiants, the triple alpha reaction rate is proportional to T^{40}; these strong dependencies imply that stars usually shine with constancy over times of thousands and millions of years, since any change in temperature would be damped by a very efficient feedback mechanism. (Of course, there are exceptions: variable stars get brighter and darker with periods of a few days; the Sun shows small oscillations in the minute range; and some stars change in brightness every few years.)

How can the Sun's surface have a temperature of 6000 K, whereas the corona around it, the thin gas emanating from the Sun, reaches two million Kelvin? In the latter part of the twentieth century it was shown, using satellites, that the magnetic field of the Sun is the cause; through the violent flows in the Sun's matter, magnetic energy is transferred to the corona in those places were flux tubes form knots, above the bright spots in the left of Figure 71 or above the dark spots in the right photograph. As a result, the particles of the corona are accelerated and heat the whole corona.

When the Sun erupts, as shown in the lower left corner in Figure 71, matter is ejected far into space. When this matter reaches the Earth,* after being diluted by the journey, it affects the environment. Solar storms can deplete the higher atmosphere and can thus possibly trigger usual Earth storms. Other effects of the Sun are the formation of auroras and the loss of orientation of birds during their migration; this happens during exceptionally strong solar storms, as the magnetic field of the Earth is disturbed in these situations. The most famous effect of a solar storm was the loss of electricity in large parts of Canada in March of 1989. The flow of charged solar particles triggered large induced currents in the power lines, blew fuses and destroyed parts of the network, shutting down the power system. Millions of Canadians had no electricity, and in the most remote places it took two weeks to restore the electricity supply. Due to the coldness of the winter and a train accident resulting from the power loss, over 80 people died. In the meantime the network has been redesigned to withstand such events.

The proton cycle and the CNO cycles are not the only options. Heavier and older stars than the Sun can also shine through other fusion reactions. In particular, when hydrogen is consumed, such stars run *helium burning*:

$$3\,{}^{4}\mathrm{He} \rightarrow {}^{12}\mathrm{C}\,. \tag{60}$$

This fusion reaction is of low probability, since it depends on three particles being at the same point in space at the same time. In addition, small amounts of carbon disappear rapidly via the reaction $\alpha + {}^{12}\mathrm{C} \rightarrow {}^{16}\mathrm{O}$. Nevertheless, since ${}^{8}\mathrm{Be}$ is unstable, the reaction with 3 alpha particles is the only way for the universe to produce carbon. All these negative odds are countered only by one feature: carbon has an excited state at 7.65 MeV, which is 0.3 MeV above the sum of the alpha particle masses; the excited state resonantly enhances the low probability of the three particle reaction. Only in this way the universe is able to produce the atoms necessary for pigs, apes and people. The prediction of this resonance by Fred Hoyle is one of the few predictions in physics that used the simple experimental observation that humans exist. The story has lead to a huge outflow of metaphysical speculations, most of which are unworthy of being even mentioned. Page 244

* It might even be that the planets affect the solar wind; the issue is not settled and is still under study.

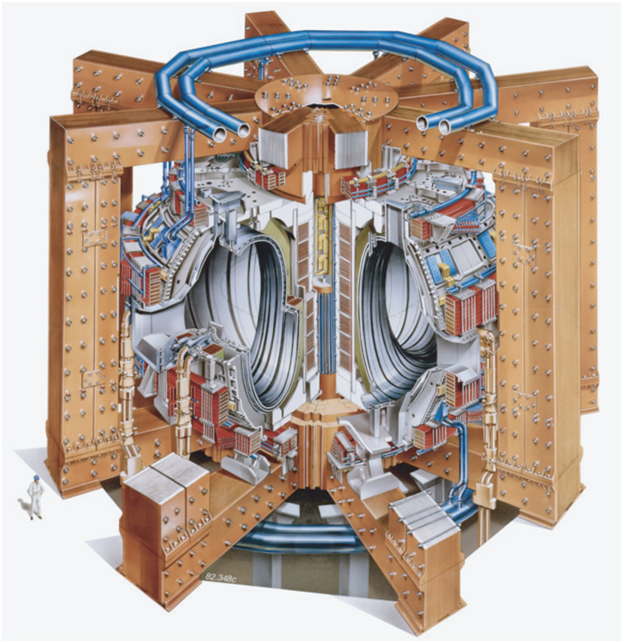


FIGURE 72 A simplified drawing of the Joint European Torus in operation at Culham, showing the large toroidal chamber and the magnets for the plasma confinement (© EFDA-JET)

Why are fusion reactors not common yet?

Across the world, for over 50 years, a large number of physicists and engineers have tried to build fusion reactors. Fusion reactors try to copy the mechanism of energy release used by the Sun. The first machine that realized macroscopic energy production was the Joint European Torus* (JET for short) located in Culham in the United Kingdom.

Ref. 156 The idea of JET is to produce an extremely hot plasma that is as dense as possible. At high enough temperature and density, fusion takes place; the energy is released as a particle flux that is transformed (like in a fission reactor) into heat and then into electricity. To achieve ignition, JET used the fusion between deuterium and tritium, because this reaction has the largest cross section and energy gain:

$$\mathrm{D} + \mathrm{T} \rightarrow \mathrm{He}^{4} + \mathrm{n} + 17.6\,\mathrm{MeV}\ . \tag{61}$$

Because tritium is radioactive, most research experiments are performed with the much less efficient deuterium–deuterium reactions, which have a lower cross section and a lower energy gain:

$$\begin{aligned} \mathrm{D} + \mathrm{D} &\rightarrow \mathrm{T} + \mathrm{H} + 4\,\mathrm{MeV} \\ \mathrm{D} + \mathrm{D} &\rightarrow \mathrm{He}^{3} + \mathrm{n} + 3.3\,\mathrm{MeV}\ . \end{aligned} \tag{62}$$

Fusion takes place when deuterium and tritium (or deuterium) collide at high energy. The high energy is necessary to overcome the electrostatic repulsion of the nuclei. In

* See www.jet.edfa.org.

other words, the material has to be hot. To release energy from deuterium and tritium, one therefore first needs energy to heat it up. This is akin to the ignition of wood: in order to use wood as a fuel, one first has to heat it with a match.

Following the so-called Lawson criterion, published in 1957 by the English engineer John Lawson, (but already known to Russian researchers) a fusion reaction releases energy only if the triple product of density n, reaction (or containment) time τ and temperature T exceeds a certain value. Nowadays this criterion is written as

Ref. 157

$$n\tau T > 3 \cdot 10^{28}\ \mathrm{s\,K/m^3}\ . \qquad (63)$$

In order to realize the Lawson criterion, JET uses temperatures of 100 to 200 MK, particle densities of 2 to $3 \cdot 10^{20}\ \mathrm{m^{-3}}$, and confinement times of 1 s. The temperature is much higher than the 20 MK at the centre of the Sun, because the densities and the confinement times are lower for JET.

Matter at these temperatures is in form of plasma: nuclei and electrons are completely separated. Obviously, it is impossible to pour a plasma at 100 MK into a container: the walls would instantaneously evaporate. The only option is to make the plasma float in a vacuum, and to avoid that the plasma touches the container wall. The main challenge of fusion research in the past has been to find a way to keep a hot gas mixture of deuterium and tritium suspended in a chamber so that the gas never touches the chamber walls. The best way is to suspend the gas using a magnetic field. This works because in the fusion plasma, charges are separated, so that they react to magnetic fields. The most successful geometric arrangement was invented by the famous Russian physicists Igor Tamm and Andrei Sakharov: the *tokamak*. Of the numerous tokamaks around the world, JET is the largest and most successful. Its concrete realization is shown in Figure 72. JET manages to keep the plasma from touching the walls for about a second; then the situation becomes unstable: the plasma touches the wall and is absorbed there. After such a *disruption*, the cycle consisting of gas injection, plasma heating and fusion has to be restarted. As mentioned, JET has already achieved ignition, that is the state were more energy is released than is added for plasma heating. However, so far, no sustained commercial energy production is planned or possible, because JET has no attached electrical power generator.

The successor project, *ITER*, an international tokamak built with European, Japanese, US-American and Russian funding, aims to pave the way for commercial energy generation. Its linear reactor size will be twice that of JET; more importantly, ITER plans to achieve 30 s containment time. ITER will use superconducting magnets, so that it will have extremely cold matter at 4 K only a few metres from extremely hot matter at 100 MK. In other words, ITER will be a high point of engineering. The facility will be located in Cadarache in France and is planned to start operation in the year 2016.

Like many large projects, fusion started with a dream: scientists spread the idea that fusion energy is safe, clean and inexhaustible. These three statements are still found on every fusion website across the world. In particular, it is stated that fusion reactors are not dangerous, produce much lower radioactive contamination than fission reactors, and use water as basic fuel. 'Solar fusion energy would be as clean, safe and limitless as the Sun.' In reality, the only reason that we do not feel the radioactivity of the Sun is that we are far away from it. Fusion reactors, like the Sun, are highly radioactive. The management of radioactive fusion reactors is much more complex than the management of radioactive fission reactors.

Fusion fuels are almost inexhaustible: deuterium is extracted from water and the tritium – a short-lived radioactive element not found in nature in large quantities – is produced from lithium. The lithium must be enriched, but since material is not radioactive, this is not problematic. However, the production of tritium from lithium is a dirty process that produces large amounts of radioactivity. Fusion energy is thus inexhaustible, but not safe and clean.

In short, of all technical projects ever started by mankind, fusion is by far the most challenging and ambitious. Whether fusion will ever be successful – or whether it ever should be successful – is another issue.

Where do our atoms come from?

People consist of electrons and various nuclei. Electrons, hydrogen and helium nuclei Page 219 are formed during the big bang. All other nuclei are formed in stars. Young stars run Ref. 158 hydrogen burning or helium burning; heavier and older stars run neon-burning or even silicon-burning. These latter processes require high temperatures and pressures, which are found only in stars with a mass at least eight times that of the Sun. However, all fusion processes are limited by photodissociation and will not lead to nuclei heavier than ^{56}Fe.

Heavier nuclei can only be made by neutron capture. There are two main processes; the s-process (for 'slow') runs inside stars, and gradually builds up heavy elements until the most heavy, lead, from neutron flying around. The rapid r-process occurs in stellar explosions. Many stars die this violent death. Such an explosion has two main effects: on one hand it distributes most of the matter of the star, such as carbon, nitrogen or oxygen, into space in the form of neutral atoms. On the other hand, new elements are synthesized during the explosion. The abundances of the elements in the solar system can be precisely measured. These several hundred data points correspond exactly with what is expected from the material ejected by a (type II) supernova explosion. In other words, the solar system formed from the remnants of a supernova, as did, somewhat later, life on Earth.* We all are recycled stardust.

Curiosities about the stars

It is still not clear whether the radiation of the Sun is constant over long time scales. There is an 11 year periodicity, the famous solar cycle, but the long term trend is still unknown. Precise measurements cover only the years from 1978 onwards, which makes only about 3 cycles. A possible variation of the solar constant might have important consequences for climate research; however, the issue is still open.

* *

The sun is not a static object. An impressive way to experience the violent processes it contains is to watch the film shown in Figure 73, showing the evolution of a solar flare.

* *

* By chance, the composition ratios between carbon, nitrogen and oxygen inside the Sun are the same as inside the human body.

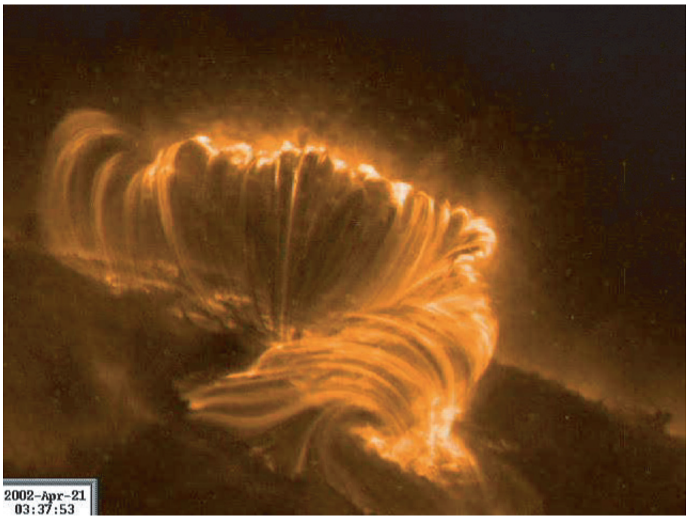


FIGURE 73 The evolution of a solar flare, as observed by the TRACE satellite (mpg film courtesy NASA)

In the 1960s and 70s, it was discovered that the Sun pulsates with a frequency of 5 minutes. The amplitude is small, only 3 kilometres out of 1.4 million; nevertheless, it is measurable. In the meantime, helioseismologists have discovered numerous additional oscillations of the Sun, and in 1993, even on other stars. Such oscillations allow to study what is happening inside stars, even separately in each of the layers they consist of.

* *

Some stars shine like a police siren: their luminosity increases and decreases regularly. Such stars, called *Cepheids*, are important because their period depends on their average (absolute) brightness. Measuring their period and their brightness on Earth thus allows astronomers to determine their distance.

* *

The first human-made hydrogen bomb explosion took place the Bikini atoll. But nature is much better at this. The most powerful nuclear explosions take place on the surface of neutron stars in X-ray binaries. The matter falling into such a neutron star from the companion star, mostly hydrogen, will heat up until the temperature allows fusion. The resulting explosions can be observed in telescopes as light or X-ray flashes of about 10 s duration; the explosions are millions of times more powerful that those of human-made hydrogen bombs.

* *

Nucleosynthesis is mainly regulated by the strong interaction. However, if the electromagnetic interaction would be much stronger or much weaker, stars would either produce too little oxygen or too little carbon, and we would not exist This famous argument is due to Fred Hoyle. Can you fill in the details?

Challenge 123 d

Summary on stars

Stars and the Sun burn because of nuclear fusion. When stars at the end of their lifetime explode, they distribute nuclei around them. In the distant past, such nuclei recollected because of gravity and formed the Sun, the Earth and humans.

Chapter 7

THE STRONG INTERACTION

Both radioactivity and medical images show that nuclei are composed. But quantum theory predicts more: also protons and neutrons must be composed. There are two reasons: first, nucleons have a finite size and second, their magnetic moments do not match the value predicted for point particles.

The prediction of components inside the protons was confirmed in the late 1960s
Ref. 159 when Kendall, Friedman and Taylor shot high energy electrons into hydrogen atoms. They found that a proton contains *three* constituents with spin 1/2, which they called *partons*. The experiment was able to 'see' the constituents through large angle scattering of electrons, in the same way that we see objects through large angle scattering of photons. These constituents correspond in number and (most) properties to the so-called *quarks*
Ref. 160 predicted in 1964 by George Zweig and also by Murray Gell-Mann.*

The feeble side of the strong interaction

It turns out that the interaction keeping the protons together in a nucleus, which was first described by Yukawa Hideki,** is only a feeble shadow of the interaction that keeps quarks together in a proton. Both are called by the same name. The two cases correspond somewhat to the two cases of electromagnetism found in atomic matter. Neon atoms

* The physicist George Zweig (b. 1937 Moscow , d.) proposed the quark idea – he called them *aces* – in 1963, with more clarity than Gell-Mann. Zweig stressed the reality of aces, whereas Gell-Mann, in the beginning, did not believe in the existence of quarks. Zweig later moved on to a more difficult field: neurobiology.

Murray Gell-Mann (b. 1929 New York, d.) received the Nobel Prize for physics in 1969. He is the originator of the term 'quark'. The term has two origins: officially, it is said to be taken from *Finnegans Wake*, a novel by James Joyce; in reality, Gell-Mann took it from a Yiddish and German term meaning 'lean soft
Ref. 161 cheese' and used figuratively in those languages to mean 'silly idea'.

Gell-Mann was the central figure of particle physics in the 20th century; he introduced the concept of strangeness, the renormalization group, the flavour SU(3) symmetry and quantum chromodynamics. A disturbing story is that he took the idea, the data, the knowledge, the concepts and even the name of the V−A theory of the weak interaction from the bright physics student George Sudarshan and published it, together with Richard Feynman, as his own. The wrong attribution is still found in many textbooks.

Gell-Mann is also known for his constant battle with Feynman about who deserved to be called the most arrogant physicist of their university. A famous anecdote is the following. Newton's once used a common saying of his time in a letter to Hooke: 'If I have seen further than you and Descartes, it is by standing upon the shoulders of giants.' Gell-Mann is known for saying: 'If I have seen further than others, it is because I am surrounded by dwarfs.'

** Yukawa Hideki (1907–1981), important Japanese physicist specialized in nuclear and particle physics. He founded the journal Progress of Theoretical Physics and together with his class mate Tomonaga he was an example to many scientists in Japan. He received the 1949 Nobel Prize for physics for this theory of mesons.

TABLE 14 The properties of the nucleons: proton and neutron (source: pdg.web.cern.ch)

Property	Proton	Neutron
Mass	$1.672\,621\,637(83) \cdot 10^{-27}$ kg	$1.674\,927\,29(28) \cdot 10^{-27}$ kg
	$0.150\,327\,7359(75)$ nJ	$0.150\,534\,9505(75)$ nJ
	$938, 272\,013(23)$ MeV	$939, 565\,346(23)$ MeV
	$1.007\,276\,466\,77(10)$ u	$1.008\,664\,915\,97(43)$ u
	$1836.152\,6675(39) \cdot m_e$	$1838.683\,6605(11) \cdot m_e$
Spin	1/2	1/2
P parity	+1	+1
Electric charge	1 e	0
Charge radius	0.88(1) fm	$0.12(1)\ \mathrm{fm}^2$
Electric dipole moment	$< 5.4 \cdot 10^{-27}$ e · m	$< 2.9 \cdot 10^{-28}$ e · m
Electric polarizability	$1.20(6) \cdot 10^{-3}\ \mathrm{fm}^3$	$1.16(15) \cdot 10^{-3}\ \mathrm{fm}^3$
Magnetic moment	$1.410\,606\,662(37) \cdot 10^{-26}$ J/T	$-0.966\,236\,41(23) \cdot 10^{-26}$ J/T
g-factor	5.585 694 701(56)	-3.826 0854(10)
	$2.792\,847\,351\,(28) \cdot \mu_N$	$-1.913\,0427(5) \cdot \mu_N$
Gyromagnetic ratio	0.267 522 205(23) 1/nsT	
Magnetic polarizability	$1.9(5) \cdot 10^{-4}\ \mathrm{fm}^3$	$3.7(20) \cdot 10^{-4}\ \mathrm{fm}^3$
Mean life (free particle)	$> 2 \cdot 10^{29}$ a	885.7(8) s
Shape (quadrupole moment)	oblate	oblate
Excited states	more than ten	more than ten

show the two cases most clearly: the strongest aspect of electromagnetism is responsible for the attraction of the electrons to the neon nuclei; its feeble 'shadow', the Van-der-Waals interaction, is responsible for the attraction of neon atoms in liquid neon and for processes like evaporation. Both attractions are electromagnetic, but the strengths differ markedly. Similarly, the strongest aspect of the strong interaction leads to the formation of the proton and the neutron through quark binding; the feeble aspect leads to the formation of nuclei and to alpha decay. Obviously, most information can be gathered by studying the strongest aspect.

Bound motion, the particle zoo and the quark model

Physicists are simple people. To understand the constituents of matter, and of nuclei in particular, they had no better idea than to take all particles they could get hold of and

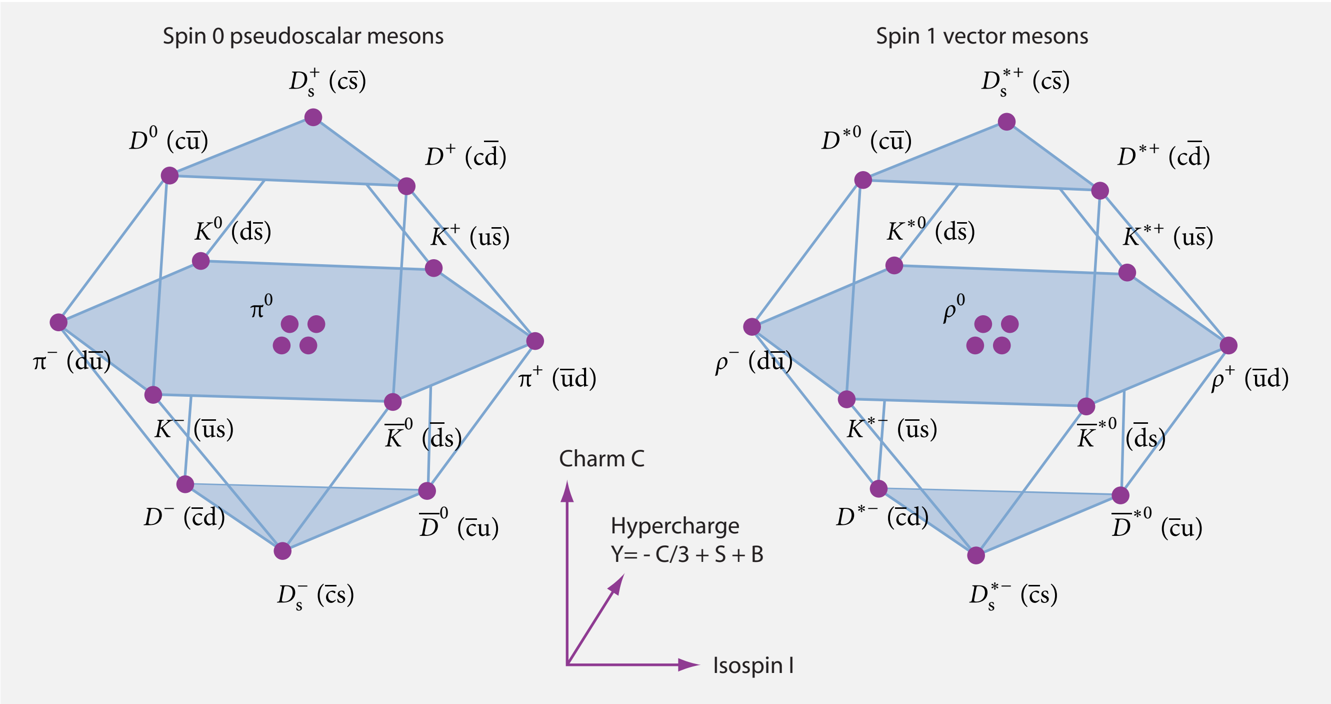


FIGURE 74 The least massive pseudoscalar and vector mesons that can be built as $q\bar{q}$ composites of the first four quark flavours

Ref. 162 to smash them into each other. Many played this game for several decades. (Obviously, this is not a fair comment; in fact, quantum theory forbids any other method. Can you
Challenge 124 s explain why?)

Imagine that you want to study how cars are built just by crashing them into each other. Before you get a list of all components, you must perform and study a non-negligible number of crashes. Most give the same result, and if you are looking for a particular part, you might have to wait for a long time. If the part is tightly attached to others, the crashes have to be especially energetic. Compared to car crashes, quantum theory adds the possibility for debris to transform, to react, to bind and to get excited. Therefore the required diligence and patience is even greater than for car crashes. Despite these difficulties, researchers have collected an ever increasing number of debris, also called *hadrons*, for many decades. The list, part of which is given in Appendix B, is overwhelmingly long; the official full list, several hundred pages of fine print, is found at pdg.web.cern.ch and contains hundreds of hadrons.

Then came the quark model. Using the ingenuity of many experimentalists and theoreticians, the model explained the whole catalogue as a consequence of only 6 types of bound quarks. Typically, a large part of the catalogue could be ordered in tables such as the ones given in Figure 74 and Figure 75. These tables were the beginning of the end of high energy physics. They explained all quantum numbers of the debris, and allowed to understand their mass ratios as well as their decays.

Debris were divided into two types: *mesons* consist of a quark and an antiquark; *baryons* consist of three quarks. In particular, the proton and the neutron are seen as

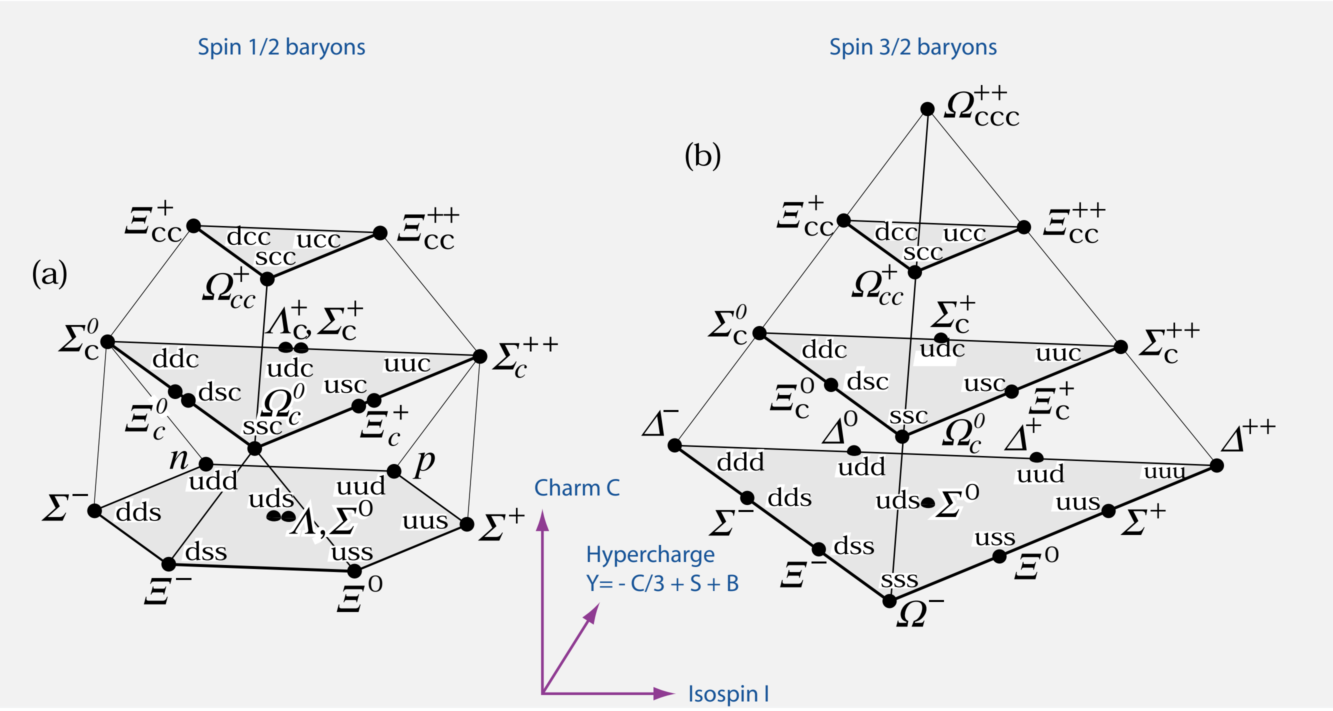


FIGURE 75 A selection of baryons and their classification as bound states of quarks (from Ref. 163)

combinations of two quark types, called *up* (u) and *down* (d): the proton is a *uud* state, the neutron a *udd* state. The discovery of other hadrons lead to the addition of four additional types of quarks. Their names are somewhat confusing: they are called *strange* (s), *charm* (c), *bottom* (b) – also called 'beauty' in the old days – and *top* (t) – called 'truth' in the past. The quark types are called *flavours*; in total, there are thus 6 quark flavours in nature.

All quarks have spin one half; their electric charges are multiples of 1/3 of the electron charge. In addition, quarks carry a strong charge, called, again confusingly, *colour*. In contrast to electromagnetism, which has only positive, negative, and neutral charges, the strong interaction has red, blue, green quarks on one side, and anti-red, anti-blue and anti-green on the other. The neutral state is called 'white'. All baryons and mesons are white, in the same way that all atoms are neutral.

The theory describing the bound states of quarks is called *quantum chromodynamics*, or QCD, and was formulated in its final form in 1973 by Fritzsch, Gell-Mann and Ref. 164 Leutwyler. In the same way that in atoms, electrons and protons are held together by the exchange of virtual photons, in protons, quarks are held together by the exchange of virtual gluons. *Gluons* are the quanta of the strong interaction, and correspond to photons, the quanta of the electromagnetic interactions.

TABLE 15 The quarks

Quark	Mass m (see text)	Spin J parity P	Possible colours; possible weak behaviour	Charge Q, isospin I, strangeness S, charm C, beauty B', topness T	Lepton number L, baryon number B
Down d	3.5 to 6 MeV/c^2	$\frac{1}{2}^+$	red, green, blue; singlet, doublet	$-\frac{1}{3}, -\frac{1}{2}, 0, 0, 0, 0$	$0, \frac{1}{3}$
Up u	1.5 to 3.3 MeV/c^2	$\frac{1}{2}^+$	red, green, blue; singlet, doublet	$+\frac{2}{3}, +\frac{1}{2}, 0, 0, 0, 0$	$0, \frac{1}{3}$
Strange s	70 to 130 MeV/c^2	$\frac{1}{2}^+$	red, green, blue; singlet, doublet	$-\frac{1}{3}, 0, -1, 0, 0, 0$	$0, \frac{1}{3}$
Charm c	1.27(11) GeV/c^2	$\frac{1}{2}^+$	red, green, blue; singlet, doublet	$+\frac{2}{3}, 0, 0, +1, 0, 0$	$0, \frac{1}{3}$
Bottom b	4.20(17) GeV/c^2	$\frac{1}{2}^+$	red, green, blue; singlet, doublet	$-\frac{1}{3}, 0, 0, 0, -1, 0$	$0, \frac{1}{3}$
Top t	171.2(2.1) GeV/c^2	$\frac{1}{2}^+$	red, green, blue; singlet, doublet	$+\frac{2}{3}, 0, 0, 0, 0, +1$	$0, \frac{1}{3}$

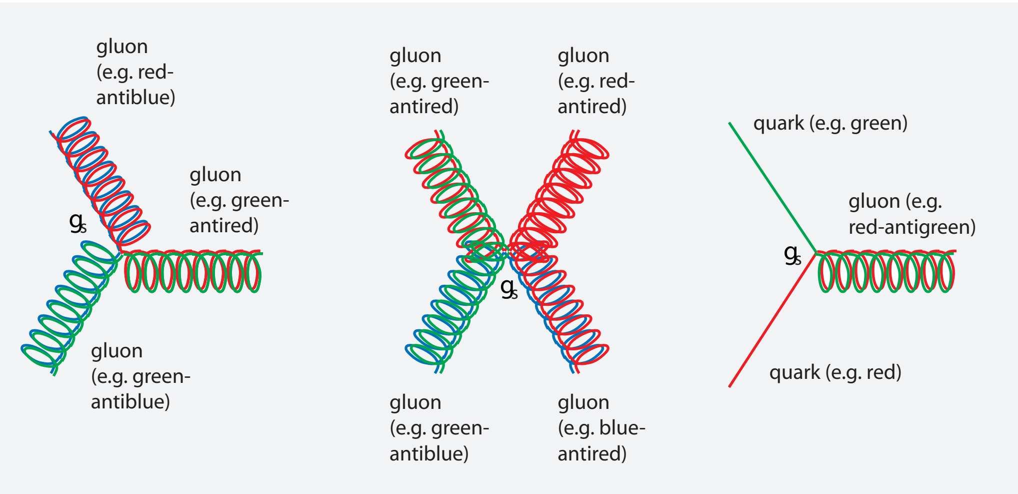


FIGURE 76 The essence of the QCD Lagrangian

The Lagrangian of quantum chromodynamics

Ref. 165 Quantum chromodynamics describes all motion due to the strong interaction with the three fundamental processes shown in Figure 76: two gluons can scatter, a gluon can emit or absorb another, and a quark can emit or absorb a gluon. In electrodynamics, only the last diagram is possible; in the strong interaction, the first two appear as well. Among others, the first two diagrams are responsible for the confinement of quarks, and thus for the lack of free quarks in nature.

QCD is a gauge theory: the fields of the strong interaction show gauge invariance under the Lie group SU(3). In the case of electrodynamics, the gauge group is U(1), and Abelian, or commutative. In contrast, SU(3) is non-Abelian; QCD is a non-abelian gauge theory, a so-called *Yang-Mills theory*.

Due to the SU(3) gauge symmetry, there are 8 gluons; they are called red-antigreen, blue-antired, etc. Since SU(3) is non-abelian, gluons interact among themselves, as shown in the first two processes in Figure 76. Out of the combinations red-antired, blue-antiblue, green-antigreen, only two gluons are linearly independent, thus giving a total of $3^2 - 1 = 8$ gluons.

The three fundamental processes of the strong interaction, together with its SU(3) gauge symmetry and the observed number of six quarks, completely determine the Lagrangian density of the strong interaction. Indeed, the Lagrangian density of the strong interaction can be seen as a complicated rewriting of Figure 76.

The Lagrangian density of quantum chromodynamics is

$$\mathcal{L}_{QCD} = -\tfrac{1}{4}F^{(a)}_{\mu\nu}F^{(a)\mu\nu} - c^2\sum_q m_q \overline{\psi}^k_q \psi_{qk} + i\hbar c \sum_q \overline{\psi}^k_q \gamma^\mu (D_\mu)_{kl}\psi^l_q \qquad (64)$$

where the gluon field strength and the gauge covariant derivative are

$$F^{(a)}_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f_{abc} A^b_\mu A^c_\nu$$

$$(D_\mu)_{kl} = \delta_{kl}\partial_\mu - i\frac{g_s}{2}\sum_a \lambda^a_{k,l} A^a_\mu \,.$$

We remember from the section on the principle of least action that Lagrangians are always sums of scalar products; this is clearly seen in the expression. Page 225 The index $a = 1\dots8$ numbers the eight types of gluons and the index $k = 1, 2, 3$ numbers the three colours, all due to SU(3). The index $q = 1\dots6$ numbers the six quark flavours. The fields $A^a_\mu(x)$ are the eight gluon fields, represented by the coiled lines in Figure 76. The fields $\psi^k_q(x)$ are those of the quarks of flavour q and colour k, represented by the straight line in the figure. The six times three quark fields, like those of any elementary fermion, are 4-component Dirac spinors with masses m_q.*

The Lagrangian (64) is that of a *local* field theory: observables are functions of position. In other words, QCD can be applied in the same way usual quantum theory, and be

* In their simplest form, the matrices γ_μ can be written as

$$\gamma_0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \quad \text{and} \quad \gamma_n = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad \text{for } n = 1, 2, 3 \qquad (65)$$

Page 197 where the σ^i are the Pauli spin matrices.

compared to experiment in the same way.

The first term of the Lagrangian (64) represents the kinetic energy of the radiation (gluons), the second or mass term the kinetic energy of the matter particles (the quarks) and the third term the interaction between the two.

The mass term in the Lagrangian is the only term that spoils or *breaks flavour symmetry*, i.e., the symmetry under exchange of quark types. (In particle physics, this symmetry is also called *chiral symmetry*, for historical reasons.) Obviously, the mass term also breaks space-time conformal symmetry.

The interaction term thus corresponds to the third diagram in Figure 76. The strength of the strong interaction is described by the *strong coupling constant* g_s. The constant is independent of flavour and colour, as observed in experiment. The Interaction term does not mix different quarks; as observed in experiments, flavour is conserved in the strong interaction, as is baryon number. The strong interaction also conserves spatial parity P and charge conjugation parity C. The strong interaction does not transform matter.

In QCD, the eight gluons are massless; also this property is taken from experiment. Therefore no gluon mass term appears in the Lagrangian. It is easy to see that massive Challenge 125 ny gluons would spoil gauge invariance. In contrast to electromagnetism, where the gauge group U(1) is Abelian, the gauge group SU(3) of the strong interactions is non-Abelian. As a consequence, the colour field itself is charged, i.e., carries colour, and thus the index a appears on the fields A and F. As a result, gluons can interact with each other, in contrast to photons, which pass each other undisturbed. The first two diagrams of Figure 76 are thus reflected in the somewhat complicated definition of the field $F^{(a)}_{\mu\nu}$. In contrast to electrodynamics, the definition has an extra term that is *quadratic* in the fields A; it is described by the so-called structure constants f_{abc} and the interaction strength g_s. These are the structure constants of the SU(3) Lie algebra. The structure constants of SU(3) describe the detailed interaction between the quarks and the gluons and the interaction between gluons themselves.

The behaviour of the gauge transformations and of the gluon field is described by the eight matrices $\lambda^a_{k,l}$. They are a fundamental, 3-dimensional representation of the generators of the SU(3) algebra and can be associated to the eight gluon types. The matrices λ_a, $a = 1..8$, and the structure constants f_{abc} obey the relations

$$
\begin{aligned}
[\lambda_a, \lambda_b] &= 2i f_{abc} \lambda_c \\
\{\lambda_a, \lambda_b\} &= 4/3 \delta_{ab} I + 2 d_{abc} \lambda_c
\end{aligned}
\qquad (66)
$$

where I is the unit matrix. The structure constants f_{abc}, which are odd under permutation of any pair of indices, and d_{abc}, which are even, are

abc	f_{abc}	abc	d_{abc}	abc	d_{abc}
123	1	118	$1/\sqrt{3}$	355	$1/2$
147	$1/2$	146	$1/2$	366	$-1/2$
156	$-1/2$	157	$1/2$	377	$-1/2$
246	$1/2$	228	$1/\sqrt{3}$	448	$-1/(2\sqrt{3})$
257	$1/2$	247	$-1/2$	558	$-1/(2\sqrt{3})$
345	$1/2$	256	$1/2$	668	$-1/(2\sqrt{3})$
367	$-1/2$	338	$1/\sqrt{3}$	778	$-1/(2\sqrt{3})$
458	$\sqrt{3}/2$	344	$1/2$	888	$-1/\sqrt{3}$
678	$\sqrt{3}/2$				

(67)

All other elements vanish. A fundamental 3-dimensional representation of the generators λ_a is given, for example, by the set of the *Gell-Mann matrices*

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$
$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} . \qquad (68)$$

There are eight matrices, one for each gluon type, with 3×3 elements, corresponding to the three colours of the strong interactions. There is no ninth gluon, because that gluon would be colourless, or 'white'.

Only quarks and gluons appear in the Lagrangian of QCD, because only quarks and gluons interact via the strong force. This can be also expressed by saying that only quarks and gluons carry colour; *colour* is the source of the strong force in the same way that electric charge is the source of the electromagnetic field. In the same way as electric charge, colour charge is *conserved* in all interactions. Electric charge comes in two types, positive and negative; in contrast, colour comes in three types, called *red*, *green* and *blue*. The neutral state, with no colour charge, is called *white*. Protons and neutrons, but also electrons or neutrinos, are thus 'white', thus *neutral* for the strong interaction.

Experimental consequences of the quark model

How can we pretend that quarks and gluons exist, even though they are never found alone? There are a number of arguments in favour.

* *

The quark model explains the non-vanishing magnetic moment of the neutron and explains the magnetic moments μ of the baryons. By describing the proton as a uud state and the neutron a udd state with no orbital angular momentum and using the precise wave functions, we get

$$\mu_u = \frac{1}{5}(4\mu_p + \mu_n) \qquad \text{and} \qquad \mu_d = \frac{1}{5}(4\mu_n + \mu_p) \; . \tag{69}$$

Assuming that $m_u = m_d$ and that that the quark magnetic moment is proportional to their charge, the quark model predicts a ratio of the magnetic moments of the proton and the neutron of

$$\frac{\mu_p}{\mu_n} = -\frac{3}{2} \; . \tag{70}$$

This prediction differs from measurements only by 3 %. Furthermore, using the same values for the magnetic moment of the quarks, magnetic moment values of over half a dozen of other baryons can be predicted. The results typically deviate from measurements by around 10 %. In particular, the sign of the resulting baryon magnetic moment is always correctly calculated.

* *

The quark model describes all quantum numbers of mesons and baryons. P-parity, C-parity, and the absence of certain meson parities are all reproduced. The observed conservation of electric charge, baryon number, isospin, strangeness etc. is reproduced. Diagrams such as those shown in Figure 74 and in Figure 75 describe all existing states completely; the states not listed are not observed. The quark model thus produces a complete and correct classification of all hadrons as bound states of quarks.

* *

The quark model also explains the mass spectrum of mesons and baryons. The best predictions are made by QCD lattice calculations. With months of computer time, researchers were able to reproduce the masses of proton and neutron to within a few per cent. Interestingly, if one sets the u and d quark masses to zero, the resulting proton and neutron mass differ from experimental values only by 10 %. The mass of protons and neutrons is almost completely due to the binding.

Ref. 166 Ref. 167

* *

The number of colours of quarks must be taken into account to get correspondence of theory and calculation. For example, the measured decay time of the neutral pion is 83 as. The calculation without colour gives 750 as; if each quark is assumed to appear in 3 colours the value must be divided by 9, and then matches the measurement.

* *

In particle colliders, collisions of electrons and positrons sometimes lead to the production of hadrons. The calculated production rates also fit experiments only if quarks have three colours. In more detail, if one compares the ratio of muon–antimuon production

Challenge 126 s and of hadron production, a simple estimate relates them to their charges:

$$R = \frac{\sum q_{\text{hadrons}}}{\sum q_{\text{muons}}} \qquad (71)$$

Between 2 and 4 GeV, when only three quarks can appear, this argument thus predicts $R = 2$ if colours exist, or $R = 2/3$ if they don't. Experiments yield a value of $R = 2.2$, thus confirming the number of colours. Many other such *branching ratios* can be calculated in this way. They agree with experiments only if the number of colours is three.

* *

In 1979, the first clear decays of gluons have been observed at the PETRA particle collider in Hamburg. The occurrence of certain events, called *gluon jets*, are due to the decay of high-energy gluons into narrow beams of particles. Gluon jets appear in coplanar three-jet events. The observed rate and the other properties of these events confirmed the predictions of QCD. Experiments at PETRA also determined the spin $S = 1$ of the gluon and the running of the strong coupling constant. The hero of those times was the project manager Gustav-Adolf Voss, who completed the accelerator on budget and six months ahead of schedule. Ref. 168

Confinement of quarks – and elephants

Many of the particlesquark confinement which are not part of the diagrams of Figure 74 and Figure 75 can be explained as *rotational excitations* of the fundamental mesons. The idea of rotational excitations leads to quantitative predictions, as shown by Tullio Regge in 1957. Regge assumed that mesons and baryons are quarks connected by strings, like rubber bands, and that the force or tension k between the quarks is thus constant over distance.

We assume that the strings, whose length we call $2r_0$, rotate around their centre of mass as rapidly as possible, as shown in Figure 78. Then we have

$$v(r) = c\frac{r}{r_0}\,. \qquad (72)$$

The quark masses are assumed negligible. For the total energy this implies the relation

$$E = mc^2 = 2\int_0^{r_0} \frac{k}{\sqrt{1 - v(r)/c^2}}\mathrm{d}r = kr_0\pi \qquad (73)$$

and for angular momentum the relation

$$J = \frac{2}{\hbar c^2}\int_0^{r_0} \frac{krv(r)}{\sqrt{1 - v(r)/c^2}}\mathrm{d}r = \frac{kr_0^2}{2\hbar c}\,. \qquad (74)$$

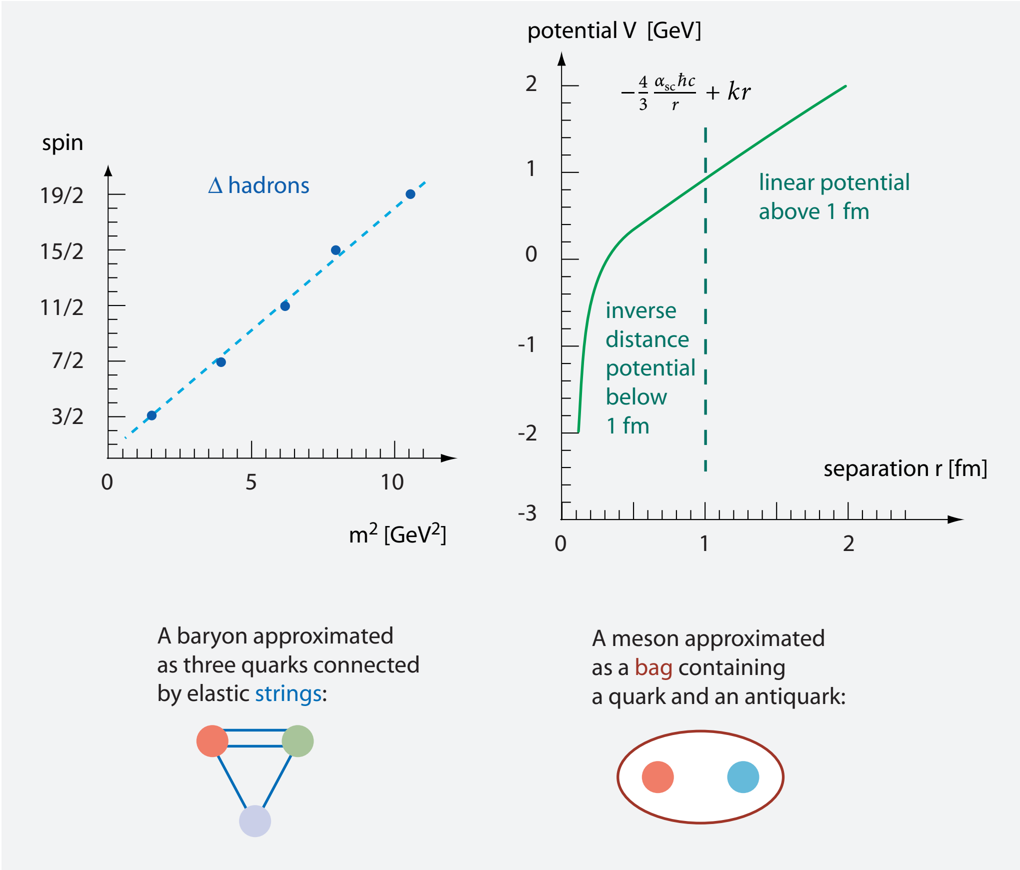


FIGURE 77 A Regge trajectory, or Chew–Frautschi plot, due to the confinement of quarks, the quark confinement potential, and two approximate ways to describe the confinement: the string model and the bag model of hadrons

Including the spin of the quarks, we thus get

$$J = \alpha_0 + \alpha' m^2 \quad \text{where} \quad \alpha' = \frac{c^3}{2\pi k \hbar} . \tag{75}$$

Regge thus deduced a simple expression that relates the mass m of excited hadrons to their total spin J. For bizarre historical reasons, this relation is called a *Regge trajectory*.

The value of the constant α' is predicted to be of the quark–antiquark pairing. A few years later, as shown in Figure 77, such linear relations were found in experiments: the *Chew-Frautschi plots*. For example, the three lowest lying states of Δ are the spin 3/2 Δ(1232) with m^2 of 1.5 GeV^2, the spin 7/2 Δ(1950) with m^2 of 3.8 GeV^2, and the spin 11/2 Δ(2420) with m^2 of 5.9 GeV^2. The value of the constant α' is found experimentally to be around 0.93 GeV^{-2} for almost all mesons and baryons, whereas the value for α_0 varies from particle to particle. The string tension is thus found to be
Ref. 165 k =0.87 GeV/fm=0.14 MN. In other words, two quarks in a hadron attract each other with a force equal to the weight of two elephants: 15 tons.

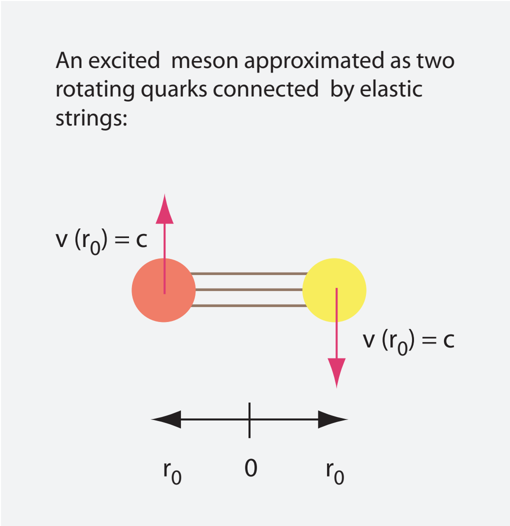


FIGURE 78 Calculating masses of excited hadrons

Experiments are thus clear: the observed Chew-Frautschi plots, as well as several other observations not discussed here, are best described by a quark–quark potential that grows linearly with distance above 1 fm. As a result, quarks never appear as free particles: quarks are always *confined* in hadrons. This is in contrast with QED, where the force between charges goes to zero for large distances; electric charges are thus not confined, but can exist as free particles. At low distances, the potential decreases with the inverse of the distance. In total, experiments lead to a potential given by

Ref. 162

$$V = -\frac{4}{3}\frac{\alpha_{sc}\hbar c}{r} + kr \qquad (76)$$

where k is the mentioned 0.87 GeV/fm, α_{sc} is 0.2, and $\hbar c$ is 0.1975 GeV/fm. It is shown in Figure 77

Even though experiments are clear, theoreticians face a problem. So far, neither the quark-quark potential nor the quark bound states can be deduced from the QCD Lagrangian with a *simple* approximation method. Nevertheless, complicated non-perturbative calculations show that the QCD Lagrangian does predict a force between two coloured particles that levels off at a constant value (corresponding to a linearly increasing potential). These calculations show that the old empirical approximations of hadrons as quarks connected by strings or a quarks in bags, shown in Figure 77, can indeed be deduced from the QCD Lagrangian. However, the calculations are too complex to be summarized in a few lines. Independently, the constant force value has also been reproduced in computer calculations, in which one simplifies space-time to a lattice and then approximates QCD by so-called *lattice QCD* or *lattice gauge theory*. Lattice calculations have also reproduced the masses of most mesons and baryons with reasonable accuracy. Using the most powerful computers available, these calculations have given predictions of the mass of the proton and other baryons within a few per cent. Discussing these complex and fascinating calculations lies outside the scope of this

Ref. 169

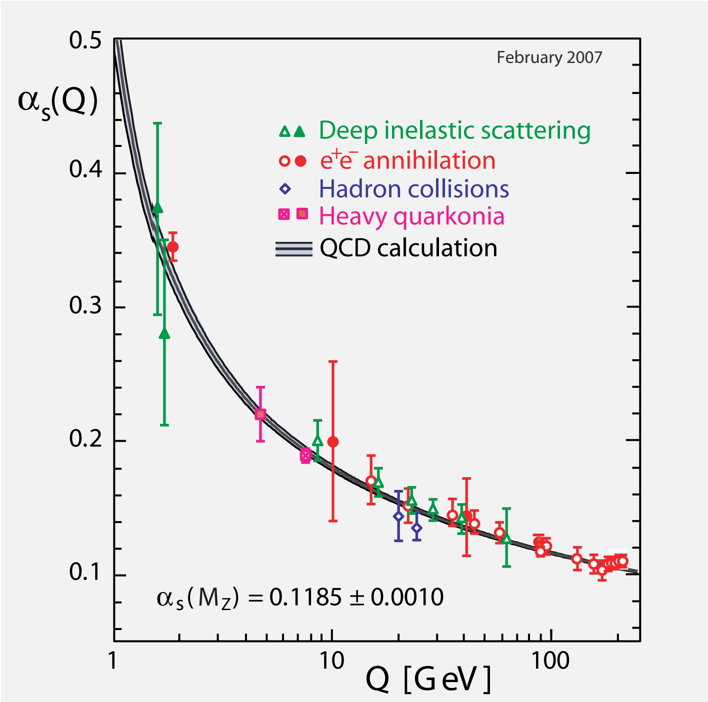


FIGURE 79 The measured and the calculated variation of the strong coupling with energy, showing the precision of the QCD Lagrangian and the asymptotic freedom of the strong interaction (© Siegfried Bethke, updated from Ref. 170)

text, however.

In fact, the challenge of explaining confinement in simple terms is so difficult that the brightest minds have been unable to solve it yet. This is not a surprise, as its solution probably requires the unification of the interactions and, most probably, also the unification with gravity. We therefore leave this issue for later in our adventure.

Asymptotic freedom

QCD has another property that sets it apart form QED: the behaviour of its coupling with energy. In fact, there are three equivalent ways to describe the strong coupling. The quantity appearing in the QCD Lagrangian, g_s, is often used to define the equivalent quantity $\alpha_s = g_s^2/4\pi$. Both α_s and g_s depend on the energy Q of the experiment. If they are known for one energy, they are known for all of them. Presently, the best experimental value is $\alpha_s(M_Z) = 0.1185 \pm 0.0010$.

Ref. 165, Ref. 163 The energy dependence of the strong coupling can be calculated with the standard renormalization procedures and is expected to be

$$\alpha_s(Q^2) = \frac{12\pi}{33 - 2n_f}\frac{1}{L}\left(1 - \frac{(918 - 114n_f)\ln L}{(33 - 2n_f)^2 L} + ...\right) \quad \text{where} \quad L = \ln\frac{Q^2}{\Lambda^2(n_f)} \qquad (77)$$

where n_f is the number of quarks with mass below the energy scale Q, thus a number between 3 and 6. (The expression has been expanded to many additional terms with help of computer algebra.) The third way to describe the strong coupling is thus the energy parameter $\Lambda(n_f)$. Experiments yield $\Lambda(3)$ =230(60) GeV, $\Lambda(4)$ =180(50) GeV and $\Lambda(5)$ =120(30) GeV.

The accelerator experiments that measure the coupling are extremely involved, and hundreds of people across the world have worked for many years to gather the relevant data. The comparison of QCD and experiment, shown in Figure 79, does not show any contradiction between the two.

Figure 79 and expression (77) also illustrate what is called *asymptotic freedom*: α_s decreases at high energies. In other words, at high energies quarks are *freed* from the strong interaction; they behave as free particles.* As a result of asymptotic freedom, in QCD, a perturbation expansion can be used only at energies much larger than Λ. Historically, the discovery of asymptotic freedom was essential to establish QCD as a theory of the strong interaction.

Asymptotic freedom can be understood qualitatively if the situation is compared to QED. The electron coupling increases at small distances, because the screening due to the virtual electron-positron pairs has less and less effect. In QCD, the effective colour coupling also changes at small distances, due to the smaller number of virtual quark-antiquark pairs. However, the gluon properties lead to the opposite effect, an *antiscreening*: the effective coupling decreases at small distances.

The sizes and masses of quarks

The size of quarks, like that of all elementary particles, is predicted to vanish by qcd, as in all quantum field theory. So far, no experiment has found any effect due to a finite quark size. Measurements show that quarks are surely smaller than 10^{-19} m. No size conjecture has been given by any hypothetical theory. Quarks are assumed point-like, or at most Planck-sized, in all descriptions so far. Ref. 171

We noted in several places that a neutral compound of charged particles is always less massive than its components. But when the mass values for quarks are looked up in most tables, the masses of u and d quarks are only of the order of a few MeV/c^2, whereas the proton's mass is 938 MeV/c^2. What is the story here?

It turns out that the definition of the mass is more involved for quarks than for other particles. Quarks are never found as free particles, but only in bound states. As a result, the concept of quark mass depends on the theoretical framework one is using.

Due to *asymptotic freedom*, quarks behave almost like free particles only at high energies. The mass of such a 'free' quark is called the *current quark mass*; for the light quarks it is only a few MeV/c^2, as shown in Table 15.

At low energy, for example inside a proton, quarks are *not* free, but must carry along a large amount of energy due to the confinement process. As a result, bound quarks have a much larger effective, so-called *constituent quark mass*, which takes into account this confinement energy. To give an idea of the values, take a proton; the indeterminacy relation for a particle inside a sphere of radius 0.9 fm gives a momentum indeterminacy of around 190 MeV/c. In three dimensions this gives an energy of $\sqrt{3}$ times that value, or an effective, constituent quark mass of about 330 MeV/c^2. Three confined quarks are thus

* Asymptotic freedom was discovered in 1972 by Gerard 't Hooft; since he had received the Nobel Prize already, the 2004 Prize was then given to the next people who highlighted it: David Gross, David Politzer and Frank Wilczek, who studied it extensively in 1973.

heavier than a proton, whose mass is 938 MeV/c^2; we can thus still say that a compound proton is less massive than its constituents.

In short, the mass of the proton and the neutron is (almost exclusively) the kinetic energy of the quarks inside them, as their rest mass is almost negligible. As Frank Wilczek says, some people put on weight even though they never eat anything heavy.

Ref. 172

But also the small current quark mass values for the up, down, strange and charmed quarks that appear in the QCD Lagrangian are framework dependent. The values of Table 15 are those for a renormalization scale of 2 GeV. For half that energy, the mass values increase by 35%. The heavy quark masses are those used in the so-called $\bar{M}S$ scheme, a particular way to perform perturbation theory.

Ref. 163

The mass, shape and colour of protons

Frank Wilczek mentions that one of the main results of QCD, the theory of strong interactions, is to explain mass relations such as

Ref. 172

$$m_{\text{proton}} \sim e^{-k/\alpha} m_{\text{Planck}} \quad \text{and} \quad k = 11/2\pi\ ,\ \alpha_{\text{unif}} = 1/25\ . \tag{78}$$

Here, the value of the coupling constant α_{unif} is taken at the unifying energy, a factor of 1000 below the Planck energy. (See the section of unification below.) In other words, a general understanding of masses of bound states of the strong interaction, such as the proton, requires almost purely a knowledge of the unification energy and the coupling constant at that energy. The approximate value $\alpha_{\text{unif}} = 1/25$ is an extrapolation from the low energy value, using experimental data. The proportionality factor k in expression (78) is not easy to calculate. It is usually determined on computers using lattice QCD.

Page 196

But the mass is not the only property of the proton. Being a cloud of quarks and gluons, it also has a shape. Surprisingly, it took a long time before people started to become interested in this aspect. The proton, being made of two up quarks and one down quark, resembles a ionized H_2^+ molecule, where one electron forms a cloud around two protons. Obviously, the H_2^+ molecule is elongated, or prolate.

Is the proton prolate? There is no spectroscopically measurable non-sphericty – or quadrupole moment – of the proton. However, the proton has an *intrinsic* quadrupole moment. The quadrupole moments of the proton and of the neutron are predicted to be positive in all known calculation methods, implying an *prolate* shape. Recent measurements at Jefferson Laboratories confirm this prediction. A prolate shape is predicted for all $J = 1/2$ baryons, in contrast to the oblate shape predicted for the $J = 3/2$ baryons. The spin 0 pseudoscalar mesons are predicted to be prolate, whereas the spin 1 vector mesons are expected to be oblate.

Ref. 173

The shape of any molecule will depend on whether other molecules surround it. Recent research showed that similarly, both the size and the shape of the proton in nuclei is slightly variable; both seem to depend on the nucleus in which the proton is built-in.

Ref. 174

Apart from shapes, molecules also have a colour. The colour of a molecule, like that of any object, is due to the energy absorbed when it is irradiated. For example, the H_2^+ molecule can absorb certain light frequencies by changing to an excited state. Molecules change mass when they absorb light; the excited state is heavier than the ground state. In the same way, protons and neutrons can be excited. In fact, their excited states have

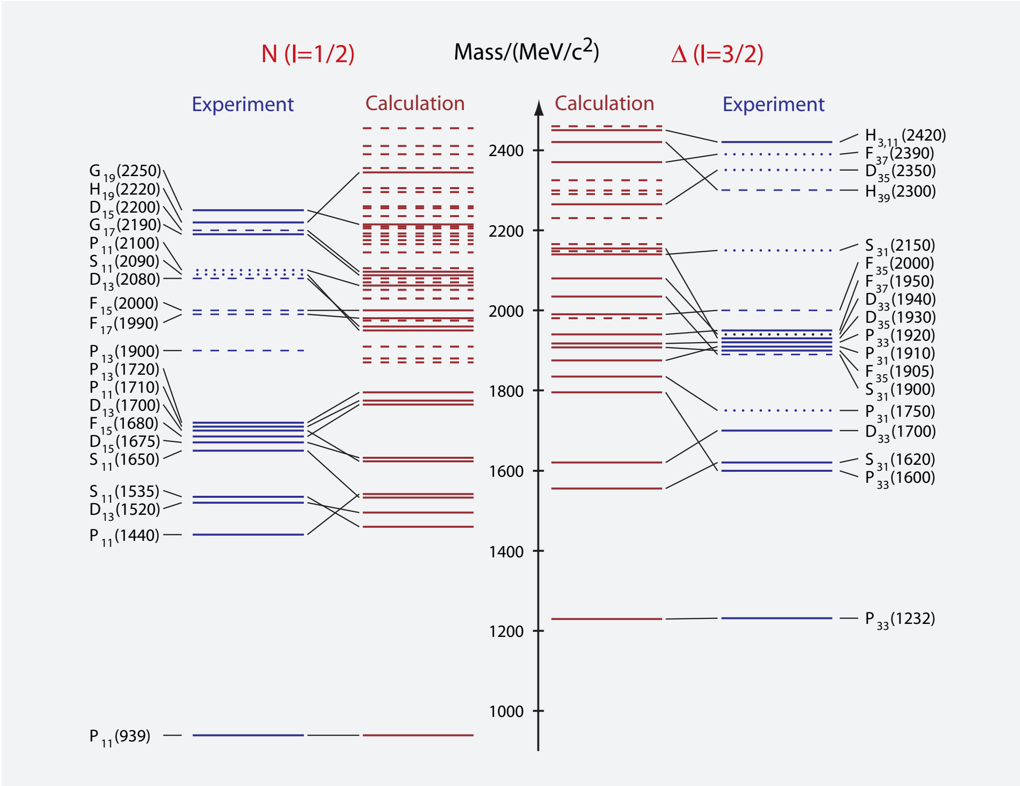


FIGURE 80 The mass spectrum of the excited states of the proton: experimental and calculated values (from Ref. 163)

Ref. 163 been studied in detail; a summary, also showing the limitation of the approach, is shown in Figure 80. Many excitations can be explained as excited quarks states, but many more are predicted. The calculated masses agree with observations to about 10%. The quark model and QCD thus structure and explain a large part of the baryon spectrum; but the agreement is not yet perfect.

Obviously, in our everyday environment the energies necessary to excite nucleons do not appear – in fact, they do not even appear inside the Sun – and these excited states can be neglected. They only appear in particle accelerators and in cosmic radiation. In a sense, we can say that in our corner of the universe, the colour of protons usually is not visible.

Curiosities about the strong interactions

The computer calculations necessary to extract particle data from the Lagrangian of quantum chromodynamics are among the most complex calculations ever performed. They beat weather forecasts, fluid simulations and the like by orders of magnitude. Nobody knows whether this will be necessary also in the future: the race for a simple approximation method for finding solutions is still open.

* *

Even though gluons are massless, like photons and gravitons, there is no colour radiation. Gluons carry colour and couple to themselves; as a result, free gluons were predicted to directly decay into quark–antiquark pairs. This decay has indeed been observed in experiments at particle accelerators.

* *

Something similar to colour radiation, but still stranger might have been found in 1997. It might be that a scalar meson with a mass of 1.5 GeV/c^2 is a *glueball*. This is a hypothetical meson composed of gluons only. Numerical results from lattice gauge theory seem to confirm the possibility of a glueball in that mass range. Ref. 175

* *

There is a growing consensus that most light scalar mesons below 1 GeV/c^2, are tetraquarks. In 2003, experiments provided also candidates for heavier tetraquarks, namely the X(3872), Ds(2317) and Ds(2460). The coming years will show whether this interpretation is correct. Ref. 176

* *

Do particles made of five quarks, so-called *pentaquarks*, exist? So far, they seem to exist only in a few laboratories in Japan, whereas in other laboratories across the world they are not seen. The issue is still open, though most researchers do not believe in them any more. Ref. 177

* *

Whenever we look at a periodic table of the elements, we look at a manifestation of the strong interaction. The Lagrangian of the strong interaction describes the origin and properties of the presently known 115 elements.

Nevertheless one central aspect of nuclei is determined by the electromagnetic interaction. Why are there around 115 different elements? Because the electromagnetic coupling constant α is around 1/137.036. In more detail, the answer is the following. If the charge of a nucleus were much higher than around 130, the electric field around nuclei would lead to spontaneous electron–positron pair generation; the generated electron would fall into the nucleus and transform one proton into a neutron, thus inhibiting a larger proton number. The finite number of the elements is thus due to the electromagnetic interaction.

* *

The instability of the vacuum also yields a (trivial) limit on the fine structure constant. The fine structure constant value of around 1/137.036 cannot be explained by quantum electrodynamics. However, it can be deduced that it must be lower than 1 to lead to a consistent theory. Indeed, if its value were larger than 1, the vacuum would become unstable and would spontaneously generate electron-positron pairs. Ref. 178

* *

To know more about radioactivity, its effects, its dangers and what a government can do about it, see the English and German language site of the Federal Office for Radiation Protection at www.bfs.de.

TABLE 16 Correspondence between QCD and superconductivity

QCD	Superconductivity
Quark	magnetic monopole
Colour force non-linearities	Electron–lattice interaction
Chromoelectric flux tube	magnetic flux tube
Gluon-gluon attraction	electron–electron attraction
Glueballs	Cooper pairs
Instability of bare vacuum	instability of bare Fermi surface
Discrete centre symmetry	continuous U(1) symmetry
High temperature breaks symmetry	low temperature breaks symmetry

* *

From the years 1990 onwards, it has regularly been claimed that extremely poor countries are building nuclear weapons. Why is this highly unlikely? Challenge 127 s

* *

Historically, nuclear reactions provided the first test of the relation $E = \gamma mc^2$. This was achieved in 1932 by Cockcroft and Walton. They showed that by shooting protons into lithium one gets the reaction

$$^{7}_{3}\mathrm{Li} + ^{1}_{1}\mathrm{H} \rightarrow ^{8}_{4}\mathrm{Be} \rightarrow ^{4}_{2}\mathrm{He} + ^{4}_{2}\mathrm{He} + 17\,\mathrm{MeV}\ . \tag{79}$$

The measured energy on the right is exactly the value that is derived from the differences in total mass of the nuclei on both sides.

* *

A large minority of researchers say that QCD is defined by two parameters. Apart from the coupling constant, they count also the strong CP parameter. Indeed, it might be that the strong interaction violates CP invariance. This violation would be described by a second term in the Lagrangian; its strength would be described by a second parameter, a phase usually called θ_{CP}. However, many high-precision experiments have been performed to search for this effect, and no CP violation in the strong interaction has ever been detected.

* *

In a well-known analogy, QCD can be compared to superconductivity. Table 16 gives an overview of the correspondence.

A summary of QCD

Quantum chromodynamics, the non-abelian gauge theory based on the SU(3) symmetric Lagrangian, describes the properties of gluons and quarks, the properties of the proton, the neutron and all other hadrons, the properties of atomic nuclei, the working of the stars and the origin of the atoms inside us and around us. Without the strong interaction, we would not have flesh and blood. And all these aspects of nature follow from a single number, the strong coupling constant.

QCD and experiment agree wherever comparisons have been made. The limitations of QCD are only conceptual. Like in all of quantum field theory, also in the case of QCD the mathematical form of the Lagrangian is almost uniquely defined by requiring renormalizability, Lorentz invariance, and gauge invariance – SU(3) in this case. We say 'almost', because the Lagrangian contains a few parameters that remain unexplained:

— The number, 6, and the masses m_q of the quarks are not explained by QCD.
— The coupling constant of the strong interaction g_s, or equivalently, α_s or Λ, is unexplained. QCD predicts its energy dependence, but not its absolute value.
— Experimentally, the strong interaction is found to be CP conserving. This is not obvious; the QCD Lagrangian assumes that any possible CP-violating term vanishes, even though there exist CP-violating Lagrangian terms that are Lorentz-invariant, gauge-invariant and renormalizable.
— The properties of space-time, in particular its Lorentz invariance, its continuity and the number of its dimensions are obviously all unexplained in QCD and assumed from the outset.
— It is also not known how QCD has to be modified in strong gravity, thus in strongly curved space-time.

Before we explore ways to overcome these limits, we have a look at the other nuclear interaction.

Chapter 8

THE WEAK NUCLEAR INTERACTION AND THE HANDEDNESS OF NATURE

The weirdest interaction in nature is the weak interaction. The weak interaction transforms elementary particles into each other, has radiation particles that have mass, and violates parity. In short, we do not experience the weak interaction in our everyday life, and its properties violate much of what we experience in everyday life. This makes the weak interaction also the most fascinating interaction.

Transformation of elementary particles

Radioactivity, in particular the so-called β decay, is a bizarre phenomenon. Experiments in the 1910s showed that when beta sources emit electrons, atoms change from one chemical element to another. For example, one observed that tritium decays into Helium as

$$^{3}_{1}\mathrm{H} \rightarrow {}^{3}_{2}\mathrm{He} + e^{-} + \overline{\nu}_{\mathrm{e}} \,. \qquad (80)$$

In the 1930s it became clear that this process is due to a neutron in the nucleus changing into a proton (and more):

$$n \rightarrow p + e^{-} + \overline{\nu}_{\mathrm{e}} \,. \qquad (81)$$

In the 1960s, the quark model showed that beta decay is due to an up quark changing to a down quark:

$$d \rightarrow u + e^{-} + \overline{\nu}_{\mathrm{e}} \,. \qquad (82)$$

Elementary particles are thus not immutable. The dream of Democritus and Leucippus about immutable basic building blocks is definitely *not* realized in nature.

Experiments show that quark transformation is not possible to achieve with electromagnetic fields, nor with gluon fields, nor with gravitation. There must be another type of radiation in nature, and thus another interaction. Fortunately, such transformations are rare, otherwise we would not be running around: the interaction is *weak*.

The weak nuclear interaction transforms quarks into each other. But where does the energy go in beta decay? Measurements in 1911 showed that the energy spectrum of the emitted electron is continuous. Then, in 1930, Wolfgang Pauli had the courage and genius to explain this observation in with a daring hypothesis: the energy of the decay is split between the electron and a new, truly astonishing particle, the *neutrino* – more precisely, the electron anti-neutrino $\bar{\nu}_{\mathrm{w}}$. Experiments showed that the neutrino must be uncharged, must not interact strongly, and must be of very low mass. As a result, neutrinos interact

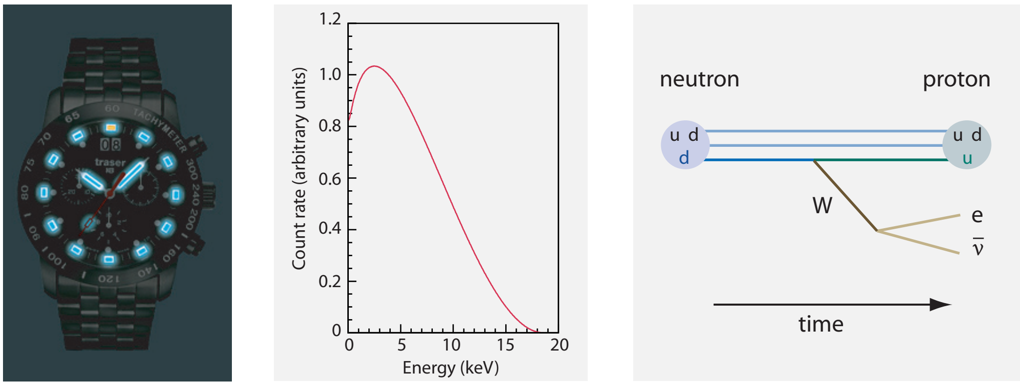


FIGURE 81 Beta decay in tritium: a modern, tritium-powered illuminated watch, the continuous energy spectrum of the emitted electrons from tritium, and the process occurring in the tritium nucleus (© www.traser.com, Katrin)

TABLE 17 The leptons: the three neutrinos and the three charged leptons (antiparticles have opposite charge Q and parity P)

Neutrino	Mass m (see text)	Spin J parity P	Colour; possible weak behaviour	Charge Q, isospin I, strangeness S, charm C, beauty B', topness T	Lepton number L, baryon number B
Electron neutrino ν_e	$< 2\,\mathrm{eV}/c^2$	$\frac{1}{2}^+$	white; singlet, doublet	0, 0, 0, 0, 0, 0	1, 0
Muon neutrino ν_e	$< 2\,\mathrm{eV}/c^2$	$\frac{1}{2}^+$	white; singlet, doublet	0, 0, 0, 0, 0, 0	1, 0
Tau neutrino ν_e	$< 2\,\mathrm{eV}/c^2$	$\frac{1}{2}^+$	white; singlet, doublet	0, 0, 0, 0, 0, 0	1, 0
Electron e	0.510 998 910(13) MeV/c^2	$\frac{1}{2}^+$	white; singlet, doublet	−1, 0, 0, 0, 0, 0	1, 0
Muon μ	105.658 367(4) MeV/c^2	$\frac{1}{2}^+$	white; singlet, doublet	−1, 0, 0, 0, 0, 0	1, 0
Tau τ	1.776 84(17) GeV/c^2	$\frac{1}{2}^+$	white; singlet, doublet	−1, 0, 0, 0, 0, 0	1, 0

with ordinary matter only extremely rarely, and usually fly through the Earth without being affected. This property makes detection very difficult, but not impossible; the particle was detected in 1952.

The weakness of the weak nuclear interaction

'Weak radiation' consists of massive particles; there are two types, the neutral Z boson with a mass of 91.2 GeV – that is the mass of a silver atom – and the electrically charged

TABLE 18 The intermediate vector bosons of the weak interaction (the Z boson is its own antiparticle; the W boson has an antiparticle of opposite charge)

Boson	Mass m	Spin J	Colour; weak behaviour	Charge Q, isospin I, strangeness S, charm C, beauty B', topness T	Lepton number L, baryon number B
Z boson	91.1876(21) GeV/c^2	1	white; 'triplet'	0, 0, 0, 0, 0, 0	0, 0
W boson	80.398(25) GeV/c^2	1	white; 'triplet'	1, 0, 0, 0, 0, 0	0, 0

W boson with a mass of 80.4 GeV. The masses are so large that free radiation exists only for an extremely short time, about 0.1 ys; then the particles decay. The large mass is the reason that the weak interaction is extremely short range and weak; any exchange of virtual particles scales with the negative exponential of the intermediate particle's mass. In fact, the weak interaction is so weak that neutrinos, particles which interact only weakly, have an overwhelming probability to fly through the Sun without any interaction.

The existence of a massive charged intermediate vector boson, today called the W, was already deduced in the 1940s; but theoretical physicists did not accept the idea until the Dutch physicist Gerard 't Hooft proved that it was possible to have such a mass without having problems in the rest of the theory. For this proof he later received the Nobel Prize in Physics.

Experimentally, the Z boson was first observed as a virtual particle in 1973 at CERN in Geneva. In 1983, CERN groups produced and detected the first real W and Z bosons. This experiment was a five-year effort by thousands of people working together. The energetic manager of the project, Carlo Rubbia, famous for changing one secretary every three weeks, and the chief technologist, Simon van der Meer, received the 1984 Nobel Prize in Physics for the discovery.

In the same way that photons are emitted by accelerated charges, W and Z bosons are emitted by accelerated weak charges. The W and the Z are observed to be elementary.
Ref. 163 For example, the W gyromagnetic ratio is as predicted for elementary particles.

Distinguishing left from right

The next weird characteristic of the weak interaction is the non-conservation of parity P under spatial inversion. The weak interaction distinguishes between mirror systems, in contrast to everyday life, gravitation, electromagnetism, and the strong interactions. Parity non-conservation had been predicted by 1956 by Lee Tsung-Dao and Yang Chen Ning in order to explain the ability of K^0 mesons to decay either into 2 pions, which have
Ref. 179 even parity, or into 3 pions, which have odd parity.

Lee and Yang suggested an experiment to Wu Chien-Shiung* The experiment she

* Wu Chien-Shiung (1912–1997) was called 'madame Wu' by everybody. She was a bright and driven physi-

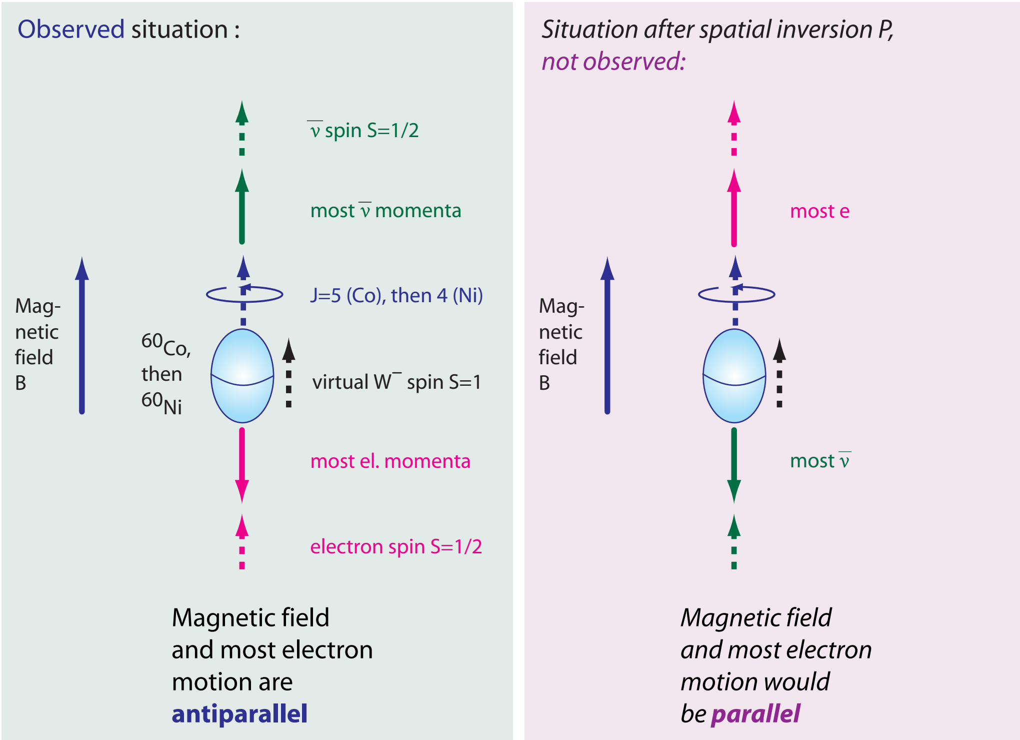


FIGURE 82 The measured behaviour of β-decay, and its imagined, but unobserved behaviour under spatial inversion P (corresponding to a mirror reflection plus subsequent rotation by π around an axis perpendicular to the mirror plane)

FIGURE 83 Wu Chien-Shiung (1912–1997)

performed with her team is shown schematically in Figure 82. A few months after the first meetings with Lee and Yang, the Wu and her team found that in the β-decay of cobalt nuclei aligned along a magnetic field, the emitted electrons are emitted mostly *against* the spin of the nuclei. In the parity-inversed experiment, the electrons would be emitted *along* the spin direction; however, this case is not observed. Parity is violated. This earned Lee and Yang a Nobel Prize in 1957.

cist born in China. She worked also on nuclear weapons; later in life she was president of the American Physical Society.

Parity is violated in the weak interaction. This is not only the case for β-decay. The result has been confirmed for muon decay and every other weak process studied so far.

Ref. 180 The weak interaction is not parity invariant. In particular, when two electrons collide, those collisions that occur through the weak interaction behave differently in a mirror experiment. The number of experiments showing this increases from year to year. In 2004, two polarized beams of electrons – one left-handed and one right-handed – were shot at a matter target and the reflected electrons were counted. The difference was 0.175 parts per million – small, but measurable. The experiment also confirmed the predicted weak charge of −0.046 of the electron.

A beautiful consequence of parity violation is its influence on the *colour* of certain atoms. This prediction was made in 1974 by Bouchiat and Bouchiat. (Ref. 181) The weak interaction is triggered by the weak charge of electrons and nuclei; therefore, electrons in atoms do not exchange only virtual photons with the nucleus, but also virtual Z particles. The chance for this latter process is extremely small, around 10^{-11} times smaller than for exchange of virtual photons. But since the weak interaction is not parity conserving, this process allows electron transitions which are impossible by purely electromagnetic effects. In 1984, measurements confirmed that certain optical transitions of caesium atoms that are impossible via the electromagnetic interaction, are allowed when the weak interaction is taken into account. (Ref. 182) Several groups have improved these results and have been able to confirm the calculations based on the weak interaction properties, including the weak charge of the nucleus, to within a few per cent. (Ref. 183)

The weak interaction thus allows one to distinguish left from right. Nature contains processes who differ from their mirror version. This is in full contrast to everyday life. In short, particle physics has shown that nature is weakly left-handed.

The handedness of nature is to be taken literally: further experiments confirmed to central statements on the weak interaction that can be already guessed from Figure 82. (Challenge 128 e) First, the weak interaction only couples to left-handed particles and to right-handed antiparticles. Parity is *maximally violated* in the weak interaction. Secondly, all neutrinos observed so far are left-handed, and that all antineutrinos are right-handed. (This can only be true if their mass vanishes or is negligibly small.) These statements define several aspects of the Lagrangian of the weak interaction and of the standard model of particle physics.

Distinguishing particles and antiparticles, CP violation

In the weak interaction, the observation that only right-handed particles and left-handed antiparticles are affected has an important consequence: it implies a violation of charge conjugation parity C. (Challenge 129 e) Observations of muons into electrons shows this most clearly: antimuon decay differs from muon decay. The weak interaction distinguishes particles from antiparticles. In fact, C parity, like P parity, is *maximally violated* in the weak interaction. Also this effect has been confirmed in all subsequent observations ever performed on the weak interaction.

But that is not all. The weak interaction also violates the combination of parity inversion with particle-antiparticle symmetry, the so-called CP invariance. In contrast to P violation and C violation, which are maximal, CP violation is a tiny effect. Its observation was made first in 1964 by Val Fitch and James Cronin in the decay of the neutral K

FIGURE 84 The essence of the electroweak interaction Lagrangian

mesons. This experiment earned them the Nobel Prize in 1980. CP violation has also been observed in neutral B mesons, in several different processes and reactions. The search for other manifestations of CP violation, such as in non-vanishing electric dipole moments of elementary particles, is an intense research field. The search is not simple because CP violation is a small effect in an already very weak interaction; this tends to makes experiments large and expensive.

Since the weak interaction violates CP invariance, it also violates motion (or time) reversal T. But like all gauge theories, the weak interaction is invariant under the combined CPT transformation. If CPT would be violated, the masses. lifetimes and magnetic moments of particles and antiparticles would differ. That is not observed.

Weak charge, mixings and symmetry breaking

All weak interaction processes can be described by the Feynman diagrams in Figure 84. But a few remarks are necessary. First of all, the W and Z act only on left-handed fermion and on right-handed anti-fermions. Secondly, the weak interaction conserves a *weak charge*, or *weak isospin* T_3. The three quarks u, c and t, as well as the neutrinos, have weak isospin $T_3 = 1/2$; the other three quarks and the charged leptons have weak isospin $T_3 = -1/2$. In an idealized, SU(2) world, the three vector bosons W^+, W^0, W^- would have weak isospin values $1, 0$ and -1 and be massless. However, three aspects complicate the issue.

First of all, it turns out that the quarks appearing in Figure 84 are not those of the strong interaction: there is a slight difference, due to quark mixing. Secondly, also neutrinos mix. And thirdly, the vector bosons are massive and break the SU(2) symmetry of the imagined idealized world; the Lie group SU(2) is not an exact symmetry of the weak interaction, and the famous Higgs boson has mass.

1. Surprisingly, the weak interaction eigenstates of the quarks are not the same as the mass eigenstates. This is described by the so-called Cabibbo–Kobayashi–Maskawa

or CKM mixing matrix. It is defined by

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = (V_{ij}) \begin{pmatrix} d \\ s \\ b \end{pmatrix} . \tag{83}$$

where, by convention, the states of the +2/3 quarks (u, c, t) are unmixed. In its standard parametrization, the CKM matrix reads

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13} \end{pmatrix} \tag{84}$$

where $c_{ij} = \cos\theta_{ij}$, $s_{ij} = \sin\theta_{ij}$ and i and j label the generation ($1 \leqslant i, j \leqslant 3$). In the limit $\theta_{23} = \theta_{13} = 0$, i.e., when only two generations mix, the only remaining parameter is the angle θ_{12}, called the *Cabibbo angle*, which was introduced when only the first two generations of fermions were known. The phase δ_{13}, lying between 0 and 2π, is different from zero in nature, and expresses the fact that CP invariance is violated in the case of the weak interactions. It appears in the third column and shows that CP violation is related to the existence of (at least) three generations.

The CKM mixing matrix is predicted to be unitary in the standard model. This is confirmed by all experiments so far. The 90 % confidence upper and lower limits for the *magnitude* of the complex CKM matrix are given by

Ref. 163

$$V = \begin{pmatrix} 0.97419(22) & 0.2257(10) & 0.00359(16) \\ 0.2256(10) & 0.97334(23) & 0.0415(11) \\ 0.00874(37) & 0.0407(10) & 0.999133(44) \end{pmatrix} \tag{85}$$

The CP violating phase δ_{13} is usually expressed with the *Jarlskog invariant*, defined as $J = \sin\theta_{12}\sin\theta_{13}\sin\theta_{23}^2\cos\theta_{12}\cos\theta_{13}\cos\theta_{23}\sin\delta_{13}$. This expression is independent of the definition of the phase angles; it was discovered by Cecilia Jarlskog, an important Swedish particle physicist. Its measured value is $J = 3.05(20) \cdot 10^{-5}$.

Ref. 184

2. Also neutrinos mix, in the same way as the d, s and b quarks. The determination of the matrix elements is not as complete as for the quark case. This is an intense research field. Like for quarks, also for neutrinos the mass eigenstates and the flavour eigenstates differ. There is a dedicated neutrino mixing matrix, with 6 angles, and many experiments are trying to measure the parameters. Only a few are known so far, but no surprises are expected.

3. Finally, the electroweak interaction does not show a SU(2) symmetry. Electromagnetic and the weak processes lead to a single interaction, the so-called *electroweak interaction*. The electroweak interaction is described by a *coupling constant* g and by the *weak mixing angle* θ_W. The mixing angle describes the strength of the breaking of the SU(2) symmetry.

Ref. 163

The usual electromagnetic coupling constant e is related to the electroweak coupling by

$$e = g \sin \theta_{\mathrm{w}} , \tag{86}$$

which at low four-momentum transfers is the fine structure constant with the value 1/137.036. The electroweak coupling constant g also defines the historically defined *Fermi constant* G_{F} by

$$G_{\mathrm{F}} = \frac{g^2 \sqrt{2}}{8 M_W^2} . \tag{87}$$

The broken SU(2) symmetry implies that in the *real* world, in contrast to the *ideal* SU(2) world, the intermediate vector bosons are

— the massless, neutral photon, given as $A = B \cos \theta_W + W^3 \sin \theta_W$;
— the massive neutral Z boson, given as $Z = -B \sin \theta_W + W^3 \cos \theta_W$;
— the massive charged W bosons, given as $W^{\pm} = (W^1 \mp i W^2)/\sqrt{2}$.

This implies that the electroweak interaction relates the electromagnetic coupling, the weak coupling and the intermediate boson masses by the impressive relation

$$\left(\frac{m_W}{m_Z}\right)^2 + \left(\frac{e}{g}\right)^2 = 1 \tag{88}$$

The relation is well verified by experiments.

The electroweak interaction also suggests the existence a scalar, elementary Higgs boson. The Higgs boson maintains the unitarity of longitudinal boson scattering at energies of a few TeV and gives mass to all other particles. The Higgs boson is currently being searched in a large experiment at CERN.

The Lagrangian of the electroweak interaction

If we combine the observed properties of the weak interaction mentioned above, namely its observed Feynman diagrams, its particle transforming ability, P and C violation, quark mixing, neutrino mixing and symmetry breaking, we arrive at the full Lagrangian density. It is given by:

$$
\begin{aligned}
\mathcal{L}_{\mathrm{EW}} = \; & \textstyle\sum_k \overline{\psi}_k (i\partial\!\!\!/ - m_k - \frac{g m_k H}{2 m_W}) \psi_k && \text{1. fermion mass terms}\\
& -e \textstyle\sum_k q_k \overline{\psi}_k \gamma^\mu \psi_k A_\mu && \text{2. e.m. interaction}\\
& -\tfrac{g}{2\sqrt{2}} \textstyle\sum_k \overline{\psi}_k \gamma^\mu (1-\gamma^5)(T^+ W^+_\mu + T^- W^-_\mu) \psi_k && \text{3. charged weak currents}\\
& -\tfrac{g}{2\cos\theta_W} \textstyle\sum_k \overline{\psi}_k \gamma^\mu (g^k_V - g^k_A \gamma^5) \psi_k Z_\mu && \text{4. neutral weak currents}\\
& -\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} && \text{5. electromagnetic field}\\
& -\tfrac{1}{2} W^+_{\mu\nu} W^{-\mu\nu} - \tfrac{1}{4} Z_{\mu\nu} Z^{\mu\nu} && \text{6. weak W and Z fields}\\
& +m^2_W W^+ W^- + \tfrac{1}{2} m^2_Z Z^2 && \text{7. W and Z mass terms}\\
& -gWWA - gWWZ && \text{8. cubic interaction}\\
& -\tfrac{g^2}{4}(W^4 + Z^4 + W^2 F^2 + Z^2 F^2) && \text{9. quartic interaction}\\
& +\tfrac{1}{2}(\partial^\mu H)(\partial_\mu H) - \tfrac{1}{2} m^2_H H^2 && \text{10. Higgs boson mass}\\
& -\tfrac{g m^2_H}{4 m_W} H^3 - \tfrac{g^2 m^2_H}{32 m^2_W} H^4 && \text{11. Higgs self-interaction}\\
& +(g m_W H + \tfrac{g^2}{4} H^2)(W^+_\mu W^{-\mu} + \tfrac{1}{2\cos^2\theta_w} Z_\mu Z^\mu) && \text{12. Higgs–W and Z int.}
\end{aligned}
\tag{89}
$$

The terms in the Lagrangian are easily associated to the Feynman diagrams of Figure 84:

1. this term describes the inertia of every object around us, yields the motion of fermions, and represents the kinetic energy of the quarks and leptons, as it appears in the usual Dirac equation, modified by the so-called *Yukawa coupling* to the Higgs field H and possibly by a Majorana term for the neutrinos (not shown);
2. the second term describes the well-known interaction of matter and electromagnetic radiation, and explains practically all material properties and colours observed in daily life;
3. the term is the so-called *charged current interaction*, due to exchange of virtual W bosons, that is responsible for the beta decay and for the fact that the Sun is shining;
4. this term is the *neutral current interaction*, the '$V - A$ theory' of George Sudarshan, that explains the elastic scattering of neutrinos in matter;
5. this term represents the kinetic energy of photons and yields the evolution of the electromagnetic field in vacuum, thus the basic Maxwell equations;
6. this term represents the kinetic energy of the weak radiation field and gives the evolution of the intermediate W and Z bosons of the weak interaction;
7. this term is the kinetic energy of the vector bosons;
8. this term represents the triple vertex of the self-interaction of the vector boson;
9. this term represents the quadruple vertex of the self-interaction of the vector boson;
10. this term is predicted to be the kinetic energy of Higgs boson;
11. this term is predicted to be the self-interaction of the Higgs;
12. the last term is predicted to represent the interaction of the vector bosons with the Higgs boson that restoring unitarity at high energies.

Let us look into the formal details. The quantities appearing in the Lagrangian are:

— The wave functions $\psi_k = (\nu'_k \;\; l^-_k)$ for leptons and $(u_k \;\; d'_k)$ for quarks are the left-handed fermion fields of the k-th fermion generation; every component is a spinor. The index $k = 1, 2, 3$ numbers the generation: the value 1 corresponds to (u d $\nu_e e^-$), the second generation is (c s $\nu_\mu \mu^-$) and the third (t b $\nu_\tau \tau^-$). The ψ_k transform as doublets under SU(2); the right handed fields are SU(2) singlets.

In the doublets, one has

$$d'_k = \sum_l V_{kl} d_l \;, \tag{90}$$

where V_{kl} is the Cabibbo–Kobayashi–Maskawa mixing matrix, d'_k are the quark flavour eigenstates and d_k are the quark mass eigenstates. A similar expression holds for the mixing of the neutrinos:

$$\nu'_k = \sum_l P_{kl} \nu_l \;, \tag{91}$$

where P_{kl} is the Pontecorvo–Maki–Nakagawa–Sakata mixing matrix, ν'_l the neutrino flavour eigenstates and ν_l the neutrino mass eigenstates.

— For radiation, A^μ and $F^{\mu\nu}$ is the field of the massless vector boson of the electromagnetic field, the *photon* γ.

$W^\pm_\mu$ are the massive charged gauge vector bosons of the weak interaction; the corresponding particles, W^+ and W^-, are each other's antiparticles.

Z_μ is the field of the massive neutral gauge vector boson of the weak interactions; the neutral vector boson itself is usually called Z^0.

— H is the field of the neutral scalar Higgs boson H^0, the only elementary scalar particle in the standard model. It is not yet discovered, so that all terms involving H still need final confirmation.

— Two charges appear, one for each interaction. The number q_k is the well-known *electric charge* of the particle ψ_k in units of the positron charge. The number $t_{3L}(k)$ is the *weak isospin*, or *weak charge*, of fermion k, whose value is $+1/2$ for u_k and ν_k and is $-1/2$ for d_k and l_k. These two charges together define the so-called *vector coupling*

$$g^k_V = t_{3L}(k) - 2q_k \sin^2\theta_W \tag{92}$$

and the axial coupling

$$g^k_A = t_{3L}(k) \;. \tag{93}$$

The combination $g^k_V - g^k_A$, or $V - A$ for short, expresses the maximal violation of P and C parity in the weak interaction.

— The operators T^+ and T^- are the weak isospin raising and lowering operators. Their action on a field is given e.g. by $T^+ l^-_k = \nu_k$ and $T^- u_k = d_k$.

We see that the Lagrangian indeed contains all the ideas developed above. The electroweak Lagrangian is *unique*: it could not have a different mathematical form, because both the electromagnetic terms and the weak terms are fixed by the requirements of Lorentz invariance, U(1) and broken SU(2) gauge invariance, permutation symmetry and renormalizability.

Curiosities about the weak interaction

The Lagrangian of the weak interaction has been checked by thousands of experiments. Many experiments have been design specifically to probe it to the highest precision possible. In all these cases, no contradictions between observation and calculation has ever been found. Ref. 163

Nevertheless, the weak interaction, with its breaking of parity and its elusive neutrino, exerts a deep fascination on all those who have explored it. Let us explore this fascination a bit more. Ref. 185

* *

The weak interaction is responsible for the burning of hydrogen to helium in the early universe. Without helium, there would be no path to make still heavier elements. Thus we owe our own existence to the weak interaction.

* *

The weak interaction is required to have an excess of matter over antimatter. Without the parity violation of the weak interactions, there would be no matter at all in the universe. Also this property of the weak interaction is necessary for our own existence.

* *

Through the emitted neutrinos, the weak interaction helps to get the energy out of a supernova. If that were not the case, black holes would form in almost every supernova, heavier elements – of which we are made – would not have been spread out into space, and we would not exist.

* *

The fascination of the Higgs boson is underlined by the fact that it is the only fundamental observable field which bears the name of a physicist.

* *

The weak interaction is also responsible for the heat produced inside the Earth. This heat keeps the magma liquid. As a result, the weak interaction, despite its weakness, is responsible for most earthquakes, tsunamis and volcanic eruptions.

* *

Ref. 186 The paper by Peter Higgs on the boson named after him is only 79 lines long, and has only five equations.

* *

Beta decay, due to the weak interaction, separates electrons and protons. Only in 2005 people have managed to propose practical ways to use this effect to build long-life batteries that could be used in satellites. Future will tell whether the method will be successful.

* *

Already in 1957, the great physicist Bruno Pontecorvo imagined that travelling neutrinos could spontaneously change into their own antiparticles. Today, it is known experimentally that travelling neutrinos can change generation, and one speaks of *neutrino oscillations*.

* *

Ref. 187 Every second around 10^{16} neutrinos fly through our body. They have five sources:

- *Solar neutrinos* arrive on Earth at $6 \cdot 10^{14}\,/\mathrm{m^2s}$, with an energy from 0 to 0.42 MeV; they are due to the p-p reaction in the sun; a tiny is due to the ^{8}B reaction and has energies up to 15 MeV.
- *Atmospheric neutrinos* are products of cosmic rays hitting the atmosphere, consist of 2/3 of muon neutrinos and one third of electron neutrinos, and have energies mainly between 100 MeV and 5 GeV.
- Page 137 *Earth neutrinos* from the radioactivity that keeps the Earth warm form a flux of $6 \cdot 10^{10}\,/\mathrm{m^2s}$.
- *Fossil neutrinos* from the big bang, with a temperature of 1.95 K are found in the universe with a density of $300\,\mathrm{cm^{-3}}$, corresponding to a flux of $10^{15}\,/\mathrm{m^2s}$.
- *Man-made neutrinos* are produced in nuclear reactors (at 4 MeV) and as neutrino beams in accelerators, using pion and kaon decay. A standard nuclear plant produces $5 \cdot 10^{20}$ neutrinos per second. Neutrino beams are produced, for example, at the CERN in Geneva. They are routinely sent 700 km across the Earth to central Italy, where they are detected.

  They are mainly created in the atmosphere by cosmic radiation, but also coming directly from the background radiation and from the centre of the Sun. Nevertheless, during our whole life – around 3 thousand million seconds – we have only a 10 % chance that one of them interacts with one of the $3 \cdot 10^{27}$ atoms of our body. The reason is that the weak interaction is felt only over distances less than 10^{-17} m, about 1/100th of the diameter of a proton. The weak interaction is indeed weak.

* *

In the years 1993 and 1994 an intense marketing campaign was carried out across the United States of America by numerous particle physicists. They sought funding for the 'superconducting supercollider', a particle accelerator with a circumference of 80 km. This should have been the largest machine ever built, with a planned cost of more than twelve thousand million dollars, aiming at finding the Higgs boson before the Europeans would do so at a fraction of that cost. The central argument brought forward was the following: since the Higgs boson was the basis of mass, it was central to US science to know about it first. Apart from the issue of the relevance of the conclusion, the worst is that the premise is wrong.

Page 170 We have seen above that 99 % of the mass of protons, and thus of the universe, is due to confinement; this part of mass appears even if the quarks are approximated as massless. The Higgs boson is *not* responsible for the origin of mass itself; it just might shed some light on the issue. Ref. 172 In particular, discovering the Higgs boson will not allow to calculate or understand the mass of any particle. The whole campaign was a classic case of disinformation, and many people involved have shown their lack of honesty.* In the end, the project was stopped, mainly for financial reasons.

> Difficile est saturam non scribere.**
>
> Juvenal, *Saturae* 1, 30.

* *

There is no generally accepted name for the quantum field theory of the weak interaction. Expressions such as *quantum asthenodynamics* (QAD) – from the Greek word for 'weak' – have not yet been universally adopted.

* *

The weak interaction is so weak that a neutrino–antineutrino annihilation – which is only possible by producing a massive intermediate Z boson – has never been observed up to this day.

* *

Only one type of particles interacts (almost) only weakly: neutrinos. Neutrinos carry no electric charge, no colour charge and almost no gravitational charge (mass). To get an impression of the weakness of the weak interaction, it is usually said that the probability of a neutrino to be absorbed by a lead screen of the thickness of one light-year is less than 50 %. The universe is thus essentially empty for neutrinos. Is there room for bound states of neutrinos circling masses? How large would such a bound state be? Can we imagine bound states, which would be called neutrinium, of neutrinos and antineutrinos circling each other? The answer depends on the mass of the neutrino. Bound states of massless particles do not exist. They could and would decay into two free massless particles.***

Since neutrinos are massive, a neutrino–antineutrino bound state is possible in principle. Challenge 130 ny How large would it be? Does it have excited states? Can they ever be detected? These issues are still open.

* *

Do ruminating *cows* move their jaws equally often in clockwise and anticlockwise direction? In 1927, the theoretical physicists Pascual Jordan and Ralph de Laer Kronig pub-

* We should not be hypocrites. The supercollider lie is negligible when compared to other lies. The biggest lie in the world is probably the one that states that to ensure its survival, the USA government need to spend more on the military than all other countries in the world combined. This lie is, every single year, around 40 times as big as the once-only supercollider lie. Many other governments devote even larger percentages of their gross national product to their own version of this lie. As a result, the defence spending lie is directly responsible for most of the poverty in all the countries that use it.

** 'It is hard not to be satirical.'

*** In particular, this is valid for photons bound by gravitation; this state is not possible.

Ref. 188 lished a study showing that in Denmark the two directions are almost equally distributed. The rumination direction of cows is thus not related to the weak interaction.

* *

The weak interaction plays an important part in daily life. First of all, the Sun is shining. The fusion of two protons to deuterium, the first reaction of the hydrogen cycle, implies Page 148 that one proton changes into a neutron. This transmutation and the normal beta decay Challenge 131 e have the same first-order Feynman diagram. The weak interaction is thus essential for the burning of the Sun. The weakness of the process is one of the guarantees that the Sun will continue burning for quite some time.

* *

Of course, the weak interaction is responsible for radioactive beta decay, and thus for part of the radiation background that leads to mutations and thus to biological evolution.

* *

What would happen if the Sun suddenly stopped shining? Obviously, temperatures would fall by several tens of degrees within a few hours. It would rain, and then all water would freeze. After four or five days, all animal life would stop. After a few weeks, the oceans would freeze; after a few months, air would liquefy.

* *

Not everything about the Sun is known. For example, the neutrino flux from the Sun oscillates with a period of 28.4 days. That is the same period with which the magnetic field of the Sun oscillates. The connections are still being studied.

* *

The energy carried away by neutrinos is important in supernovas; if neutrinos would not carry it away, supernovas would collapse instead of explode. That would have prevented the distribution of heavier elements into space, and thus our own existence.

* *

Even earlier on in the history of the universe, the weak interaction is important, as it prevents the symmetry between matter and antimatter, which is required to have an excess of one over the other in the universe.

* *

Due to the large toll it placed on society, research in nuclear physics, like poliomyelitis, has almost disappeared from the planet. Like poliomyelitis, nuclear research is kept alive only in a few highly guarded laboratories around the world, mostly by questionable figures, in order to build dangerous weapons. Only a small number of experiments carried on by a few researchers are able to avoid this involvement and continue to advance the topic.

* *

Interesting aspects of nuclear physics appear when powerful lasers are used. In 1999, a

British team led by Ken Ledingham observed laser induced uranium fission in ^{238}U nuclei. In the meantime, this has even be achieved with table-top lasers. The latest feat, in 2003, was the transmutation of ^{129}I to ^{128}I with a laser. This was achieved by focussing a 360 J laser pulse onto a gold foil; the ensuing plasma accelerates electrons to relativistic speed, which hit the gold and produce high energy γ rays that can be used for the transmutation.

Ref. 189

A summary of the electroweak interaction

The electroweak interaction is described by a non-abelian gauge theory based on a broken SU(2) gauge group for weak processes and an unbroken U(1) group for electrodynamic processes. This description matches the observed properties of neutrinos, of the W and Z boson, of parity violation, of β decay, of the heat production inside the Earth, of several important reactions in the Sun, and of the origin of matter in the universe. The weak interaction might be weak, but it is a bit everywhere.

Theory and experiment disagree only on the Higgs boson, which has not yet been observed, due to the large investments that were needed to do so. In all other cases, theory and experiment agree whenever comparisons have been made.

The limitations of the theory are only conceptual. Like in all of quantum field theory, also in the case of the weak interaction the mathematical form of the Lagrangian is almost uniquely defined by requiring renormalizability, Lorentz invariance, and (broken) gauge invariance – SU(2) in this case. We say again 'almost', as we did for the case of the strong interaction, because the Lagrangian contains a few parameters that remain unexplained:

— The two coupling constants g and g' of the electroweak interaction are unexplained. (They define weak mixing angle $\theta_W = \arctan(g'/g)$.)
— The mass $M_Z = 91\,\mathrm{GeV}/c^2$ of the neutral Z boson is unexplained.
— The mass M_H of the neutral scalar Higgs boson is unknown, since the particle has not been discovered yet; the present limit is $m_H > 110\,\mathrm{GeV}/c^2$.
— The number $n = 3$ of generations is unexplained.
— The masses of the six leptons and the six quarks are unexplained.
— The four parameters of the Cabibbo–Kobayashi–Maskawa quark mixing matrix and the six parameters of the neutrino mixing matrix are unexplained, including the CP violating phases.
— The properties of space-time, in particular its Lorentz invariance, its continuity and the number of its dimensions are obviously all unexplained and assumed from the outset.
— It is also not known how the weak interaction behaves in strong gravity, thus in strongly curved space-time.

Before exploring how to overcome these limitations, we summarize all results so far in the so-called *standard model of particle physics*.

Chapter 9

THE STANDARD MODEL OF ELEMENTARY PARTICLE PHYSICS – AS SEEN ON TELEVISION

The following table lists the known and predicted elementary particles. The list has not changed since the mid-1970s, mainly because of the inefficient use that was made of the relevant research budgets across the world since then, and possibly also because the list is not far from final. The table contains *all knowledge* about matter and radiation. It is the basis for materials science, geology, chemistry, biology, medicine, the neurosciences and psychology. For this reasons, it regularly features on television.

TABLE 19 The elementary particles

Radiation	electromagnetic interaction	weak interaction	strong interaction
	γ	W^+, W^- Z^0	g_1 … g_8
	photon	intermediate vector bosons	gluons
	Radiation particles are bosons with spin 1. W^- is the antiparticle of W^+; all others are their own antiparticles.		
Matter	generation 1	generation 2	generation 3
Leptons	e	μ	τ
	ν_e	ν_μ	ν_τ
Quarks	d	s	t
(each in three colours)	u	c	b
	Matter particles are fermions with spin 1/2; all have a corresponding antiparticle.		
Hypothetical matter			
Higgs boson	H	predicted to have spin 0 and to be elementary	

The next table lists *all* properties of the elementary particles. For reasons of space, colour and weak isospin are not mentioned explicitly. Also the decay modes of the unstable particles are not given in detail; they are found in the standard references. The table that follows is fascinating. It allows us to give a *complete* characterization of the intrinsic properties of any *composed* moving entity, be it an object or an image. The aim from the beginning of our study of motion, namely to have a complete list of the *intrinsic properties* of moving entities, is thus achieved.

Ref. 163

Page 27

The other aim that we formulated at the same time was to have a complete list of all *state properties*. This aim is also achieved, namely by the wave function. Were it not for the possibility of space-time curvature, we would be at the end of our exploration. In short, the previous and the following table lists everything about motion in flat space-time.

TABLE 20 Elementary particle properties

Particle	Mass m[a]	Lifetime τ or energy width,[b] main decay modes	Isospin I, spin J,[c] parity P, charge parity C	Charge, isospin, strangeness,[c] charm, beauty, topness: $QISCBT$	Lepton & baryon[e] numbers LB
Elementary radiation (bosons)					
photon γ	0 ($<10^{-53}$ kg)	stable	$I(J^{PC}) = 0, 1(1^{--})$	000000	0, 0
$W^{\pm}$	80.398(25) GeV/c^2	2.124(41) GeV 67.60(27)% hadrons, 32.12(36)% $l^{+}\nu$	$J = 1$	±100000	0, 0
Z	91.1876(21) GeV/c^2	2.4952(23) GeV/c^2 69.91(6)% hadrons 10.0974(69)% $l^{+}l^{-}$	$J = 1$	000000	0, 0
gluon	0	stable	$I(J^{P}) = 0(1^{-})$	000000	0, 0
Elementary matter (fermions): leptons					
electron e	9.109 381 88(72) · 10^{-31} kg = 81.871 0414(64) pJ/c^2 = 0.510 998 910(13) MeV/c^2 = 0.000 548 579 909 43(23) u; gyromagnetic ratio $\mu_e/\mu_B = -1.001\,159\,652\,1811(7)$; electric dipole moment $d = (0.7 \pm 0.7) \cdot 10^{-29}\,e\,$m[f]	$> 13 \cdot 10^{30}$ s	$J = \frac{1}{2}$	−100 000	1, 0
muon μ	0.188 353 109(16) yg = 105.658 376(4) MeV/c^2 = 0.113 428 9256(29) u	2.197 03(4) µs 99 % $e^{-}\bar{\nu}_e\nu_\mu$	$J = \frac{1}{2}$	−100000	1, 0

Particle	Mass m [a]	Lifetime τ or energy width, [b] main decay modes	Isospin I, spin J, [c] parity P, charge parity C	Charge, isospin, strangeness, [c] charm, beauty, topness: $QISCBT$	Lepton & baryon [e] numbers $L\,B$
	gyromagnetic ratio $\mu_\mu/(e\hbar/2m_\mu) = -1.001\,165\,9208(6)$				
	electric dipole moment $d = (3.7 \pm 3.4) \cdot 10^{-21}\,e\,\text{m}$				
tau τ	$1.776\,84(17)\,\text{GeV}/c^2$	$290.6(1.0)$ fs	$J = \frac{1}{2}$	-100000	$1, 0$
el. neutrino ν_e	$< 2\,\text{eV}/c^2$		$J = \frac{1}{2}$		$1, 0$
muon neutrino ν_μ	$< 2\,\text{eV}/c^2$		$J = \frac{1}{2}$		$1, 0$
tau neutrino ν_τ	$< 2\,\text{eV}/c^2$		$J = \frac{1}{2}$		$1, 0$
Elementary matter (fermions): quarks [g]					
up u	1.5 to 3.3 MeV/c^2	see proton	$I(J^P) = \frac{1}{2}(\frac{1}{2}^+)$	$+\frac{2}{3}+\frac{1}{2}0000$	$0, \frac{1}{3}$
down d	3.5 to 6 MeV/c^2	see proton	$I(J^P) = \frac{1}{2}(\frac{1}{2}^+)$	$-\frac{1}{3}-\frac{1}{2}0000$	$0, \frac{1}{3}$
strange s	70 to 130 MeV/c^2		$I(J^P) = 0(\frac{1}{2}^+)$	$-\frac{1}{3}0-1000$	$0, \frac{1}{3}$
charm c	$1.27(11)\,\text{GeV}/c^2$		$I(J^P) = 0(\frac{1}{2}^+)$	$+\frac{2}{3}00+100$	$0, \frac{1}{3}$
bottom b	$4.20(17)\,\text{GeV}/c^2$	$\tau = 1.33(11)$ ps	$I(J^P) = 0(\frac{1}{2}^+)$	$-\frac{1}{3}000-10$	$0, \frac{1}{3}$
top t	$171.2(2.1)\,\text{GeV}/c^2$		$I(J^P) = 0(\frac{1}{2}^+)$	$+\frac{2}{3}0000+1$	$0, \frac{1}{3}$
Hypothetical elementary matter (boson)					
Higgs [h] H	$> 114\,\text{GeV}/c^2$		$J = 0$		

Notes:

a. See also the table of SI prefixes on page 348. About the eV/c^2 mass unit, see page 252.

b. The *energy width* Γ of a particle is related to its lifetime τ by the indeterminacy relation $\Gamma\tau = \hbar$. There is a difference between the *half-life* $t_{1/2}$ and the lifetime τ of a particle: they are related by $t_{1/2} = \tau \ln 2$, where $\ln 2 \approx 0.693\,147\,18$; the half-life is thus shorter than the lifetime. The unified *atomic mass unit* u is defined as 1/12 of the mass of a carbon 12 atom at rest and in its ground state. One has $1\,\text{u} = \frac{1}{12}m(^{12}\text{C}) = 1.660\,5402(10)$ yg.

c. To keep the table short, the header does not explicitly mention *colour*, the charge of the strong interactions. This has to be added to the list of basic object properties. Quantum numbers containing the word 'parity' are multiplicative; all others are additive. Time parity T (not to be confused with topness 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 $G = (-1)^{IC}$.

The header also does not mention the *weak charge* of the particles. The details on weak charge g, or, more precisely, on the *weak isospin*, a quantum number assigned to all left-handed fermions (and right-handed anti-fermions), but to no right-handed fermion (and no left-handed antifermion), are given in the section on the weak interactions. Page 178

d. 'Beauty' is now commonly called *bottomness*; similarly, 'truth' is now commonly called *topness*. The signs

of the quantum numbers S, I, 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.
e. If supersymmetry exists, R-parity must be added to this column. R-parity is a multiplicative quantum number 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 R-odd.
Ref. 190, Ref. 191 *f*. The electron radius is less than 10^{-22} m. It is possible to store single electrons in traps for many months.
g. See page 170 for the precise definition and meaning of the quark masses.
h. Currently a hypothetical particle. It is also known to be lighter than 185 GeV. In addition, in 2010, the range between 158 GeV and 175 GeV has been excluded with 95% probability.

To complete the standard model, apart from the above two tables, we need the Lagrangians of the electroweak and strong interaction. The combination of the Lagrangians, based on the U(1), SU(3) and broken SU(2) gauge groups, is possible only in one particular way. The resulting Lagrangian is a bit complex; since it does not contain more information than already discussed, it is not written down here. All the information on the standard model Lagrangian is contained in the Feynman diagrams of the electroweak and strong interactions, taking into account the mixing of quarks and neutrinos.

In summary, the standard model includes a minimum action, a maximum speed, electric charge quantization and conservation, colour conservation and weak charge conservation. With these basic ideas, the standard model describes *every* observation ever made in flat space-time.

Summary and open questions

The standard model of particle physics clearly distinguishes *elementary* from *composed* particles. It provides the full list of properties that characterizes a particle – and thus any moving object and image. The properties are: mass, spin, charge, colour, weak isospin, parity, charge parity, isospin, strangeness, charm, topness, beauty, lepton number and baryon number.

The standard model also describes electromagnetic and nuclear interactions as gauge theories, i.e., as exchange of virtual radiation particles. The standard model describes the three types of radiation that are observed in nature at all experimentally accessible energies. As a result, the standard model describes the structure of the atoms, their formation in the history of the universe, the properties of matter and the mechanisms of life.

In short, the standard model realizes the dream of Leucippus and Democritus, plus a bit more: we know the bricks that compose all of matter and radiation, and in addition we know precisely how they move, interact and transform.

But in addition, we also know what we still do *not* know:

— we have not yet observed the Higgs boson;
— we do not know the origin of the coupling constants;
— we do not know the origin of the masses of the particles;
— we do not know the origin of the mixing and CP violation parameters;
— we do not know the origin of the gauge groups;
— we do not know the origin of the three generations;
— we do not know why positrons and protons have the same charge;
— we do not know whether the particle concept survives at high energy;

— we do not know what happens in curved space-time.

To study these issues, the simplest way is to explore nature at particle energies that are as high as possible. There are two methods: building large experiments and exploring hypothetical models. Both are useful.

Chapter 10

DREAMS OF UNIFICATION

> Materie ist geronnenes Licht.*
>
> Albertus Magnus

Is there a common origin to the three particle interactions? We have seen in the preceding chapters that the Lagrangian densities of the three gauge interactions are determined almost uniquely by two types of requirements: to possess a certain gauge symmetry, and to possess mathematical consistency (Lorentz invariance and renormalizability). The search for *unification* of the interactions thus seems to
Challenge 132 s
require the identification of the one, unified symmetry of nature. (Do you agree?)

In the past decades, several candidate models have fuelled the hope to achieve unification through higher symmetry: grand unification, supersymmetry, conformal invariance and coupling constant duality. We start with the first, which is conceptually the simplest.

Grand unification

At all measured energies up to the year 2009, thus below about 1 TeV, there are no contradictions between the Lagrangian of the standard model and observation. On the other hand, the Lagrangian itself can be seen as a low energy approximation. It should thus be possible – attention, this a belief – to find a unifying symmetry that *contains* the symmetries of the electroweak and strong interactions as subgroups and thus as different aspects of a single, unified interaction; we can then examine the physical properties that follow and compare them with observation. This approach, called *grand unification*, attempts the unified description of all types of matter. All known elementary particles are seen as fields which appear in a Lagrangian determined by a single gauge symmetry group.

Like for each gauge theory described so far, also the grand unified Lagrangian is fixed by the symmetry group, the representation assignments for each particle, and the corresponding coupling constant. A general search for the symmetry group starts with all
Ref. 192
those (semisimple) Lie groups which contain $U(1) \times SU(2) \times SU(3)$. The smallest groups with these properties are SU(5), SO(10) and E(6); they are defined in Appendix C. For each of these candidate groups, the predicted consequences of the model must be studied
Ref. 193
and compared with experiment.

* 'Matter is coagulated light.' Albertus Magnus (b. *c.* 1192 Lauingen, d. 1280 Cologne) was the most important thinker of his time.

The data

Grand unification models make several predictions that can be matched with experiment. First of all, any grand unified model predicts relations between the quantum numbers of quarks and those of leptons. In particular, grand unification successfully explains why the electron charge is exactly the opposite of the proton charge.

Grand unification models predict a value for the weak mixing angle θ_{w}; its value is not fixed by the standard model. The most frequently predicted value,

$$\sin^2\theta_{w,th} = 0.2 \tag{94}$$

is close to the measured value of

$$\sin^2\theta_{w,ex} = 0.231(1) , \tag{95}$$

which is not a good match, but might be correct.

All grand unified models predict the existence of magnetic monopoles, as was shown by Gerard 't Hooft. However, despite extensive searches, no such particles have been found yet. Monopoles are important even if there is only one of them in the whole universe: the existence of a single monopole implies that electric charge is quantized. If monopoles were found, grand unification would explain why electric charge appears in multiples of a smallest unit.

Ref. 194

Grand unification predicts the existence of heavy intermediate vector bosons, called *X bosons*. Interactions involving these bosons do not conserve baryon or lepton number, but only the difference $B - L$ between baryon and lepton number. To be consistent with experiment, the X bosons must have a mass of the order of 10^{16} GeV. However, this mass is outside the range of experiments, so that the prediction cannot be tested directly.

Most spectacularly, the X bosons of grand unification imply that the *proton decays*. This prediction was first made by Pati and Salam in 1974. If protons decay, means that neither coal nor diamond* – nor any other material – would be for ever. Depending on the precise symmetry group, grand unification predicts that protons decay into pions, electrons, kaons or other particles. Obviously, we know 'in our bones' that the proton lifetime is rather high, otherwise we would die of leukaemia; in other words, the low level of cancer in the world already implies that the lifetime of the proton is larger than about 10^{16} years.

Detailed calculations for the proton lifetime τ_p using the gauge group SU(5) yield the expression

$$\tau_p \approx \frac{1}{\alpha_G^2(M_X)} \frac{M_X^4}{M_p^5} \approx 10^{31\pm1}\ \mathrm{a} \tag{96}$$

where the uncertainty is due to the uncertainty of the mass M_X of the gauge bosons involved and to the exact decay mechanism. Several large experiments aim to measure this lifetime. So far, the result is simple but clear. Not a single proton decay has ever been

* As is well known, diamond is not stable, but metastable; thus diamonds are not for ever, but coal might be, as long as protons do not decay.

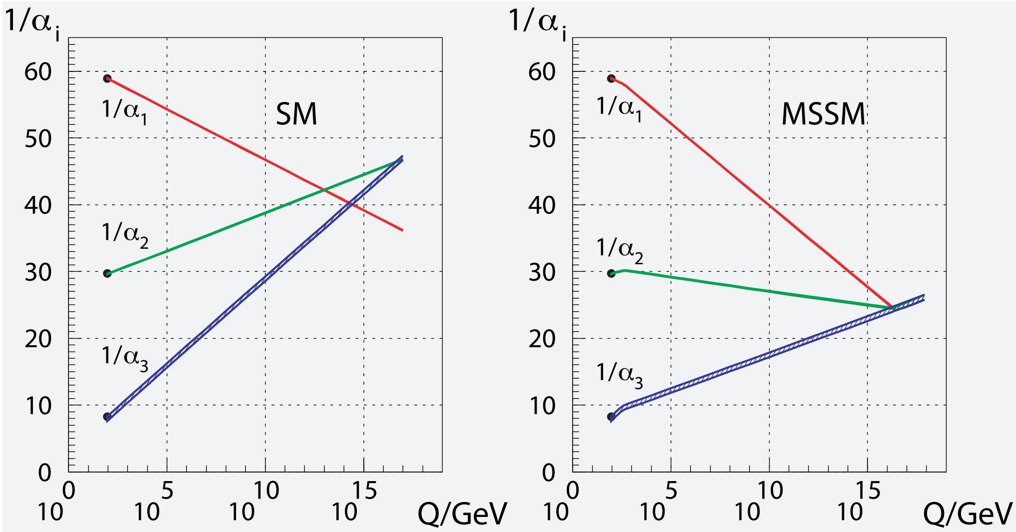


FIGURE 85 The behaviour of the three coupling constants with energy, for simple grand unification (left) and for the minimal supersymmetric model (right); the graph shows the constants $\alpha_1 = \frac{5}{3}\alpha_{em}/\cos^2\theta_W$ for the electromagnetic interaction (the factor 5/3 appears in GUTs), $\alpha_2 = \alpha_{em}/\sin^2\theta_W$ for the weak interaction and the strong coupling constant $\alpha_3 = \alpha_s$ (© Wim de Boer)

Ref. 195 observed. The data can be summarized by

$$\begin{aligned} \tau(\mathrm{p} \rightarrow e^+ \pi^0) &> 5 \cdot 10^{33}\,\mathrm{a} \\ \tau(\mathrm{p} \rightarrow K^+ \bar{\nu}) &> 1.6 \cdot 10^{33}\,\mathrm{a} \\ \tau(\mathrm{n} \rightarrow e^+ \pi^-) &> 5 \cdot 10^{33}\,\mathrm{a} \\ \tau(\mathrm{n} \rightarrow K^0 \bar{\nu}) &> 1.7 \cdot 10^{32}\,\mathrm{a} \end{aligned} \qquad (97)$$

These values are higher than the prediction by SU(5) models. For the other gauge group candidates the situation is not settled yet.

The state of grand unification

To settle the issue of grand unification definitively, one last prediction of grand unification remains to be checked: the unification of the coupling constants. Most estimates of the grand unification energy are near the Planck energy, the energy at which gravitation starts to play a role even between elementary particles. As grand unification does not take gravity into account, for a long time there was a doubt whether something was lacking in the approach. This doubt changed into certainty when the precision measurements of the Ref. 196 coupling constants became available. This happened in 1991, when these measurements were put into the diagram of Figure 85. The GUT prediction of the way the constants evolve with energy implies that the three constants do *not* meet at the grand unification energy. Simple grand unification by SU(5), SU(10) or E_6 is thus ruled out by experiment.

This state of affairs is changed if *supersymmetry* is taken into account. Supersymmetry is a hypothesis on the way to take into account low-energy effects of gravitation in the particle world. Supersymmetry predicts new particles that change the curves at intermediate energies, so that they all meet at a grand unification energy of about 10^{16} GeV. (The line thicknesses in Figure 85 represent the experimental errors.) The inclusion of supersymmetry also puts the proton lifetime prediction back to a value higher (but not by much) than the present experimental bound and predicts the correct value of the mixing angle. With supersymmetry, we can thus retain the advantages of grand unification (charge quantization, one coupling constant) without being in contradiction with experiments. The predicted particles, not yet found, are in a region accessible to the LHC Vol. V, page 199 collider at CERN in Geneva. We will explore supersymmetry in a bit more detail below.

Pure grand unification is thus in contradiction with experiments. This is not a surprise, as its goal, to unify the description of matter, *cannot* been achieved in this way. Indeed, the gauge group must be introduced at the very beginning, because grand unification cannot deduce it from a general principle. Neither does grand unification tell us completely which elementary particles exist in nature. In other words, grand unification only *shifts* the open questions of the standard model to the next level, while keeping most of them unanswered. The name 'grand unification' is ridiculous.

The story of grand unification is a first hint that looking at higher energies using only low-energy concepts is not the way to solve the mystery of motion. We definitively need to continue our adventure.

Searching for higher symmetries

Since we want to reach the top of Motion Mountain, we go on. We have seen in the preceding sections that the symmetry properties are the main ingredients of the Lagrangian that describes nature. The discovery of the correct symmetry, together with mathematical consistency, usually restricts the possible choices for a Lagrangian down to a limited number (to one in the best case), and then allows to make experimental predictions.

The history of particle physics has also shown that progress was always coupled to the discovery of *larger* symmetries, in the sense that the newly discovered symmetries always included the old ones as a subgroup. Therefore, in the twentieth century, researchers searched for the largest possible symmetry that is consistent with experiments on one hand and with gauge theories on the other hand. Since grand unification is a failure, a better approach is to search directly for a symmetry that includes gravity.

Supersymmetry

In the search for possible symmetries of a Lagrangian describing a gauge theory, one way to proceed is to find general theorems which restrict the symmetries that a Lagrangian Ref. 197 can possibly have.

A well-known theorem by Coleman and Mandula states that if the symmetry transformations transform fermions into fermions and bosons into bosons, no quantities other than the following can be conserved:

- the energy momentum tensor $T^{\mu\nu}$, a consequence of the *external* Poincaré space-time symmetry, and
- the internal quantum numbers, all scalars, associated with each gauge group generator – such as electric charge, colour, etc. – and consequences of the *internal* symmetries

of the Lagrangian.

But, and here comes a way out, if transformations that *mix* fermions and bosons are considered, new conserved quantities become possible. This family of symmetries includes gravity and came to be known as *supersymmetry*. Its conserved quantities are not scalars but spinors. Therefore, *classical* supersymmetry does not exist; it is a purely quantum-mechanical symmetry. The study of supersymmetry is a vast research field in its own. Supersymmetry generalizes gauge theory to super-gauge theory. The possible super-gauge groups have been completely classified.

Supersymmetry can be extended to incorporate gravitation by changing it into a local gauge theory; in that case it is called *supergravity*. Supergravity is based on the idea that coordinates can be fermionic as well as bosonic. Supergravity thus makes specific statements on the behaviour of space-time at small distances. Supergravity predicts N additional conserved, spinorial charges. The number N lies between 1 and 8; each value leads to a different candidate Lagrangian. The simplest case is called $N = 1$ supergravity.

In short, supersymmetry is an option to unify matter and radiation at low energies. Supersymmetry, and in particular $N = 1$ supergravity, might be an approximation to reality.

Supersymmetric models make a number of predictions that can be tested by experiment.

— Supersymmetry predicts partners to the usual elementary particles, called *sparticles*. Fermions are predicted to have boson partners, and vice versa. For example, supersymmetry predicts a *photino* as fermionic partner of the photon, *gluinos* as partner of the gluons, a *selectron* as partner of the electron, etc. However, none of these particles have been observed yet; this might be due to their large mass.
— Supersymmetry allows for the unification of the coupling constants in a way compatible with the data, as shown already above. Vol. V, page 198
— Supersymmetry slows down the proton decay rates predicted by grand unified theories. The slowed-down rates are compatible with observation.
— Supersymmetry predicts electric dipole moments for the neutron (and other elementary particles). The largest predicted values, $10^{-30} e$ m, are in contradiction with observations; the smallest predictions have not yet been reached by experiment. In comparison, the values expected from the standard model are at most $10^{-33} e$ m. This is a vibrant experimental research field.

In summary, the experimental situation of supersymmetry is weak. Is supersymmetry is an ingredient to the unified theory? The safe answer is: nobody knows yet. The optimistic answer is: supersymmetry might be discovered in future collider experiments at CERN in Geneva. The pessimistic answer is: supersymmetry has not been confirmed by experiment, but is a belief system made up to correct the failings of grand unified theories. Time will tell which answer is correct.

Dualities – the most incredible symmetries of nature

One of the great discoveries of theoretical physics took place in 1977, when Claus Montonen and David Olive proved that the standard concept of symmetry could be expanded dramatically in a different and new way.

The standard class of symmetry transformations, which turns out to be only the first class, acts on fields. This class encompasses gauge symmetries, space-time symmetries, motion reversal, parities, flavour symmetries and supersymmetry.

The second, new class is quite different. If one takes all coupling constants of nature, collected in the table on page 239, one can imagine that they are members of a continuous space of all possible coupling constants, called the *parameter space*.* Montonen and Olive showed that there are transformations in parameter space that leave nature invariant. These transformations thus form a new, second class of symmetries of nature.

In fact, we already encountered a member of this class: renormalization symmetry. But Olive and Montonen expanded the symmetry class considerably by the discovery of Vol. III, page 77 *electromagnetic duality*. Electromagnetic duality is a discrete symmetry exchanging

$$e \leftrightarrow \frac{4\pi\hbar c}{e} \tag{98}$$

where the right hand turns out to be the unit of magnetic charge. Electro-magnetic duality thus relates the electric charge e and the magnetic charge m

$$Q_{\text{el}} = me \quad \text{and} \quad Q_{\text{mag}} = ng = 2\pi\hbar c/e \tag{99}$$

and puts them on equal footing. In other words, the transformation exchanges

$$\alpha \leftrightarrow \frac{1}{\alpha} \quad \text{or} \quad \frac{1}{137.04} \leftrightarrow 137.04\,, \tag{100}$$

and thus exchanges weak and strong coupling. In other words, electromagnetic duality relates a regime where particles make sense (the low coupling regime) with one where particles do not make sense (the strong coupling regime). It is the most mind-boggling symmetry ever conceived.

Dualities are among the deepest connections of physics. They contain $\hbar$ and are thus intrinsically quantum. They do not exist in classical physics and show that quantum theory is more fundamental than classical physics. More clearly stated, dualities are intrinsically non-classical symmetries. Dualities confirm that quantum theory stands on its own.

Obviously, if we want to understand the values of unexplained parameters such as coupling constants, the obvious thing to do is to study *all* possible symmetries in parameter space, thus all possible symmetries of the second class, possibly combining them with those of the first symmetry class. (Indeed, the combination of duality with supersymmetry is studied in string theory.) Page 141

These investigations showed that there are several types of dualities, which all are *non-perturbative* symmetries: Ref. 198

— S duality, the generalization of electromagnetic duality for all interactions;
— T duality, also called space-time duality, a mapping between small and large lengths and times following $l \leftrightarrow l_{\text{Pl}}^2/l$;**

* The space of solutions for all value of the parameters is called the *moduli space*.
** Space-time duality, the transformation between large and small sizes, leads one to ask whether there is

— infrared dualities.

So far, research into dualities has not led to experimental predictions; however, they have led to a deeper understanding of field theory. Dualities play an important role in string theory. Page 141

Collective aspects of quantum field theory

For many decades, mathematicians asked physicists: What is the essence of quantum field theory? Despite intensive research, this question has yet to be answered precisely.

The first half of the answer is given by the usual definition given by physicists: QFT is the most general known way to describe quantum mechanically continuous systems with a finite number of types of quanta but with an infinite number of degrees of freedom. (Of course, this definition implies that the Lagrangian must be relativistically invariant and must be described by a gauge theory.) However, this half of the answer is already sufficient to spell trouble. We will show in the next part of our ascent that neither space-time nor physical systems are continuous; we will discover that nature does not have infinite numbers of degrees of freedom. In other words, *quantum field theory is an effective theory*; this is the modern way to say that it is approximate, or more bluntly, that it is wrong.

The second, still partly unknown half of the answer would specify which (mathematical) conditions a physical system, i.e., a Lagrangian, actually needs to realize in order to become a quantum field theory. Despite the work of many mathematicians, no complete list of conditions is known yet. It is known that the list includes at least two conditions. First of all, a quantum field theory must be *renormalizable*. Secondly, a quantum field theory must be *asymptotically free*; in other words, the coupling must go to zero when the energy goes to infinity. This condition ensures that interactions are defined properly. Only a subset of renormalizable Lagrangians obey this condition.

In four dimensions, the *only* known theories with these two properties are the non-Abelian gauge theories. These Lagrangians have several general aspects which are not directly evident when we arrive at them through the usual way, i.e., by generalizing naive wave quantum mechanics. This standard approach, the historical one, emphasizes the *perturbative* aspects: we think of elementary fermions as field quanta and of interactions as exchanges of virtual bosons, to various orders of perturbation.

On the other hand, all field theory Lagrangians also show two other configurations, apart from particles, which play an important role. These mathematical solutions appear when a non-perturbative point of view is taken; they are *collective* configurations.

— Quantum field theories show solutions which are static and of finite energy, created by non-local field combinations, called *solitons*. In quantum field theories, solitons are usually magnetic monopoles and dyons; also the famous *skyrmions* are solitons. In this approach to quantum field theory, it is assumed that the actual equations of nature are non-linear at high energy. Like in liquids, one then expects stable, localized and propagating solutions, the solitons. These solitons could be related to the observed particles. Page 249

Page 94 an inside and an outside to particles. (We encountered this question already in our study of gloves.) The issue has not been addressed in research yet. Sometimes one has to be patient.

— Quantum field theories show self-dual or anti-self dual solutions, called *instantons*. Instantons play a role in QCD, and could also play a role in the fundamental Lagrangian of nature.

Ref. 199 These fascinating topics have been explored in detail by mathematical physicists. Even though this has deepened the understanding of gauge theories, all the results have not helped in the path towards unification.

A summary on higher symmetries

No research program trying to unravel higher, i.e., more general symmetries than those of the standard model has been successful so far. This includes many topics not mentioned here. For example, the topic of *quantum groups* was very popular around the year 2000. Non-commutative space-time was another fashion in elementary particle physics around the same time.

Despite hundreds of extremely smart people exploring potential higher symmetries of all kinds, their effort has not been successful for the description of nature. This leads to a question: Did researchers rely on incorrect assumptions on the structure of particles or of space-time? Or is the search for higher symmetry the wrong way to go? Before we explore these two questions, we take a break.

Chapter 11

BACTERIA, FLIES AND KNOTS

> La première et la plus belle qualité de la nature est le mouvement qui l'agite sans cesse ; mais ce mouvement n'est qu'une suite perpétuelle de crimes ; ce n'est que par des crimes qu'elle le conserve.
> Donatien de Sade, *Justine, ou les malheurs de la vertu.**

Wobbly entities, in particular jellyfish or amoebas, open up a fresh vision of the world of motion, if we allow to be led by the curiosity to study them in detail. We have missed many delightful insights by leaving them aside. In particular, wobbly entities yield surprising connections between shape change and motion that will be of great use in the last part of our mountain ascent. Instead of continuing to look at the smaller and smaller, we now take a second look at everyday motion and its mathematical description.

To enjoy this chapter, we change a dear habit. So far, we always described any general example of motion as composed of the motion of *point particles*. This worked well in classical physics, in general relativity and in quantum theory; we based the approach on the silent assumption that during motion, each point of a complex system can be followed separately. We will soon discover that this assumption is *not* realized at smallest scales. Therefore the most useful description of motion of *extended* bodies uses methods that do not require that body parts be followed piece by piece. We explore these methods in this chapter; doing so is a lot of fun in its own right.

If we describe elementary particles as extended entities – as we soon will have to – a particle moving through space is similar to a dolphin swimming through water or to a bee flying through air. Let us explore how these animals do this.

Bumblebees and other miniature flying systems

If a butterfly passes by during our mountain ascent, we can stop a moment to appreciate a simple fact: a butterfly flies, and it is rather small. If we leave some cut fruit in the kitchen until it rots, we observe the even smaller fruit flies (*Drosophila melanogaster*), just around one millimetre in size. If you have ever tried to build small model aeroplanes, or if you

* 'The primary and most beautiful of nature's qualities is motion, which agitates her at all times; but this motion is simply a perpetual consequence of crimes; she conserves it by means of crimes only.' Donatien Alphonse François de Sade (1740–1814) is the intense French writer from whom the term 'sadism' was deduced.

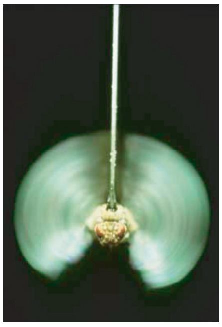


FIGURE 86 A flying fruit fly, tethered to a string

FIGURE 87 Vortices around a butterfly wing (© Robert Srygley/Adrian Thomas)

even only compare these insects to paper aeroplanes (probably the smallest man-made flying thing you ever saw) you start to get a feeling for how well evolution optimized flying insects.

Compared to paper planes, insects also have engines, flapping wings, sensors, navigation systems, gyroscopic stabilizers, landing gear and of course all the features due to life, reproduction and metabolism, built into an incredibly small volume. Evolution really is an excellent engineering team. The most incredible flyers, such as the common house fly (*Musca domestica*), can change flying direction in only 30 ms, using the stabilizers that nature has given them by reshaping the original second pair of wings. Human engineers are getting more and more interested in the technical solutions evolution has developed; Ref. 200 many engineers trying to achieve similar miniaturization. The topic of miniature flying systems is extremely vast, so that we will pick out only a few examples.

How does a bumblebee (Bombus terrestris) fly? The lift mg generated by a *fixed* wing Page 274 (as explained before) follows the empirical relation

$$mg = f\,A\,v^2\,\rho \qquad (101)$$

where A is the surface of the wing, v is the speed of the wing in the fluid of density ρ. The factor f is a pure number, usually with a value between 0.2 and 0.4, that depends on the Ref. 201 angle of the wing and its shape; here we use the average value 0.3. For a Boeing 747, the Page 36 surface is 511 m^2, the top speed at sea level is 250 m/s; at an altitude of 12 km the density of air is only a quarter of that on the ground, thus only 0.31 kg/m^3. We deduce (correctly) Challenge 133 e that a Boeing 747 has a mass of about 300 ton. For bumblebees with a speed of 3 m/s and a wing surface of 1 cm^2, we get a lifted mass of about 35 mg, much less than the weight of the bee, namely about 1 g. The mismatch is even larger for fruit flies. In other words, an insect cannot fly if it keeps its wings *fixed*. It could not fly with fixed wings even if it had tiny propellers attached to them!

Due to the limitations of fixed wings at small dimensions, insects and small birds must *move* their wings, in contrast to aeroplanes. They must do so not only to take off or to gain height, but also to simply remain airborne in horizontal flight. In contrast, aeroplanes generate enough lift with *fixed* wings. Indeed, if you look at flying animals,

FIGURE 88 Examples of the three larger wing types in nature, all optimized for rapid flows: turkey vulture (*Cathartes aura*), ruby-throated hummingbird (*Archilochus colubris*) and a dragonfly (© S.L. Brown, Pennsylvania Game Comission/Joe Kosack and nobodythere)

such as the ones shown in Figure 88, you note that the larger they are, the less they need to move their wings (at cruising speed).

Challenge 134 s Can you deduce from equation (101) that birds or insects can fly but people cannot? Conversely, the formula also (partly) explains why human-powered aeroplanes must be so large.*

But *how* do insects, small birds, flying fish or bats have to move their wings in order to fly? This is a tricky question and the answer has been uncovered only recently. The main point is that insect wings move in a way to produce eddies at the front edge which in turn thrust the insect upwards. Aerodynamic studies of butterflies – shown in Figure 87 – and studies of enlarged insect models moving in oil instead of in air explore the precise way insects make use of vortices. At the same time, more and more 'mechanical birds' and 'model aeroplanes' that use flapping wings for their propulsion are being built around the world. The field is literally in full swing.** Researchers are especially interested in understanding how vortices allow change of flight direction at the small dimensions typical for insects. One aim is to reduce the size of flying machines. However, none of the human-built systems is yet small enough that it actually *requires* wing motion to fly, as is the case for insects. (Ref. 202, Ref. 203)

The expression (101) for the lift of wings also shows what is necessary for safe take-off and landing. The lift of all wings *decreases* for smaller speeds. Thus both animals and aeroplanes *increase* their wing surface in these occasions. Many birds also vigorously increase the flapping of wings in these situations. But even strongly flapping, enlarged wings often are insufficient for take-off. Many flying animals, such as swallows, therefore avoid landing completely. For flying animals which do take off from the ground, nature most commonly makes them hit the wings against each other, over their back, so that when the wings separate again, the low pressure between them provides the first lift. This method is used by insects and many birds, including pheasants. As every hunter knows, pheasants make a loud 'clap' when they take off. The clap is due to the low pressure region thus created.

* Another part of the explanation requires some aerodynamics, which we will not study here. Aerodynamics shows that the power consumption, and thus the resistance of a wing with given mass and given cruise speed, is inversely proportional to the square of the wingspan. Large wingspans with long slender wings are thus of advantage in (subsonic) flying, especially when energy is scarce.

** The website www.aniprop.de presents a typical research approach and the sites ovirc.free.fr and www.ornithopter.org give introductions into the way to build such systems for hobbyists.

Both wing use and wing construction depend on size. There are four types of wings in nature.

1. First of all, all large flying objects, such aeroplanes and large birds, fly using *fixed* wings, except during take-off and landing.
2. Second, common size birds use *flapping* wings. (Hummingbirds can have over 50 wing beats per second.) These first two types of wings have a thickness of about 10 to 15 % of the wing depth.
3. At smaller dimensions, a third wing type appears, as seen in dragonflies and other insects. At these scales, at Reynolds numbers of around 1000 and below, thin *membrane* wings are the most efficient. The *Reynolds number* measures the ratio between inertial and viscous effects in a fluid. It is defined as

$$R = \frac{l v \rho}{\eta} \qquad (102)$$

where l is a typical length of the system, v the speed, ρ the density and η the dynamic *viscosity* of the fluid.* A Reynolds number much larger than one is typical for rapid air flow and fast moving water. In fact, the Reynolds numbers specifies what is meant by a 'rapid' or 'fluid' flow on one hand, and a 'slow' or 'viscous' flow on the other. The first three wing types, shown in Figure 88, are all for rapid flows.
4. The fourth type of wings is found at the smallest possible dimensions, for insects smaller than one millimetre; their wings are not membranes at all. Typical are the cases of thrips and of parasitic wasps, which can be as small as 0.3 mm. All these small insects have wings which consist of a central *stalk* surrounded by hair. In fact, Figure 89 shows that some species of thrips have wings which look like miniature toilet brushes.

At even smaller dimensions, corresponding to Reynolds number below 10, nature does not use wings any more, though it still makes use of air transport. In principle, at the smallest Reynolds numbers gravity plays no role any more, and the process of flying merges with that of swimming. However, air currents are too strong compared with the speeds that such a tiny system could realize. No active navigation is then possible any more. At these small dimensions, which are important for the transport through air of spores and pollen, nature uses the air currents for passive transport, making use of special, but fixed shapes.

* The viscosity is the resistance to flow a fluid poses. It is defined by the force F necessary to move a layer of surface A with respect to a second, parallel one at distance d; in short, the (coefficient of) *dynamic viscosity* is defined as $\eta = d\,F/A\,v$. The unit is 1 kg/s m or 1 Pa s or 1 N s/m^2, once also called 10 P or 10 poise. In other words, given a horizontal tube, the viscosity determines how strong the pump needs to be to pump the fluid through the tube at a given speed. The viscosity of air 20°C is 1.8×10^{-5} kg/s m or 18 μPa s and increases with temperature. In contrast, the viscosity of liquids decreases with temperature. (Why?) The viscosity of water at 0°C is 1.8 mPa s, at 20°C it is 1.0 mPa s (or 1 cP), and at 40°C is 0.66 mPa s. Hydrogen has a viscosity smaller than 10 μPa s, whereas honey has 25 Pa s and pitch 30 MPa s.

Challenge 135 ny

Physicists also use a quantity ν called the *kinematic viscosity*. It is defined with the help of the mass density of the fluid as $\nu = \eta/\rho$ and is measured in m^2/s, once called 10^4 stokes. The kinematic viscosity of water at 20°C is 1 mm^2/s (or 1 cSt). One of the smallest values is that of acetone, with 0.3 mm^2/s; a larger one is glycerine, with 2000 mm^2/s. Gases range between 3 mm^2/s and 100 mm^2/s.

FIGURE 89 The wings of a few types of insects smaller than 1 mm (thrips, *Encarsia*, *Anagrus*, *Dicomorpha*) (HortNET)

We can summarize that active flying is only possible through shape change. Only two types of shape changes are possible for active flying: that of propellers (or turbines) and Ref. 204 that of wings. Engineers are studying with intensity how these shape changes have to take Ref. 205 place in order to make flying most effective. Interestingly, the same challenge is posed by swimming.

Swimming

Swimming is a fascinating phenomenon. The Greeks argued that the ability of fish to swim is a proof that water is made of atoms. If atoms would not exist, a fish could not advance through it. Indeed, swimming is an activity that shows that matter cannot be continuous. Studying swimming can thus be quite enlightening. But how exactly do fish swim?

Whenever dolphins, jellyfish, submarines or humans *swim*, they take water with their fins, body, propellers, hands or feet and push it backwards. Due to momentum conservation they then move forward.* In short, people swim in the same way that fireworks or rockets fly: by throwing matter behind them. This is macroscopic swimming. Does all swimming work in this way? In particular, do *small* organisms advancing through the molecules of a liquid use the same method? No. They use a different, *microscopic* way of swimming.

Small organisms such as bacteria do *not* have the capacity to propel or accelerate water against their surroundings. Indeed, the water remains attached around a microorganism without ever moving away from it. Physically speaking, in these cases of swimming the kinetic energy of the water is negligible. In order to swim, unicellular beings thus need to use other effects. In fact, their only possibility is to change their body shape in controlled ways. From far away, the swimming of microorganisms thus resembles the motion of particles through vacuum. Like microorganisms, also particles have nothing to throw behind them.

A good way to distinguish macroscopic from microscopic swimming is provided by scallops. *Scallops* are molluscs up to a few cm in size. Scallops have a double shell connected by a hinge that they can open and close. If they close it *rapidly*, water is expelled

Page 81 * Fish could use propellers, as the arguments against wheels we collected at the beginning of our walk do not apply for swimming. But propellers with blood supply would be a weak point in the construction, and thus in the defence of a fish.

FIGURE 90 A swimming scallop (here from the genus *Chlamys*) (© Dave Colwell)

and the mollusc is accelerated; the scallop then can glide for a while through the water. Then the scallop opens the shell again, this time *slowly*, and repeats the feat. When swimming, the larger scallops look like clockwork false teeth. If we reduce the size of the scallop by a thousand times to the size of single cells we get a simple result: such a tiny scallop *cannot* swim.

The origin of the lack of scalability of swimming methods is the changing ratio between inertial and dissipative effects at different scales. This ratio is measured by the Reynolds number. For the scallop the Reynolds number is about 100, which shows that when it swims, inertial effects are much more important than dissipative, viscous effects. For a bacterium the Reynolds number is much smaller than 1, so that inertial effects effectively play no role. There is no way to accelerate water *away* from a bacterial-sized scallop, and thus no way to glide. But this is not the only problem microorganism face when they want to swim.

A famous theorem states that no cell-sized being can move if the shape change is the same in the two halves of the motion (opening and closing). Such a shape change would simply make it move back and forward. There is also a mathematical theorem, the so-called *scallop theorem*, that states that no microscopic system can swim if it uses movable parts with only *one* degree of freedom. Thus it is impossible to move at cell dimensions using the method the scallop uses on centimetre scale.

Ref. 206

In order to swim, microorganisms thus need to use a more evolved, *two-dimensional motion* of their shape. Indeed, biologists found that all microorganisms use one of the following three swimming styles:

1. Microorganisms of compact shape of diameter between 20 μm and about 20 mm, use *cilia*. Cilia are hundreds of little hairs on the surface of the organism. The organisms move the cilia in waves wandering around their surface, and these surface waves make the body advance through the fluid. All children watch with wonder *Paramecium*, the unicellular animal they find under the microscope when they explore the water in which some grass has been left for a few hours. *Paramecium*, which is between 100 μm and 300 μm in size, as well as many plankton species* use cilia for its motion. The cilia and their motion are clearly visible in the microscope. A similar swimming method is even used by some large animals; you might have seen similar waves on the borders of certain ink fish; even the motion of the manta (partially) belongs into this class. Ciliate motion is an efficient way to change the shape of a body making use of two dimensions and thus avoiding the scallop theorem.

Ref. 207

* See the www.liv.ac.uk/ciliate website for an overview.

figure to be added

FIGURE 91 Ciliated and flagellate motion

2. Sperm and eukaryote microorganisms whose sizes are in the range between 1 μm and 50 μm swim using an (eukaryote) flagellum.* Flagella, Latin for 'small whips', work like flexible oars. Even though their motion sometimes appears to be just an oscillation, flagella get a kick only during one half of their motion, e.g. at every swing to the left. Flagella are indeed used by the cells like miniature oars. Some cells even twist their flagellum in a similar way that people rotate an arm. Some microorganisms, such as *Chlamydomonas*, even have two flagella which move in the same way Ref. 209 as people move their legs when they perform the breast stroke. Most cells can also change the sense in which the flagellum is kicked, thus allowing them to move either Ref. 210 forward or backward. Through their twisted oar motion, bacterial flagella avoid retracing the same path when going back and forward. As a result, the bacteria avoid the scallop theorem and manage to swim despite their small dimensions. The flexible oar motion they use is an example of a non-adiabatic mechanism; an important fraction of the energy is dissipated.
3. The smallest swimming organisms, bacteria with sizes between 0.2 μm and 5 μm, Ref. 211 swim using *bacterial* flagella. These flagella, also called prokaryote flagella, are different from the ones just mentioned. Bacterial flagella move like turning corkscrews. They are used by the famous *Escherichia coli* bacterium and by all bacteria of the genus *Salmonella*. This type of motion is one of the prominent exceptions to the non-existence of wheels in nature; we mentioned it in the beginning of our walk. Page 81 Corkscrew motion is an example of an adiabatic mechanism.

   A Coli bacterium typically has a handful of flagella, each about 30 nm thick and of corkscrew shape, with up to six turns; the turns have a 'wavelength' of 2.3 μm. Each flagellum is turned by a sophisticated rotation motor built into the cell, which the cell can control both in rotation direction and in angular velocity. For Coli bacteria, the Ref. 212 range is between 0 and about 300 Hz.

   A turning flagellum does not propel a bacterium like a propeller; as mentioned, the velocities involved are much too small, the Reynolds number being only about 10^{-4}. At these dimensions and velocities, the effect is better described by a corkscrew turning in honey or in cork: a turning corkscrew produces a motion against the material around it, in the direction of the corkscrew axis. The flagellum moves the bacterium in the same way that a corkscrew moves the turning hand with respect to the cork.

Note that still smaller bacteria do not swim at all. Indeed, each bacterium faces a min- Ref. 213 imum swimming speed requirement: is must outpace diffusion in the liquid it lives in. Slow swimming capability makes no sense; numerous microorganisms therefore do not

Ref. 208 * The largest sperm, of 6 cm length, are produced by the 1.5 mm sized *Drosophila bifurca* fly, a relative of the famous *Drosophila melanogaster*. Even when thinking about the theory of motion, it is impossible to avoid thinking about sex.

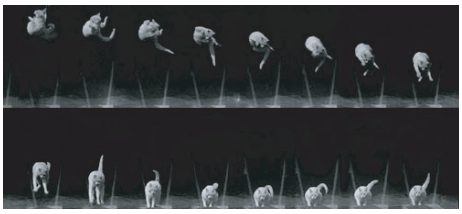


FIGURE 92 Cats can turn themselves, even with *no* initial angular momentum (photographs by Etienne-Jules Marey, 1894)

manage or do not try to swim at all. Some microorganisms are specialized to move along liquid–air interfaces. Others attach themselves to solid bodies they find in the liquid. Some of them are able to *move* along these solids. The amoeba is an example for a microorganism moving in this way. Also the smallest active motion mechanisms known, Page 19 namely the motion of molecules in muscles and in cell membranes, work this way.

Let us summarize these observations in a different way. All known active motion, or self-propulsion, (in flat space) takes place in fluids – be it air or liquids. All active motion requires shape change. In order that shape change leads to motion, the environment, e.g. the water, must itself consist of moving components always pushing onto the swimming entity. The motion of the swimming entity can then be deduced from the particular shape change it performs. To test your intuition, you may try the following puzzle: is microscopic swimming possible in two spatial dimensions? In four? Challenge 136 ny

Rotation, falling cats and the theory of shape change

At small dimensions, flying and swimming takes place through phase change. In the last decades, the description of shape change has changed from a fashionable piece of research to a topic whose results are both appealing and useful. There are many studies, both experimental and theoretical, about the exact way small systems move in water and air, about the achievable and achieved efficiency, and much more. The focus is on motion through translation.

But shape change can also lead to a *rotation* of a body. In this case, the ideas are not restricted to microscopic systems, but apply at all scales. In particular, the theory of shape change is useful in explaining how falling cats manage to fall always on their feet. Cats are not born with this ability; they have to learn it. But the feat has fascinated people for centuries, as shown in the ancient photograph given in Figure 92. In fact, cats confirm Page 104 in three dimensions what we already knew for two dimensions: a deformable body can change its own orientation in space without outside help. Also humans can perform the feat: simply observe the second, lateral rotation of the diver in Figure 93. Astronauts in the space station and passengers of parabolic 'zero-gravity' flights regularly do the same, as do many artificial satellites sent into space.

In the 1980s, following the pioneering work by Michael Berry, Wilczek and Zee as well as Shapere and Wilczek made the point that all motion due to shape change is described by a gauge theory. The equivalence is shown in Table 21. A simple and beautiful example for these ideas has been given by Putterman and Raz and is shown in Figure 94. Imagine four spheres on perfect ice, all of the same mass and size, connected by four rods forming a parallelogram. Now imagine that this parallelogram can change length along one side, called a, and that it can also change the angle θ between the sides. Putterman and Raz call

Ref. 214
Ref. 215

FIGURE 93 Humans can turn themselves in mid air like cats: see the second, lateral rotation of Artem Silchenko, at the 2006 cliff diving world championship (© World High Diving Federation)

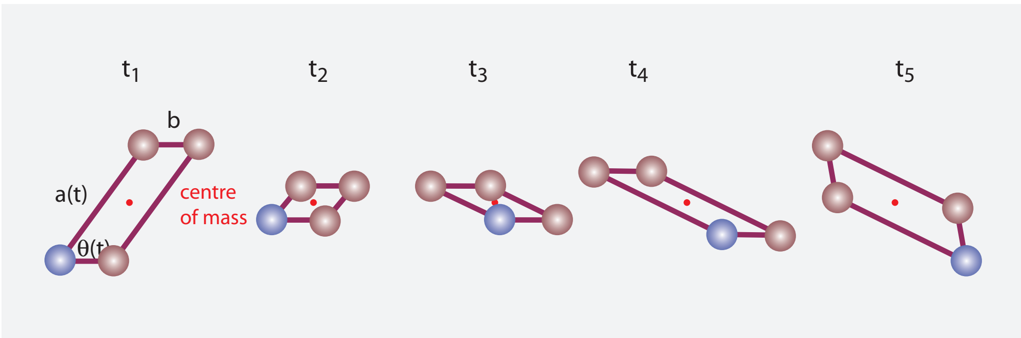


FIGURE 94 The square cat: in free space, or on perfect ice, a deformable body made of four masses that can change one body angle and one extension is able to rotate itself

this the *square cat*. The figure shows that the square cat can change its own orientation on the ice while, obviously, keeping its centre of mass at rest. The figure also shows that this only works because the two motions that the cat can perform, the stretching and the angle change, do *not* commute.

The rotation of the square cat occurs in *strokes*; large rotations are achieved by repeating strokes, similar to the situation of swimmers. If the square cat would be swimming

TABLE 21 The correspondence between shape change and gauge theory

Concept	Shape change	Gauge theory
System	deformable body	matter–field combination
Gauge freedom	freedom of description of body orientation and position	freedom to define vector potential
Gauge-dependent quantity	shape's angular orientation and position	vector potential, phase
	orientation and position change along an open path	vector potential and phase change along open path
Gauge transformation	changes angular orientation and position	changes vector potential
Gauge-independent quantities	orientation and position after full stroke	phase difference on closed path, integral of vector potential along a closed path
	deformations	field strengths
Gauge group	e.g. possible rotations SO(3) or motions E(3)	U(1), SU(2), SU(3)

in a liquid, the cat could thus rotate itself – though it could not advance.

When the cat rotates itself, each stroke results in a rotation angle that is independent of the speed of the stroke. (The same experience can be made when rotating oneself on an office chair by rotating the arm above the head: the chair rotation angle after arm turn is independent of the arm speed.) This leads to a puzzle: what is the largest angle that a cat can turn in one stroke? Challenge 137 d

Rotation in strokes has a number of important implications. First of all, the number of strokes is a quantity that all observers agree upon: it is observer-invariant. Secondly, the orientation change after a *complete* stroke is also observer-invariant. Thirdly, the orientation change for *incomplete* strokes is observer-dependent: it depends on the way that orientation is defined. For example, if orientation is defined by the direction of the body diagonal through the blue mass (see Figure 94), it changes in a certain way during a stroke. If the orientation is defined by the direction of the fixed bar attached to the blue mass, it changes in a different way during a stroke. Only when a full stroke is completed do the two values coincide. Mathematicians say that the choice of the definition and thus the value of the orientation is *gauge-dependent*, but that the value of the orientation change at a full stroke is *gauge-invariant*.

In summary, the square cat shows two interesting points. First, the orientation of a deformable body can change if the deformations it can perform are *non-commuting*. Secondly, such deformable bodies are described by *gauge theories*: certain aspects of the bodies are gauge-invariant, others are gauge-dependent. This summary leads to a question: can we use these ideas to increase our understanding of the gauge theories of the electromagnetic, weak and strong interaction? We will find out later on. In fact, shape change bears even more surprises.

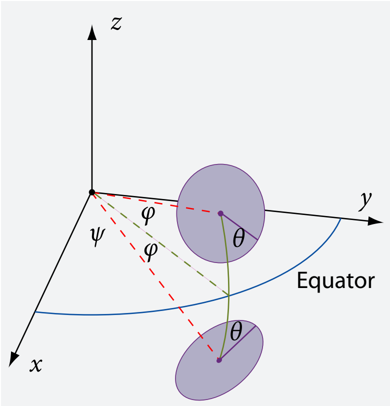


FIGURE 95 Swimming on a curved surface using two discs

Swimming in curved space

In flat space it is not possible to move through shape change. Only orientation changes are possible. Surprisingly, if space is *curved*, motion does become possible. A simple example was published in 2003 by Jack Wisdom. He found that cyclic changes in the shape of a body can lead to net translation, a rotation of the body, or both. Ref. 216

There is a simple system that shows the main idea. We know from Galilean physics that on a frictionless surface we cannot move, but that we can change orientation. This is true only for a flat surface. On a curved surface, we can use the ability to turn and translate it into motion.

Take to massive discs that lie on the surface of a frictionless, spherical planet, as shown in Figure 95. Consider the following four steps: 1. the disc separation φ is increased by the angle $\Delta\varphi$, 2. the discs are rotated oppositely about their centres by the angle $\Delta\theta$, 3. their separation is decreased by $-\Delta\varphi$, and 4. they are rotated back by $-\Delta\theta$. Due to the conservation of angular momentum, the two-disc system changes its longitude $\Delta\psi$ as Challenge 138 ny

$$\Delta\psi = \frac{1}{2}\gamma^2 \Delta\theta \Delta\varphi , \qquad (103)$$

where γ is the angular radius of the discs. This cycle can be repeated over and over. The cycle it allows a body, located on the surface of the Earth, to swim along the surface. Unfortunately, for a body of size of one metre, the motion for each swimming cycle is only around 10^{-27} m.

Wisdom showed that the same procedure also works in curved space, thus in the presence of gravitation. The mechanism thus allows a falling body to swim away from the path of free fall. Unfortunately, the achievable distances for everyday objects are negligibly small. Nevertheless, the effect exists.

In other words, there is a way to swim through curved space that looks similar to swimming at low Reynolds numbers, where swimming results of simple shape change. Does this tell us something about fundamental descriptions of motion? The last part of our ascent will tell.

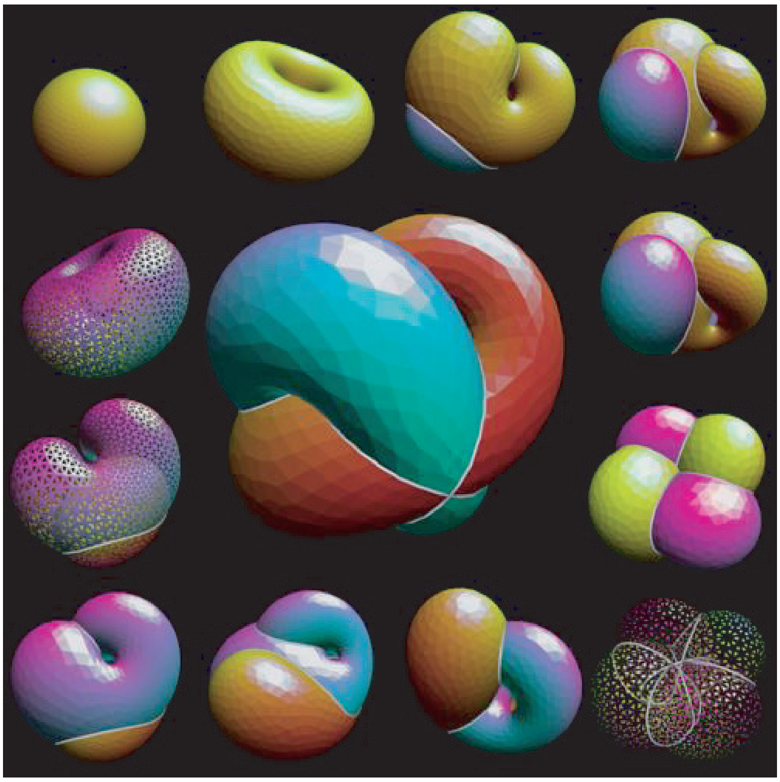


FIGURE 96 A way to turn a sphere inside out, with intermediate steps ordered clockwise (© John Sullivan)

Turning a sphere inside out

> "A text should be like a lady's dress; long enough to cover the subject, yet short enough to keep it interesting.

Exploring the theme of motion of wobbly entities, a famous example cannot be avoided. Ref. 217 In 1957, the mathematician Stephen Smale proved that a sphere can be turned inside out. The discovery brought him the Fields medal in 1966, the highest prize for discoveries in mathematics. Mathematicians call his discovery the *eversion* of the sphere.

To understand the result, we need to describe more clearly the rules of mathematical eversion. First of all, it is assumed that the sphere is made of a thin membrane which has the ability to stretch and bend without limits. Secondly, the membrane is assumed to be able to *intersect* itself. Of course, such a ghostly material does not exist in everyday life; but in mathematics, it can be imagined. A third rule requires that the moves must be performed in such a way that the membrane is not punctured, ripped nor creased; in short, everything must happen *smoothly* (or differentiably, as mathematicians like to say).

Even though Smale proved that eversion is possible, the first way to actually perform it Ref. 218, Ref. 219 was discovered by the blind topologist Bernard Morin in 1961, based on ideas of Arnold Shapiro. After him, several additional methods have been discovered.

Ref. 220 Several computer videos of sphere eversions are now available.* The most famous ones are Outside in, which shows an eversion due to William P. Thurston, and The Optiverse, which shows the most efficient method known so far, discovered by a team led by John Sullivan and shown in Figure 96.

Why is sphere eversion of interest to physicists? If elementary particles were extended and at the same time were of spherical shape, eversion might be a symmetry of particles. To make you think a little, we mention the effects of eversion on the whole surrounding space, not only on the sphere itself. The final effect of eversion is the transformation

$$(x, y, z) \rightarrow \frac{(x, y, -z)\, R^2}{r^2} \qquad (104)$$

where R is the radius of the sphere and r is the length of the coordinate vector (x, y, z), thus $r = \sqrt{x^2 + y^2 + z^2}$. Due to the minus sign in the z-coordinate, eversion is thus different from inversion, but not by too much. As we will find out shortly, a transformation Page 105 similar to eversion, space-time duality, is a fundamental symmetry of nature.

Clouds

Clouds are another important class of wobbly objects. The lack of a definite boundary makes them even more fascinating than amoebas, bacteria or falling cats. We can observe the varieties of clouds from an aeroplane.

Page 147 The common cumulus or cumulonimbus in the sky, like all the other types, are vapour and water droplet clouds. Galaxies are clouds of stars. Stars are clouds of plasma. The atmosphere is a gas cloud. Atoms are clouds of electrons. Nuclei are clouds of protons and neutrons, which in turn are clouds of quarks. Comparing different cloud types is illuminating and fun.

Clouds of all types can be described by a shape and a size, even though in theory they have no bound. An *effective* shape and size can be defined by that region in which the cloud density is only, say, 1 % of the maximum density; slightly different procedures can also be used. All clouds are described by *probability densities* of the components making up the cloud. All clouds show *conservation* of the number of their constituents.

Whenever we see a cloud, we can ask why it does not collapse. Every cloud is an aggre- Page 207 gate; all aggregates are kept from collapse in only *three* ways: through rotation, through pressure, or through the Pauli principle, i.e., the quantum of action. For example, galaxies are kept from collapsing by rotation. Most stars, the atmosphere and rain clouds are kept from collapsing by gas pressure. Neutron stars, the Earth, atomic nuclei, protons or the electron clouds of atoms are kept apart by the quantum of action.

A rain cloud is a method to keep several thousand tons of water suspended in the air. Can you explain what keeps it afloat, and what else keeps it from continuously diffusing Challenge 139 ny into a thinner and thinner structure?

* Summaries of the videos can be seen at the www.geom.umn.edu/docs/outreach/oi website, which also has a good pedagogical introduction. Another simple eversion and explanation is given by Erik de Neve on the www.xs4all.nl/~alife/sphere1.htm website. It is even possible to run the film software at home; see the www.cslub.uwaterloo.ca/~mjmcguff/eversion website. Figure 96 is from the new.math.uiuc.edu/optiverse website.

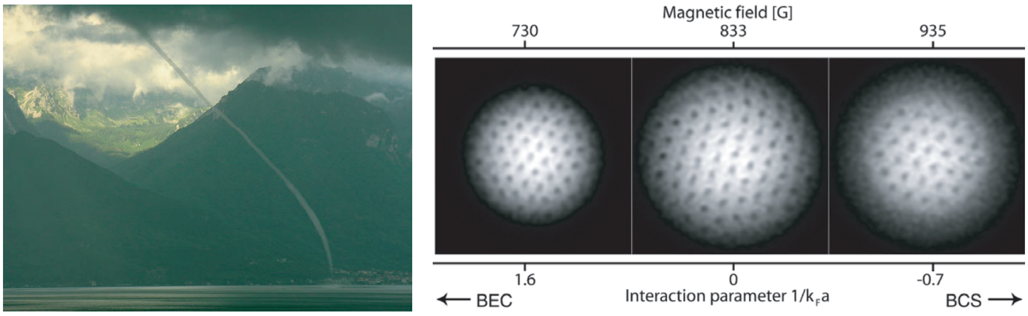


FIGURE 97 Vortices: a waterspout and a vortex lattice in cold Lithium gas (© Zé Nogueira, Andre Schirotzek)

Two rain clouds can merge. So can two atomic electron clouds. So can galaxies. But only atomic clouds are able to *cross* each other. We remember that a normal atom can be inside a Rydberg atom and leave it again without change. In contrast, rain clouds, stars, galaxies or other macroscopic clouds cannot cross each other. When their paths cross, they can only merge or be ripped into pieces. Due to this lack of crossing ability, only microscopic clouds can be counted. In the macroscopic cases, there is no real way to define a 'single' cloud in an accurate way. If we aim for full precision, we are unable to claim that there is more than one rain cloud, as there is no clear-cut boundary between them. Electronic clouds are different. True, in a piece of solid matter we can argue that there is only a single electronic cloud throughout the object; however, when the object is divided, the cloud is divided in a way that makes the original atomic clouds reappear. We thus can speak of 'single' electronic clouds.

If one wants to be strict, galaxies, stars and rain clouds can be seen as made of localized particles. Their cloudiness is only apparent. Could the same be true for electron clouds? And what about space itself? Let us explore some aspects of these questions.

Vortices and the Schrödinger equation

Fluid dynamics is a topic with many interesting aspects. Take the *vortex* that can be observed in any deep, emptying bath tub: it is an extended, one-dimensional 'object', it is deformable, and it is observed to wriggle around. Larger vortices appear as tornadoes on Earth and on other planets, as *waterspouts*, and at the ends of wings or propellers of all kinds. Smaller vortices appear in superfluids. Some examples are shown in Figure 97.

Vortices, also called *vortex tubes* or *vortex filaments*, are thus wobbly entities. Now, a beautiful result from the 1960s states that a vortex filament in a rotating liquid is described by the one-dimensional Schrödinger equation. Let us see how this is possible. Ref. 221

Any deformable linear vortex, as illustrated in Figure 98, is described by a continuous set of position vectors $\boldsymbol{r}(t, s)$ that depend on time t and on a single parameter s. The parameter s specifies the relative position along the vortex. At each point on the vortex, there is a unit tangent vector $\boldsymbol{e}(t, s)$, a unit normal curvature vector $\boldsymbol{n}(t, s)$ and a unit

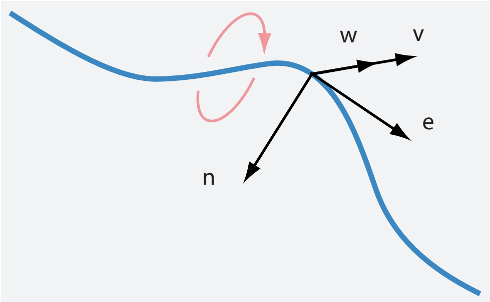


FIGURE 98 The mutually perpendicular tangent $\boldsymbol{e}$, normal $\boldsymbol{n}$, torsion $\boldsymbol{w}$ and velocity $\boldsymbol{v}$ of a vortex in a rotating fluid

torsion vector $\boldsymbol{w}(t, s)$. The three vectors, shown in Figure 98, are defined as usual as

$$\begin{aligned} \boldsymbol{e} &= \frac{\partial \boldsymbol{r}}{\partial s} , \\ \kappa \boldsymbol{n} &= \frac{\partial \boldsymbol{e}}{\partial s} , \\ \tau \boldsymbol{w} &= -\frac{\partial (\boldsymbol{e} \times \boldsymbol{n})}{\partial s} , \end{aligned} \tag{105}$$

where κ specifies the value of the curvature and τ specifies the value of the torsion. In general, both numbers depend on time and on the position along the line.

In the simplest possible case the rotating environment induces a local velocity $\boldsymbol{v}$ for the vortex that is proportional to the curvature κ, perpendicular to the tangent vector $\boldsymbol{e}$ and perpendicular to the normal curvature vector $\boldsymbol{n}$:

$$\boldsymbol{v} = \eta \kappa (\boldsymbol{e} \times \boldsymbol{n}) , \tag{106}$$

Ref. 221 where η is the so-called *coefficient of local self-induction* that describes the coupling between the liquid and the vortex motion. This is the evolution equation of the vortex.

We now assume that the vortex is deformed only *slightly* from the straight configuration. Technically, we are thus in the *linear* regime. For such a linear vortex, directed along the x-axis, we can write

$$\boldsymbol{r} = (x, y(x, t), z(x, t)) . \tag{107}$$

Slight deformations imply $\partial s \approx \partial x$ and therefore

$$\begin{aligned} \boldsymbol{e} &= (1, \frac{\partial y}{\partial x}, \frac{\partial z}{\partial x}) \approx (1, 0, 0) , \\ \kappa \boldsymbol{n} &\approx (0, \frac{\partial^2 y}{\partial x^2}, \frac{\partial^2 z}{\partial x^2}) , \text{ and} \\ \boldsymbol{v} &= (0, \frac{\partial y}{\partial t}, \frac{\partial z}{\partial t}) . \end{aligned} \tag{108}$$

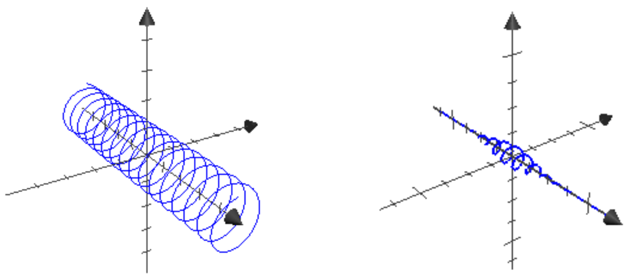


FIGURE 99 Motion of a vortex: the fundamental helical solution and a moving helical 'wave packet'

We can thus rewrite the evolution equation (106) as

$$(0, \frac{\partial y}{\partial t}, \frac{\partial z}{\partial t}) = \eta\,(0, -\frac{\partial^2 z}{\partial x^2}, \frac{\partial^2 y}{\partial x^2})\ . \tag{109}$$

This equation is well known; if we drop the first coordinate and introduce complex numbers by setting $\Phi = y + iz$, we can rewrite it as

$$\frac{\partial \Phi}{\partial t} = i\eta \frac{\partial^2 \Phi}{\partial x^2}\ . \tag{110}$$

Ref. 222 This is the one-dimensional Schrödinger equation for the evolution of a free wave function! The complex function Φ specifies the transverse deformation of the vortex. In other words, we can say that the Schrödinger equation in one dimension describes the evolution of the deformation for an almost linear vortex in a rotating liquid. We note that there is no constant $\hbar$ in the equation, as we are exploring a *classical* system.

Schrödinger's equation is linear in Φ. Therefore the fundamental solution is

$$\Phi(x, y, z, t) = a\ \mathrm{e}^{i(\tau x - \omega t)} \quad \text{with} \quad \omega = \eta\tau^2 \quad \text{and} \quad \kappa = a\tau^2\ , \tag{111}$$

The amplitude a and the wavelength or pitch $b = 1/\tau$ can be freely chosen, as long as the approximation of small deviation is fulfilled; this condition translates as $a \ll b$.* In the present interpretation, the fundamental solution corresponds to a vortex line that is deformed into a *helix*, as shown in Figure 99. The angular speed ω is the rotation speed around the axis of the helix.

* The curvature is given by $\kappa = a/b^2$, the torsion by $\tau = 1/b$. Instead of $a \ll b$ one can thus also write $\kappa \ll \tau$.

Challenge 140 ny A helix moves along the axis with a speed given by

$$v_{\text{helix along axis}} = 2\eta\tau \,. \tag{112}$$

In other words, for extended entities following evolution equation (106), rotation and translation are coupled.* The momentum p can be defined using $\partial\Phi/\partial x$, leading to

$$p = \tau = \frac{1}{b} \,. \tag{113}$$

Momentum is thus inversely proportional to the helix wavelength or pitch, as expected. The energy E is defined using $\partial\Phi/\partial t$, leading to

$$E = \eta\tau^2 = \frac{\eta}{b^2} \,. \tag{114}$$

Energy and momentum are connected by

$$E = \frac{p^2}{2\mu} \quad \text{where} \quad \mu = \frac{1}{2\eta} \,. \tag{115}$$

In other words, a vortex with a coefficient η – describing the coupling between environment and vortex – is thus described by a number μ that behaves like an effective mass. We can also define the (real) quantity $|\Phi| = a$; it describes the *amplitude* of the deformation. Page 218

In the Schrödinger equation (110), the second derivative implies that the deformation 'wave packet' has tendency to spread out over space. Can you confirm that the wavelength–frequency relation for a vortex wave group leads to something like the indeterminacy relation (however, without a $\hbar$ appearing explicitly)? Challenge 142 ny

In summary, the complex amplitude Φ for a linear vortex in a rotating liquid behaves like the one-dimensional wave function of a non-relativistic free particle. In addition, we found a suggestion for the reason that complex numbers appear in the Schrödinger equation of quantum theory: they could be due to the intrinsic rotation of an underlying substrate. Is this suggestion correct? We will find out in the last part of our adventure. Vol. VI, page 164

Fluid space-time

General relativity shows that space can move and oscillate: space is a wobbly entity. Is space more similar to clouds, to fluids, or to solids?

An intriguing approach to space-time as a fluid was published in 1995 by Ted Jacobson. He explored what happens if space-time, instead of assumed to be continuous, is assumed to be the statistical average of numerous components moving in a disordered fashion. Ref. 223

The standard description of general relativity describes space-time as an entity similar to a flexible mattress. Jacobson studied what happens if the mattress is assumed to be Page 130

Challenge 141 ny * A wave packet moves along the axis with a speed given by $v_{\text{packet}} = 2\eta\tau_0$, where τ_0 is the torsion of the helix of central wavelength.

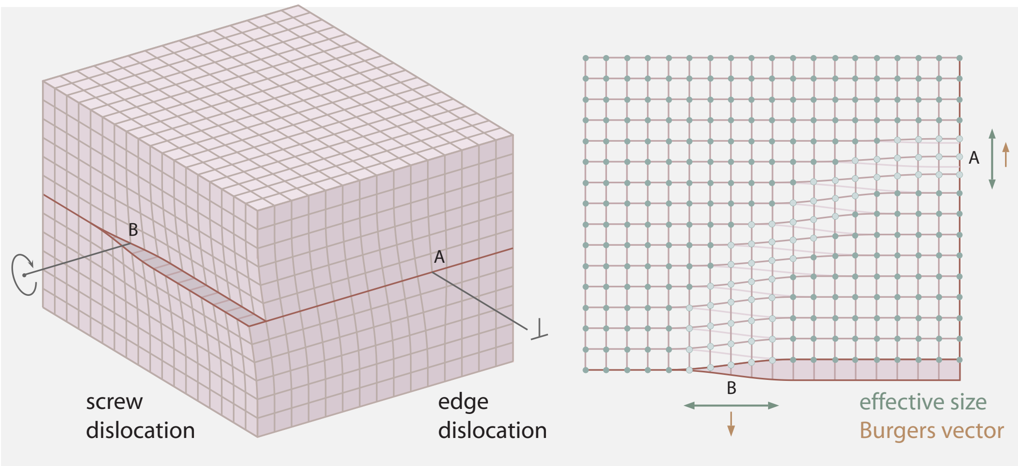


FIGURE 100 The two pure dislocation types, edge and screw dislocations, seen from the outside of a cubic crystal (left) and the mixed dislocation – a quarter of a dislocation loop – joining them in a horizontal section of the same crystal (right) (© Ulrich Kolberg)

made of a fluid. A fluid is a collection of (undefined) components moving randomly and described by a temperature varying from place to place.

Page 88
Jacobson started from the Fulling–Davies–Unruh effect and assumed that the local fluid temperature is given by a multiple of the local gravitational acceleration. He also used the proportionality – correct on horizons – between area and entropy. Since the energy flowing through a horizon can be called heat, one can thus translate the expression $\delta Q = T\delta S$ into the expression $\delta E = a\delta A(c^2/4G)$, which describes the behaviour of space-time at horizons. As we have seen, this expression is fully equivalent to general relativity.
Page 29

In other words, imagining space-time as a fluid is a powerful analogy that allows to deduce general relativity. Does this mean that space-time actually *is* similar to a fluid? So far, the analogy is not sufficient to answer the question and we have to wait for the last part of our adventure to settle it. In fact, just to confuse us a bit more, there is an old argument for the opposite statement.

Dislocations and solid space-time

General relativity tells us that space behaves like a deformable mattress; space thus behaves like a solid. There is a second argument that underlines this point and that exerts a continuing fascination. This argument is connected to a famous property of the motion of dislocations.

Dislocations are one-dimensional construction faults in crystals, as shown in Figure 100. A general dislocation is a mixture of the two pure dislocation types: *edge dislocations* and *screw dislocations*. Both are shown in Figure 100.

If one explores how the atoms involved in dislocations can rearrange themselves, one finds that edge dislocations can only move perpendicularly to the added plane. In contrast, screw dislocations can move in all directions.* An important case of general, mixed
Challenge 143 e

* See the uet.edu.pk/dmems/edge_dislocation.htm, uet.edu.pk/dmems/screw_dislocation.htm and uet.edu.

dislocations, i.e., of mixtures of edge and screw dislocations, are closed *dislocation rings*. On such a dislocation ring, the degree of mixture changes continuously from place to place.

Any dislocation is described by its *strength* and by its effective *size*; they are shown, respectively, in red and blue in Figure 100. The strength of a dislocation is measured by the so-called *Burgers vector*; it measures the misfits of the crystal around the dislocation. More precisely, the Burgers vector specifies by how much a section of perfect crystal needs to be displaced, after it has been cut open, to produce the dislocation. Obviously, the strength of a dislocation is quantized in multiples of a minimal Burgers vector. In fact, dislocations with large Burgers vectors can be seen as composed of dislocations of minimal Burgers vector, so that one usually studies only the latter.

The size or *width* of a dislocation is measured by an *effective width* w. Also the width is a multiple of the lattice vector. The width measures the size of the deformed region of the crystal around the dislocation. Obviously, the size of the dislocation depends on the elastic properties of the crystal, can take continuous values and is direction-dependent. The width is thus related to the energy content of a dislocation.

A general dislocation can move, though only in directions which are both perpendicular to its own orientation and to its Burgers vector. Screw dislocations are simpler: they can move in any direction. Now, the motion of screw dislocations has a peculiar property. We call c the speed of sound in a pure (say, cubic) crystal. As Frenkel and Kontorowa found in 1938, when a screw dislocation moves with velocity v, its width w changes as

Ref. 224

$$w = \frac{w_0}{\sqrt{1 - v^2/c^2}} . \tag{116}$$

In addition, the energy of the moving dislocation obeys

$$E = \frac{E_0}{\sqrt{1 - v^2/c^2}} . \tag{117}$$

A screw dislocation thus cannot move faster than the speed of sound c in a crystal and its width shows a speed-dependent contraction. (Edge dislocations have similar, but more complex behaviour.) The motion of screw dislocations in solids is thus described by the same effects and formulae that describe the motion of bodies in special relativity; the speed of sound is the limit speed for dislocations in the same way that the speed of light is the limit speed for objects.

Does this mean that elementary particles are dislocations of space or even of space-time, maybe even dislocation rings? The speculation is appealing, even though it supposes that space-time is a solid crystal, and thus contradicts the model of space or space-time as a fluid. Worse, we will soon encounter other reasons to reject modelling space-time as a lattice; maybe you can find a few arguments already by yourself. Still, expressions (116) and (117) for dislocations continue to fascinate.

Page 69
Challenge 144 s

At this point, we are confused. Space-time seems to be solid and liquid at the same

pk/dmems/mixed_dislocation.htm web pages to watch a moving dislocation.

time. Despite this contrast, the discussion somehow gives the impression that there is something waiting to be discovered. But what? We will find out in the last part of our adventure.

Polymers

Ref. 225 The study of polymers is both economically important and theoretically fascinating. Polymers are materials built of long and flexible macromolecules that are sequences of many ('poly' in Greek) similar monomers. These macromolecules are thus wobbly entities.

Polymers form *solids*, like rubber or plexiglas, *melts*, like those used to cure teeth, and many kinds of *solutions*, like glues, paints, eggs, or people. Polymer gases are of lesser importance.

All the material properties of polymers, such as their elasticity, their viscosity, their electric conductivity or their unsharp melting point, depend on the number of monomers and the topology of their constituent molecules. In many cases, this dependence can be calculated. Let us explore an example.

If L is the contour length of a free, ideal, unbranched polymer molecule, the *average* end-to-end distance R is proportional to the square root of the length L:

$$R = \sqrt{Ll} \sim \sqrt{L} \quad \text{or} \quad R \sim \sqrt{N} \qquad (118)$$

where N is the number of monomers and l is an effective monomer length describing the scale at which the polymer molecule is effectively stiff. R is usually much smaller than L; this means that free, ideal polymer molecules are usually in a *coiled* state.

Obviously, the end-to-end distance R varies from molecule to molecule, and follows a Gaussian distribution for the probability P of a end-to-end distance R:

$$P(R) \sim e^{\frac{-3R^2}{2Nl^2}} . \qquad (119)$$

The average end-to-end distance mentioned above is the root-mean-square of this distribution. Non-ideal polymers are polymers which have, like non-ideal gases, interactions with neighbouring molecules or with solvents. In practice, polymers follow the ideal behaviour quite rarely: polymers are ideal only in certain solvents and in melts.

If a polymer is *stretched*, the molecules must rearrange. This changes their entropy and produces an elastic force f that tries to inhibit the stretching. For an ideal polymer, the force is not due to molecular interactions, but is entropic in nature. Therefore the force can be deduced from the *free energy*

$$F \sim -T \ln P(R) \qquad (120)$$

of the polymer: the force is then simply given as $f = \partial F(R)/\partial R$. For an ideal polymer, using its probability distribution, the force turns out to be proportional to the stretched length. Thus the spring constant k can be introduced, given by (Challenge 145 e)

$$k = \frac{f}{R} = \frac{3T}{Ll} . \qquad (121)$$

We thus deduced a material property, the spring constant k, from the simple idea that polymers are made of long, flexible molecules. The proportionality to temperature T is a result of the entropic nature of the force; the dependence on L shows that longer molecules are more easy to stretch. For a real, non-ideal polymer, the calculation is more complex, but the procedure is the same. Indeed, this is the mechanism at the basis of the elasticity of rubber.

Using the free energy of polymer conformations, we can calculate the material properties of macromolecules in many other situations, such as their reaction to compression, their volume change in the melt, their interactions in solutions, the effect of branched molecules, etc. This is a vast field of knowledge on its own, which we do not pursue here. Modern research topics include the study of knotted polymers and the study of polymer mixtures. Extensive computer calculations and experiments are regularly compared.

Knots and links

> "Don't touch this, or I shall tie your fingers into knots!"
>
> (Surprisingly efficient child education technique.)

Knots and their generalization are central to the study of wobbly object motion. A *(mathematical) knot* is a closed piece of rubber string, i.e., a string whose ends have been glued together, which cannot be deformed into a circle or a simple loop. The simple loop is also called the *trivial knot*.

Knots are of importance in the context of this chapter as they visualize the limitations of the motion of wobbly entities. In addition, we will discover other reasons to study knots later on. In this section, we just have a bit of fun.* Ref. 226

In 1949, Schubert proved that every knot can be decomposed in a unique way as sum of prime knots. Knots thus behave similarly to integers. Ref. 227

If prime knots are ordered by their crossing numbers, as shown in Figure 101, the trivial knot (0_1) is followed by the trefoil knot (3_1) and by the figure-eight knot (4_1). The figure only shows *prime knots*, i.e., knots that cannot be decomposed into two knots that are connected by two parallel strands. In addition, the figure only shows one of the often possible two mirror versions.

Together with the search for invariants, the tabulation of knots is a modern mathematical sport. Flat knot diagrams are usually ordered by the minimal number of crossings as done in Figure 101. There is 1 knot with zero, 1 with three and 1 with four crossings (not counting mirror knots); there are 2 knots with five and 3 with six crossings, 7 knots with seven, 21 knots with eight, 41 with nine, 165 with ten, 552 with eleven, 2176 with twelve, 9988 with thirteen, 46 972 with fourteen, 253 293 with fifteen and 1 388 705 knots with sixteen crossings. Ref. 227

The mirror image of a knot usually, but not always, is different from the original. If you want a challenge, try to show that the trefoil knot, the knot with three crossings, is different from its mirror image. The first mathematical proof was by Max Dehn in 1914.

* Beautiful illustrations and detailed information about knots can be found on the Knot Atlas website at katlas.math.toronto.edu and at the KnotPlot website at www.knotplot.com.

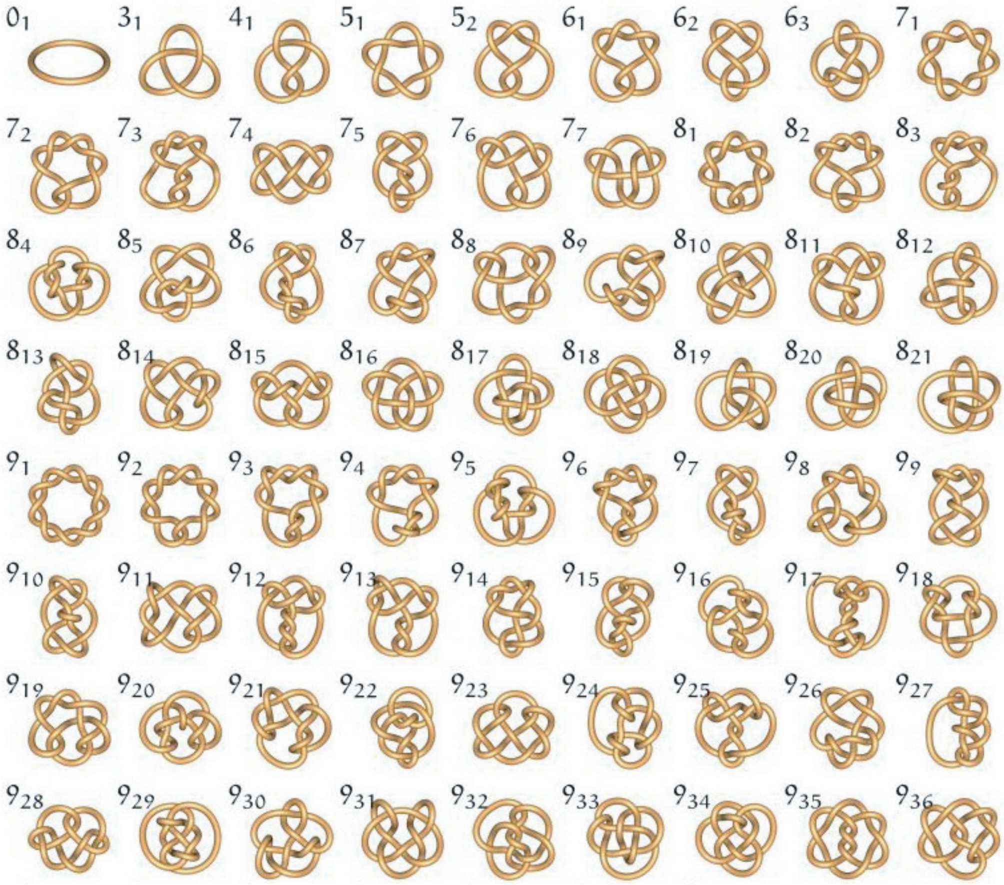


FIGURE 101 The knot diagrams for the simplest prime knots (© Robert Scharein)

Antiknots do not exist. An *antiknot* would be a knot on a rope that cancels out the corresponding knot when the two are made to meet along the rope. It is easy to prove that this is impossible. We take an infinite sequence of knots and antiknots on a string, Ref. 228 $K - K + K - K + K - K....$ On one hand, we could make them disappear in this way $K-K+K-K+K-K... = (K-K)+(K-K)+(K-K)... = 0$. On the other hand, we could do the same thing using $K-K+K-K+K-K... = K(-K+K)+(-K+K)+(-K+K)... = K$. The only knot K with an antiknot is thus the unknot $K = 0$.*

How do we describe such a knot through the telephone? Mathematicians have spent a lot of time to figure out smart ways to achieve it. The obvious way is to flatten the knot onto a plane and to list the position and the type (below or above) of the crossings. But what is the *simplest* way to describe knots by the telephone? The task is not completely finished, but the end is in sight. Mathematicians do not talk about 'telephone messages', they talk about *knot invariants*, i.e., about quantities that do not depend on the precise shape of the knot. At present, the best description of knots use polynomial invariants.

* This proof does *not* work when performed with numbers; we would be able to deduce $1 = 0$ by setting
Challenge 146 s K=1. Why is this proof valid with knots but not with numbers?

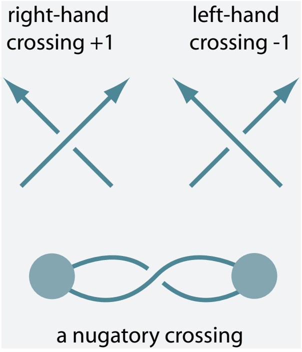


FIGURE 102 Crossing types in knots

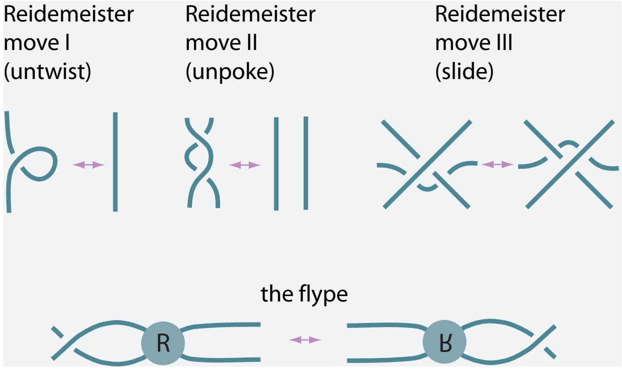


FIGURE 103 The Reidemeister moves and the flype

Most of them are based on a discovery by Vaughan Jones in 1984. However, though the Jones polynomial allows to uniquely describe most simple knots, it fails to do so for more complex ones. But the Jones polynomial finally allowed to prove that a diagram which is alternating and eliminates nugatory crossings (i.e. if it is 'reduced') is indeed one which has minimal number of crossings. The polynomial also allows to show that any two reduced alternating diagrams are related by a sequence of flypes.

In short, the simplest way to describe a knot through the telephone is to give its Kauffman polynomial, together with a few other polynomials.

Since knots are stable in time, a knotted line in three dimensions is equivalent to a knotted surface in space-time. When thinking in higher dimensions, we need to be careful. Every knot (or knotted line) can be untied in four or more dimensions. However, there is no surface *embedded* in four dimensions which has as $t = 0$ slice a knot, and as $t = 1$ slice the circle. Such a surface embedding needs at least five dimensions.

In higher dimensions, knots are thus possible *only* if n-spheres are tied instead of circles; for example, as just said, 2-spheres can be tied into knots in 4 dimensions, 3-spheres in 5 dimensions and so forth.

The hardest open problems that you can tell your grandmother

Even though mathematicians have achieved good progress in the classification of knots, surprisingly, they know next to nothing about the *shapes* of knots. Here are a few problems that are still open today:

— Take a piece of rope, and tie a knot into it. Pull the rope as tight as possible. By how many diameters did the ends of the rope come closer? Mathematicians call this the *ropelength* of an open knot (or of a long knot). Of course, there are computer approximations for the value – though only a few. The ropelength for the open granny/trefoil knot is about 10.1 diameters, and for the open figure-eight knot it is around 13.7 diameters. But *no* formula giving these numbers is known. Ref. 229, Ref. 230

— For mathematical knots, i.e., *closed* knots, the problem is equally unsolved. For example: the ropelength of the tight trefoil knot is known to be around 16.37 diameters,

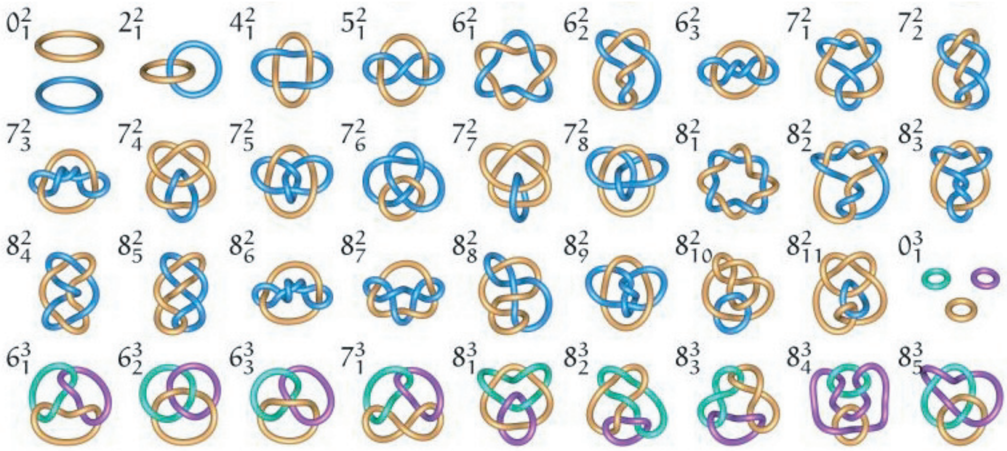


FIGURE 104 The diagrams for the simplest links with two and three components (© Robert Scharein)

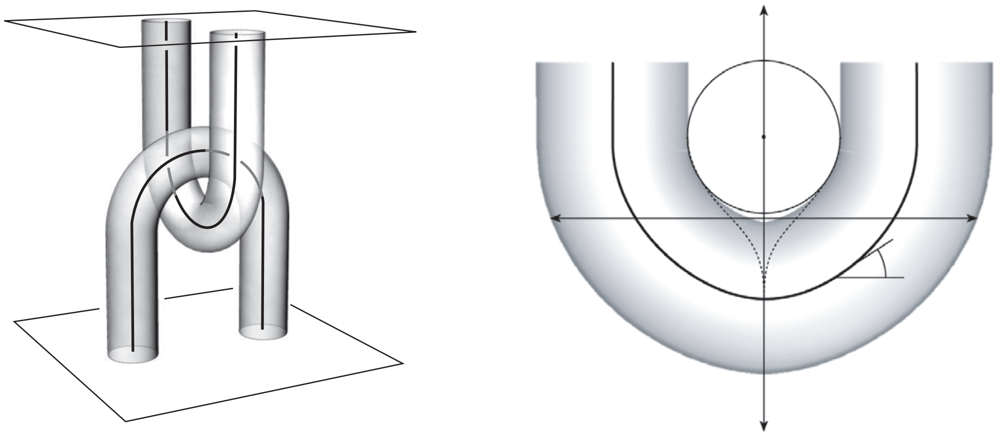


FIGURE 105 The ropelength problem for the simple clasp, and the candidate configuration that probably minimizes ropelenth, leaving a gap between the two ropes (© Jason Cantarella)

Ref. 230 and that of the figure-eight knot about 21.04 diameters. For beautiful visualizations of the tightening process, see the animations on the website www.jasoncantarella.com/movs. But what is the formula giving the ropelength values? Nobody knows, because the precise shape of the trefoil knot – or of any other knot – is unknown. Lou Kauffman has a simple comment for the situation: 'It is a scandal of mathematics!'

— Mathematicians also study more general structures than knots. *Links* are the generalization of knots to several closed strands. *Braids* and *long links* are the generalization of links to *open* strands. Now comes the next surprise, illustrated in Figure 105. Even for two ropes that form a simple *clasp*, i.e., two linked letters 'U', the ropelength
Ref. 231 problem is unsolved – and there is not even a knot involved! In fact, in 2004, Jason Cantarella and his colleagues have presented a candidate for the shape that minimizes ropelength. Astonishingly, the candidate configuration leaves a small gap between the two ropes, as shown in Figure 105.

FIGURE 106 A hagfish tied into a knot (© Christine Ortlepp)

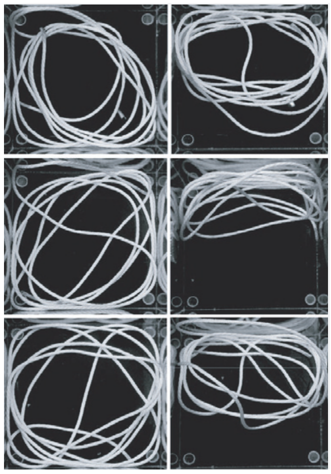


FIGURE 107 How apparent order for long rope coils (left) changes over time when shaking the container (right) (© 2007 PNAS)

In short, the shape of knots is a research topic that has barely taken off. Therefore we have to leave these questions for a future occasion.

Curiosities and fun challenges on knots and wobbly entities

Proteins, the molecules that make up many cell structures, are chains of aminoacids. It seems that very few proteins are knotted, and that most of these form trefoil knots. However, a figure-eight knotted protein has been discovered in 2000 by William Taylor. Ref. 232

* *

Knots form also in other polymers. They seem to play a role in the formation of radicals in carbohydrates. Research on knots in polymers is presently in full swing.

* *

This is the simplest unsolved knot problem: Imagine an ideally wobbly rope, that is, a rope that has the same radius everywhere, but whose curvature can be changed as one prefers. Tie a trefoil knot into the rope. By how much do the ends of the rope get nearer? In 2006, there are only numerical estimates for the answer: about 10.1 radiuses. There is no formula yielding the number 10.1. Alternatively, solve the following problem: what is the rope length of a closed trefoil knot? Also in this case, only numerical values are known – about 16.33 radiuses – but no exact formula. The same is valid for any other knot, of course. Challenge 147 r

* *

A famous type of eel, the *knot fish Myxine glutinosa*, also called hagfish or slime eel, is able to make a knot in his body and move this knot from head to tail. It uses this motion Ref. 233

to cover its body with a slime that prevents predators from grabbing it; it also uses this motion to escape the grip of predators, to get rid of the slime after the danger is over, and to push against a prey it is biting in order to extract a piece of meat. All studied knot fish form only left handed trefoil knots, by the way; this is another example of chirality in nature.

* *

One of the most incredible discoveries of recent years is related to knots in DNA molecules. The *DNA molecules* inside cell nuclei can be hundreds of millions of base pairs long; they regularly need to be packed and unpacked. When this is done, often the same happens as when a long piece of rope or a long cable is taken out of a closet.

It is well known that you can roll up a rope and put it into a closet in such a way that it looks orderly stored, but when it is pulled out at one end, a large number of knots is suddenly found. In 2007, this effects was finally explored in detail. Strings of a few metres in length were put into square boxes and shaken, in order to speed up the effect. The result, shown partly in Figure 107, was astonishing: almost every imaginable knot – up to a certain complexity that depends on the length and flexibility of the string – was formed in this way.

Ref. 234

To make a long story short, this also happens to nature when it unpacks DNA in cell nuclei. Life requires that DNA molecules move inside the cell nucleus without hindrance. So what does nature do? Nature takes a simpler approach: when there are unwanted crossings, it cuts the DNA, moves it over and puts the ends together again. In cell nuclei, there are special enzymes, the so-called topoisomerases, which perform this process. The details of this fascinating process are still object of modern research.

* *

The great mathematician Carl-Friedrich Gauß was the first person to ask what would happen when an electrical current I flows along a wire A linked with a wire B. He discovered a beautiful result by calculating the effect of the magnetic field of one wire onto the other. Gauss found the expression

Ref. 235

$$\frac{1}{4\pi I}\int_A \mathrm{d}\boldsymbol{x}_A \cdot \boldsymbol{B}_B = \frac{1}{4\pi}\int_A \mathrm{d}\boldsymbol{x}_A \cdot \int_B \mathrm{d}\boldsymbol{x}_B \times \frac{(\boldsymbol{x}_A - \boldsymbol{x}_B)}{|\boldsymbol{x}_A - \boldsymbol{x}_B|^3} = n \,, \qquad (122)$$

where the integrals are performed along the wires. Gauss found that the number n does not depend on the precise shape of the wires, but only on the way they are linked. Deforming the wires does not change it. Mathematicians call such a number a *topological invariant*. In short, Gauss discovered a physical method to calculate a mathematical invariant for links; the research race to do the same for other invariants, also for knots and braids, is still going on today.

In the 1980s, Edward Witten was able to generalize this approach to include the nuclear interactions, and to define more elaborate knot invariants, a discovery that brought him the Fields medal.

* *

Knots are also of importance at Planck scales, the smallest dimensions possible in nature.

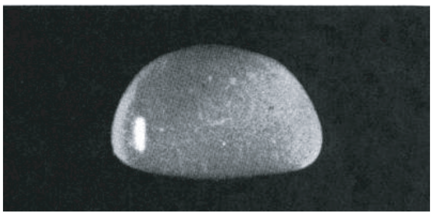


FIGURE 108 A large raindrop falling downwards

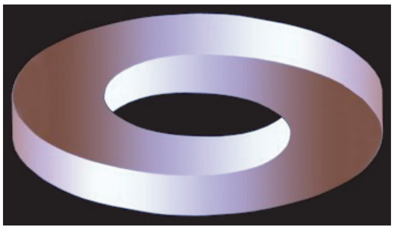


FIGURE 109 Is this possible?

We will soon explore how knots and the structure of elementary particles are related.

* *

Challenge 148 r Knots appear rarely in nature. For example, tree roots do not seem to grow many knots during the lifetime of a plant. How do plants avoid this? In other words, why are there no knotted bananas in nature?

* *

If we move along the knot and count the crossings where we stay above and subtract the number of crossings where we pass below, we get a number called the *writhe* of the knot. It is not an invariant, but usually a tool in building them. The writhe is not necessarily invariant under one of the three Reidemeister moves. Can you see which one? However, the writhe is invariant under flypes. Challenge 149 ny

* *

Modern knot research is still a topic with many open questions. A recent discovery is the *quantization of three-dimensional writhe* in tight knots. Many discoveries are still expected in this domain. Ref. 236

* *

Challenge 150 s What is the shape of raindrops? Try to picture it. However, use your reason, not your prejudice! By the way, it turns out that there is a maximum size for raindrops, with a value of about 4 mm. The shape of such a large raindrop is shown in Figure 108. Can you imagine where the limit comes from? Ref. 237

For comparison, the drops in clouds, fog or mist are in the range of 1 to 100 μm, with a peak at 10 to 15 μm. In situations where all droplets are of similar size and where light is scattered only once by the droplets, one can observe coronae, glories or fogbows. Page 129 Page 108

* *

Challenge 151 s What is the entity shown in Figure 109 – a knot, a braid or a link?

* *

Challenge 152 d Can you find a way to classify tie knots?

* *

FIGURE 110 A flying snake, *Chrysopelea paradisii*, performing the feat that gave it its name (QuickTime film © Jake Socha)

Challenge 153 s Are you able to find a way to classify the way shoe laces can be threaded?

* *

A striking example of how wobbly entities can behave is given in Figure 110. There is indeed a family of snakes that like to jump off a tree and sail through the air to a neighbouring tree. Both the jump and the sailing technique have been studied in recent years. The website www.flyingsnake.org by Jake Socha provides additional films. His fascinating publications tell more about these intriguing reptiles. Ref. 238

* *

Page 258 Challenge 154 e When a plane moves at supersonic speed through humid air, sometimes a conical cloud forms and moves with the plane. How does this cloud differ from the ones studied above?

* *

Challenge 155 ny One of the biggest challenges about clouds: is it possible to make rain on demand? So far, there are almost no positive results. Inventing a method, possibly based on hygroscopic salt injection, will be a great help to mankind.

Summary on wobbly objects

We can sum up the possible motions of extended systems in a few key themes. In earlier chapters we studied the first: waves, solitons and interpenetration. In this chapter we explored the way to move through shape change, explored eversion and duality, studied polymers, knots and their rearrangement, and looked at the relation between clouds and extension. As an interesting application, we explored some possible analogies for the Schrödinger equation and for space-time. Page 249

The motion of wobbly objects is often a neglected topic in textbooks on motion. Research is progressing at full speed; it is expected that many beautiful analogies will be discovered in the near future. For example, in this chapter we have not explored any possible analogy for the motion of light; similarly, including quantum theory into the description of wobbly bodies' motion remains a fascinating issue for anybody aiming to publish in a new field.

Challenge 156 r

Chapter 12

QUANTUM PHYSICS IN A NUTSHELL – AGAIN

Compared to classical physics, quantum theory is remarkably more complex. The basic idea however, is simple: in nature there is a minimum change, i.e., a minimum action $\hbar$. The minimum action leads to all the strange observations made in the microscopic domain, such as wave behaviour of matter, tunnelling, indeterminacy relations, randomness in measurements, quantization of angular momentum, pair creation, decay, indistinguishability and particle reactions.

The essence of quantum theory is thus the lack of the infinitely small. The mathematics of quantum theory is often disturbingly involved. Was this part of our walk worth the effort? It was; the accuracy is excellent and the results profound. We give an overview of both and then turn to the list of questions that are still left open.

Achievements in precision

Quantum theory improved the accuracy of predictions from the few – if any – digits common in classical mechanics to the full number of digits – sometimes thirteen – that can be measured today. The limited precision is usually *not* given by the inaccuracy of 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 22 shows this in more detail.

TABLE 22 Selected comparisons between classical physics, quantum theory and experiment

Observable	Classical prediction	Prediction of quantum theory[a]	Measurement	Cost estimate[b]
Simple motion of bodies				
Indeterminacy	0	$\Delta x \Delta p \geqslant \hbar/2$	$(1 \pm 10^{-2})\,\hbar/2$	10 k€
Matter wavelength	none	$\lambda p = 2\pi\hbar$	$(1 \pm 10^{-2})\,\hbar$	10 k€
Tunnelling rate in alpha decay	0	$1/\tau$ is finite	$(1 \pm 10^{-2})\,\tau$	5 k€
Compton wavelength	none	$\lambda_c = h/m_e c$	$(1 \pm 10^{-3})\,\lambda$	20 k€
Pair creation rate	0	σE	agrees	100 k€
Radiative decay time in hydrogen	none	$\tau \sim 1/n^3$	(1 ± 10^{-2})	5 k€

Observable	Classical prediction	Prediction of quantum theory[a]	Measurement	Cost[b] estimate
Smallest angular momentum	0	$\hbar/2$	$(1 \pm \pm 10^{-6})\,\hbar/2$	10 k€
Casimir effect/pressure	0	$p = (\pi^2 \hbar c)/(240 r^4)$	(1 ± 10^{-3})	30 k€
Colours of objects				
Spectrum of hot objects	diverges	$\lambda_{max} = hc/(4.956\,kT)$	$(1 \pm 10^{-4})\ \Delta\lambda$	10 k€
Lamb shift	none	$\Delta\lambda = 1057.86(1)$ MHz	$(1 \pm 10^{-6})\,\Delta\lambda$	50 k€
Rydberg constant	none	$R_\infty = m_e c \alpha^2/2h$	$(1 \pm 10^{-9})\,R_\infty$	50 k€
Stefan–Boltzmann constant	none	$\sigma = \pi^2 k^4/60\hbar^3 c^2$	$(1 \pm 3 \cdot 10^{-8})\,\sigma$	20 k€
Wien's displacement constant	none	$b = \lambda_{max} T$	$(1 \pm 10^{-5})\,b$	20 k€
Refractive index of water	none	1.34	a few %	1 k€
Photon-photon scattering	0	from QED: finite	agrees	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 yg	1 M€
Proton lifetime	≈ 1 μs	∞	$> 10^{35}$ a	100 M€
Composite matter properties				
Atom lifetime	≈ 1 μs	∞	$> 10^{20}$ a	1 €
Molecular size	none	from QED	within 10^{-3}	20 k€
Von Klitzing constant	∞	$h/e^2 = \mu_0 c/2\alpha$	$(1 \pm 10^{-7})\,h/e^2$	1 M€
AC Josephson constant	0	$2e/h$	$(1 \pm 10^{-6})\,2e/h$	5 M€
Heat capacity of metals at 0 K	25 J/K	0	$< 10^{-3}$ J/K	10 k€
Heat capacity of diatomic gas at 0 K	25 J/K	0	$< 10^{-3}$ J/K	10 k€
Water density	none	1000.00 kg/m³ at 4 °C	agrees	10 k€
Minimum electr. conductivity	0	$G = 2e^2/\hbar$	$G(1 \pm 10^{-3})$	3 k€
Ferromagnetism	none	exists	exists	2 €
Superfluidity	none	exists	exists	200 k€
Bose–Einsein condensation	none	exists	exists	2 M€
Superconductivity (metal)	none	exists	exists	100 k€

Observable	Classical prediction	Prediction of quantum theory[a]	Measurement	Cost[b] estimate
Superconductivity (high T)	none	*none yet*	exists	100 k€

Page 247
a. All these predictions are calculated from the fundamental quantities given in Appendix A.
b. Sometimes the cost for the calculation of the prediction is higher than that of the experimental observation. (Can you spot the examples?) The sum of the two is given.
Challenge 157 s

We notice that the values predicted by quantum theory do not differ from the measured ones. In contrast, classical physics does not allow to calculate any of the observed values. This shows the progress that quantum physics has brought.

In short, 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. 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 23.
Page 239

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* $\hbar = 1.1 \cdot 10^{-34}$ 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.*

In fact, the second statement results from the first. The existence of a smallest action in nature directly leads to the main lesson we learned about motion in the quantum part of our adventure:

▷ *If it moves, it is made of quantons, or quantum 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 *quantons*. Stones, water waves, light, sound waves, earthquakes, gelatine and everything else we can interact with is made of quantum particles.

Once we asked: what is matter and what are interactions? Now we know: they are composites of elementary quantum particles.

Page 191
To be clear, an *elementary quantum particle* is a countable entity, smaller than its own Compton wavelength, described by energy–momentum, mass, spin, C, P and T parity. As we will see in the next volume however, this is not the complete list of intrinsic particle properties.

Challenge 158 e
All moving entities are made of quantum 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.

Quantum particle motion

Quantons, or quantum particles, differ from everyday particles: quantum particles *interfere*: they behave like a mixture of particles and waves. This property follows directly from the existence of $\hbar$, the smallest action in nature. From the existence of $\hbar$, quantum theory deduces all its statements about quantum particle motion. We summarize the main ones.

There is *no rest* for microscopic particles. All objects obey the indeterminacy principle, which states that the indeterminacies in position x and momentum p follow

$$\Delta x \Delta p \geqslant \hbar/2 \quad \text{with} \quad \hbar = 1.1 \cdot 10^{-34}\,\text{Js} \tag{123}$$

and 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 minimum action value through the interactions with the baths in the environment of any system.

Large number of identical particles with the same momentum behave like *waves*. The associated de Broglie wavelength λ is given by the momentum p of a single particle through

$$\lambda = \frac{h}{p} = \frac{2\pi\hbar}{p} \tag{124}$$

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. Quantum theory states that all waves interfere, refract, disperse, dampen, can be dampened and can be polarized. This applies to electrons, atoms, photons and molecules. All waves being made of quantum 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. The interference and diffraction of wave particles is observed daily in the electron microscope.

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.

All particles and all particle beams can be rotated. Particles possess an intrinsic 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 low-energy fermion. Solids are impenetrable because of the fermion character of its electrons in the atoms.

Identical particles are *indistinguishable*. Radiation is made of indistinguishable particles called 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. Perfect copying machines do not exist.

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.

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; they can be created and annihilated only in pairs. Apart from neutrinos, elementary fermions have non-vanishing mass and move slower than light.

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.

The appearance of Planck's constant $\hbar$ implies that *length scales* and *time 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 metre 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.

Ref. 239 Quantum physics leaves no room for cold fusion, astrology, teleportation, telekinesis, supernatural phenomena, multiple universes, or faster than light phenomena – the EPR paradox notwithstanding.

Results of quantum field theory

Quantum field theory is that part of quantum theory that includes the process of transformation of particles into each other. The possibility of transformation results from the existence of a minimum action and of a maximum speed. Particle transformations have important consequences. Page 27

Radioactivity, the working of the sun, the history of the composite matter we are made of is due to particle transformations.

Objects are composed of quantum particles. Quantum field theory provides a complete list of the intrinsic properties which make up what is called an 'object' in everyday Page 192 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.

Composite matter is separable because of the finite interaction energies of the constituents. Atoms are made of a nucleus made of quarks surrounded by electrons. Their interactions provide an effective minimal length scale to all everyday matter.

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 Page 191 the photon, the eight gluons and the two weak interaction bosons.

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, an object of mass m can be localized only within intervals of the Compton wavelength

$$\lambda_C = \frac{h}{mc} = \frac{2\pi\hbar}{mc} , \tag{125}$$

where c is the speed of light. 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 non-linearities thus appearing produce small departures from the superposition principle for electromagnetic fields, resulting in photon-photon scattering.

Quantum chromodynamics, the field theory of the strong interactions, explains the Page 148 masses of mesons and baryons through its descriptions as bound quark states. At fundamental scales, the strong interaction is mediated by eight elementary gluons. At femtometre scales, the strong interaction effectively acts through the exchange of spin 0 pions, is strongly attractive, and leads to the formation of atomic nuclei.

The theory of electroweak interactions describes electromagnetism and the weak interactions. It includes the Higgs mechanism, massive Vector bosons, quark mixings and neutrino mixings.

Quantum theory explains the origin of material properties and 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 materials 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.

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.

The essence of quantum theory

Generalizing even more, we can summarize quantum physics with a simple statement: *quantum physics is the description of matter and radiation* without *the concept of infinitely small*. Matter and radiation are described by finite quantities.

We had already eliminated the infinitely large in our exploration of relativity. Quantum theory eliminates the infinitely small from the description of matter and radiation. However, some types of infinities remain. We had to retain the infinitely small in the description of space or time, and in topics related to them, such as renormalization. Therefore, we did not manage to eliminate all infinities yet; we are not yet at the end of our quest. Surprisingly, we shall soon find out that a completely finite description of all of nature is equally impossible. Let us have a look at the path that still remains to be followed.

What is unexplained by quantum theory and general relativity?

The material gathered in this quantum part of our mountain ascent, together with the earlier summary of general relativity, allows us to describe all observed phenomena connected to motion. Therefore, we are now able to provide a complete list of the *unexplained* properties of nature. Whenever we ask 'why?' about an observation and continue doing so after each answer, we arrive at one of the points listed in Table 23. The table lists all issues that were unexplained in the year 2000, so that we can call it the *millennium list* of open problems.

Vol. II, page 259

TABLE 23 The millennium list: *everything* the standard model of particle physics and general relativity do *not* explain; thus, also the list of *the only* experimental data available to test the unified description of motion.

Observable	Property unexplained in the year 2000
Local quantities, from quantum field theory: particle properties	
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 value of the strong coupling constant at one specific energy value
m_{q}	the values of the 6 quark masses
m_{l}	the values of 6 lepton masses
m_{W}	the value of the mass of the W vector boson

TABLE 23 (Continued) *Everything* the standard model and general relativity do *not* explain.

Observable	Property unexplained in the year 2000
m_H	the value of the mass of the scalar Higgs boson
θ_w	the value of the weak mixing angle
$\theta_{12}, \theta_{13}, \theta_{23}$	the value of the three quark mixing angles
δ	the value of the CP violating phase for quarks
$\theta^\nu_{12}, \theta^\nu_{13}, \theta^\nu_{23}$	the value of the three neutrino mixing angles
$\delta^\nu, \alpha_1, \alpha_2$	the value of the three CP violating phases for neutrinos
$3 \cdot 4$	the number of fermion generations and of particles in each generation
P, C, etc.	the origin of all quantum numbers of each fermion and each boson
Local mathematical structures, from quantum field theory	
$c, \hbar, k$	the origin of the invariant Planck units of quantum field theory
$3 + 1$	the number of dimensions of physical space and time
SO(3,1)	the origin of Lorentz and Poincaré symmetry (i.e., of spin, position, energy, momentum)
$S(n)$	the origin of particle identity, i.e., of permutation symmetry
Gauge symmetry	the origin of the gauge groups, in particular:
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 and its breaking
SU(3)	the origin of strong interaction gauge group
Ren. group	the origin of renormalization properties
$\delta W = 0$	the origin of the least action principle in quantum theory
$W = \int L_{SM}\, dt$	the origin of the Lagrangian of the standard model of particle physics
Global quantities, from general relativity: vacuum and energy properties	
$3 + 1$	the number of dimensions of physical space and time
0	the observed flatness, i.e., vanishing curvature, of the universe
$1.2(1) \cdot 10^{26}$ m	the distance of the horizon, i.e., the 'size' of the universe (if it makes sense)
$\rho_{de} = \Lambda c^4/(8\pi G)$ $= 0.5\,\mathrm{nJ/m^3}$	the value and nature of the observed vacuum energy density, dark energy or cosmological constant
$(5 \pm 4) \cdot 10^{79}$	the number of baryons in the universe (if it makes sense), i.e., the average visible matter density in the universe
$f_0(1, ..., c.\ 10^{90})$	the initial conditions for *c.* 10^{90} particle fields in the universe (if or as long as they make sense), including the homogeneity and isotropy of matter distribution, and the density fluctuations at the origin of galaxies
ρ_{dm}	the density and nature of dark matter
Global mathematical structures, from general relativity	
c, G	the origin of the invariant Planck units of general relativity
$\delta \int L_{GR}\, dt = 0$	the origin of the least action principle and the Lagrangian of general relativity
$R \times S^3$	the observed topology of the universe

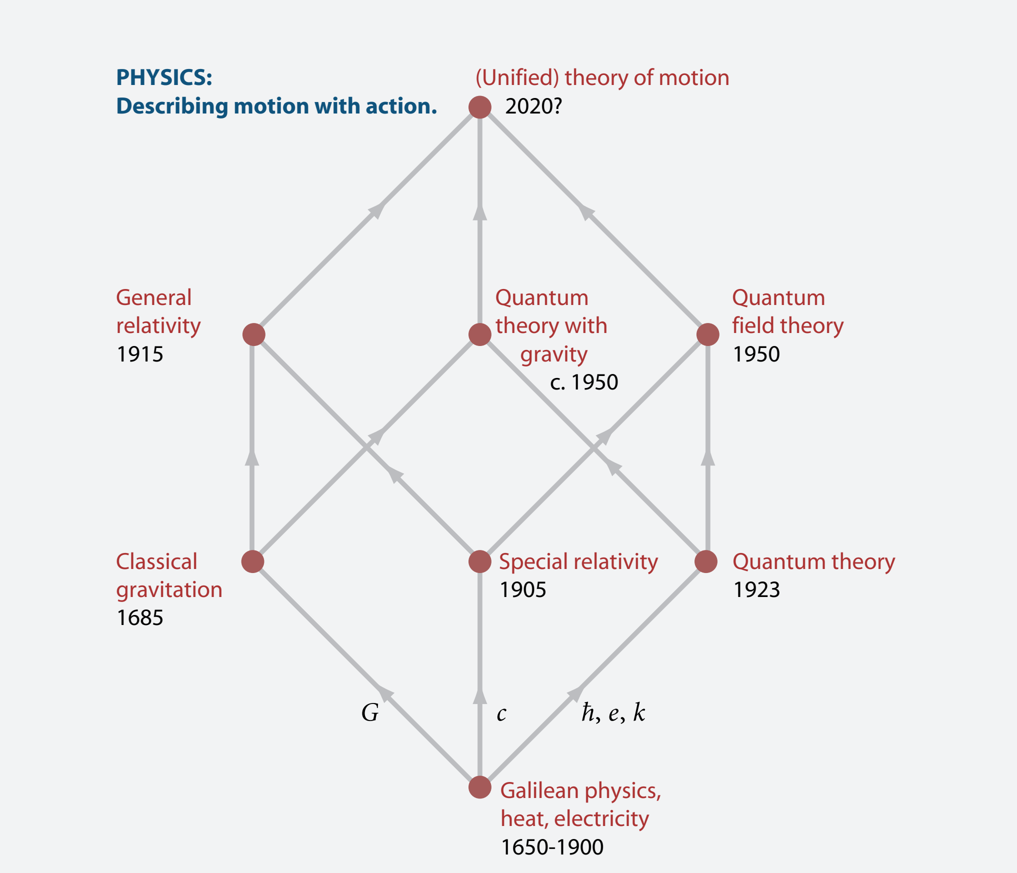


FIGURE 111 A simplified history of the description of motion in physics, by giving the limits to motion included in each description

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 millennium list, Table 23, contains extremely *different* concepts. That means that at this point of our walk there is *a lot* we do not understand. Finding the answers will not be easy, but will require effort.

On the other hand, the millennium list of unexplained properties of nature is also *short*. The description of nature our adventure has produced so far is concise and precise. No discrepancies from experiments are known. In other words, we have a good description of motion *in practice*. Going further is 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, as the millennium list shows, even though we have *no* known discrepancies with experiments, we are *not* at the top of Motion Mountain.

The physics cube

Another summary of the progress and open issues of physics, already given in the introduction, is shown in Figure 111. From the lowest corner of a cube, representing Galilean physics and related topics, three edges – labelled c, G and $\hbar$, e, k – lead to classical gravity, special relativity and quantum theory. Each constant implies a limit to motion; in the corresponding theory, this *one* limit is taken into account. From these first level theories, corresponding parallel edges lead upwards to general relativity, quantum field theory and quantum theory in gravity; in each of theses second level theories, *two* of the limits.*

From the second level theories, all edges lead to the last missing corner: the (unified) theory of motion. The theory of motion takes into account all limits found so far. Only this theory is a complete or unified description of nature. The important point is that we already know all limits to motion. To arrive at the last point, no new experiments are necessary. No new knowledge is required. We only have to advance in the right direction, with careful thinking. Reaching the final theory of motion is the topic of the last volume of our adventure.

How to delude oneself that one has reached the top of Motion Mountain

Nowadays it is sometimes deemed chic to pretend that the adventure is over at the stage we have just reached.** The reasoning is as follows. If we change the values of the unexplained constants from the millennium list of Table 23 only ever so slightly, nature would look completely different from what it does. The consequences have been studied in great detail; Table 24 gives an overview of the results.

Ref. 241

TABLE 24 A selection of the consequences of changing the properties of nature

Observable	Change	Result
Local quantities, from quantum theory		
α_{em}	smaller:	only short lived, smaller and hotter stars; no Sun

* Of course, Figure 111 gives a simplified view of the history of physics. A more precise diagram would use three different arrows for $\hbar$, e and k, making the figure a five-dimensional cube. However, not all of its corners would have dedicated theories (can you confirm this?). The diagram would be much less appealing; but most of all, the conclusions mentioned in the text would not change.

Challenge 159 e

** Actually this attitude is not new. Only the arguments have changed. Maybe the greatest physicist ever, James Clerk Maxwell, already fought against this attitude over a hundred years ago: '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. 240

Observable	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
	$G_F m_e^2 \neq \sqrt{G m_e^2}$:	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
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
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

Observable	Change	Result
Jupiter mass	larger:	too little comet impacts on Earth; no Moon; no dinosaur extinction
Oort cloud object number	smaller:	no comets; no irregular asteroids; no Moon; still dinosaurs
galaxy centre distance	smaller:	irregular planet motion; supernova dangers
initial cosmic speed	+0.1%:	1000 times faster universe expansion
	−0.0001%:	universe recollapses after 10 000 years
vacuum energy density	change by 10^{-55}:	no flatness
3 + 1 dimensions	different:	no atoms, no planetary systems
Local structures, from quantum theory		
permutation symmetry	none:	no matter
Lorentz symmetry	none:	no communication possible
U(1)	different:	no Huygens principle, no way to *see* anything
SU(2)	different:	no radioactivity, no Sun, no life
SU(3)	different:	no stable quarks and nuclei
Global structures, from general relativity		
topology	other:	unknown; possibly correlated gamma ray bursts or star images at the antipodes

Some researchers even speculate that the whole Table 24 can be condensed into a single sentence: if any parameter in nature is changed, the universe would either have too many or too few black holes. However, the proof of this condensed summary is not complete yet.

Ref. 242

Challenge 160 r

In fact, Table 24, on the effects of changing nature, is overwhelming. It shows that 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. The three most fashionable beliefs are that the universe is *created* or *designed*, that the universe is *designed for people*, or that the values are random, as 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 or from the simple request that statements be based on observations. In 2004,

Page 238

Page 244

this belief has even become fashionable among older string theorists. The third belief, namely *multiple universes*, is a minority view, but also sells well.

Stopping our mountain ascent with a belief at the present point is not different from doing so directly at the beginning. This choice was taken in societies which lacked the passion for rational investigation, and still is 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 24 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, but *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 is swept away by large wave. Our walk has repeatedly reminded us of this smallness and dependence, and overwhelmingly does so again at this point.

Having faced this powerful experience, everybody has to make up his own mind on whether to proceed with the adventure or not. Of course, there is no obligation to do so. Challenge 161 s

What awaits us?

The shortness of the list of unexplained aspects of nature means that *no additional experimental data* are 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 or astronomical experiments with the topology issue.

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: this is not an adventure of action, but an adventure of the mind. And it is an incredible one, as we shall soon find out. To provide a feeling of what awaits us, we rephrase the remaining issues in five simple challenges.

1 – What determines *colours*? In other words, what relations of nature fix the famous fine structure constant? Like the hero of Douglas Adams' books, physicists know the answer to the greatest of questions: it is 137.036. But they do not know the question.

2 – 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?

3 – 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 the existence of space-time, and neither *predicts* them. Even worse, both theories completely fail to predict the existence of *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.

4 – Was Democritus *wrong*? It is often said that the standard model has only about twenty unknown parameters; this common mistake negates about 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

$$N = n\, v\, d\, p\, f \qquad (126)$$

where n is the number of particles, v is the number of variables (position, momentum, spin), d is the number of different values each of them can take (limited by the maximum of 61 decimal digits), p is the number of visible space-time points (about 10^{183}) and f is a factor expressing how many of all these initial conditions are actually independent of each other. We thus have the following number of possibilities

$$N = 10^{92} \cdot 8 \cdot 10^{61} \cdot 10^{183} \cdot f = 10^{336} \cdot f \qquad (127)$$

from which the 10^{93} actual initial conditions have to be explained. There is a small problem that we know nothing whatsoever about f. Its value could be 0, if all data were interdependent, or 1, if none were. Worse, above we noted that initial conditions cannot be defined for the universe at all; thus f should be undefined and not be a number at all! Whatever the case, we need to understand how all the visible particles get their 10^{93} states assigned from this range of options.

Page 143

5 – 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, whereas 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 at this point of our walk, as you might want to check by yourself.

Page 334
Challenge 162 e

The answers to these five 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. 243

APPENDIX A

UNITS, MEASUREMENTS AND CONSTANTS

Measurements are comparisons with standards. Standards are based on a *unit*. Many different systems of units have been used throughout the world. Most standards confer power to the organization in charge of them. Such power can be misused; this is the case today, for example in the computer industry, and was so in the distant past. The solution is the same in both cases: organize an independent and global standard. For units, this happened in the eighteenth century: to avoid misuse by authoritarian institutions, to eliminate problems with differing, changing and irreproducible standards, and – this is not a joke – to simplify tax collection, a group of scientists, politicians and economists agreed on a set of units. It is called the *Système International d'Unités*, abbreviated *SI*, and is defined by an international treaty, the 'Convention du Mètre'. The units are maintained by an international organization, the 'Conférence Générale des Poids et Mesures', and its daughter organizations, the 'Commission Internationale des Poids et Mesures' and the 'Bureau International des Poids et Mesures' (BIPM), which all originated in the times just before the French revolution. Ref. 244

SI units

All SI units are built from seven *base units*, whose official definitions, translated from French into English, are given below, together with the dates of their formulation:

- 'The *second* is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom.' (1967)*
- 'The *metre* is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second.' (1983)
- 'The *kilogram* is the unit of mass; it is equal to the mass of the international prototype of the kilogram.' (1901)*
- 'The *ampere* is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 metre apart in vacuum, would produce between these conductors a force equal to $2 \cdot 10^{-7}$ newton per metre of length.' (1948)
- 'The *kelvin*, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.' (1967)*
- 'The *mole* is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12.' (1971)*
- 'The *candela* is the luminous intensity, in a given direction, of a source that emits

monochromatic radiation of frequency $540 \cdot 10^{12}$ hertz and has a radiant intensity in that direction of (1/683) watt per steradian.' (1979)*

Note that both time and length units are defined as certain properties of a standard example of motion, namely light. In other words, also the Conférence Générale des Poids et Mesures makes the point that the observation of motion is a *prerequisite* for the definition and construction of time and space. *Motion is the fundament each observation and measurements.* By the way, the use of light in the definitions had been proposed already in 1827 by Jacques Babinet.*

From these basic units, all other units are defined by multiplication and division. Thus, all SI units have the following properties:

- SI units form a system with *state-of-the-art precision*: all units are defined with a precision that is higher than the precision of commonly used measurements. Moreover, the precision of the definitions is regularly being improved. The present relative uncertainty of the definition of the second is around 10^{-14}, for the metre about 10^{-10}, for the kilogram about 10^{-9}, for the ampere 10^{-7}, for the mole less than 10^{-6}, for the kelvin 10^{-6} and for the candela 10^{-3}.
- SI units form an *absolute* system: all units are defined in such a way that they can be reproduced in every suitably equipped laboratory, independently, and with high precision. This avoids as much as possible any misuse by the standard-setting organization. (The kilogram, still defined with the help of an artefact, is the last exception to this requirement; extensive research is under way to eliminate this artefact from the definition – an international race that will take a few more years. There are two approaches: counting particles, or fixing $\hbar$. The former can be achieved in crystals, the latter using any formula where $\hbar$ appears, such as the formula for the de Broglie wavelength or that of the Josephson effect.)
- SI units form a *practical* system: the base units are quantities of everyday magnitude. Frequently used units have standard names and abbreviations. The complete list includes the seven base units, the supplementary units, the derived units and the admitted units.

The *supplementary* SI units are two: the unit for (plane) angle, defined as the ratio of arc length to radius, is the *radian* (rad). For solid angle, defined as the ratio of the subtended area to the square of the radius, the unit is the *steradian* (sr).

The *derived* units with special names, in their official English spelling, i.e., without capital letters and accents, are:

* The respective symbols are s, m, kg, A, K, mol and cd. The international prototype of the kilogram is a platinum–iridium cylinder kept at the BIPM in Sèvres, in France. For more details on the levels of the caesium atom, consult a book on atomic physics. The Celsius scale of temperature 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. SI also states: 'When the mole is used, the elementary entities must be specified and may be atoms, molecules, ions, electrons, other particles, or specified groups of such particles.' In the definition of the mole, it is understood that the carbon 12 atoms are unbound, at rest and in their ground state. In the definition of the candela, the frequency of the light corresponds to 555.5 nm, i.e., green colour, around the wavelength to which the eye is most sensitive.
Vol. I, page 88
Ref. 245

* Jacques Babinet (1794–1874), French physicist who published important work in optics.

Name	Abbreviation
hertz	Hz = 1/s
pascal	$\mathrm{Pa = N/m^2 = kg/m\,s^2}$
watt	$\mathrm{W = kg\,m^2/s^3}$
volt	$\mathrm{V = kg\,m^2/As^3}$
ohm	$\Omega = \mathrm{V/A = kg\,m^2/A^2s^3}$
weber	$\mathrm{Wb = Vs = kg\,m^2/As^2}$
henry	$\mathrm{H = Vs/A = kg\,m^2/A^2s^2}$
lumen	lm = cd sr
becquerel	Bq = 1/s
sievert	$\mathrm{Sv = J/kg = m^2/s^2}$

Name	Abbreviation
newton	$\mathrm{N = kg\,m/s^2}$
joule	$\mathrm{J = Nm = kg\,m^2/s^2}$
coulomb	C = As
farad	$\mathrm{F = As/V = A^2s^4/kg\,m^2}$
siemens	$\mathrm{S} = 1/\Omega$
tesla	$\mathrm{T = Wb/m^2 = kg/As^2 = kg/Cs}$
degree Celsius	°C (see definition of kelvin)
lux	$\mathrm{lx = lm/m^2 = cd\,sr/m^2}$
gray	$\mathrm{Gy = J/kg = m^2/s^2}$
katal	kat = mol/s

We note that in all definitions of units, the kilogram only appears to the powers of 1, 0 and −1. The final explanation for this fact appeared only recently. Can you try to formulate the reason?

Challenge 163 ny

The *admitted* non-SI units are *minute*, *hour*, *day* (for time), *degree* $1° = \pi/180$ rad, *minute* $1' = \pi/10\,800$ rad, *second* $1'' = \pi/648\,000$ rad (for angles), *litre* and *tonne*. All other units are to be avoided.

All SI units are made more practical by the introduction of standard names and abbreviations for the powers of ten, the so-called *prefixes*:*

Power	Name	
10^{1}	deca	da
10^{2}	hecto	h
10^{3}	kilo	k
10^{6}	Mega	M
10^{9}	Giga	G
10^{12}	Tera	T
10^{15}	Peta	P

Power	Name	
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f

Power	Name	
10^{18}	Exa	E
10^{21}	Zetta	Z
10^{24}	Yotta	Y
unofficial:		
10^{27}	Xenta	X
10^{30}	Wekta	W
10^{33}	Vendekta	V
10^{36}	Udekta	U

Ref. 246

Power	Name	
10^{-18}	atto	a
10^{-21}	zepto	z
10^{-24}	yocto	y
10^{-27}	xenno	x
10^{-30}	weko	w
10^{-33}	vendeko	v
10^{-36}	udeko	u

▪ SI units form a *complete* system: they cover in a systematic way the complete set of observables of physics. Moreover, they fix the units of measurement for all other sciences

* Some of these names are invented (yocto to sound similar to Latin *octo* 'eight', zepto to sound similar to Latin *septem*, yotta and zetta to resemble them, exa and peta to sound like the Greek words ἑξάκις and πεντάκις for 'six times' and 'five times', the unofficial ones to sound similar to the Greek words for nine, ten, eleven and twelve); some are from Danish/Norwegian (atto from *atten* 'eighteen', femto from *femten* 'fifteen'); some are from Latin (from *mille* 'thousand', from *centum* 'hundred', from *decem* 'ten', from *nanus* 'dwarf'); some are from Italian (from *piccolo* 'small'); some are Greek (micro is from μικρός 'small', deca/deka from δέκα 'ten', hecto from ἑκατόν 'hundred', kilo from χίλιοι 'thousand', mega from μέγας 'large', giga from γίγας 'giant', tera from τέρας 'monster').

Translate: I was caught in such a traffic jam that I needed a microcentury for a picoparsec and that my car's fuel consumption was two tenths of a square millimetre.

Challenge 164 e

as well.

▪ SI units form a *universal* system: they can be used in trade, in industry, in commerce, at home, in education and in research. They could even be used by extraterrestrial civilizations, if they existed.

▪ SI units form a *coherent* system: the product or quotient of two SI units is also an SI unit. This means that in principle, the same abbreviation, e.g. 'SI', could be used for every unit.

The SI units are not the only possible set that could fulfil all these requirements, but they are the only existing system that does so.*

Since every measurement is a comparison with a standard, any measurement requires *matter* to realize the standard (even for a speed standard), and *radiation* to achieve the comparison. The concept of measurement thus assumes that matter and radiation exist and can be clearly separated from each other.

Challenge 165 e

Planck's natural units

Since the exact form of many equations depends on the system of units used, theoretical physicists often use unit systems optimized for producing simple equations. The chosen units and the values of the constants of nature are related. In microscopic physics, the system of *Planck's natural units* is frequently used. They are defined by setting $c = 1$, $\hbar = 1$, $G = 1$, $k = 1$, $\varepsilon_0 = 1/4\pi$ and $\mu_0 = 4\pi$. Planck units are thus defined from combinations of fundamental constants; those corresponding to the fundamental SI units are given in Table 26.** The table is also useful for converting equations written in natural units back to SI units: just substitute every quantity X by $X/X_{\rm Pl}$.

Challenge 166 e

TABLE 26 Planck's (uncorrected) natural units

Name	Definition			Value	
Basic units					
the Planck length	$l_{\rm Pl}$	=	$\sqrt{\hbar G/c^3}$	=	$1.616\,0(12)\cdot10^{-35}$ m
the Planck time	$t_{\rm Pl}$	=	$\sqrt{\hbar G/c^5}$	=	$5.390\,6(40)\cdot10^{-44}$ s
the Planck mass	$m_{\rm Pl}$	=	$\sqrt{\hbar c/G}$	=	21.767(16) µg
the Planck current	$I_{\rm Pl}$	=	$\sqrt{4\pi\varepsilon_0 c^6/G}$	=	$3.479\,3(22)\cdot10^{25}$ A

* Apart from international units, there are also *provincial* units. Most provincial units still in use are of Roman origin. The mile comes from *milia passum*, which used to be one thousand (double) strides of about 1480 mm each; today a nautical mile, once defined as minute of arc on the Earth's surface, is exactly 1852 m). The inch comes from *uncia/onzia* (a twelfth – now of a foot). The pound (from *pondere* 'to weigh') is used as a translation of *libra* – balance – which is the origin of its abbreviation lb. Even the habit of counting in dozens instead of tens is Roman in origin. These and all other similarly funny units – like the system in which all units start with 'f', and which uses furlong/fortnight as its unit of velocity – are now officially defined as multiples of SI units.

** 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 $h = 2\pi\hbar$. The electromagnetic units can also be defined with other factors than $4\pi\varepsilon_0$ in the expressions: for example, using $4\pi\varepsilon_0\alpha$, with the *fine structure constant* α, gives $q_{\rm Pl} = e$. For the explanation of the numbers between brackets, the standard deviations, see below.

Name	Definition				Value
the Planck temperature	T_{Pl}	=	$\sqrt{\hbar c^5/Gk^2}$	=	$1.417\,1(91) \cdot 10^{32}$ K
Trivial units					
the Planck velocity	v_{Pl}	=	c	=	0.3 Gm/s
the Planck angular momentum	L_{Pl}	=	$\hbar$	=	$1.1 \cdot 10^{-34}$ Js
the Planck action	S_{aPl}	=	$\hbar$	=	$1.1 \cdot 10^{-34}$ Js
the Planck entropy	S_{ePl}	=	k	=	13.8 yJ/K
Composed units					
the Planck mass density	ρ_{Pl}	=	$c^5/G^2\hbar$	=	$5.2 \cdot 10^{96}$ kg/m^3
the Planck energy	E_{Pl}	=	$\sqrt{\hbar c^5/G}$	=	2.0 GJ = $1.2 \cdot 10^{28}$ eV
the Planck momentum	p_{Pl}	=	$\sqrt{\hbar c^3/G}$	=	6.5 Ns
the Planck power	P_{Pl}	=	c^5/G	=	$3.6 \cdot 10^{52}$ W
the Planck force	F_{Pl}	=	c^4/G	=	$1.2 \cdot 10^{44}$ N
the Planck pressure	p_{Pl}	=	$c^7/G\hbar$	=	$4.6 \cdot 10^{113}$ Pa
the Planck acceleration	a_{Pl}	=	$\sqrt{c^7/\hbar G}$	=	$5.6 \cdot 10^{51}$ m/s^2
the Planck frequency	f_{Pl}	=	$\sqrt{c^5/\hbar G}$	=	$1.9 \cdot 10^{43}$ Hz
the Planck electric charge	q_{Pl}	=	$\sqrt{4\pi\varepsilon_0 c\hbar}$	=	1.9 aC = 11.7 e
the Planck voltage	U_{Pl}	=	$\sqrt{c^4/4\pi\varepsilon_0 G}$	=	$1.0 \cdot 10^{27}$ V
the Planck resistance	R_{Pl}	=	$1/4\pi\varepsilon_0 c$	=	30.0 Ω
the Planck capacitance	C_{Pl}	=	$4\pi\varepsilon_0\sqrt{\hbar G/c^3}$	=	$1.8 \cdot 10^{-45}$ F
the Planck inductance	L_{Pl}	=	$(1/4\pi\varepsilon_0)\sqrt{\hbar G/c^7}$	=	$1.6 \cdot 10^{-42}$ H
the Planck electric field	E_{Pl}	=	$\sqrt{c^7/4\pi\varepsilon_0\hbar G^2}$	=	$6.5 \cdot 10^{61}$ V/m
the Planck magnetic flux density	B_{Pl}	=	$\sqrt{c^5/4\pi\varepsilon_0\hbar G^2}$	=	$2.2 \cdot 10^{53}$ T

The natural units are important for another reason: whenever a quantity is sloppily called 'infinitely small (or large)', the correct expression is 'as small (or as large) as the corresponding corrected Planck unit'. As explained throughout the text, and especially in the Page 33 final part, this substitution is possible because almost all Planck units provide, within a correction factor of order 1, the extremal value for the corresponding observable – some an upper and some a lower limit. Unfortunately, these correction factors are not yet widely known. The exact extremal value for each observable in nature is obtained when G is substituted by $4G$ and $4\pi\varepsilon_0$ by $4\pi\varepsilon_0\alpha$ in all Planck quantities. These extremal values, or *corrected Planck units*, are the *true natural units*. To exceed the extremal values Challenge 167 s is possible only for some extensive quantities. (Can you find out which ones?)

Other unit systems

A central aim of research in high-energy physics is the calculation of the strengths of all interactions; therefore it is not practical to set the gravitational constant G to unity, as in the Planck system of units. For this reason, high-energy physicists often only set

$c = \hbar = k = 1$ and $\mu_0 = 1/\varepsilon_0 = 4\pi$,* leaving only the gravitational constant G in the equations.

In this system, only one fundamental unit exists, but its choice is free. Often a standard length is chosen as the fundamental unit, length being the archetype of a measured quantity. The most important physical observables are then related by

$$\begin{aligned} 1/[l^2] &= [E]^2 = [F] = [B] = [E_{\text{electric}}] ,\\ 1/[l] &= [E] = [m] = [p] = [a] = [f] = [I] = [U] = [T] ,\\ 1 &= [v] = [q] = [e] = [R] = [S_{\text{action}}] = [S_{\text{entropy}}] = \hbar = c = k = [\alpha] ,\\ [l] &= 1/[E] = [t] = [C] = [L] \quad \text{and}\\ [l]^2 &= 1/[E]^2 = [G] = [P] \end{aligned} \tag{128}$$

where we write $[x]$ for the unit of quantity x. Using the same unit for time, capacitance and inductance is not to everybody's taste, however, and therefore electricians do not use this system.**

Often, in order to get an impression of the energies needed to observe an effect under study, a standard energy is chosen as fundamental unit. In particle physics the most common energy unit is the *electronvolt* (eV), defined as the kinetic energy acquired by an electron when accelerated by an electrical potential difference of 1 volt ('protonvolt' would be a better name). Therefore one has $1\,\text{eV} = 1.6 \cdot 10^{-19}\,\text{J}$, or roughly

$$1\,\text{eV} \approx \tfrac{1}{6}\,\text{aJ} \tag{129}$$

which is easily remembered. The simplification $c = \hbar = 1$ yields $G = 6.9 \cdot 10^{-57}\,\text{eV}^{-2}$ and allows one to use the unit eV also for mass, momentum, temperature, frequency, time and length, with the respective correspondences $1\,\text{eV} \equiv 1.8 \cdot 10^{-36}\,\text{kg} \equiv 5.4 \cdot 10^{-28}\,\text{Ns} \equiv 242\,\text{THz} \equiv 11.6\,\text{kK}$ and $1\,\text{eV}^{-1} \equiv 4.1\,\text{fs} \equiv 1.2\,\mu\text{m}$. Challenge 168 e

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 belongs to a cosmic ray with an energy of $3 \cdot 10^{20}$ eV or 48 J. Ref. 248 Down here on the Earth, an accelerator able to produce an energy of about 105 GeV or 17 nJ for electrons and antielectrons has been built, and one able to produce an energy of 14 TeV or 2.2 μJ for protons will be finished soon. 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

* 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. Ref. 247

** In the list, l is length, E energy, F force, E_{electric} the electric and B the magnetic field, m mass, p momentum, a acceleration, f frequency, I electric current, U voltage, T temperature, v speed, q charge, R resistance, P power, G the gravitational constant.

The web page www.chemie.fu-berlin.de/chemistry/general/units_en.html provides a tool to convert various units into each other.

Researchers in general relativity often use another system, in which the *Schwarzschild radius* $r_S = 2Gm/c^2$ is used to measure masses, by setting $c = G = 1$. In this case, mass and length have the *same* dimension, and $\hbar$ has the dimension of an area.

Ref. 249 held inside a special cooling system. The interior of that cryostat may even be the coolest point in the whole universe. 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.

Curiosities and fun challenges about units

Not using SI units can be expensive. In 1999, NASA lost a satellite on Mars because some software programmers had used provincial units instead of SI units in part of the code. As a result, the Mars Climate Orbiter crashed into the planet, instead of orbiting it; the loss was around 100 million euro.*

* *

A *gray* is the amount of radioactivity that deposits 1 J on 1 kg of matter. A *sievert* is the unit of radioactivity adjusted to humans by weighting each type of human tissue with a factor representing the impact of radiation deposition on it. 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 implies an irradiation
Ref. 250 of 1 mSv; a CAT scan 8 mSv.

* *

The Planck length is roughly the de Broglie wavelength $\lambda_B = h/mv$ of a man walking
Ref. 251 comfortably ($m = 80$ kg, $v = 0.5$ m/s); this motion is therefore aptly called the 'Planck stroll.'

* *

The Planck mass is equal to the mass of about 10^{19} protons. This is roughly the mass of a human embryo at about ten days of age.

* *

The most precisely measured quantities in nature are the frequencies of certain millisec-
Ref. 252 ond pulsars, the frequency of certain narrow atomic transitions, and the Rydberg constant of *atomic* hydrogen, which can all be measured as precisely as the second is defined. The caesium transition that defines the second has a finite line width that limits the achievable precision: the limit is about 14 digits.

* *

The most precise clock ever built, using microwaves, had a stability of 10^{-16} during a
Ref. 253 running time of 500 s. For longer time periods, the record in 1997 was about 10^{-15}; but
Ref. 254 values around 10^{-17} seem 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; it usually lies between 1 ms for optical clocks and 5000 s for masers. Pulsars are the only type of clock

* This story revived an old (and false) urban legend that states that only three countries in the world do not use SI units: Liberia, the USA and Myanmar.

for which this region is not known yet; it certainly lies at more than 20 years, the time elapsed at the time of writing since their discovery.

* *

Ref. 255 Challenge 169 s The shortest times measured are the lifetimes of certain 'elementary' particles. In particular, the lifetime of certain D mesons have been measured at less than 10^{-23} s. Such times are measured using a bubble chamber, where the track is photographed. Can you estimate how long the track is? (This is a trick question – if your length cannot be observed with an optical microscope, you have made a mistake in your calculation.)

* *

Ref. 256 The longest times encountered in nature are the lifetimes of certain radioisotopes, over 10^{15} years, and the lower limit of certain proton decays, over 10^{32} years. These times are thus much larger than the age of the universe, estimated to be fourteen thousand million years.

Precision and accuracy of measurements

Measurements are the basis of physics. Every measurement has an *error*. Errors are due to lack of precision or to lack of accuracy. *Precision* means how well a result is reproduced when the measurement is repeated; *accuracy* is the degree to which a measurement corresponds to the actual value. Lack of precision is due to accidental or *random errors*; they are best measured by the *standard deviation*, usually abbreviated σ; it is defined through

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 , \qquad (130)$$

Challenge 170 s where $\bar{x}$ is the average of the measurements x_i. (Can you imagine why $n-1$ is used in the formula instead of n?)

For most experiments, the distribution of measurement values tends towards a normal distribution, also called *Gaussian distribution*, whenever the number of measurements is increased. The distribution, shown in Figure 112, is described by the expression

$$N(x) \approx \mathrm{e}^{-\frac{(x-\bar{x})^2}{2\sigma^2}} . \qquad (131)$$

Challenge 171 e The square σ^2 of the standard deviation is also called the *variance*. For a Gaussian distribution of measurement values, 2.35σ is the full width at half maximum.

Ref. 257 Lack of accuracy is due to *systematic errors*; usually these can only be estimated. This estimate is often added to the random errors to produce a *total* experimental error, sometimes also called *total uncertainty*.

Ref. 258 The tables below give the values of the most important physical constants and particle properties in SI units and in a few other common units, as published in the standard references. The values are the world averages of the best measurements made up to the present. As usual, experimental errors, including both random and estimated systematic errors, are expressed by giving the standard deviation in the last digits; e.g. 0.31(6) means

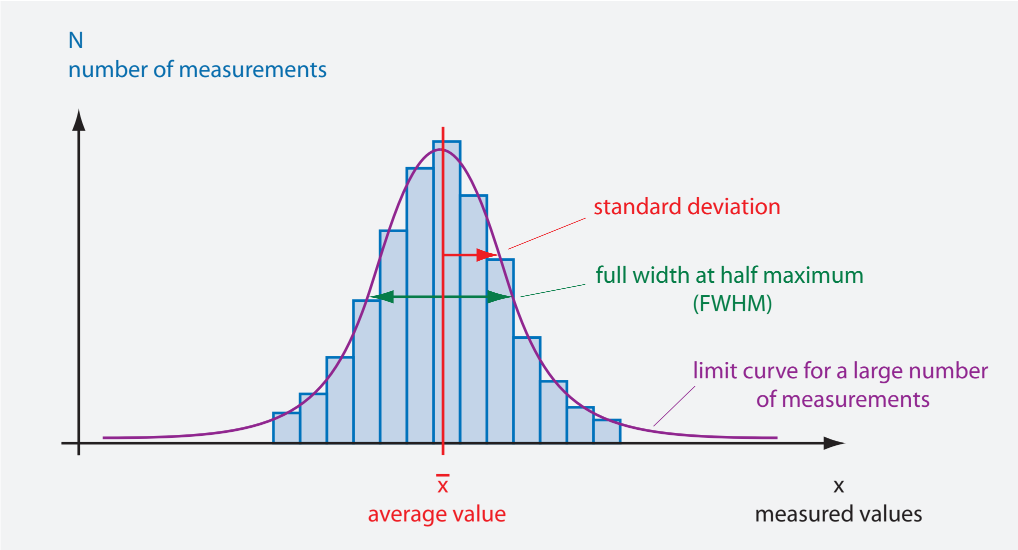


FIGURE 112 A precision experiment and its measurement distribution

– roughly speaking – 0.31 ± 0.06. In fact, behind each of the numbers in the following Ref. 259 tables there is a long story which is worth telling, but for which there is not enough room here.

Limits to precision

What are the limits to accuracy and precision? There is no way, even in principle, to measure a length x to a *precision* higher than about 61 digits, because the ratio between the largest and the smallest measurable length is $\Delta x/x > l_{Pl}/d_{horizon} = 10^{-61}$. (Is this Challenge 172 e ratio valid also for force or for volume?) In the final volume of our text, studies of clocks Vol. VI, page 87 and metre bars strengthen this theoretical limit.

But it is not difficult to deduce more stringent practical limits. No imaginable machine can measure quantities with a higher precision than measuring the diameter of the Earth within the smallest length ever measured, about 10^{-19} m; that is about 26 digits of precision. Using a more realistic limit of a 1000 m sized machine implies a limit of 22 digits. If, as predicted above, time measurements really achieve 17 digits of precision, then they are nearing the practical limit, because apart from size, there is an additional practical restriction: cost. Indeed, an additional digit in measurement precision often means an additional digit in equipment cost.

Physical constants

Ref. 258 In principle, all quantitative properties of matter can be calculated with quantum theory. For example, colour, density and elastic properties can be predicted using the values of the following constants using the equations of the standard model of high-energy Page 191 physics.

TABLE 27 Basic physical constants

Quantity	Symbol	Value in SI units	Uncert.[a]
number of space-time dimensions		3 + 1	0[b]
vacuum speed of light[c]	c	299 792 458 m/s	0
vacuum permeability[c]	μ_0	$4\pi \cdot 10^{-7}$ H/m	0
		= 1.256 637 061 435 ... μH/m	0
vacuum permittivity[c]	$\varepsilon_0 = 1/\mu_0 c^2$	8.854 187 817 620 ... pF/m	0
original Planck constant	h	$6.626\,068\,76(52) \cdot 10^{-34}$ Js	$7.8 \cdot 10^{-8}$
reduced Planck constant	$\hbar$	$1.054\,571\,596(82) \cdot 10^{-34}$ Js	$7.8 \cdot 10^{-8}$
positron charge	e	0.160 217 646 2(63) aC	$3.9 \cdot 10^{-8}$
Boltzmann constant	k	$1.380\,650\,3(24) \cdot 10^{-23}$ J/K	$1.7 \cdot 10^{-6}$
gravitational constant	G	$6.673(10) \cdot 10^{-11}$ Nm²/kg²	$1.5 \cdot 10^{-3}$
gravitational coupling constant	$\kappa = 8\pi G/c^4$	$2.076(3) \cdot 10^{-43}$ s²/kg m	$1.5 \cdot 10^{-3}$
fine structure constant,[d]	$\alpha = \frac{e^2}{4\pi\varepsilon_0 \hbar c}$	1/137.035 999 76(50)	$3.7 \cdot 10^{-9}$
e.m. coupling constant	$= \alpha_{em}(m_e^2 c^2)$	= 0.007 297 352 533(27)	$3.7 \cdot 10^{-9}$
Fermi coupling constant,[d]	$G_F/(\hbar c)^3$	$1.166\,39(1) \cdot 10^{-5}$ GeV^{-2}	$8.6 \cdot 10^{-6}$
weak coupling constant	$\alpha_w(M_Z) = g_w^2/4\pi$	1/30.1(3)	$1 \cdot 10^{-2}$
weak mixing angle	$\sin^2 \theta_W(\overline{MS})$	0.231 24(24)	$1.0 \cdot 10^{-3}$
weak mixing angle	$\sin^2 \theta_W$ (on shell)	0.2224(19)	$8.7 \cdot 10^{-3}$
	$= 1 - (m_W/m_Z)^2$		
strong coupling constant[d]	$\alpha_s(M_Z) = g_s^2/4\pi$	0.118(3)	$25 \cdot 10^{-3}$

a. Uncertainty: standard deviation of measurement errors.
b. Only down to 10^{-19} m and up to 10^{26} m.
c. Defining constant.
d. All coupling constants depend on the 4-momentum transfer, as explained in the section on renormalization. *Fine structure constant* is the traditional name for the electromagnetic coupling constant α in the case of a 4-momentum transfer of $Q^2 = m_e^2 c^2$, which is the smallest one possible. At higher momentum transfers it has larger values, e.g., $\alpha_{em}(Q^2 = M_W^2 c^2) \approx 1/128$. In contrast, the strong coupling constant has lover values at higher momentum transfers; e.g., $\alpha_s(34\,\text{GeV}) = 0.14(2)$.

Page 89

Why do all these constants have the values they have? For any constant *with a dimension*, such as the quantum of action $\hbar$, the numerical value has only historical meaning. It is $1.054 \cdot 10^{-34}$ Js because of the SI definition of the joule and the second. The question why the value of a dimensional constant is not larger or smaller therefore always requires one to understand the origin of some dimensionless number giving the ratio between the constant and the corresponding natural unit that is defined with c, G, $\hbar$ and α. Understanding the sizes of atoms, people, trees and stars, the duration of molecular and atomic processes, or the mass of nuclei and mountains, implies understanding the ratios between these values and the corresponding natural units. The key to understanding nature is thus the understanding of all ratios, and thus of all dimensionless constants. The quest of understanding all ratios, all dimensionless constants, including the fine structure constant α itself, is completed only in the final volume of our adventure.

The basic constants yield the following useful high-precision observations.

TABLE 28 Derived physical constants

Quantity	Symbol	Value in SI units	Uncert.
Vacuum wave resistance	$Z_0 = \sqrt{\mu_0/\varepsilon_0}$	376.730 313 461 77... Ω	0
Avogadro's number	N_A	$6.022\,141\,99(47) \cdot 10^{23}$	$7.9 \cdot 10^{-8}$
Rydberg constant [a]	$R_\infty = m_e c\alpha^2/2h$	$10\,973\,731.568\,549(83)\ \mathrm{m}^{-1}$	$7.6 \cdot 10^{-12}$
conductance quantum	$G_0 = 2e^2/h$	77.480 916 96(28) μS	$3.7 \cdot 10^{-9}$
magnetic flux quantum	$\varphi_0 = h/2e$	2.067 833 636(81) pWb	$3.9 \cdot 10^{-8}$
Josephson frequency ratio	$2e/h$	483.597 898(19) THz/V	$3.9 \cdot 10^{-8}$
von Klitzing constant	$h/e^2 = \mu_0 c/2\alpha$	25 812.807 572(95) Ω	$3.7 \cdot 10^{-9}$
Bohr magneton	$\mu_B = e\hbar/2m_e$	9.274 008 99(37) yJ/T	$4.0 \cdot 10^{-8}$
cyclotron frequency of the electron	$f_c/B = e/2\pi m_e$	27.992 4925(11) GHz/T	$4.0 \cdot 10^{-8}$
classical electron radius	$r_e = e^2/4\pi\varepsilon_0 m_e c^2$	2.817 940 285(31) fm	$1.1 \cdot 10^{-8}$
Compton wavelength	$\lambda_c = h/m_e c$	2.426 310 215(18) pm	$7.3 \cdot 10^{-9}$
of the electron	$\lambda\!\!\!^{-}_c = \hbar/m_e c = r_e/\alpha$	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	$3.7 \cdot 10^{-9}$
nuclear magneton	$\mu_N = e\hbar/2m_p$	$5.050\,783\,17(20) \cdot 10^{-27}$ J/T	$4.0 \cdot 10^{-8}$
proton–electron mass ratio	m_p/m_e	1 836.152 667 5(39)	$2.1 \cdot 10^{-9}$
Stefan–Boltzmann constant	$\sigma = \pi^2 k^4/60\hbar^3 c^2$	$56.704\,00(40)\ \mathrm{nW/m^2K^4}$	$7.0 \cdot 10^{-6}$
Wien's displacement constant	$b = \lambda_{max} T$	2.897 768 6(51) mmK	$1.7 \cdot 10^{-6}$
bits to entropy conversion const.		10^{23} bit = 0.956 994 5(17) J/K	$1.7 \cdot 10^{-6}$
TNT energy content		3.7 to 4.0 MJ/kg	$4 \cdot 10^{-2}$

a. For infinite mass of the nucleus.

Some useful properties of our local environment are given in the following table.

TABLE 29 Astronomical constants

Quantity	Symbol	Value
tropical year 1900 [a]	a	31 556 925.974 7 s
tropical year 1994	a	31 556 925.2 s
mean sidereal day	d	$23^h 56' 4.090\,53''$
astronomical unit [b]	AU	149 597 870.691(30) km
light year	al	9.460 528 173 ... Pm
parsec	pc	30.856 775 806 Pm = 3.261 634 al
Earth's mass	$M_♁$	$5.973(1) \cdot 10^{24}$ kg
Geocentric gravitational constant	GM	$3.986\,004\,418(8) \cdot 10^{14}\ \mathrm{m^3/s^2}$
Earth's gravitational length	$l_♁ = 2GM/c^2$	8.870 056 078(16) mm
Earth's equatorial radius [c]	$R_{♁eq}$	6378.1366(1) km
Earth's polar radius [c]	$R_{♁p}$	6356.752(1) km

TABLE 29 (Continued) Astronomical constants

Quantity	Symbol	Value
Equator–pole distance [c]		10 001.966 km (average)
Earth's flattening [c]	$e_♁$	1/298.25642(1)
Earth's av. density	$\rho_♁$	5.5 Mg/m^3
Earth's age	$T_♁$	4.50(4) Ga = 142(2) Ps
Moon's radius	$R_{☾v}$	1738 km in direction of Earth
Moon's radius	$R_{☾h}$	1737.4 km in other two directions
Moon's mass	$M_☾$	$7.35 \cdot 10^{22}$ kg
Moon's mean distance [d]	$d_☾$	384 401 km
Moon's distance at perigee [d]		typically 363 Mm, historical minimum 359 861 km
Moon's distance at apogee [d]		typically 404 Mm, historical maximum 406 720 km
Moon's angular size [e]		average 0.5181° = 31.08′, minimum 0.49°, maximum - shortens line 0.55°
Moon's average density	$\rho_☾$	3.3 Mg/m^3
Jupiter's mass	$M_♃$	$1.90 \cdot 10^{27}$ kg
Jupiter's radius, equatorial	$R_♃$	71.398 Mm
Jupiter's radius, polar	$R_♃$	67.1(1) Mm
Jupiter's average distance from Sun	$D_♃$	778 412 020 km
Sun's mass	$M_\odot$	$1.988\,43(3) \cdot 10^{30}$ kg
Sun's gravitational length	$l_\odot = 2GM_\odot/c^2$	2.953 250 08 km
Sun's luminosity	$L_\odot$	384.6 YW
Solar equatorial radius	$R_\odot$	695.98(7) Mm
Sun's angular size		0.53° average; minimum on fourth of July (aphelion) 1888″, maximum on fourth of January (perihelion) 1952″
Sun's average density	$\rho_\odot$	1.4 Mg/m^3
Sun's average distance	AU	149 597 870.691(30) km
Sun's age	$T_\odot$	4.6 Ga
Solar velocity around centre of galaxy	$v_{\odot g}$	220(20) km/s
Solar velocity against cosmic background	$v_{\odot b}$	370.6(5) km/s
Distance to Milky Way's centre		8.0(5) kpc = 26.1(1.6) kal
Milky Way's age		13.6 Ga
Milky Way's size		*c.* 10^{21} m or 100 kal
Milky Way's mass		10^{12} solar masses, *c.* $2 \cdot 10^{42}$ kg
Most distant galaxy cluster known	SXDF-XCLJ 0218-0510	$9.6 \cdot 10^9$ al

a. Defining constant, from vernal equinox to vernal equinox; it was once used to define the second. (Remember: π seconds is about a nanocentury.) The value for 1990 is about 0.7 s less, corresponding to a slowdown of roughly 0.2 ms/a. (Watch out: why?) There is even an empirical formula for the change of the length of the year over time.
Challenge 173 s
Ref. 260
b. Average distance Earth–Sun. The truly amazing precision of 30 m results from time averages of signals sent from Viking orbiters and Mars landers taken over a period of over twenty years.
c. The shape of the Earth is described most precisely with the World Geodetic System. The last edition dates from 1984. For an extensive presentation of its background and its details, see the www.wgs84.com website. The International Geodesic Union refined the data in 2000. The radii and the flattening given here are those for the 'mean tide system'. They differ from those of the 'zero tide system' and other systems by about 0.7 m. The details constitute a science in itself.
d. Measured centre to centre. To find the precise position of the Moon at a given date, see the www.fourmilab.ch/earthview/moon_ap_per.html page. For the planets, see the page www.fourmilab.ch/solar/solar.html and the other pages on the same site.
e. Angles are defined as follows: 1 degree = $1^\circ = \pi/180$ rad, 1 (first) minute = $1' = 1^\circ/60$, 1 second (minute) = $1'' = 1'/60$. The ancient units 'third minute' and 'fourth minute', each 1/60th of the preceding, are not in use any more. ('Minute' originally means 'very small', as it still does in modern English.)

Some properties of nature at large are listed in the following table. (If you want a challenge, can you determine whether any property of the universe itself is listed?)
Challenge 174 s

TABLE 30 Astrophysical constants

Quantity	Symbol	Value
gravitational constant	G	$6.672\,59(85) \cdot 10^{-11}\ \mathrm{m^3/kg\,s^2}$
cosmological constant	Λ	*c.* $1 \cdot 10^{-52}\ \mathrm{m^{-2}}$
age of the universe [a]	t_0	$4.333(53) \cdot 10^{17}\ \mathrm{s} = 13.73(0.17) \cdot 10^9\ \mathrm{a}$
(determined from space-time, via expansion, using general relativity)		
age of the universe [a]	t_0	over $3.5(4) \cdot 10^{17}\ \mathrm{s} = 11.5(1.5) \cdot 10^9\ \mathrm{a}$
(determined from matter, via galaxies and stars, using quantum theory)		
Hubble parameter [a]	H_0	$2.3(2) \cdot 10^{-18}\ \mathrm{s^{-1}} = 0.73(4) \cdot 10^{-10}\ \mathrm{a^{-1}}$
		$= h_0 \cdot 100\ \mathrm{km/s\,Mpc} = h_0 \cdot 1.0227 \cdot 10^{-10}\ \mathrm{a^{-1}}$
reduced Hubble parameter [a]	h_0	$0.71(4)$
deceleration parameter	$q_0 = -(\ddot{a}/a)_0/H_0^2$	$-0.66(10)$
universe's horizon distance [a]	$d_0 = 3ct_0$	$40.0(6) \cdot 10^{26}\ \mathrm{m} = 13.0(2)\ \mathrm{Gpc}$
universe's topology		trivial up to $10^{26}\ \mathrm{m}$
number of space dimensions		3, for distances up to $10^{26}\ \mathrm{m}$
critical density of the universe	$\rho_c = 3H_0^2/8\pi G$	$h_0^2 \cdot 1.878\,82(24) \cdot 10^{-26}\ \mathrm{kg/m^3}$
		$= 0.95(12) \cdot 10^{-26}\ \mathrm{kg/m^3}$
(total) density parameter [a]	$\Omega_0 = \rho_0/\rho_c$	$1.02(2)$
baryon density parameter [a]	$\Omega_{B0} = \rho_{B0}/\rho_c$	$0.044(4)$
cold dark matter density parameter [a]	$\Omega_{CDM0} = \rho_{CDM0}/\rho_c$	$0.23(4)$
neutrino density parameter [a]	$\Omega_{\nu 0} = \rho_{\nu 0}/\rho_c$	0.001 to 0.05
dark energy density parameter [a]	$\Omega_{X0} = \rho_{X0}/\rho_c$	$0.73(4)$
dark energy state parameter	$w = p_X/\rho_X$	$-1.0(2)$
baryon mass	m_b	$1.67 \cdot 10^{-27}\ \mathrm{kg}$

TABLE 30 (Continued) Astrophysical constants

Quantity	Symbol	Value
baryon number density		$0.25(1)\,/\mathrm{m}^3$
luminous matter density		$3.8(2) \cdot 10^{-28}\,\mathrm{kg/m}^3$
stars in the universe	n_s	$10^{22\pm1}$
baryons in the universe	n_b	$10^{81\pm1}$
microwave background temperature [b]	T_0	$2.725(1)\,\mathrm{K}$
photons in the universe	n_γ	10^{89}
photon energy density	$\rho_\gamma = \pi^2 k^4/15T_0^4$	$4.6 \cdot 10^{-31}\,\mathrm{kg/m}^3$
photon number density		$410.89\,/\mathrm{cm}^3$ or $400\,/\mathrm{cm}^3 (T_0/2.7\,\mathrm{K})^3$
density perturbation amplitude	$\sqrt{S}$	$5.6(1.5) \cdot 10^{-6}$
gravity wave amplitude	$\sqrt{T}$	$< 0.71\sqrt{S}$
mass fluctuations on 8 Mpc	σ_8	$0.84(4)$
scalar index	n	$0.93(3)$
running of scalar index	$\mathrm{d}n/\mathrm{d}\ln k$	-0.03(2)
Planck length	$l_{Pl} = \sqrt{\hbar G/c^3}$	$1.62 \cdot 10^{-35}\,\mathrm{m}$
Planck time	$t_{Pl} = \sqrt{\hbar G/c^5}$	$5.39 \cdot 10^{-44}\,\mathrm{s}$
Planck mass	$m_{Pl} = \sqrt{\hbar c/G}$	$21.8\,\mu\mathrm{g}$
instants in history [a]	t_0/t_{Pl}	$8.7(2.8) \cdot 10^{60}$
space-time points inside the horizon [a]	$N_0 = (R_0/l_{Pl})^3 \cdot (t_0/t_{Pl})$	$10^{244\pm1}$
mass inside horizon	M	$10^{54\pm1}\,\mathrm{kg}$

a. The index 0 indicates present-day values.
b. The radiation originated when the universe was 380 000 years old and had a temperature of about 3000 K; the fluctuations ΔT_0 which led to galaxy formation are today about $16 \pm 4\,\mu\mathrm{K} = 6(2) \cdot 10^{-6}\,T_0$.

Vol. II, page 206

Useful numbers

Ref. 261

π	$3.14159\,26535\,89793\,23846\,26433\,83279\,50288\,41971\,69399\,37510_5$
e	$2.71828\,18284\,59045\,23536\,02874\,71352\,66249\,77572\,47093\,69995_9$
γ	$0.57721\,56649\,01532\,86060\,65120\,90082\,40243\,10421\,59335\,93992_3$
ln 2	$0.69314\,71805\,59945\,30941\,72321\,21458\,17656\,80755\,00134\,36025_5$
ln 10	$2.30258\,50929\,94045\,68401\,79914\,54684\,36420\,76011\,01488\,62877_2$
$\sqrt{10}$	$3.16227\,76601\,68379\,33199\,88935\,44432\,71853\,37195\,55139\,32521_6$

Appendix B

COMPOSITE PARTICLE PROPERTIES

Page 191
Page 354

The following table lists the most important composite particles. The list has not changed much recently, mainly because of the vast progress that was achieved in the middle of the twentieth century. In principle, using the standard model of particle physics, together with the fundamental constants, *all* properties of composite matter and radiation can be deduced. In particular, all properties of objects encountered in everyday life follow. (Can you explain how the size of an apple follows from the standard model?) The most important examples of composites are grouped in the following table.

Challenge 175 s

TABLE 31 Properties of selected composites

Composite	Mass m, quantum numbers	Lifetime τ, main decay modes	Size (diam.)
mesons (hadrons, bosons) (selected from over 130 known types)			
pion $\pi^0(u\bar{u}-d\bar{d})/\sqrt{2}$	134.976 4(6) MeV/c^2	84(6) as, 2γ 98.798(32)%	∼ 1 fm
	$I^G(J^{PC}) = 1^-(0^{-+})$, $S = C = B = 0$		
pion $\pi^+(u\bar{d})$	139.569 95(35) MeV/c^2	26.030(5) ns,	∼ 1 fm
		$\mu^+\nu_\mu$ 99.987 7(4)%	
	$I^G(J^P) = 1^-(0^-)$, $S = C = B = 0$		
kaon K^0_S	$m_{K^0_S}$	89.27(9) ps	∼ 1 fm
kaon K^0_L	$m_{K^0_S}$ + 3.491(9) μeV/c^2	51.7(4) ns	∼ 1 fm
kaon $K^\pm$ $(u\bar{s}, \bar{u}s)$	493.677(16) MeV/c^2	12.386(24) ns,	∼ 1 fm
		$\mu^+\nu_\mu$ 63.51(18)%	
		$\pi^+\pi^0$ 21.16(14)%	
kaon K^0 $(d\bar{s})$ (50 % K_S, 50 % K_L)	497.672(31) MeV/c^2	n.a.	∼ 1 fm
all kaons $K^\pm, K^0, K^0_S, K^0_L$	$I(J^P) = \frac{1}{2}(0^-)$, $S = \pm1$, $B = C = 0$		
baryons (hadrons, fermions) (selected from over 100 known types)			
proton p or N^+ (uud)	1.672 621 58(13) yg	$\tau_{total} > 1.6 \cdot 10^{25}$ a,	0.89(1) fm
	= 1.007 276 466 88(13) u	$\tau(p \rightarrow e^+\pi^0) > 5.5 \cdot 10^{32}$ a	Ref. 262
	= 938.271 998(38) MeV/c^2		
	$I(J^P) = \frac{1}{2}(\frac{1}{2}^+)$, $S = 0$		
	gyromagnetic ratio μ_p/μ_N = 2.792 847 337(29)		

Composite	Mass m, quantum numbers	Lifetime τ, main decay modes	Size (diam.)
	electric dipole moment $d = (-4 \pm 6) \cdot 10^{-26} e$ m		
	electric polarizability $\alpha_e = 12.1(0.9) \cdot 10^{-4}$ fm^3		
	magnetic polarizability $\alpha_m = 2.1(0.9) \cdot 10^{-4}$ fm^3		
neutron n or N^0 (udd)	1.674 927 16(13) yg	887.0(2.0) s, $pe^- \bar{\nu}_e$ 100 %	~ 1 fm
	$= 1.008\,664\,915\,78(55)$ u $= 939.565\,330(38)$ MeV$/c^2$		
	$I(J^P) = \frac{1}{2}(\frac{1}{2}^+)$, $S = 0$		
	gyromagnetic ratio $\mu_n/\mu_N = -1.913\,042\,72(45)$		
	electric dipole moment $d_n = (-3.3 \pm 4.3) \cdot 10^{-28} e$ m		
	electric polarizability $\alpha = 0.98(23) \cdot 10^{-3}$ fm^3		
omega Ω^- (sss)	1672.43(32) MeV$/c^2$	82.2(1.2) ps, ΛK^- 67.8(7)%, $\Xi^0 \pi^-$ 23.6(7)%	~ 1 fm
	gyromagnetic ratio $\mu_\Omega/\mu_N = -1.94(22)$		
composite radiation: glueballs			
glueball $f_0(1500)$	1503(11) MeV, $I^G(J^{PC}) = 0^+(0^{++})$	full width 120(19) MeV	~ 1 fm
atoms (selected from 114 known elements with over 2000 known nuclides) Ref. 263			
hydrogen (^{1}H) [lightest]	1.007 825 032(1) u = 1.6735 yg		2 · 53 pm
antihydrogen	1.007 u = 1.67 yg		2 · 53 pm
helium (^{4}He) [smallest]	4.002 603250(1) u = 6.6465 yg		2 · 31 pm
carbon (^{12}C)	12 u = 19.926 482(12) yg		2 · 77 pm
bismuth (^{209}Bi*) [shortest living and rarest]	209 u	0.1 ps Ref. 264	
tantalum (^{180m}Ta) [second longest living radioactive]	180 u	$> 10^{15}$ a Ref. 265	
bismuth (^{209}Bi) [longest living radioactive]	209 u	$1.9(2)10^{19}$ a Ref. 264	
francium (^{223}Fr) [largest]	223 u	22 min	2 · 0.28 nm
atom 116 (^{289}Uuh) [heaviest]	289 u	0.6 ms	
molecules (selected from over 10^7 known types)			
hydrogen (H_2)	~ 2 u	$> 10^{25}$ a	
water (H_2O)	~ 18 u	$> 10^{25}$ a	
ATP (adenosinetriphosphate)	507 u	$> 10^{10}$ a	*c.* 3 nm
human Y chromosome	$70 \cdot 10^6$ base pairs	$> 10^6$ a	*c.* 50 mm (uncoiled)
other composites			
blue whale nerve cell	~ 1 kg	~ 50 a	20 m

Composite	Mass m, quantum numbers	Lifetime τ, main decay modes	Size (diam.)
cell (red blood)	0.1 ng	7 plus 120 days	~ 10 μm
cell (sperm)	10 pg	not fecundated: ~ 5 d	length 60 μm, head 3 μm × 5 μm
cell (ovule)	1 μg	fecundated: over 4000 million years	~ 120 μm
cell (*E. coli*)	1 pg	4000 million years	body: 2 μm
adult human	35 kg < m < 350 kg	$\tau \approx 2.5 \cdot 10^9$ s Ref. 266 ≈ 600 million breaths ≈ 2 500 million heartbeats < 122 a, 60 % H_2O and 40 % dust	~ 1.7 m
heaviest living thing: colony of aspen trees	$6.6 \cdot 10^6$ kg	> 130 a	> 4 km
larger composites	See the table on page 209.		

Page 193 Notes (see also those of the previous table):

- G-parity is defined only for mesons and given by $G = (-1)^{L+S+I} = (-1)^I C$.
- Neutrons bound in nuclei have a lifetime of at least 10^{20} years.
- The $f_0(1500)$ resonance is a candidate for the glueball ground state and thus for a *radiation composite*.
- The $Y(3940)$ resonance is a candidate for a hybrid meson, a composite of a gluon and a quark–antiquark pair. This prediction of 1980 seems to have been confirmed in 2005.
- Ref. 267 In 2002, the first evidence for the existence of tetra-neutrons was published by a French group. However, more recent investigations seem to have refuted the claim.
- The number of existing molecules is several orders of magnitude larger than the number of molecules that have been analysed and named.
- Some nuclei have not yet been observed; in 2006 the known nuclei ranged from 1 to 116, but 113 and 115 were still missing.
- The first *anti-atoms*, made of antielectrons and antiprotons, were made in January 1996 at CERN in Geneva. All properties of antimatter checked so far are consistent with theoretical predictions. Ref. 268
- The charge parity C is defined only for certain neutral particles, namely those that 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 $\boldsymbol{r} \to -\boldsymbol{r}$. For mesons, it is related to the orbital angular momentum L through $P = (-1)^{L+1}$.
- The electric polarizability, defined on page 59, is predicted to vanish for all elementary particles.

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 usefully set out in the so-called *periodic table*, which groups together atoms with similar properties in rows and columns. It is given in Table 32 and results from the various ways in which protons, neutrons and electrons can combine to form aggregates.

Comparable to the periodic table of the atoms, there are tables for the mesons (made of two quarks) and the baryons (made of three quarks). Neither the meson nor the baryon table is included here; they can both be found in the *Review of Particle Physics* at pdg.web.cern.ch. 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.

TABLE 32 The periodic table of the elements known in 2006, with their atomic numbers

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 Ds	111 Rg	112 Uub	113 Uut	114 Uuq	115 Uup	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

The atomic number gives the number of protons (and electrons) found in an atom of a given element. This number determines the chemical behaviour of an element. Most – but not all – elements up to 92 are found on Earth; the others can be produced in lab-

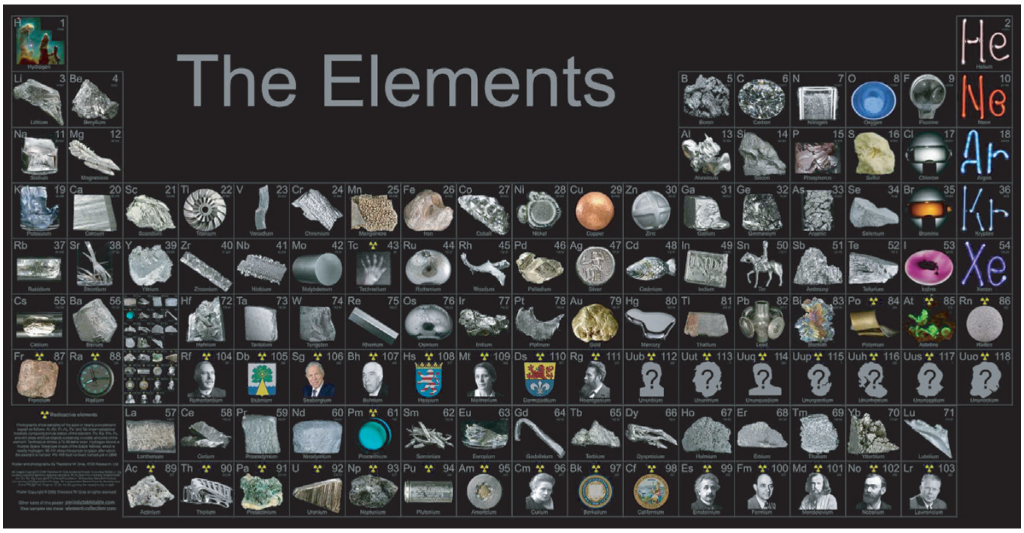


FIGURE 113 A modern table of the elements (© Theodore Gray, for sale at www.theodoregray.com)

oratories. The highest element discovered is element 116. (In a famous case of research fraud, a scientist in the 1990s tricked two whole research groups into claiming to have made and observed elements 116 and 118. Element 116 was independently made and observed by another group later on.) Nowadays, extensive physical and chemical data are
Ref. 269 available for every element. Photographs of the pure elements are shown in Figure 113.

Elements in the same *group* behave similarly in chemical reactions. The *periods* define the repetition of these similarities. More elaborate periodic tables can be found on the chemlab.pc.maricopa.edu/periodic website. The most beautiful of them all can be found on page 47 of this text.

Group 1 are the alkali metals (though hydrogen is a gas), group 2 the Earth-alkali metals. Actinoids, lanthanoids and groups 3 to 13 are metals; in particular, groups 3 to 12 are transition or 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 inert *noble gases*, which form (almost) no chemical compounds. The groups 13, 14 and 15 contain metals, semimetals, a liquid and gases; they have no special name. Groups 1 and 13 to 17 are central for the chemistry of life; in fact, 96 % of living matter is made of C, O, N, H;* almost 4 % of P, S, Ca, K, Na, Cl; trace elements such as Mg, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, Pb, Sn, Li, Mo, Se, Si, I, F, As, B form the rest. Over 30 elements are known to be essential for animal life. The full list is not yet known; candidate elements to extend this list are Al, Br, Ge and W.

Many elements exist in versions with different numbers of neutrons in their nucleus, and thus with different mass; these various *isotopes* – so called because they are found at the *same place* in the periodic table – behave identically in chemical reactions. There are
Ref. 263, Ref. 270 over 2000 of them.

* The ‘average formula’ of life is approximately $C_5H_{40}O_{18}N$.

TABLE 33 The elements, with their atomic number, average mass, atomic radius and main properties

Name	Symbol	At. n.	Aver. mass[a] in u (error), longest lifetime	Atomic[e] radius in pm	Main properties, (naming)[h] discovery date and use
Actinium[b]	Ac	89	(227.0277(1)) 21.77(2) a	(188)	highly radioactive metallic rare Earth (Greek *aktis* ray) 1899, used as alpha-emitting source
Aluminium	Al	13	26.981 538 (8) stable	118c, 143m	light metal (Latin *alumen* alum) 1827, used in machine construction and living beings
Americium[b]	Am	95	(243.0614(1)) 7.37(2) ka	(184)	radioactive metal (Italian *America* from Amerigo) 1945, used in smoke detectors
Antimony	Sb	51	121.760(1)[f] stable	137c, 159m, 205v	toxic semimetal (via Arabic from Latin *stibium*, itself from Greek, Egyptian for one of its minerals) antiquity, 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 and 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	Earth-alkali metal (Greek *bary* heavy) 1808, used in vacuum tubes, paint, oil industry, pyrotechnics and X-ray diagnosis
Berkelium[b]	Bk	97	(247.0703(1)) 1.4(3) ka	n.a.	made in lab, probably metallic (Berkeley, US town) 1949, no use because rare
Beryllium	Be	4	9.012 182(3) stable	106c, 113m	toxic Earth-alkali metal (Greek *beryllos*, a mineral) 1797, used in light alloys, in nuclear industry as moderator
Bismuth	Bi	83	208.980 40(1) stable	170m, 215v	diamagnetic metal (Latin via German *weisse Masse* white mass) 1753, used in magnets, alloys, fire safety, cosmetics, as catalyst, nuclear industry
Bohrium[b]	Bh	107	(264.12(1)) 0.44 s[g]	n.a.	made in lab, probably metallic (after Niels Bohr) 1981, found in nuclear reactions, no use
Boron	B	5	10.811(7)[f] stable	83c	semimetal, semiconductor (Latin *borax*, from Arabic and Persian for brilliant) 1808, used in glass, bleach, pyrotechnics, rocket fuel, medicine

Name	Symbol	At. n.	Aver. mass[a] in u (error), longest lifetime	Atomic[e] radius in pm	Main properties, (naming)[h] discovery date and use
Bromine	Br	35	79.904(1) stable	120c, 185v	red-brown liquid (Greek *bromos* strong odour) 1826, fumigants, photography, water purification, dyes, medicines
Cadmium	Cd	48	112.411(8)[f] stable	157m	heavy metal, cuttable and screaming (Greek *kadmeia*, a zinc carbonate mineral where it was discovered) 1817, electroplating, solder, batteries, TV phosphors, dyes
Caesium	Cs	55	132.905 4519(2) stable	273m	alkali metal (Latin *caesius* sky blue) 1860, getter in vacuum tubes, photoelectric cells, ion propulsion, atomic clocks
Calcium	Ca	20	40.078(4)[f] stable	197m	Earth-alkali metal (Latin *calcis* chalk) antiquity, pure in 1880, found in stones and bones, reducing agent, alloying
Californium[b]	Cf	98	(251.0796(1)) 0.90(5) ka	n.a.	made in lab, probably metallic, strong neutron emitter (Latin *calor* heat and *fornicare* have sex, the land of hot sex :-) 1950, used as neutron source, for well logging
Carbon	C	6	12.0107(8)[f] stable	77c	makes up coal and diamond (Latin *carbo* coal) antiquity, used to build most life forms
Cerium	Ce	58	140.116(1)[f] stable	183m	rare Earth metal (after asteroid Ceres, Roman goddess) 1803, cigarette lighters, incandescent gas mantles, glass manufacturing, self-cleaning ovens, carbon-arc lighting in the motion picture industry, catalyst, metallurgy
Chlorine	Cl	17	35.453(2)[f] stable	102c, 175v	green gas (Greek *chloros* yellow-green) 1774, drinking water, polymers, paper, dyes, textiles, medicines, insecticides, solvents, paints, rubber
Chromium	Cr	24	51.9961(6) stable	128m	transition metal (Greek *chromos* colour) 1797, hardens steel, makes steel stainless, alloys, electroplating, green glass dye, catalyst
Cobalt	Co	27	58.933 195(5) stable	125m	ferromagnetic transition metal (German *Kobold* goblin) 1694, part of vitamin B_{12}, magnetic alloys, heavy-duty alloys, enamel dyes, ink, animal nutrition

Name	Sym-bol	At. n.	Aver. mass[a] in u (error), longest lifetime	Atomic[e] radius in pm	Main properties, (naming)[h] discovery date and use
Copper	Cu	29	63.546(3)[f] stable	128m	red metal (Latin *cuprum* from Cyprus island) antiquity, part of many enzymes, electrical conductors, bronze, brass and other alloys, algicides, etc.
Curium[b]	Cm	96	(247.0704(1)) 15.6(5) Ma	n.a.	highly radioactive, silver-coloured (after Pierre and Marie Curie) 1944, used as radioactivity source
Darmstadtium[b]	Ds	110	(271) 1.6 min[g]	n.a.	(after the German city) 1994, no use
Dubnium[b]	Db	105	(262.1141(1)) 34(5) s	n.a.	made in lab in small quantities, radioactive (Dubna, Russian city) 1967, no use (once known as hahnium)
Dysprosium	Dy	66	162.500(1)[f] stable	177m	rare Earth metal (Greek *dysprositos* difficult to obtain) 1886, used in laser materials, as infrared source material, and in nuclear industry
Einsteinium[b]	Es	99	(252.0830(1)) 472(2) d	n.a.	made in lab, radioactive (after Albert Einstein) 1952, no use
Erbium	Er	68	167.259(3)[f] stable	176m	rare Earth metal (Ytterby, Swedish town) 1843, used in metallurgy and optical fibres
Europium	Eu	63	151.964(1)[f] stable	204m	rare Earth metal (named after the continent) 1901, used in red screen phosphor for TV tubes
Fermium[b]	Fm	100	(257.0901(1)) 100.5(2) d	n.a.	(after Enrico Fermi) 1952, no use
Fluorine	F	9	18.998 4032(5) stable	62c, 147v	gaseous halogen (from fluorine, a mineral, from Greek *fluo* flow) 1886, used in polymers and toothpaste
Francium[b]	Fr	87	(223.0197(1)) 22.0(1) min	(278)	radioactive metal (from France) 1939, no use
Gadolinium	Gd	64	157.25(3)[f] stable	180m	(after Johan Gadolin) 1880, used in lasers and phosphors
Gallium	Ga	31	69.723(1) stable	125c, 141m	almost liquid metal (Latin for both the discoverer's name and his nation, France) 1875, used in optoelectronics
Germanium	Ge	32	72.64(1) stable	122c, 195v	semiconductor (from Germania, as opposed to gallium) 1886, used in electronics
Gold	Au	79	196.966 569(4) stable	144m	heavy noble metal (Sanskrit *jval* to shine, Latin aurum) antiquity, electronics, jewels

Name	Sym-bol	At. n.	Aver. mass[a] in u (error), longest lifetime	Atomic[e] radius in pm	Main properties, (naming)[h] discovery date and use
Hafnium	Hf	72	178.49(2)[c] stable	158m	metal (Latin for Copenhagen) 1923, alloys, incandescent wire
Hassium[b]	Hs	108	(277) 16.5 min[g]	n.a.	radioactive element (Latin form of German state Hessen) 1984, no use
Helium	He	2	4.002 602(2)[f] stable	(31n)	noble gas (Greek *helios* Sun) where it was discovered 1895, used in balloons, stars, diver's gas and cryogenics
Holmium	Ho	67	164.930 32(2) stable	177m	metal (Stockholm, Swedish capital) 1878, alloys
Hydrogen	H	1	1.007 94(7)[f] stable	30c	reactive gas (Greek for water-former) 1766, used in building stars and universe
Indium	In	49	114.818(3) stable	141c, 166m	soft metal (Greek *indikon* indigo) 1863, used in solders and photocells
Iodine	I	53	126.904 47(3) stable	140c, 198v	blue-black solid (Greek *iodes* violet) 1811, used in photography
Iridium	Ir	77	192.217(3) stable	136m	precious metal (Greek *iris* rainbow) 1804, electrical contact layers
Iron	Fe	26	55.845(2) stable	127m	metal (Indo-European *ayos* metal, Latin ferrum) antiquity, used in metallurgy
Krypton	Kr	36	83.798(2)[f] stable	(88n)	noble gas (Greek *kryptos* hidden) 1898, used in lasers
Lanthanum	La	57	138.905 47(7)[c,f] stable	188m	reactive rare Earth metal (Greek *lanthanein* to be hidden) 1839, used in lamps and in special glasses
Lawrencium[b]	Lr	103	(262.110 97(1)) 3.6(3) h	n.a.	appears in reactions (after Ernest Lawrence) 1961, no use
Lead	Pb	82	207.2(1)[c,f] stable	175m	poisonous, malleable heavy metal (Latin *plumbum*) antiquity, used in car batteries, radioactivity shields, paints
Lithium	Li	3	6.941(2)[f] stable	156m	light alkali metal with high specific heat (Greek *lithos* stone) 1817, used in batteries, anti-depressants, alloys and many chemicals
Lutetium	Lu	71	174.967(1)[f] stable	173m	rare Earth metal (Latin *Lutetia* for Paris) 1907, used as catalyst
Magnesium	Mg	12	24.3050(6) stable	160m	light common alkaline Earth metal (from Magnesia, a Greek district in Thessalia) 1755, used in alloys, pyrotechnics, chemical synthesis and medicine, found in chlorophyll

Name	Symbol	At. n.	Aver. mass[a] in u (error), longest lifetime	Atomic[e] radius in pm	Main properties, (naming)[h] discovery date and use
Manganese	Mn	25	54.938 045(5) stable	126m	brittle metal (Italian *manganese*, a mineral) 1774, used in alloys, colours amethyst and permanganate
Meitnerium[b]	Mt	109	(268.1388(1)) 0.070 s[g]	n.a.	appears in nuclear reactions (after Lise Meitner) 1982, no use
Mendelevium[b]	Md	101	(258.0984(1)) 51.5(3) d	n.a.	appears in nuclear reactions (after Дмитрии Иванович Менделеев Dmitriy Ivanovich Mendeleyev) 1955, no use
Mercury	Hg	80	200.59(2) stable	157m	liquid heavy metal (Latin god Mercurius, Greek *hydrargyrum* liquid silver) antiquity, used in switches, batteries, lamps, amalgam alloys
Molybdenum	Mo	42	95.94(2)[f] stable	140m	metal (Greek *molybdos* lead) 1788, used in alloys, as catalyst, in enzymes and lubricants
Neodymium	Nd	60	144.242(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, used in lamps, lasers and cryogenics
Neptunium[b]	Np	93	(237.0482(1)) 2.14(1) Ma	n.a.	radioactive metal (planet Neptune, after Uranus in the solar system) 1940, appears in nuclear reactors, used in neutron detection and by the military
Nickel	Ni	28	58.6934(2) stable	125m	metal (German *Nickel* goblin) 1751, used in coins, stainless steels, batteries, as catalyst
Niobium	Nb	41	92.906 38(2) stable	147m	ductile metal (Greek Niobe, mythical daughter of Tantalos) 1801, used in arc welding, alloys, jewellery, superconductors
Nitrogen	N	7	14.0067(2)[f] stable	70c, 155v	diatomic gas (Greek for nitre-former) 1772, found in air, in living organisms, Viagra, fertilizers, explosives
Nobelium[b]	No	102	(259.1010(1)) 58(5) min	n.a.	(after Alfred Nobel) 1958, no use
Osmium	Os	76	190.23(3)[f] stable	135m	heavy metal (from Greek *osme* odour) 1804, used for fingerprint detection and in very hard alloys

Name	Symbol	At. n.	Aver. mass[a] in u (error), longest lifetime	Atomic[e] radius in pm	Main properties, (naming)[h] discovery date and use
Oxygen	O	8	15.9994(3)[f] stable	66c, 152v	transparent, diatomic gas (formed from Greek to mean 'acid former') 1774, used for combustion, blood regeneration, to make most rocks and stones, in countless compounds, colours auroras red
Palladium	Pd	46	106.42(1)[f] stable	138m	heavy metal (from asteroid Pallas, after the Greek goddess) 1802, used in alloys, white gold, catalysts, for hydride storage
Phosphorus	P	15	30.973 762(2) stable	109c, 180v	poisonous, waxy, white solid (Greek *phosphoros* light bearer) 1669, fertilizers, glasses, porcelain, steels and alloys, living organisms, bones
Platinum	Pt	78	195.084(9) stable	139m	silvery-white, ductile, noble heavy metal (Spanish *platina* little silver) pre-Columbian, again in 1735, used in corrosion-resistant alloys, magnets, furnaces, catalysts, fuel cells, cathodic protection systems for large ships and pipelines; being a catalyst, a fine platinum wire glows red hot when placed in vapour of methyl alcohol, an effect used in hand warmers
Plutonium	Pu	94	(244.0642(1)) 80.0(9) Ma	n.a.	extremely toxic alpha-emitting metal (after the planet) synthesized 1940, found in nature 1971, used as nuclear explosive, and to power space equipment, such as satellites and the measurement equipment brought to the Moon by the Apollo missions
Polonium	Po	84	(208.9824(1)) 102(5) a	(140)	alpha-emitting, volatile metal (from Poland) 1898, used as thermoelectric power source in space satellites, as neutron source when mixed with beryllium; used in the past to eliminate static charges in factories, and on brushes for removing dust from photographic films
Potassium	K	19	39.0983(1) stable	238m	reactive, cuttable light metal (German *Pottasche*, Latin *kalium* from Arabic *quilyi*, a plant used to produce potash) 1807, part of many salts and rocks, essential for life, used in fertilizers, essential to chemical industry

Name	Symbol	At. n.	Aver. mass[a] in u (error), longest lifetime	Atomic[e] radius in pm	Main properties, (naming)[h] discovery date and use
Praeseodymium	Pr	59	140.907 65(2) stable	183m	white, malleable rare Earth metal (Greek *praesos didymos* green twin) 1885, used in cigarette lighters, material for carbon arcs used by the motion picture industry for studio lighting and projection, glass and enamel dye, darkens welder's goggles
Promethium[b]	Pm	61	(144.9127(1)) 17.7(4) a	181m	radioactive rare Earth metal (from the Greek mythical figure of Prometheus) 1945, used as beta source and to excite phosphors
Protactinium	Pa	91	(231.035 88(2)) 32.5(1) ka	n.a.	radioactive metal (Greek *protos* first, as it decays into actinium) 1917, found in nature, no use
Radium	Ra	88	(226.0254(1)) 1599(4) a	(223)	highly radioactive metal (Latin *radius* ray) 1898, no use any more; once used in luminous paints and as radioactive source and in medicine
Radon	Rn	86	(222.0176(1)) 3.823(4) d	(130n)	radioactive noble gas (from its old name 'radium emanation') 1900, no use (any more), found in soil, produces lung cancer
Rhenium	Re	75	186.207(1)[c] stable	138m	(Latin *rhenus* for Rhine river) 1925, used in filaments for mass spectrographs and ion gauges, superconductors, thermocouples, flash lamps, and as catalyst
Rhodium	Rh	45	102.905 50(2) stable	135m	white metal (Greek *rhodon* rose) 1803, used to harden platinum and palladium alloys, for electroplating, and as catalyst
Roentgenium[b]	Rg	111	(272.1535(1)) 1.5 ms[g]	n.a.	1994, no use
Rubidium	Rb	37	85.4678(3)[f] stable	255m	silvery-white, reactive alkali metal (Latin *rubidus* red) 1861, used in photocells, optical glasses, solid electrolytes
Ruthenium	Ru	44	101.107(2)[f] stable	134m	white metal (Latin *Rhuthenia* for Russia) 1844, used in platinum and palladium alloys, superconductors, as catalyst; the tetroxide is toxic and explosive
Rutherfordium[b]	Rf	104	(261.1088(1)) 1.3 min[g]	n.a.	radioactive transactinide (after Ernest Rutherford) 1964, no use

Name	Symbol	At. n.	Aver. mass[a] in u (error), longest lifetime	Atomic[e] radius in pm	Main properties, (naming)[h] discovery date and use
Samarium	Sm	62	150.36(2)[c,f] stable	180m	silver-white rare Earth metal (from the mineral samarskite, after Wassily Samarski) 1879, used in magnets, optical glasses, as laser dopant, in phosphors, in high-power light sources
Scandium	Sc	21	44.955 912(6) stable	164m	silver-white metal (from Latin *Scansia* Sweden) 1879, the oxide is used in high-intensity mercury vapour lamps, a radioactive isotope is used as tracer
Seaborgium[b]	Sg	106	266.1219(1) 21 s[g]	n.a.	radioactive transurane (after Glenn Seaborg) 1974, no use
Selenium	Se	34	78.96(3)[f] stable	120c, 190v	red or black or grey semiconductor (Greek *selene* Moon) 1818, used in xerography, glass production, photographic toners, as enamel dye
Silicon	Si	14	28.0855(3)[f] stable	105c, 210v	grey, shiny semiconductor (Latin *silex* pebble) 1823, Earth's crust, electronics, sand, concrete, bricks, glass, polymers, solar cells, essential for life
Silver	Ag	47	107.8682(2)[f] stable	145m	white metal with highest thermal and electrical conductivity (Latin *argentum*, Greek *argyros*) antiquity, used in photography, alloys, to make rain
Sodium	Na	11	22.989 769 28(2) stable	191m	light, reactive metal (Arabic *souwad* soda, Egyptian and Arabic *natrium*) component of many salts, soap, paper, soda, salpeter, borax, and essential for life
Strontium	Sr	38	87.62(1)[f] stable	215m	silvery, spontaneously igniting light metal (Strontian, Scottish town) 1790, used in TV tube glass, in magnets, and in optical materials
Sulphur	S	16	32.065(5)[f] stable	105c, 180v	yellow solid (Latin) antiquity, used in gunpowder, in sulphuric acid, rubber vulcanization, as fungicide in wine production, and is essential for life; some bacteria use sulphur instead of oxygen in their chemistry
Tantalum	Ta	73	180.947 88(2) stable	147m	heavy metal (Greek Tantalos, a mythical figure) 1802, used for alloys, surgical instruments, capacitors, vacuum furnaces, glasses

Name	Symbol	At. n.	Aver. mass[a] in u (error), longest lifetime	Atomic[e] radius in pm	Main properties, (naming)[h] discovery date and use
Technetium[b]	Tc	43	(97.9072(1)) 6.6(10) Ma	136m	radioactive (Greek *technetos* artificial) 1939, used as radioactive tracer and in nuclear technology
Tellurium	Te	52	127.60(3)[f] stable	139c, 206v	brittle, garlic-smelling semiconductor (Latin *tellus* Earth) 1783, used in alloys and as glass component
Terbium	Tb	65	158.925 35(2) stable	178m	malleable rare Earth metal (Ytterby, Swedish town) 1843, used as dopant in optical material
Thallium	Tl	81	204.3833(2) stable	172m	soft, poisonous heavy metal (Greek *thallos* branch) 1861, used as poison and for infrared detection
Thorium	Th	90	232.038 06(2)[d,f] 14.0(1) Ga	180m	radioactive (Nordic god Thor, as in 'Thursday') 1828, found in nature, heats Earth, used as oxide in gas mantles for campers, in alloys, as coating, and in nuclear energy
Thulium	Tm	69	168.934 21(2) stable	175m	rare Earth metal (Thule, mythical name for Scandinavia) 1879, found in monazite
Tin	Sn	50	118.710(7)[f] stable	139c, 210v, 162m	grey metal that, when bent, allows one to hear the 'tin cry' (Latin *stannum*) antiquity, used in paint, bronze and superconductors
Titanium	Ti	22	47.867(1) stable	146m	metal (Greek hero Titanos) 1791, alloys, fake diamonds
Tungsten	W	74	183.84(1) stable	141m	heavy, highest-melting metal (Swedish *tung sten* heavy stone, German name *Wolfram*) 1783, lightbulbs
Ununbium[b]	Uub	112	(285) 15.4 min[g]	n.a.	1996, no use
Ununtrium	Uut	113		n.a.	2004, no use
Ununquadium[b]	Uuq	114	(289) 30.4 s[g]	n.a.	1999, no use
Ununpentium	Uup	115		n.a.	2004, no use
Ununhexium[b]	Uuh	116	(289) 0.6 ms[g]	n.a.	2000 (earlier claim was false), no use
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 and of high density (planet Uranus, after the Greek sky god) 1789, found in pechblende and other minerals, used for nuclear energy

Name	Sym-bol	At. n.	Aver. mass[a] in u (error), longest lifetime	Atomic[e] radius in pm	Main properties, (naming)[h] discovery date and use
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, used in lasers
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, nuclear industry

a. The atomic mass unit is defined as $1\,\mathrm{u} = \frac{1}{12} m(^{12}\mathrm{C})$, making $1\,\mathrm{u} = 1.660\,5402(10)\,\mathrm{yg}$. For elements found on Earth, the *average* atomic mass for the naturally occurring isotope mixture is given, with the error in the
Ref. 270 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 is customary in this domain.

b. The element is not found on Earth because of its short lifetime.

c. The element has at least one radioactive isotope.

d. The element has no stable isotopes.

e. Strictly speaking, the *atomic radius* does not exist. Because atoms are clouds, they have no boundary. Several approximate definitions of the 'size' of atoms are possible. Usually, the radius is defined in such a way as to be useful for the estimation of distances between atoms. This distance is different for different bond
Ref. 271 types. In the table, radii for metallic bonds are labelled m, radii for (single) covalent bonds with carbon c,
Ref. 271 and Van der Waals radii v. Noble gas radii are labelled n. Note that values found in the literature vary by about 10 %; values in brackets lack literature references.

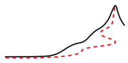
The covalent radius can be up to 0.1 nm smaller than the metallic radius for elements on the (lower) left of the periodic table; on the (whole) right side it is essentially equal to the metallic radius. In between, the
Challenge 176 s difference between the two decreases towards the right. Can you explain why? By the way, ionic radii differ considerably from atomic ones, and depend both on the ionic charge and the element itself.

All these values are for atoms in their ground state. Excited atoms can be 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 place where it was mined 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
Ref. 263 are known in atomic mass units to nine or more significant digits, and usually with one or two fewer digits
Ref. 270 in kilograms. The errors in the atomic mass are thus mainly due to the variations in isotopic composition.

g. The lifetime errors are asymmetric or not well known.

h. Extensive details on element names can be found on elements.vanderkrogt.net.



Appendix C

SPACES, ALGEBRAS AND SHAPES

Mathematicians are fond of generalizing concepts. One of the most generalized concepts of all is the concept of space. Understanding mathematical definitions and generalizations means learning to think with precision. The following pages provide a simple introduction to the types of spaces that are of importance in physics.

VECTOR SPACES

Vector spaces, also called linear spaces, are mathematical generalizations of certain aspects of the intuitive three-dimensional space. A set of elements any two of which can be added together and any one of which can be multiplied by a number 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*, for which a vector addition and a *scalar multiplication* is defined, such that for all vectors a, b, c and for all numbers s and r from K one has

$$\begin{aligned}
(a+b)+c = a+(b+c) &= a+b+c && \textit{associativity of vector addition}\\
n+a &= a && \textit{existence of null vector}\\
(-a)+a &= n && \textit{existence of negative vector}\\
1a &= a && \textit{regularity of scalar multiplication}\\
(s+r)(a+b) &= sa+sb+ra+rb && \textit{complete distributivity of scalar multiplication}
\end{aligned} \tag{132}$$

If the field K, whose elements are called *scalars* in this context, is taken to be the real (or complex, or quaternionic) numbers, one speaks of a real (or complex, or quaternionic) vector space. Vector spaces are also called *linear vector spaces* or simply *linear spaces*.

The complex numbers, the set of all real functions defined on the real line, the set of all polynomials, the set of matrices with a given number of rows and columns, all form vector spaces. In mathematics, a vector is thus a more general concept than in physics.
Challenge 177 ny (What is the simplest possible mathematical vector space?)

In physics, the term 'vector' is reserved for elements of a more specialized type of vector space, namely normed inner product spaces. To define these, we first need the concept of a metric space.

A *metric space* is a set with a metric, i.e., a way to define distances between elements.

A real function $d(a, b)$ between elements is called a metric if

$$\begin{aligned} d(a,b) &\geqslant 0 && \textit{positivity of metric} \\ d(a,b) + d(b,c) &\geqslant d(a,c) && \textit{triangle inequality} \\ d(a,b) = 0 \quad \text{if and only if} \quad a &= b && \textit{regularity of metric} \end{aligned} \tag{133}$$

A non-trivial example is the following. We define a special distance d between cities. If the two cities lie on a line going through Paris, we use the usual distance. In all other cases, we define the distance d by the shortest distance from one to the other travelling via Paris. This strange method defines a metric between all cities in France.

Challenge 178 s

A *normed* vector space is a linear space with a norm, or 'length', associated to each a vector. A *norm* is a non-negative number $\|a\|$ defined for each vector a with the properties

$$\begin{aligned} \|ra\| &= |r| \, \|a\| && \textit{linearity of norm} \\ \|a+b\| &\leqslant \|a\| + \|b\| && \textit{triangle inequality} \\ \|a\| = 0 \quad \text{only if} \quad a &= 0 && \textit{regularity} \end{aligned} \tag{134}$$

Challenge 179 ny

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

$$d(a,b) = \|a - b\| \tag{135}$$

so that all normed spaces are also metric spaces. This is the natural distance definition (in contrast to unnatural ones like that between French cities).

The norm is often defined with the help of an inner product. Indeed, the most special class of linear spaces are the *inner product spaces*. These are vector spaces with an *inner product*, also called *scalar product* · (not to be confused with the scalar multiplication!) which associates a number to each pair of vectors. An inner product space over ℝ satisfies

$$\begin{aligned} a \cdot b &= b \cdot a && \textit{commutativity of scalar product} \\ (ra) \cdot (sb) &= rs(a \cdot b) && \textit{bilinearity of scalar product} \\ (a+b) \cdot c &= a \cdot c + b \cdot c && \textit{left distributivity of scalar product} \\ a \cdot (b+c) &= a \cdot b + a \cdot c && \textit{right distributivity of scalar product} \\ a \cdot a &\geqslant 0 && \textit{positivity of scalar product} \\ a \cdot a = 0 \quad \text{if and only if} \quad a &= 0 && \textit{regularity of scalar product} \end{aligned} \tag{136}$$

for all vectors a, b, c 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.

An inner product space over $\mathbb{C}$ satisfies*

$$\begin{aligned}
a \cdot b = \overline{b \cdot a} = \overline{b} \cdot \overline{a} & \quad \textit{Hermitean property} \\
(ra) \cdot (sb) = r\overline{s}(a \cdot b) & \quad \textit{sesquilinearity of scalar product} \\
(a + b) \cdot c = a \cdot c + b \cdot d & \quad \textit{left distributivity of scalar product} \\
a \cdot (b + c) = a \cdot b + a \cdot c & \quad \textit{right distributivity of scalar product} \\
a \cdot a \geqslant 0 & \quad \textit{positivity of scalar product} \\
a \cdot a = 0 \quad \text{if and only if} \quad a = 0 & \quad \textit{regularity of scalar product}
\end{aligned} \qquad (137)$$

for all vectors a, b, c and all scalars r, s. A *complex* inner product space (of finite dimension) is also called a *unitary* or *Hermitean* vector space. If the inner product space is *complete*, it is called, especially in the infinite-dimensional complex case, a *Hilbert space*. Vol. IV, page 191 The space of all possible states of a quantum system forms a Hilbert space.

All inner product spaces are also metric spaces, and thus normed spaces, if the metric is defined by

$$d(a, b) = \sqrt{(a - b) \cdot (a - b)} \,. \qquad (138)$$

Only in the context of an inner product spaces we can speak about angles (or phase differences) between vectors, as we are used to in physics. Of course, like in normed spaces, inner product spaces also allows us to speak about the length of vectors and to define a *basis*, the mathematical concept necessary to define a coordinate system.

The *dimension* of a vector space is the number of linearly independent basis vectors. Can you define these terms precisely? Challenge 180 ny

A *Hilbert space* is a real or complex inner product space that is also a *complete* metric space. In other terms, in a Hilbert space, distances vary continuously and behave as naively expected. Hilbert spaces can have an infinite number of dimensions.

Which vector spaces are of importance in physics? Challenge 181 ny

ALGEBRAS

The term *algebra* is used in mathematics with three different, but loosely related, meanings. First, it denotes a part of mathematics, as in 'I hated algebra at school'. Secondly, it denotes a set of formal rules that are obeyed by abstract objects, as in the expression 'tensor algebra'. Finally – and this is the only meaning used here – an algebra denotes a *specific* type of mathematical structure.

Intuitively, an algebra is a set of vectors with a vector multiplication defined on it. More precisely, a (unital, associative) algebra is a vector space (over a field K) that is also a (unital) ring. (The concept is due to Benjamin Peirce (1809–1880), father of Charles Sanders Peirce.) A ring is a set for which an addition and a multiplication is defined – like the integers. Thus, in an algebra, there are (often) *three* types of multiplications: Vol. IV, page 190

* Two inequivalent forms of the sesquilinearity axiom exist. The other is $(ra) \cdot (sb) = \overline{r}s(a \cdot b)$. The term *sesquilinear* is derived from Latin and means for 'one-and-a-half-linear'.

— the (main) algebraic multiplication: the product of two vectors x and y is another vector $z = xy$;
— the scalar multiplication: the c-fold multiple of a vector x is another vector $y = cx$;
— if the vector space is a inner product space, the scalar product: the scalar product of two algebra elements (vectors) x and y is a scalar $c = x \cdot y$;

A precise definition of an algebra thus only needs to define properties of the (main) multiplication and to specify the number field K. An *algebra* is defined by the following axioms

$$x(y+z) = xy + xz \quad , \quad (x+y)z = xz + yz \quad \textit{distributivity of multiplication}$$
$$c(xy) = (cx)y = x(cy) \quad \textit{bilinearity} \tag{139}$$

for all vectors x, y, z and all scalars $c \in \mathrm{K}$. To stress their properties, algebras are also called *linear* algebras.

For example, the set of all linear transformations of an n-dimensional linear space (such as the translations on a plane, in space or in time) is a linear algebra, if the composition is taken as multiplication. So is the set of observables of a quantum mechanical system.*

An *associative algebra* is an algebra whose multiplication has the additional property that

$$x(yz) = (xy)z \quad \textit{associativity} \, . \tag{141}$$

Most algebras that arise in physics are associative** and unital. Therefore, in mathematical physics, a linear unital associative algebra is often simply called an algebra.

The set of multiples of the unit 1 of the algebra is called the *field of scalars* scal(A) of the algebra A. The field of scalars is also a subalgebra of A. The field of scalars and the scalars themselves behave in the same way.

We explore a few examples. The set of all polynomials in one variable (or in several
Challenge 183 e
variables) forms an algebra. It is commutative and infinite-dimensional. The constant

* *Linear transformations* are mappings from the vector space to itself, with the property that sums and scalar multiples of vectors are transformed into the corresponding sums and scalar multiples of the transformed
Challenge 182 s
vectors. Can you specify the set of all linear transformations of the plane? And of three-dimensional space? And of Minkowski space?

All linear transformations transform some special vectors, called *eigenvectors* (from the German word *eigen* meaning 'self') into multiples of themselves. In other words, if T is a transformation, e a vector, and

$$T(e) = \lambda e \tag{140}$$

where λ is a scalar, then the vector e is called an *eigenvector* of T, and λ is associated *eigenvalue*. The set of all eigenvalues of a transformation T is called the *spectrum* of T. Physicists did not pay much attention to these
Page 75
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 must be an eigenvalue of the corresponding transformation. The state of the system after the measurement is given by
Page 132
the eigenvector corresponding to the measured eigenvalue. Therefore every expert on motion must know what an eigenvalue is.

** Note that a non-associative algebra does not possess a matrix representation.

polynomials form the field of scalars.

The set of $n \times n$ matrices, with the usual operations, also forms an algebra. It is n^2-dimensional. Those diagonal matrices (matrices with all off-diagonal elements equal to zero) whose diagonal elements all have the same value form the field of scalars. How is the scalar product of two matrices defined?

Challenge 184 ny

The set of all real-valued functions over a set also forms an algebra. Can you specify the multiplication? The constant functions form the field of scalars.

Challenge 185 s

A *star algebra*, also written *∗-algebra*, is an algebra over the *complex* numbers for which there is a mapping $* : A \rightarrow A,\ x \mapsto x^*$, called an *involution*, with the properties

$$\begin{aligned} (x^*)^* &= x \\ (x+y)^* &= x^* + y^* \\ (cx)^* &= \bar{c}x^* \quad \text{for all} \quad c \in \mathbb{C} \\ (xy)^* &= y^* x^* \end{aligned} \qquad (142)$$

valid for all elements x, y of the algebra A. The element x^* is called the *adjoint* of x. Star algebras are the main type of algebra used in quantum mechanics, since quantum-mechanical observables form a ∗-algebra.

A C∗-*algebra* is a Banach algebra over the complex numbers with an involution ∗ (a function that is its own inverse) such that the norm $\|x\|$ of an element x satisfies

$$\|x\|^2 = x^* \, x \,. \qquad (143)$$

(A Banach algebra is a complete normed algebra; an algebra is *complete* if all Cauchy sequences converge.) In short, C∗-*algebra* is a nicely behaved algebra whose elements form a continuous set and a complex vector space. The name C comes from 'continuous functions'. Indeed, the *bounded* continuous functions form such an algebra, with a properly defined norm. Can you find it?

Challenge 186 s

Every C∗-algebra contains 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).

We should mention one important type of algebra used in mathematics. A *division algebra* is an algebra for which the equations $ax = b$ and $ya = b$ are uniquely solvable in x or y for all b and all $a \neq 0$. Obviously, all type of continuous numbers must be division algebras. Division algebras are thus one way to generalize the concept of a number. One of the important results of modern mathematics states that (finite-dimensional) division algebras can *only* have dimension 1, like the reals, dimension 2, like the complex numbers, dimension 4, like the quaternions, or dimension 8, like the octonions. There is thus no way to generalize the concept of (continuous) 'number' to other dimensions.

And now for some fun. Imagine a ring A which contains a number field K as a subring (or 'field of scalars'). If the ring multiplication is defined in such a way that a general ring element multiplied with an element of K is the same as the scalar multiplication, then A is a vector space, and thus an algebra – *provided* that every element of K commutes with every element of A. (In other words, the subring K must be *central*.)

For example, the quaternions $\mathbb{H}$ are a four-dimensional real division algebra, but although $\mathbb{H}$ is a two-dimensional complex vector space, it is *not* a complex algebra, because i does not commute with j (one has $ij = -ji = k$). In fact, there are no finite-dimensional complex division algebras, and the only finite-dimensional real associative division algebras are $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$.

Now, if you are not afraid of getting a headache, think about this remark: every K-algebra is also an algebra over its field of scalars. For this reason, some mathematicians prefer to define an (associative) K-algebra simply as a ring which contains K as a central subfield.

In physics, it is the algebras related to symmetries which play the most important role. We study them next.

Lie algebras

A Lie algebra is special type of algebra (and thus of vector space). Lie algebras are the most important type of non-associative algebras. A vector space L over the field $\mathbb{R}$ (or $\mathbb{C}$) with an additional binary operation [,], called *Lie multiplication* or the *commutator*, is called a real (or complex) *Lie algebra* if this operation satisfies

$$\begin{aligned} [X,Y] &= -[Y,X] && \textit{antisymmetry} \\ [aX+bY,Z] &= a[X,Z]+b[Y,Z] && \textit{(left-)linearity} \\ [X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]] &= 0 && \textit{Jacobi identity} \end{aligned} \qquad (144)$$

for all elements $X, Y, Z \in L$ and for all $a, b \in \mathbb{R}$ (or $\mathbb{C}$). (Lie algebras are named after Sophus Lie.) The first two conditions together imply bilinearity. 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.

Challenge 187 e

A Lie algebra is called a *linear* Lie algebra if its elements are linear transformations of another vector space V (intuitively, if they are 'matrices'). It turns out that every finite-dimensional Lie algebra is isomorphic to a linear Lie algebra. Therefore, there is no loss of generality in picturing the elements of finite-dimensional Lie algebras as matrices.

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 · stands for its multiplication) a Lie algebra appears when we define the *commutator* by

Page 292

$$[X,Y] = X \cdot Y - Y \cdot X \,. \qquad (145)$$

(This fact gave the commutator its name.) Lie algebras are non-associative in general; but the above definition of the commutator shows how to build one from an associative

Challenge 188 ny

* Can you explain the notation $[L,N]$? Can you define what a maximal ideal is and prove that there is only one?

algebra.

Since Lie algebras are vector spaces, the elements T_i of a *basis* of the Lie algebra always obey a relation of the form:

$$[T_i, T_j] = \sum_k c_{ij}^k T_k \ . \qquad (146)$$

The numbers c_{ij}^k are called the *structure constants* of the Lie algebra. They depend on the choice of basis. The 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 $C_{abc} = \varepsilon_{abc}$.*

Page 197

Classification of Lie algebras

Finite-dimensional Lie algebras are classified as follows. Every finite-dimensional Lie algebra is 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 yet been classified completely. 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.

Finite-dimensional semisimple Lie algebras have been completely classified. They decompose uniquely into a direct sum of *simple* Lie algebras. Simple Lie algebras can be complex

* Like 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

$$[a, a_j] = \sum_c \mathrm{ad}(a)_{cj} a_c \ . \qquad (147)$$

The definition implies that $\mathrm{ad}(a_i)_{jk} = c_{ij}^k$, 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

$$X \cdot Y = \mathrm{Tr}(\,\mathrm{ad}X \cdot \mathrm{ad}Y\,) \ . \qquad (148)$$

This scalar product is symmetric and bilinear. (Can you show that it is independent of the representation?) The corresponding bilinear form is also called the *Killing form*, after the German mathematician Wilhelm Killing (1847–1923), the discoverer of the 'exceptional' Lie groups. The Killing form is invariant under the action of any automorphism of the Lie algebra L. In a given basis, one has

$$X \cdot Y = \mathrm{Tr}(\,(\mathrm{ad}X) \cdot (\mathrm{ad}Y)) = c_{lk}^i c_{si}^k x^l y^s = g_{ls} x^l y^s \qquad (149)$$

where $g_{ls} = c_{lk}^i c_{si}^k$ is called the *Cartan metric tensor* of L.

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($n+1$) and their compact 'cousins' SU($n+1$); B_n for $n \geqslant 1$, corresponding to the Lie groups SO($2n+1$); C_n for $n \geqslant 1$, corresponding to the Lie groups Sp($2n$); and D_n for $n \geqslant 4$, corresponding to the Lie groups SO($2n$). Thus A_n is the algebra of all skew-Hermitean matrices; B_n and D_n are the algebras of the symmetric matrices; and C_n is the algebra of the traceless 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 dimensions of these algebras are $A_n : n(n+2)$; 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; their classification follows from that of the complex Lie algebras. Moreover, corresponding to each complex Lie group, there is always one compact real one. Real Lie algebras are not so important in fundamental physics. Ref. 272

Of the large number of *infinite-dimensional* Lie algebras, only few are important in physics: among them are the *Poincaré algebra*, and a few other algebras that only appear in failed attempts for unification.

TOPOLOGY – WHAT SHAPES EXIST?

> “Topology is group theory.”
> The Erlangen program

In a simplified view of topology that is sufficient for physicists, only one type of entity can possess shape: manifolds. *Manifolds* are generalized examples of pullovers: they are locally flat, can have holes and boundaries, and can often be turned inside out.

Pullovers are subtle entities. For example, can you turn your pullover inside out while your hands are tied together? (A friend may help you.) By the way, the same feat is also possible with your trousers, while your feet are tied together. Certain professors like to demonstrate this during topology lectures – of course with a carefully selected pair of underpants. Challenge 189 s

Another good topological puzzle, the handcuff puzzle, is shown in Figure 114. Which of the two situations can be untied without cutting the ropes? Ref. 273 Challenge 190 s

For a mathematician, pullovers and ropes are everyday examples of manifolds, and the operations that are performed on them are examples of deformations. Let us look at some more precise definitions. In order to define what a manifold is, we first need to define the concept of topological space.

Topological spaces

> “En Australie, une mouche qui marche au plafond se trouve dans le même sens qu'une vache chez nous.”
> Philippe Geluck, *La marque du chat.*

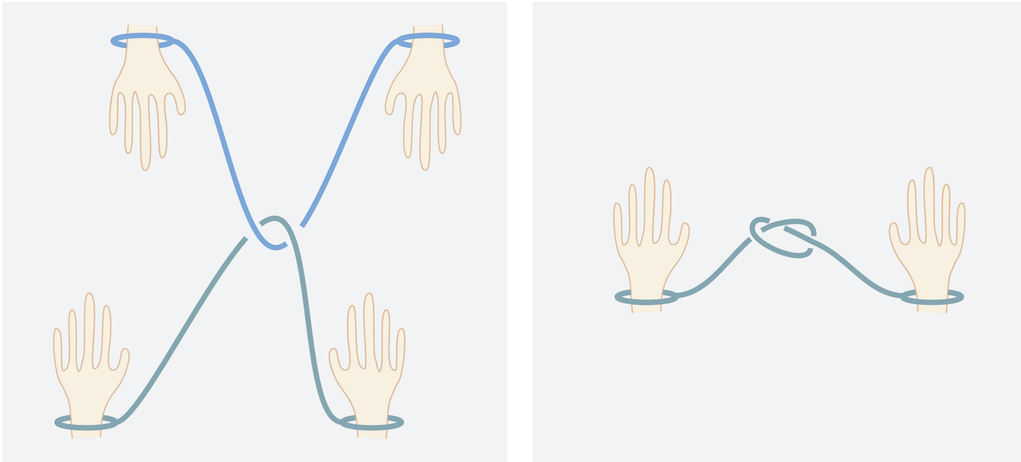


FIGURE 114 Which of the two situations can be untied without cutting?

Ref. 274 The study of shapes requires a good definition of a set made of 'points'. To be able to talk about shape, these sets must be structured in such a way as to admit a useful concept of 'neighbourhood' or 'closeness' between the elements of the set. The search for the most general type of set which allows a useful definition of neighbourhood has led to the concept of topological space. There are two ways to define a topology: one can define the concept of *open set* and then define the concept of *neighbourhood* with their help, or the other way round. We use the second option, which is somewhat more intuitive.

A *topological space* is a finite or infinite set X of elements, called *points*, together with the neighbourhoods for each point. A *neighbourhood* N of a point x is a collection of subsets Y_x of X with the properties that

— x is in every Y_x;
— if N and M are neighbourhoods of x, so is $N \cap M$;
— anything containing a neighbourhood of x is itself a neighbourhood of x.

The choice of the subsets Y_x is free. The subsets Y_x for all points x, chosen in a particular definition, contain a neighbourhood for each of their points; they are called *open sets*. (A neighbourhood and an open set usually differ, but all open sets are also neighbourhoods. Neighbourhoods of x can also be described as subsets of X that contain an open set that contains x.)

One also calls a topological space a 'set with a topology'. In effect, a *topology* specifies the systems of 'neighbourhoods' of every point of the set. 'Topology' is also the name of the branch of mathematics that studies topological spaces.

For example, the real numbers together with all open intervals form the *usual* topology of $\mathbb{R}$. Mathematicians have generalized this procedure. If one takes *all* subsets of $\mathbb{R}$ – or any other basis set – as open sets, one speaks of the *discrete* topology. If one takes *only* the full basis set and the empty set as open sets, one speaks of the *trivial* or *indiscrete* topology.

The concept of topological space allows us to define continuity. A mapping from one topological space X to another topological space Y is *continuous* if the inverse image

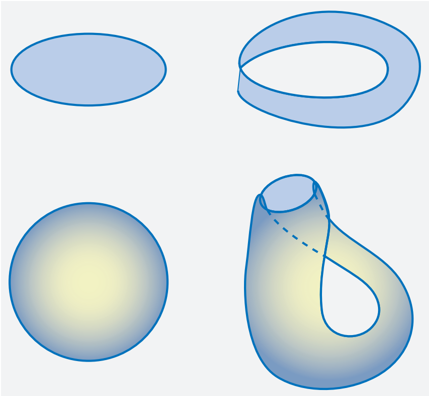


FIGURE 115 Examples of orientable and non-orientable manifolds of two dimensions: a disc, a Möbius strip, a sphere and a Klein bottle

of every open set in Y is an open set in X. You may verify that this condition is not satisfied by a real function that makes a jump. You may also check that the term 'inverse' is necessary in the definition; otherwise a function with a jump would be continuous, as such a function may still map open sets to open sets.*

Challenge 191 e

We thus need the concept of topological space, or of neighbourhood, if we want to express the idea that there are no jumps in nature. We also need the concept of topological space in order to be able to define limits.

Of the many special kinds of topological spaces that have been studied, one type is particularly important. A *Hausdorff space* is a topological space in which for any two points x and y there are disjoint open sets U and V such that x is in U and y is in V. A Hausdorff space is thus a space where, no matter how 'close' two points are, they can always be separated by open sets. This seems like a desirable property; indeed, non-Hausdorff spaces are rather tricky mathematical objects. (At Planck energy, it seems that vacuum appears to behave like a non-Hausdorff space; however, at Planck energy, vacuum is not really a space at all. So non-Hausdorff spaces play no role in physics.) A special case of Hausdorff space is well-known: the manifold.

Manifolds

In physics, the most important topological spaces are differential manifolds. Loosely speaking, a *differential manifold* – physicists simply speak of a *manifold* – is a set of points that looks like $\mathbb{R}^n$ under the microscope – at small distances. For example, a sphere and a torus are both two-dimensional differential manifolds, since they look locally like a plane. Not all differential manifolds are that simple, as the examples of Figure 115 show.

A differential manifold is called *connected* if any two points can be joined by a path lying in the manifold. (The term has a more general meaning in topological spaces. But the notions of connectedness and pathwise connectedness coincide for differential manifolds.) We focus on connected manifolds in the following discussion. A manifold is called

* The Cauchy–Weierstass definition of continuity says that a real function $f(x)$ is continuous at a point a if (1) f is defined on an open interval containing a, (2) $f(x)$ tends to a limit as x tends to a, and (3) the limit is $f(a)$. In this definition, the continuity of f is defined using the intuitive idea that the real numbers form the basic model of a set that has no gaps. Can you see the connection with the general definition given above?

Challenge 192 ny

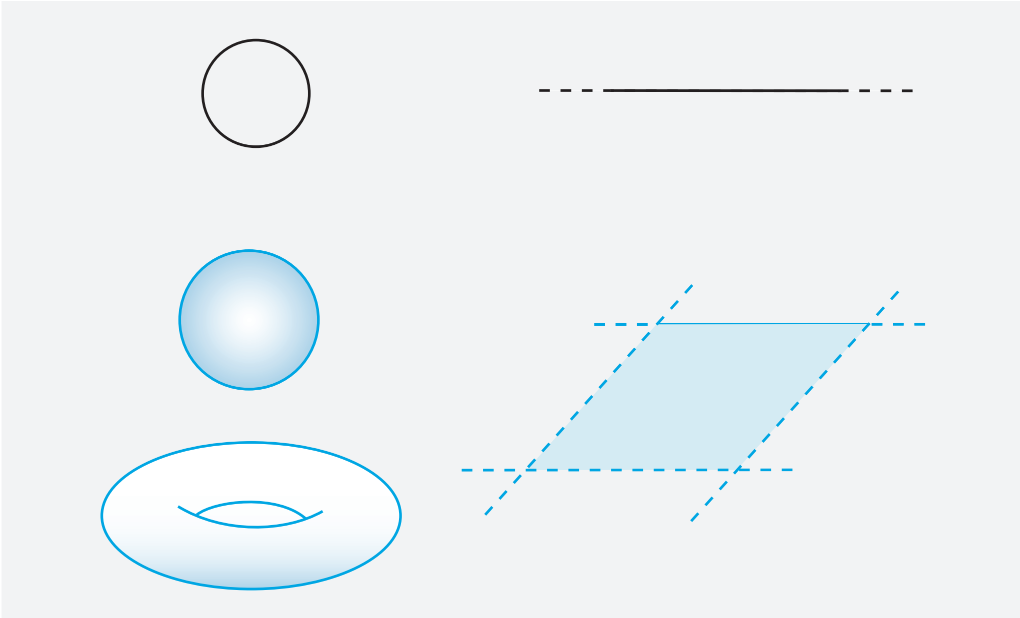


FIGURE 116 Compact (left) and non-compact (right) manifolds of various dimensions

simply connected if every loop lying in the manifold can be contracted to a point. For example, a sphere is simply connected. A connected manifold which is not simply connected is called *multiply connected*. A torus is multiply connected.

Manifolds can be *non-orientable*, as the well-known Möbius strip illustrates. Non-orientable manifolds have only one surface: they do not admit a distinction between front and back. If you want to have fun, cut a paper Möbius strip into two along a centre line. You can also try this with paper strips with different twist values, and investigate the regularities.

Challenge 193 e

In two dimensions, closed manifolds (or surfaces), i.e., surfaces that are compact and without boundary, are always of one of three types:

— The simplest type are spheres with n attached handles; they are called *n-tori* or *surfaces of genus n*. They are orientable surfaces with Euler characteristic $2 - 2n$.
— The *projective planes* with n handles attached are non-orientable surfaces with Euler characteristic $1 - 2n$.
— The Klein bottles with n attached handles are non-orientable surfaces with Euler characteristic $-2n$.

Therefore Euler characteristic and orientability describe compact surfaces up to homeomorphism (and if surfaces are smooth, then up to diffeomorphism). Homeomorphisms are defined below.

The two-dimensional compact manifolds or surfaces with boundary are found by removing one or more discs from a surface in this list. A compact surface can be embedded in $\mathbb{R}^3$ if it is orientable or if it has non-empty boundary.

In physics, the most important manifolds are space-time and Lie groups of observ-

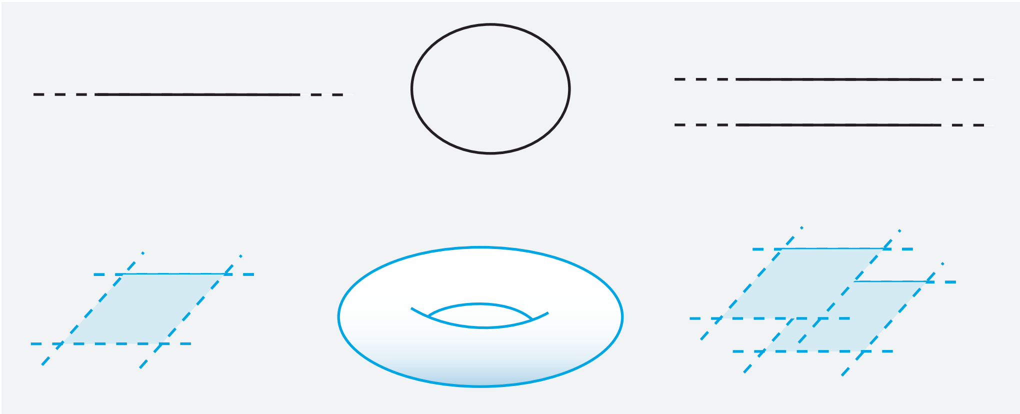


FIGURE 117 Simply connected (left), multiply connected (centre) and disconnected (right) manifolds of one (above) and two (below) dimensions

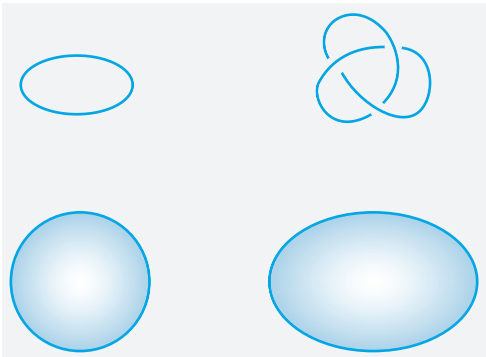


FIGURE 118 Examples of homeomorphic pairs of manifolds

ables. We study Lie groups below. Strangely enough, the topology of space-time is not known. For example, it is unclear whether or not it is simply connected. Obviously, the reason is that it is difficult to observe what happens at large distances form the Earth. However, a similar difficulty appears near Planck scales.

If a manifold is imagined to consist of rubber, connectedness and similar global properties are not changed when the manifold is deformed. This fact is formalized by saying that two manifolds are *homeomorphic* (from the Greek words for 'same' and 'shape') if between them there is a continuous, one-to-one and onto mapping with a continuous inverse. The concept of homeomorphism is somewhat more general than that of rubber deformation, as can be seen from Figure 118.

Holes, homotopy and homology

Only 'well-behaved' manifolds play a role in physics: namely those which are orientable and connected. In addition, the manifolds associated with observables, are always compact. The main non-trivial characteristic of connected compact orientable manifolds is that they contain 'holes' (see Figure 119). It turns out that a proper description of the holes of manifolds allows us to distinguish between all different, i.e., non-homeomorphic,

FIGURE 119 The first four two-dimensional compact connected orientable manifolds: 0-, 1-, 2- and 3-tori

types of manifold.

There are three main tools to describe holes of manifolds and the relations among them: homotopy, homology and cohomology. These tools play an important role in the study of gauge groups, because any gauge group defines a manifold.

In other words, through homotopy and homology theory, mathematicians can *classify* manifolds. Given two manifolds, the properties of the holes in them thus determine whether they can be deformed into each other.

Physicists are now extending these results of standard topology. Deformation is a classical idea which assumes continuous space and time, as well as arbitrarily small action. In nature, however, quantum effects cannot be neglected. It is speculated that quantum effects can transform a physical manifold into one with a *different* topology: for example, a torus into a sphere. Can you find out how this can be achieved? Challenge 194 d

Topological changes of physical manifolds happen via objects that are generalizations of manifolds. An *orbifold* is a space that is locally modelled by $\mathbb{R}^n$ modulo a finite group. Examples are the tear-drop or the half-plane. Orbifolds were introduced by Satake Ichiro in 1956; the name was coined by William Thurston. Orbifolds are heavily studied in string theory.

TYPES AND CLASSIFICATION OF GROUPS

Page 219 We introduced groups early on because groups play an important role in many parts of physics, from the description of solids, molecules, atoms, nuclei, elementary particles and forces up to the study of shapes, cycles and patterns in growth processes.

Group theory is also one of the most important branches of modern mathematics, and is still an active area of research. One of the aims of group theory is the *classification* of all groups. This has been achieved only for a few special types. In general, one distinguishes between finite and infinite groups. Finite groups are better understood.

Every *finite* group is isomorphic to a subgroup of the symmetric group S_N, for some number N. Examples of finite groups are the crystalline groups, used to classify crystal structures, or the groups used to classify wallpaper patterns in terms of their symmetries. The symmetry groups of Platonic and many other regular solids are also finite groups.

Finite groups are a complex family. Roughly speaking, a general (finite) group can be seen as built from some fundamental bricks, which are groups themselves. These fundamental bricks are called *simple* (finite) groups. One of the high points of twentieth-century mathematics was the classification of the finite simple groups. It was a collaborative effort that took around 30 years, roughly from 1950 to 1980. The complete list of finite simple groups consists of Ref. 275

1) the *cyclic groups* Z_p of prime group order;

2) the *alternating groups* A_n of degree n at least five;

3) the *classical linear groups*, $\mathrm{PSL}(n;q)$, $\mathrm{PSU}(n;q)$, $\mathrm{PSp}(2n;q)$ and $\mathrm{P}\Omega^{\varepsilon}(n;q)$;

4) the *exceptional* or *twisted groups* of Lie type ${}^3\mathrm{D}_4(q)$, $\mathrm{E}_6(q)$, ${}^2\mathrm{E}_6(q)$, $\mathrm{E}_7(q)$, $\mathrm{E}_8(q)$, $\mathrm{F}_4(q)$, ${}^2\mathrm{F}_4(2^n)$, $\mathrm{G}_2(q)$, ${}^2\mathrm{G}_2(3^n)$ and ${}^2\mathrm{B}_2(2^n)$;

5) the 26 *sporadic groups*, namely M_{11}, M_{12}, M_{22}, M_{23}, M_{24} (the Mathieu groups), J_1, J_2, J_3, J_4 (the Janko groups), Co_1, Co_2, Co_3 (the Conway groups), HS, Mc, Suz (the Co_1 'babies'), Fi_{22}, Fi_{23}, Fi'_{24} (the Fischer groups), $\mathrm{F}_1 = \mathrm{M}$ (the Monster), F_2, F_3, F_5, He ($= \mathrm{F}_7$) (the Monster 'babies'), Ru, Ly, and ON.

The classification was finished in the 1980s after over 10 000 pages of publications. The proof is so vast that a special series of books has been started to summarize and explain it. The first three families are infinite. The last family, that of the sporadic groups, is the most peculiar; it consists of those finite simple groups which do not fit into the other families. Some of these sporadic groups might have a role in particle physics: possibly even the largest of them all, the so-called *Monster* group. This is still a topic of research. (The Monster group has about $8.1 \cdot 10^{53}$ elements; more precisely, its order is 808 017 424 794 512 875 886 459 904 961 710 757 005 754 368 000 000 000 or $2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$.)

Of the *infinite* groups, only those with some finiteness condition have been studied. It is only such groups that are of interest in the description of nature. Infinite groups are divided into *discrete* groups and *continuous* groups. Discrete groups are an active area of mathematical research, having connections with number theory and topology. Continuous groups are divided into *finitely generated* and *infinitely generated* groups. Finitely generated groups can be finite-dimensional or infinite-dimensional.

The most important class of finitely generated continuous groups are the Lie groups.

Lie groups

In nature, the 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 a special type of infinite continuous group. They are named after the great Norwegian mathematician Sophus Lie (1849–1899). His name is pronounced like 'Lee'.

A (real) *Lie group* is an infinite symmetry group, i.e., a group with infinitely many elements, which is also an analytic manifold. Roughly speaking, this means that the elements of the group can be seen as points on a smooth (hyper-) surface whose shape can be described by an analytic function, i.e., by a function so smooth that it can be expressed as a power series in the neighbourhood of every point where it is defined. The points of the Lie group can be multiplied according to the group multiplication. Furthermore, the coordinates of the product have to be analytic functions of the coordinates of the factors, and the coordinates of the inverse of an element have to be analytic functions of the coordinates of the element. In fact, this definition is unnecessarily strict: it can be proved that a Lie group is just a topological group whose underlying space is a finite-dimensional, locally Euclidean manifold.

A *complex Lie group* is a group whose manifold is complex and whose group operations are holomorphic (instead of analytical) functions in the coordinates.

In short, a Lie group is a well-behaved manifold in which points can be multiplied (and technicalities). For example, the circle $\boldsymbol{T} = \{z \in \mathbb{C} : |z| = 1\}$, with the usual complex multiplication, is a real Lie group. It is Abelian. This group is also called S^1, as it is the one-dimensional sphere, or U(1), which means 'unitary group of one dimension'. The other one-dimensional Lie groups are the multiplicative group of non-zero real numbers and its subgroup, the multiplicative group of positive real numbers.

So far, in physics, only *linear* Lie groups have played a role – that is, Lie groups which act as linear transformations on some vector space. (The cover of SL(2,ℝ) or the complex compact torus are examples of non-linear Lie groups.) The important linear Lie groups for physics are the Lie subgroups of the general linear group GL(N,K), where K is a number field. This is defined as the set of all non-singular, i.e., invertible, N×N real, complex or quaternionic matrices. All the Lie groups discussed below are of this type.

Every *complex* invertible matrix A can be written in a unique way in terms of a unitary matrix U and a Hermitean matrix H:

$$A = U\mathrm{e}^{H} . \qquad (150)$$

Challenge 195 s (H is given by $H = \frac{1}{2}\ln A^{\dagger}A$, and U is given by $U = A\mathrm{e}^{-H}$.)

The simple Lie groups U(1) and SO(2,ℝ) and the Lie groups based on the real and complex numbers are Abelian (see Table 34); all others are non-Abelian.

Lie groups are manifolds. Therefore, in a Lie group one can define the distance between two points, the tangent plane (or tangent space) at a point, and the notions of integration and differentiations. Because Lie groups are manifolds, Lie groups have the same kind of structure as the objects of Figures 115, 116 and 117. Lie groups can have any number of dimensions. Like for any manifold, their global structure contains important information; let us explore it.

Connectedness

It is not hard to see that the Lie groups SU(N) are simply connected for all $N = 2, 3 \dots$; they have the topology of a 2N-dimensional sphere. The Lie group U(1), having the topology of the 1-dimensional sphere, or circle, is multiply connected.

The Lie groups SO(N) are *not* simply connected for any $N = 2, 3 \dots$. In general, SO(N,K) is connected, and GL(N,ℂ) is connected. All the Lie groups SL(N,K) are connected; and SL(N,ℂ) is simply connected. The Lie groups Sp(N,K) are connected; Sp(2N,ℂ) is simply connected. Generally, all semi-simple Lie groups are connected.

The Lie groups O(N,K), SO(N,M,K) and GL(N,ℝ) are not connected; they contain two connected components.

Note that the Lorentz group is not connected: it consists of four separate pieces. Like the Poincaré group, it is not compact, and neither is any of its four pieces. Broadly speaking, the non-compactness of the group of space-time symmetries is a consequence of the non-compactness of space-time.

Compactness

A Lie group is *compact* if it is closed and bounded when seen as a manifold. For a given parametrization of the group elements, the Lie group is compact if all parameter ranges are closed and finite intervals. Otherwise, the group is called *non-compact*. Both compact and non-compact groups play a role in physics. The distinction between the two cases is important, because representations of compact groups can be constructed in the same simple way as for finite groups, whereas for non-compact groups other methods have to be used. As a result, physical observables, which always belong to a representation of a symmetry group, have different properties in the two cases: if the symmetry group is compact, observables have *discrete* spectra; otherwise they do not.

All groups of internal gauge transformations, such as U(1) and SU(n), form compact groups. In fact, field theory *requires* compact Lie groups for gauge transformations. The only compact Lie groups are T^n, O(n), U(n), SO(n) and SU(n), their double cover Spin(n) and the Sp(n). In contrast, SL(n,$\mathbb{R}$), GL(n,$\mathbb{R}$), GL(n,$\mathbb{C}$) and all others are not compact.

Besides being manifolds, Lie groups are obviously also groups. It turns out that most of their group properties are revealed by the behaviour of the elements which are very close (as points on the manifold) to the identity.

Every element of a compact and connected Lie group has the form exp(A) for some A. The elements A arising in this way form an algebra, called the *corresponding Lie algebra*. For any linear Lie group, every element of the connected subgroup can be expressed as a finite product of exponentials of elements of the corresponding Lie algebra. Mathematically, the vector space defined by the Lie algebra is tangent to the manifold defined by the Lie group, at the location of the unit element. In short, Lie algebras express the local properties of Lie groups near the identity. That is the reason for their importance in physics.

Page 282

TABLE 34 Properties of the most important real and complex Lie groups

Lie group	Description	Properties[a]	Lie algebra	Description of Lie algebra	Dimension
1. Real groups					real
$\mathbb{R}^n$	Euclidean space with addition	Abelian, simply connected, not compact; $\pi_0 = \pi_1 = 0$	$\mathbb{R}^n$	Abelian, thus Lie bracket is zero; not simple	n
$\mathbb{R}^{\times}$	non-zero real numbers with multiplication	Abelian, not connected, not compact; $\pi_0 = \mathbb{Z}_2$, no π_1	$\mathbb{R}$	Abelian, thus Lie bracket is zero	1
$\mathbb{R}^{>0}$	positive real numbers with multiplication	Abelian, simply connected, not compact; $\pi_0 = \pi_1 = 0$	$\mathbb{R}$	Abelian, thus Lie bracket is zero	1
$S^1 = \mathbb{R}/\mathbb{Z}$ = U(1) = T = SO(2) = Spin(2)	complex numbers of absolute value 1, with multiplication	Abelian, connected, not simply connected, compact; $\pi_0 = 0$, $\pi_1 = \mathbb{Z}$	$\mathbb{R}$	Abelian, thus Lie bracket is zero	1

Lie group	Description	Properties[a]	Lie algebra	Description of Lie algebra	Dimension
$\mathbb{H}^{\times}$	non-zero quaternions with multiplication	simply connected, not compact; $\pi_0 = \pi_1 = 0$	$\mathbb{H}$	quaternions, with Lie bracket the commutator	4
S^3	quaternions of absolute value 1, with multiplication, also known as Sp(1); topologically a 3-sphere	simply connected, compact; isomorphic to SU(2), Spin(3) and to double cover of SO(3); $\pi_0 = \pi_1 = 0$	Im($\mathbb{H}$)	quaternions with zero real part, with Lie bracket the commutator; simple and semi-simple; isomorphic to real 3-vectors, with Lie bracket the cross product; also isomorphic to su(2) and to so(3)	3
GL(n, $\mathbb{R}$)	general linear group: invertible n-by-n real matrices	not connected, not compact; $\pi_0 = \mathbb{Z}_2$, no π_1	M(n, $\mathbb{R}$)	n-by-n matrices, with Lie bracket the commutator	n^2
$GL^+(n, \mathbb{R})$	n-by-n real matrices with positive determinant	simply connected, not compact; $\pi_0 = 0$, for $n = 2$: $\pi_1 = \mathbb{Z}$, for $n \geq 2$: $\pi_1 = \mathbb{Z}_2$; $GL^+(1, \mathbb{R})$ isomorphic to $\mathbb{R}^{>0}$	M(n, $\mathbb{R}$)	n-by-n matrices, with Lie bracket the commutator	n^2
SL(n, $\mathbb{R}$)	special linear group: real matrices with determinant 1	simply connected, not compact if $n > 1$; $\pi_0 = 0$, for $n = 2$: $\pi_1 = \mathbb{Z}$, for $n \geq 2$: $\pi_1 = \mathbb{Z}_2$; SL(1, $\mathbb{R}$) is a single point, SL(2, $\mathbb{R}$) is isomorphic to SU(1, 1) and Sp(2, $\mathbb{R}$)	sl(n, $\mathbb{R}$) $= A_{n-1}$	n-by-n matrices with trace 0, with Lie bracket the commutator	$n^2 - 1$
O(n, $\mathbb{R}$) = O(n)	orthogonal group: real orthogonal matrices; symmetry of hypersphere	not connected, compact; $\pi_0 = \mathbb{Z}_2$, no π_1	so(n, $\mathbb{R}$)	skew-symmetric n-by-n real matrices, with Lie bracket the commutator; so(3, $\mathbb{R}$) is isomorphic to su(2) and to $\mathbb{R}^3$ with the cross product	$n(n-1)/2$

Lie group	Description	Properties[a]	Lie algebra	Description of Lie algebra	Dimension
SO(n, $\mathbb{R}$) = SO(n)	special orthogonal group: real orthogonal matrices with determinant 1	connected, compact; for $n \geqslant 2$ not simply connected; $\pi_0 = 0$, for $n = 2$: $\pi_1 = \mathbb{Z}$, for $n \geq 2$: $\pi_1 = \mathbb{Z}_2$	so(n, $\mathbb{R}$) = $B_{\frac{n-1}{2}}$ or $D_{\frac{n}{2}}$	skew-symmetric n-by-n real matrices, with Lie bracket the commutator; for $n = 3$ and $n \geqslant 5$ simple and semisimple; SO(4) is semisimple but not simple	$n(n-1)/2$
Spin(n)	spin group; double cover of SO(n); Spin(1) is isomorphic to $\mathbb{Q}_2$, Spin(2) to S^1	simply connected for $n \geqslant 3$, compact; for $n = 3$ and $n \geqslant 5$ simple and semisimple; for $n > 1$: $\pi_0 = 0$, for $n > 2$: $\pi_1 = 0$	so(n, $\mathbb{R}$)	skew-symmetric n-by-n real matrices, with Lie bracket the commutator	$n(n-1)/2$
Sp($2n$, $\mathbb{R}$)	symplectic group: real symplectic matrices	not compact; $\pi_0 = 0$, $\pi_1 = \mathbb{Z}$	sp($2n$, $\mathbb{R}$) = C_n	real matrices A that satisfy $JA + A^T J = 0$ where J is the standard skew-symmetric matrix;[b] simple and semisimple	$n(2n+1)$
Sp(n) for $n \geqslant 3$	compact symplectic group: quaternionic $n \times n$ unitary matrices	compact, simply connected; $\pi_0 = \pi_1 = 0$	sp(n)	n-by-n quaternionic matrices A satisfying $A = -A^*$, with Lie bracket the commutator; simple and semisimple	$n(2n+1)$
U(n)	unitary group: complex $n \times n$ unitary matrices	not simply connected, compact; it is *not* a complex Lie group/algebra; $\pi_0 = 0$, $\pi_1 = \mathbb{Z}$; isomorphic to S^1 for $n = 1$	u(n)	n-by-n complex matrices A satisfying $A = -A^*$, with Lie bracket the commutator	n^2
SU(n)	special unitary group: complex $n \times n$ unitary matrices with determinant 1	simply connected, compact; it is *not* a complex Lie group/algebra; $\pi_0 = \pi_1 = 0$	su(n)	n-by-n complex matrices A with trace 0 satisfying $A = -A^*$, with Lie bracket the commutator; for $n \geqslant 2$ simple and semisimple	$n^2 - 1$
2. Complex groups[c]					complex

Lie group	Description	Properties[a]	Lie algebra	Description of Lie algebra	Dimension
$\mathbb{C}^n$	group operation is addition	Abelian, simply connected, not compact; $\pi_0 = \pi_1 = 0$	$\mathbb{C}^n$	Abelian, thus Lie bracket is zero	n
$\mathbb{C}^\times$	nonzero complex numbers with multiplication	Abelian, not simply connected, not compact; $\pi_0 = 0$, $\pi_1 = \mathbb{Z}$	$\mathbb{C}$	Abelian, thus Lie bracket is zero	1
$\mathrm{GL}(n, \mathbb{C})$	general linear group: invertible n-by-n complex matrices	simply connected, not compact; $\pi_0 = 0$, $\pi_1 = \mathbb{Z}$; for $n = 1$ isomorphic to $\mathbb{C}^\times$	$\mathrm{M}(n, \mathbb{C})$	n-by-n matrices, with Lie bracket the commutator	n^2
$\mathrm{SL}(n, \mathbb{C})$	special linear group: complex matrices with determinant 1	simply connected; for $n \geqslant 2$ not compact; $\pi_0 = \pi_1 = 0$; $\mathrm{SL}(2, \mathbb{C})$ is isomorphic to $\mathrm{Spin}(3, \mathbb{C})$ and $\mathrm{Sp}(2, \mathbb{C})$	$\mathrm{sl}(n, \mathbb{C})$	n-by-n matrices with trace 0, with Lie bracket the commutator; simple, semisimple; $\mathrm{sl}(2, \mathbb{C})$ is isomorphic to $\mathrm{su}(2, \mathbb{C}) \otimes \mathbb{C}$	$n^2 - 1$
$\mathrm{PSL}(2, \mathbb{C})$	projective special linear group; isomorphic to the Möbius group, to the restricted Lorentz group $\mathrm{SO}^+(3, 1, \mathbb{R})$ and to $\mathrm{SO}(3, \mathbb{C})$	not compact; $\pi_0 = 0$, $\pi_1 = \mathbb{Z}_2$	$\mathrm{sl}(2, \mathbb{C})$	2-by-2 matrices with trace 0, with Lie bracket the commutator; $\mathrm{sl}(2, \mathbb{C})$ is isomorphic to $\mathrm{su}(2, \mathbb{C}) \otimes \mathbb{C}$	3
$\mathrm{O}(n, \mathbb{C})$	orthogonal group: complex orthogonal matrices	not connected; for $n \geqslant 2$ not compact; $\pi_0 = \mathbb{Z}_2$, no π_1	$\mathrm{so}(n, \mathbb{C})$	skew-symmetric n-by-n complex matrices, with Lie bracket the commutator	$n(n-1)/2$

Lie group	Description	Properties[a]	Lie algebra	Description of Lie algebra	Dimension
$SO(n, \mathbb{C})$	special orthogonal group: complex orthogonal matrices with determinant 1	for $n \geqslant 2$ not compact; not simply connected; $\pi_0 = 0$, for $n = 2$: $\pi_1 = \mathbb{Z}$, for $n \geq 2$: $\pi_1 = \mathbb{Z}_2$; non-Abelian for $n > 2$, $SO(2, \mathbb{C})$ is Abelian and isomorphic to $\mathbb{C}^\times$	$so(n, \mathbb{C})$	skew-symmetric n-by-n complex matrices, with Lie bracket the commutator; for $n = 3$ and $n \geqslant 5$ simple and semisimple	$n(n-1)/2$
$Sp(2n, \mathbb{C})$	symplectic group: complex symplectic matrices	not compact; $\pi_0 = \pi_1 = 0$	$sp(2n, \mathbb{C})$	complex matrices that satisfy $JA + A^T J = 0$ where J is the standard skew-symmetric matrix;[b] simple and semi-simple	$n(2n+1)$

a. The *group of components* π_0 of a Lie group is given; the order of π_0 is the number of components of the Lie group. If the group is trivial (0), the Lie group is connected. The *fundamental group* π_1 of a connected Lie group is given. If the group π_1 is trivial (0), the Lie group is simply connected. This table is based on that in the Wikipedia, at en.wikipedia.org/wiki/Table_of_Lie_groups.
b. The standard skew-symmetric matrix J of rank $2n$ is $J_{kl} = \delta_{k,n+l} - \delta_{k+n,l}$.
c. Complex Lie groups and Lie algebras can be viewed as real Lie groups and real Lie algebras of twice the dimension.

MATHEMATICAL CURIOSITIES AND FUN CHALLENGES

Challenge 196 ny A theorem of topology says: you cannot comb a hairy football. Can you prove it?

* *

There are at least seven ways to earn a million dollars with mathematical research. The Clay Mathematics Institute at www.claymath.org offers them for major advances in seven topics:

— proving the Birch and Swinnerton–Dyer conjecture about algebraic equations;
— proving the Poincaré conjecture about topological manifolds;
— solving the Navier–Stokes equations for fluids;
— finding criteria distinguishing P and NP numerical problems;
— proving the Riemann hypothesis stating that the non-trivial zeros of the zeta function lie on a line;
— proving the Hodge conjectures;
— proving the connection between Yang–Mills theories and a mass gap in quantum field theory.

On each of these topics, substantial progress can buy you a house.

CHALLENGE HINTS AND SOLUTIONS

Challenge 1, page 9: Do not hesitate to be demanding and strict. The next edition of the text will benefit from it.

Challenge 2, page 15: A virus is an example. It has no own metabolism. (By the way, the ability of some viruses to form crystals is not a proof that they are not living beings, in contrast to what is often said.)

Challenge 3, page 16: The navigation systems used by flies are an example.

Challenge 4, page 19: The thermal energy kT is about 4 zJ and a typical relaxation time is 0.1 ps.

Challenge 5, page 23: The argument is correct.

Challenge 6, page 23: 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 7, page 23: Also identical twins count as different persons and have different fates. Imprinting in the womb is different, so that their temperament will be different. The birth experience will be different; this is the most intense experience of every human, strongly determining his fears and thus his character. A person with an old father is also quite different from that with a young father. If the womb is not that of his biological mother, a further distinction of the earliest and most intensive experiences is given.

Challenge 8, page 23: Be sure to publish your results.

Challenge 9, page 24: Life's chemicals are synthesized inside the body; the asymmetry has been inherited along the generations. The common asymmetry thus shows that all life has a common origin.

Challenge 10, page 24: Well, men are more similar to chimpanzees than to women. More seriously, the above data, even though often quoted, are wrong. Newer measurements by Roy Britten in 2002 have shown that the difference in genome between humans and chimpanzees is about 5 % (See R. J. Britten, Divergence between samples of chimpanzee and human DNA sequences is 5 %, counting indels, *Proceedings of the National Academy of Sciences* **99**, pp. 13633–13635, 15th of October, 2002.) In addition, though the difference between man and woman is smaller than one whole chromosome, the large size of the X chromosome, compared with the small size of the Y chromosome, implies that men have about 3 % less genetic material than women. However, all men have an X chromosome as well. That explains that still other measurements suggest that all humans share a pool of at least 99.9 % of common genes.

Challenge 13, page 26: 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 14, page 26: The first steps are not known at all.

Challenge 15, page 27: Since all the atoms we are made of originate from outer space, the answer is yes. But if one means that biological cells came to Earth from space, the answer is no, as most cells do not like vacuum. The same is true for DNA.

In fact, life and reproduction are properties of complex systems. In other words, asking whether life comes from outer space is like asking: 'Could car insurance have originated in outer space?'

Challenge 16, page 29: Haven't you tried yet? Physics is an experimental science.

Challenge 27, page 41: Radioactive dating methods can be said to be based on the nuclear interactions, even though the detection is again electromagnetic.

Challenge 28, page 42: All detectors of light can be called relativistic, as light moves with maximal speed. Touch sensors are not relativistic following the usual sense of the word, as the speeds involved are too small. The energies are small compared to the rest energies; this is the case even if the signal energies are attributed to electrons only.

Challenge 29, page 42: The noise is due to the photoacoustic effect; the periodic light periodically heats the air in the jam glass at the blackened surface and thus produces sound. See M. EULER, Kann man Licht hören?, *Physik in unserer Zeit* **32**, pp. 180–182, 2001.

Challenge 35, page 50: It implies that neither resurrection nor reincarnation nor eternal life are possible.

Challenge 36, page 51: You get an intense yellow colour due to the formation of lead iodide (PbI_2).

Challenge 38, page 62: With a combination of the methods of Table 5 it is possible. Indeed, using cosmic rays to search for unknown chambers in the pyramids has been already done in the
Ref. 276 1960s. The result was that no additional chambers exist.

Challenge 40, page 64: For example, a heavy mountain will push down the Earth's crust into the mantle, makes it melt on the bottom side, and thus lowers the position of the top.

Challenge 41, page 64: These developments are just starting; the results are still far from the original one is trying to copy, as they have to fulfil a second condition, in addition to being a 'copy' of original feathers or of latex: the copy has to be cheaper than the original. That is often a much tougher request than the first.

Challenge 42, page 64: About 0.2 m.

Challenge 44, page 64: Since the height of the potential is always finite, walls can always be overcome by tunnelling.

Challenge 45, page 64: 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 47, page 65: The concentrations and can be measured from polar ice caps, by measuring how the isotope concentration changes over depth. Both in evaporation and in condensation of water, the isotope ratio depends on the temperature. The measurements in Antarctica and in Greenland coincide, which is a good sign of their trustworthiness.

Challenge 51, page 77: The one somebody else has thrown away. Energy costs about 10 cents/kWh. For new lamps, the fluorescence lamp is the best for the environment, even though it is the least friendly to the eye and the brain, due to its flickering.

Challenge 52, page 81: This old dream depends on the precise conditions. How flexible does the display have to be? What lifetime should it have? The newspaper like display is many years away and maybe not even possible.

Challenge 53, page 81: The challenge here is to find a cheap way to deflect laser beams in a controlled way. Cheap lasers are already available.

Challenge 54, page 81: There is only speculation on the answer; the tendency of most researchers is to say no.

Challenge 55, page 81: No, as it is impossible because of momentum conservation, because of the no-cloning theorem.

Challenge 56, page 81: There are companies trying to do sell systems based on quantum cryptology; but despite the technical interest, the commercial success is questionable.

Challenge 57, page 81: I predicts that mass-produced goods using this technology (at least 1 million pieces sold) will not be available before 2025.

Challenge 58, page 81: Maybe, but for extremely high prices.

Challenge 59, page 84: For example, you could change gravity between two mirrors.

Challenge 60, page 84: As usual in such statements, either group or phase velocity is cited, but not the corresponding energy velocity, which is always below c.

Challenge 62, page 87: Echoes do not work once 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 63, page 88: Mirrors do not usually work for matter; in addition, if they did, matter would require much higher acceleration values.

Challenge 66, page 89: The classical radius of the electron, which is the size at which the field energy would make up the hole electron mass, is about 137 times smaller, thus much smaller, than the Compton wavelength of the electron.

Challenge 67, page 91: 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 68, page 91: A larger charge would lead to a field that spontaneously generates electron positron pairs, the electron would fall into the nucleus and reduce its charge by one unit.

Challenge 70, page 91: 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 71, page 92: The usual way to pack oranges on a table is the densest way to pack spheres.

Challenge 72, page 93: 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 73, page 93: The steps are due to the particle nature of electricity and all other moving entities.

Challenge 74, page 94: If we could apply the Banach–Tarski paradox to vacuum, it seems that Page 53 we could split, without any problem, one ball of vacuum into *two* balls of vacuum, each with the same volume as the original. In other words, one ball with vacuum energy E could not be distinguished from two balls of vacuum energy $2E$.

We used the Banach–Tarski paradox in this way to show that chocolate (or any other matter) Page 262 possesses an intrinsic length. But it is *not* clear that we can now deduce that the vacuum has an intrinsic length. Indeed, the paradox cannot be applied to vacuum for *two* reasons. First, there indeed is a maximum energy and minimum length in nature. Secondly, there is no place in nature without vacuum energy; so there is no place were we could put the second ball. We thus do not know why the Banach–Tarski paradox for vacuum cannot be applied, and thus cannot use it to deduce the existence of a minimum length in vacuum.

It is better to argue in the following way for a minimum length in vacuum. If there were no intrinsic length cut-off, the vacuum energy would be infinite. Experiments however, show that it is finite.

Challenge 75, page 94: 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 76, page 94: No. Bound states of massless particles are always unstable.

Challenge 77, page 94: The first answer is probably no, as composed systems cannot be smaller than their own compton wavelength; only elementary systems can. However, the universe is not a system, as it has no environment. As such, its length is not a precisely defined concept, as an environment is needed to measure and to define it. (In addition, gravity must be taken into account in those domains.) Thus the answer is: in those domains, the question makes no sense.

Challenge 78, page 94: Methods to move on perfect ice from mechanics:

— if the ice is perfectly flat, rest is possible only in one point – otherwise you oscillate around that point, as shown in challenge 23;
— 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 79, page 95: Methods to move on perfect ice using thermodynamics and electrodynamics:

— use the radio/TV stations nearby 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 for 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 80, page 95: 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 for the universe to contract.

Challenge 81, page 95: Methods to move on perfect ice using quantum effects:

— wait for your wave function 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 82, page 95: Methods to move on perfect ice using materials science, geophysics and 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 83, page 95: A method to move on perfect ice using self-organization, chaos theory, and biophysics:

— wait that the currents in the brain interact with the magnetic field of the Earth by controlling your thoughts.

Challenge 84, page 95: 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 86, page 100: This is a trick question: if you can say why, you can directly move to the last volume of this adventure and check your answer. The gravitational potential changes the phase of a wave function, like any other potential does; but the reason *why* this is the case will only become clear in the last volume of this series.

Challenge 90, page 102: This is easy only if the black hole size is inserted into the entropy bound by Bekenstein. A simple deduction of the black hole entropy that includes the factor 1/4 is not yet at hand; more on this in the last volume.

Challenge 91, page 103: An entropy limit implies an information limit; only a given information can be present in a given region of nature. This results in a memory limit.

Challenge 92, page 103: In natural units, the exact expression for entropy is $S = 0.25A$. If each Planck area carried one bit (degree of freedom), the entropy would be $S = \ln W = \ln(2^A) = A\ln 2 = 0.693A$. This close to the exact value.

Challenge 96, page 109: 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 102, page 111: They are carried away by the gravitational radiation.

Challenge 108, page 116: 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 112, page 123: Two stacked foils show the same effect as one foil of the same total thickness. Thus the surface plays no role.

Challenge 114, page 127: The electron is held back by the positive charge of the nucleus, if the number of protons in the nucleus is sufficient, as is the case for those nuclei we are made of.

Challenge 116, page 134: The number is small compared with the number of cells. However, it is possible that the decays are related to human ageing.

Challenge 117, page 134: There is way to conserve both energy and momentum in such a decay.

Challenge 118, page 135: By counting decays and counting atoms to sufficient precision.

Challenge 120, page 137: The radioactivity necessary to keep the Earth warm is low; lava is only slightly more radioactive than usual soil.

Challenge 122, page 150: The nuclei of nitrogen and carbon have a high electric charge which strongly repels the protons.

Challenge 123, page 155: See the paper by C. J. Hogan, Why the universe is just so, *Reviews of Modern Physics* 72, pp. 1149–1161, 2000.

Challenge 124, page 159: Touching something requires getting near it; getting near means a small time and position indeterminacy; this implies a small wavelength of the probe that is used for touching; this implies a large energy.

Challenge 126, page 166: The processes are electromagnetic in nature, thus electric charges give the frequency with which they occur.

Challenge 127, page 174: Designing a nuclear weapon is not difficult. University students can do it, and even have done so a few times. The first students who did so were two physics graduates in 1964, as told on www.guardian.co.uk/world/2003/jun/24/usa.science. It is not hard to conceive a design and even to build it. By far the hardest problem is getting or making the nuclear material. That requires either an extensive criminal activity or a vast technical effort, with numerous large factories, extensive development, and coordination of many technological activities. Most importantly, such a project requires a large financial investment, which poor countries cannot afford without great sacrifices for all the population. The problems are thus not technical, but financial.

Challenge 132, page 196: In 2008 an estimated 98% of all physicists agreed. Time will tell whether they are right.

Challenge 134, page 206: A mass of 100 kg and a speed of 8 m/s require 43 m² of wing surface.

Challenge 137, page 213: The largest rotation angle $\Delta\varphi$ that can be achieved in one stroke C is found by maximizing the integral

$$\Delta\varphi = -\int_C \frac{a^2}{a^2 + b^2} \mathrm{d}\theta \qquad (151)$$

Since the path C in shape space is closed, we can use Stokes' theorem to transform the line integral to a surface integral over the surface S enclosed by C in shape space:

$$\Delta\varphi = \int_S \frac{2ab^2}{(a^2+b^2)^2} \mathrm{d}a\,\mathrm{d}\theta \,. \tag{152}$$

The maximum angle is found by noting that θ can vary at most between 0 and π, and that a can vary at most between 0 and ∞. This yields

$$\Delta\varphi_{\max} = \int_{\theta=0}^{\pi} \int_{a=0}^{\infty} \frac{2ab^2}{(a^2+b^2)^2} \mathrm{d}a\,\mathrm{d}\theta = \pi \,. \tag{153}$$

Challenge 144, page 222: Lattices are not isotropic, lattices are not Lorentz invariant.

Challenge 146, page 225: The infinite sum is not defined for numbers; however, it is defined for a knotted string.

Challenge 147, page 228: The research race for the solution is ongoing, but the goal is still far.

Challenge 148, page 230: This is a simple but hard question. Find out.

Challenge 150, page 230: Large raindrops are pancakes with a massive border bulge. When the size increases, e.g. when a large drop falls through vapour, the drop splits, as the central membrane is then torn apart.

Challenge 151, page 230: It is a drawing; if it is interpreted as an image of a three-dimensional object, it either does not exist, or is not closed, or is an optical illusion of a torus.

Challenge 152, page 230: See T. Fink & Y. Mao, *The 85 Ways to Tie a Tie*, Broadway Books, 2000.

Challenge 153, page 231: See T. Clarke, Laces high, *Nature Science Update* 5th of December, 2002, or www.nature.com/nsu/021202/021202-4.html.

Challenge 156, page 232: In fact, nobody has even tried to do so yet. It may also be that the problem makes no sense.

Challenge 157, page 235: Most macroscopic matter properties fall in this class, such as the change of water density with temperature.

Challenge 160, page 244: Before the speculation can be fully tested, the relation between particles and black holes has to be clarified first.

Challenge 161, page 245: Never expect a correct solution for personal choices. Do what you yourself think and feel is correct.

Challenge 167, page 251: Planck limits can be exceeded for extensive observables for which many particle systems can exceed single particle limits, such as mass, momentum, energy or electrical resistance.

Challenge 169, page 254: Do not forget the relativistic time dilation.

Challenge 170, page 254: The formula with $n-1$ is a better fit. Why?

Challenge 174, page 259: No, only properties of parts of the universe are listed. The universe itself has no properties, as shown in the last volume..

Vol. VI, page 103

Challenge 173, page 259: The slowdown goes *quadratically* with time, because every new slowdown adds to the old one!

Challenge 175, page 261: The gauge coupling constants, via the Planck length, determine the size of atoms, the strength of chemical bonds and thus the size of all things.

Challenge 176, page 275: Covalent bonds tend to produce full shells; this is a smaller change on the right side of the periodic table.

Challenge 178, page 278: The metric is regular, positive definite and obeys the triangle inequality.

Challenge 182, page 280: 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 whether it is a rotation, a reflection, a dilation, a shear, or a stretch along two axes? What are the remaining possibilities?

Challenge 185, page 281: The (simplest) product of two functions is taken by point-by-point multiplication.

Challenge 186, page 281: The norm $\|f\|$ of a real function f is defined as the supremum of its absolute value:

$$\|f\| = \sup_{x \in \mathrm{R}} |f(x)| \ . \tag{154}$$

In simple terms: the maximum value taken by the absolute of the function is its norm. It is also called 'sup'-norm. Since it contains a supremum, this norm is only defined on the subspace of *bounded* continuous functions on a space X, or, if X is compact, on the space of all continuous functions (because a continuous function on a compact space must be bounded).

Challenge 189, page 284: Take out your head, then pull one side of your pullover over the corresponding arm, continue pulling it over the over arm; then pull the other side, under the first, to the other arm as well. Put your head back in. Your pullover (or your trousers) will be inside out.

Challenge 190, page 284: Both can be untied.

Challenge 194, page 289: The transformation from one manifold to another with different topology can be done with a tiny change, at a so-called *singular point*. Since nature shows a minimum action, such a tiny change cannot be avoided.

Challenge 195, page 291: The product $M^\dagger M$ is Hermitean, and has positive eigenvalues. Thus H is uniquely defined and Hermitean. U is unitary because $U^\dagger U$ is the unit matrix.

BIBLIOGRAPHY

> Gedanken sind nicht stets parat. Man schreibt auch, wenn man keine hat.*
>
> Wilhelm Busch, *Aphorismen und Reime*.

1 The use of radioactivity for breeding of new sorts of wheat, rice, cotton, roses, pineapple and many more is described by B. S. Ahloowalia & M. Maluszynski, Induced mutations – a new paradigm in plant breeding, *Euphytica* **11**, pp. 167–173, 2004. Cited on page 17.

2 The motorized screw used by viruses was described by A.A. Simpson & al., Structure of the bacteriophage phi29 DNA packaging motor, *Nature* **408**, pp. 745–750, 2000. Cited on page 20.

3 S. M. Block, Real engines of creation, *Nature* **386**, pp. 217–219, 1997. Cited on page 20.

4 Early results and ideas on molecular motors are summarised 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 page 20.

5 R. Bartussek & P. Hänggi, Brownsche Motoren, *Physikalische Blätter* **51**, pp. 506–507, 1995. Cited on page 21.

6 R. J. Cano & M. K. Borucki, Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber, *Science* **26**, pp. 1060–1064, 1995. Cited on page 24.

7 The first papers on bacteria from salt deposits were V. R. Ott & H. J. Dombrwoski, Mikrofossilien in den Mineralquellen zu Bad Nauheim, *Notizblatt des Hessischen Landesamtes für Bodenforschung* **87**, pp. 415–416, 1959, H. J. Dombrowski, Bacteria from Paleozoic salt deposits, *Annals of the New York Academy of Sciences* **108**, pp. 453–460, 1963. A recent confirmation is R. H. Vreeland, W. D. Rosenzweig & D. W. Powers, Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal, *Nature* **407**, pp. 897–899, 2000. Cited on page 24.

8 This is explained in D. Graur & T. Pupko, The permian bacterium that isn't, *Molecular Biology and Evolution* **18**, pp. 1143–1146, 2001, and also in M. B. Hebsgaard, M. J. Phillips & E. Willerslev, Geologically ancient DNA: fact or artefact?, *Trends in Microbiology* **13**, pp. 212–220, 2005. Cited on page 24.

9 Gabriele Walker, *Snowball Earth – The Story of the Great Global Catastrophe That Spawned Life as We Know It*, Crown Publishing, 2003. Cited on page 24.

* 'Thoughts are not always available. Many write even without them.'

10 The table and the evolutionary tree are taken from J. O. McInerney, M. Mullarkey, M. E. Wernecke & R. Powell, Bacteria and Archaea: Molecular techniques reveal astonishing diversity, *Biodiversity* **3**, pp. 3–10, 2002. The tree might still change a little in the coming years. Cited on page 24.

11 E. K. Costello, C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon & R. Knight, Bacterial community variation in human body habitats across space and time, *Science Express* 5 November 2009. Cited on page 26.

12 This is taken from the delightful children text Hans J. Press, *Spiel das Wissen schafft*, Ravensburger Buchverlag 1964, 2004. Cited on page 29.

13 The discovery of a specific taste for fat was published by F. Laugerette, P. Passilly-Degrace, B. Patris, I. Niot, M. Febbraio, J. P. Montmayeur & P. Besnard, CD36 involvement in orosensory detection of dietary lipids, spontaneous fat preference, and digestive secretions, *Journal of Clinical Investigation* **115**, pp. 3177–3184, 2005. Cited on page 29.

14 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 31.

The best experts are those who help others to overcome traumas. Peter A. Levine & Ann Frederick, *Waking the Tiger – Healing Trauma – The Innate Capacity to Transform Overwhelming Experiences*, North Atlantic Books, 1997. 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 realising this disentanglement.

The next step, namely full mastery in the enjoyment of life, can be found in any book written by somebody who has achieved mastery in any one topic. The topic itself is not important, only the passion is.

A. de la Garanderie, *Le dialogue pédagogique avec l'élève*, Centurion, 1984, A. de la Garanderie, *Pour une pédagogie de l'intelligence*, Centurion, 1990, A. de la Garanderie, *Réussir ça s'apprend*, Bayard, 1994. De la Garanderie explains how the results of teaching and learning depend in particular on the importance of evocation, imagination and motivation.

Plato, *Phaedrus*, Athens, 380 bce.

Françoise Dolto, *La cause des enfants*, Laffont, 1985, and her other books. Dolto (1908–1988), a child psychiatrist, is one of the world experts on the growth of the child; her main theme was that growth is only possible by giving the highest possible responsibility to every child during its development.

In the domain of art, many had the passion to achieve full pleasure. A good piece of music, a beautiful painting, an expressive statue or a good film 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 are the works 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. See also Robert Bringhurst, *The Elements of Typographic Style*, Hartley & Marks, 2004.

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.

15 The group of John Wearden in Manchester has shown by experiments with humans 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 32.

16 E. J. Zimmerman, The macroscopic nature of space-time, *American Journal of Physics* **30**, pp. 97–105, 1962. Cited on page 35.

17 The possibilities for precision timing using single-ion clocks are shown in W.H. Oskay & al., Single-atom clock with high accuracy, *Physical Review Letters* **97**, p. 020801, 2006. Cited on page 36.

18 See P.D. Pešić, The smallest clock, *European Journal of Physics* **14**, pp. 90–92, 1993. Cited on page 35.

19 D.J Morré & al., Biochemical basis for the biological clock, *Biochemistry* **41**, pp. 11941–11945, 2002. Cited on page 32.

20 The chemical clocks in our body are described in John D. Palmer, *The Living Clock*, Oxford University Press, 2002, or in A. Ahlgren & F. Halberg, *Cycles of Nature: An Introduction to Biological Rhythms*, National Science Teachers Association, 1990. See also the www.msi.umn.edu/~halberg/introd website. Cited on page 32.

21 An introduction to the sense of time as result of clocks in the brain is found in R. B. Ivry & R. Spencer, The neural representation of time, *Current Opinion in Neurobiology* **14**, pp. 225–232, 2004. The interval timer is explain in simple words in K. Wright, Times in our lives, *Scientific American* pp. 40–47, September 2002. The MRI research used is S. M. Rao, A. R. Mayer & D. L. Harrington, The evolution of brain activation during temporal processing, *Nature Neuroscience* **4**, pp. 317–323, 2001. Cited on page 34.

22 See, for example, Jan Hilgevoord, Time in quantum mechanics, *American Journal of Physics* **70**, pp. 301–306, 2002. Cited on page 34.

23 See for example ... Cited on page 37.

24 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 38.

25 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 38.

26 See, for example, R. Efron, The duration of the present, *Annals of the New York Academy of Sciences* **138**, pp. 713–729, 1967. Cited on page 38.

27 W. M. Itano, D. J. Heinzen, J. J. Bollinger & D. J. Wineland, Quantum Zeno effect, *Physical Review A* **41**, pp. 2295–2300, 1990. M. C. Fischer, B. Gutiérrez-Medina & M. G. Raizen, Observation of the Quantum Zeno and Anti-Zeno effects in an unstable system, *Physical Review Letters* **87**, p. 040402, 2001, also www-arxiv.org/abs/quant-ph/0104035. Cited on page 40.

28 See P. Facchi, Z. Hradil, G. Krenn, S. Pascazio & J. Řeháček Quantum Zeno tomography, *Physical Review A* **66**, p. 012110, 2002. Cited on page 40.

29 H. Kobayashi & S. Kohshima, Unique morphology of the human eye, *Nature* **387**, pp. 767–768, 1997. Cited on page 42.

30 James W. Prescott, Body pleasure and the origins of violence, *The Futurist* Bethseda, 1975, also available at www.violence.de/prescott/bullettin/article.html. Cited on page 42.

31 Felix Tretter & Margot Albus, *Einführung in die Psychopharmakotherapie – Grundlagen, Praxis, Anwendungen*, Thieme, 2004. Cited on page 43.

32 See the talk on these experiments by Helen Fisher on www.ted.org and the information on helenfisher.com. Cited on page 43.

33 See for example P. Pyykkö, Relativity, gold, closed-shell interactions, and CsAu.NH3, *Angewandte Chemie, International Edition* **41**, pp. 3573–3578, 2002, or L. J. Norrby, Why is mercury liquid? Or, why do relativistic effects not get into chemistry textbooks?, *Journal of Chemical Education* **68**, pp. 110–113, 1991. Cited on page 45.

34 On the internet, the 'spherical' periodic table is credited to Timothy Stowe; but there is no reference for that claim, except an obscure calendar from a small chemical company. The original table (containing a number errors) used to be found at the chemlab.pc.maricopa.edu/periodic/stowetable.html website; it is now best found by searching for images called 'stowetable' with any internet search engine. Cited on page 45.

35 For good figures of atomic orbitals, take any modern chemistry text. Or go to csi.chemie.tu-darmstadt.de/ak/immel/. Cited on page 46.

36 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 46.

37 L. Gagliardi & B. O. Roos, Quantum chemical calculations show that the uranium molecule U_2 has a quintuple bond, *Nature* **433**, pp. 848–851, 2005. B. O. Roos, A. C. Borin & L. Gagliardi, Reaching the maximum multiplicity of the covalent chemical bond, *Angewandte Chemie, International Edition* **46**, pp. 1469–1472, 2007. Cited on page 47.

38 H. -W. Fink & C. Escher, Zupfen am Lebensfaden – Experimente mit einzelnen DNS-Molekülen, *Physik in unserer Zeit* **38**, pp. 190–196, 2007. Cited on page 48.

39 L. Bindi, P. J. Steinhardt, N. Yai & P. J. Lu, Natural Quasicrystal, *Science* **324**, pp. 1306–1309, 2009. Cited on page 59.

40 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 61.

41 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 62.

42 V. F. Weisskopf, Of atoms, mountains and stars: a study in qualitative physics, *Science* **187**, p. 605, 1975. Cited on page 63.

43 K. Autumn, Y. Liang, T. Hsich, W. Zwaxh, W. P. Chan, T. Kenny, R. Fearing & R. J. Full, Adhesive force of a single gecko foot-hair, *Nature* **405**, pp. 681–685, 2000. Cited on page 64.

44 A. B. Kesel, A. Martin & T. Seidl, Getting a grip on spider attachment: an AFM approach to microstructure adhesion in arthropods, *Smart Materials and Structures* **13**, pp. 512–518, 2004. Cited on page 64.

45 R. Ritchie, M. J. Buehler & P. Hansma, Plasticity and toughness in bones, *Physics Today* **82**, pp. 41–47, 2009. This is a beautiful article. Cited on page 65.

46 H. Inaba, T. Saitou, K. -I. Tozaki & H. Hayashi, Effect of the magnetic field on the melting transition of H2O and D2O measured by a high resolution and supersensitive differential scanning calorimeter, *Journal of Applied Physics* **96**, pp. 6127–6132, 2004. Cited on page 65.

47 The graph for temperature is taken from J. Jouzel & al., Orbital and millennial Antarctic climate variability over the past 800,000 years, *Science* **317**, pp. 793–796, 2007. The graph for CO_2 is combined from six publications: D.M. Etheridge & al., Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn, *Journal of Geophysical Research* **101**, pp. 4115–4128, 1996, J.R. Petit & al., Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, *Nature* **399**, pp. 429–436, 1999, A. Indermühle & al., Atmospheric CO_2 concentration from 60 to 20 kyr BP from the Taylor Dome ic cores, Antarctica, *Geophysical Research Letters* **27**, pp. 735–738, 2000, E. Monnin & al., Atmospheric CO_2 concentration over the last glacial termination, *Science* **291**, pp. 112–114, 2001, U. Siegenthaler & al., Stable carbon cycle-climate relationship during the late pleistocene, *Science* **310**, pp. 1313–1317, 2005, D. Lüthi & al., High-resolution carbon dioxide concentration record 650,000-800,000 years before present, *Nature* **453**, pp. 379–382, 2008. Cited on page 65.

48 R. Hemley & al., presented these results at the International Conference on New Diamond Science in Tsukuba (Japan) 2005. See figures at www.carnegieinstitution.org/diamond-13may2005. Cited on page 66.

49 Ingo Dierking, *Textures of liquid crystals*, Wiley-VCH, 2003. Cited on page 67.

50 The *photon hall 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 67.

51 C. Strohm, G. L. J. A. Rikken & P. Wyder, Phenomenological evidence for the phonon Hall effect, *Physical Review Letters* **95**, p. 155901, 2005. Cited on page 67.

52 E. Chibowski, L. Hołysz, K. Terpiłowski, Effect of magnetic field on deposition and adhesion of calcium carbonate particles on different substrates, *Journal of Adhesion Science and Technology* **17**, pp. 2005–2021, 2003, and references therein. Cited on page 68.

53 K. S. Novoselov, D. Jiang, F. Schedin, T. Booth, V. V. Khotkevich, S. V. Morozov & A. K. Geim, Two Dimensional Atomic Crystals, *Proceedings of the National Academy of Sciences* **102**, pp. 10451–10453, 2005. See also the talk by Andre Geim, QED in a pencil trace, to be downloaded at online.kitp.ucsb.edu/online/graphene_m07/geim/ or A. C. Neto, F. Guinea & N. M. Peres, Drawing conclusions from graphene, *Physics World* November 2006. Cited on pages 68 and 330.

54 The first paper was J. Bardeen, L. N. Cooper & J. R. Schrieffer, Microscopic Theory of Superconductivity, *Physical Review* **106**, pp. 162–164, 1957. The full, so-called BCS theory, is then given in the masterly paper J. Bardeen, L. N. Cooper & J. R. Schrieffer, Theory of Superconductivity, *Physical Review* **108**, pp. 1175–1204, 1957. Cited on page 70.

55 A. P. Finne, T. Araki, R. Blaauwgeers, V. B. Eltsov, N. B. Kopnin, M. Krusius, L. Skrbek, M. Tsubota & G. E. Volovik, An intrinsic velocity-independent criterion for superfluid turbulence, *Nature* **424**, pp. 1022–1025, 2003. Cited on page 71.

56 The original prediction, which earned Laughlin a Nobel Prize in 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, 1983. Cited on pages 71 and 72.

57 The first experimental evidence is by V. J. Goldman & B. Su, Resonance tunnelling 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 arxiv.org/abs/cond-mat/9706307. Cited on page 71.

58 The discovery of the original quantum Hall effect, which earned von Klitzing the Nobel Prize in Physics, was published as von Klitzing, G. Dorda & M. Pepper, New method for high-accuracy determination of the fine-structure constant based on quantised Hall resistance, *Physical Review Letters* **45**, pp. 494–497, 1980. Cited on page 71.

59 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. Tsui and Stormer won the Nobel Prize in Physics with this paper. Cited on page 72.

60 K. S. Novoselov, Z. Jiang, Y. Zhang, S. V. Morozov, H. L. Stormer, U. Zeitler, J. C. Mann, G. S. Boebinger, P. Kim & A. K. Geim, Room-temperature quantum Hall effect in graphene, *Science* **315**, p. 1379, 2007. Cited on page 72.

61 See for example the overview by M. Born & T. Jüstel, Umweltfreundliche Lichtquellen, *Physik Journal* **2**, pp. 43–49, 2003. Cited on page 73.

62 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 78.

63 J. Kasparian, Les lasers femtosecondes : applications atmosph ériques, *La Recherche* pp. RE14 1–7, February 2004/ See also the www.teramobile.org website. Cited on page 78.

64 A. Takita, H. Yamamoto, Y. Hayasaki, N. Nishida & H. Misawa, Three-dimensional optical memory using a human fingernail, *Optics Express* **13**, pp. 4560–4567, 2005. They used a Ti:sapphire oscillator and a Ti:sapphire amplifier. Cited on page 78.

65 See the famous, beautiful but difficult textbook P. A. M. Dirac, *The Principles of Quantum Mechanics*, Clarendon Press, 1930, page 9. Cited on page 78.

66 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* **58**, pp. 209–231, 1986. However, in the bulk of the article the author in practice retracts the statement, e.g. on page 221. Cited on page 78.

67 G. Magyar & L. Mandel, Interference fringes produced by superposition of two independent maser light beams, *Nature* **198**, pp. 255–256, 1963. Cited on page 78.

68 L. Cser, Gy. Török, G. Krexner, I. Sharkov & B. Faragó, Holographic imaging of atoms using thermal neutrons, *Physical Review Letters* **89**, p. 175504, 2002. Cited on page 79.

69 W. Guo & H. Maris, Observations of the motion of single electrons in liquid helium, *Journal of Low Temperature Physics* **148**, pp. 199–206, 2007. Cited on page 80.

70 The story is found on many places in the internet. Cited on page 81.

71 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 82.

72 P. Ball, Popular physics myth is all at sea – does the ghostly Casimir effect really cause ships to attract each other?, online at www.nature.com/news/2006/060501/full/news060501-7.html. Cited on pages 82 and 84.

73 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 83.

74 H. B. G. Casimir, On the attraction between two perfectly conducting bodies, *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen* **51**, pp. 793–795, 1948. Cited on page 83.

75 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 83.

76 B. W. Derjaguin, I. I. Abrikosova & E. M. Lifshitz, Direct measurement of molecular attraction between solids separated by a narrow gap, *Quaterly Review of the Chemical Society* (London) **10**, pp. 295–329, 1956. Cited on page 84.

77 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 84.

78 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 84.

79 A. Lambrecht, Das Vakuum kommt zu Kräften, *Physik in unserer Zeit* **36**, pp. 85–91, 2005. Cited on page 84.

80 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 84.

81 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 84.

82 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 88.

83 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 89.

84 G. Gour & L. Sriramkumar, Will small particles exhibit Brownian motion in the quantum vacuum?, arxiv.org/abs/quant/phys/9808032 Cited on page 89.

85 P. A. M. Dirac, The requirements of fundamental physical theory, *European Journal of Physics* 5, pp. 65–67, 1984. 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 talk. Cited on page 90.

86 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 page 91.

87 Thomas Stoehlker & al., *Physical Review Letters* 9 October 2000. Cited on page 91.

88 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 91.

89 See the website by Tom Hales on 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 92.

90 For a short overview, also covering binary systems, see C. Bechinger, von Grünberg & P. Leiderer, Entropische Kräfte - warum sich repulsiv wechselwirkende Teilchen anziehen können, *Physikalische Blätter* **55**, pp. 53–56, 1999. Cited on page 92.

91 H. -W. Fink & G. Ehrlich, Direct observation of three-body interactions in adsorbed layers: Re on W(110), *Physical Review Letters* **52**, pp. 1532–1534, 1984, and H. -W. Fink & G. Ehrlich, Pair and trio interactions between adatoms: Re on W(110), *Journal of Chemical Physics* **61**, pp. 4657–4665, 1984. Cited on page 92.

92 See for example the textbook by A. Pimpinelli & J. Villain, *Physics of Crystal Growth*, Cambridge University Press, 1998. Cited on page 93.

93 Y. Furukawa, Faszination der Schneekristalle - wie ihre bezaubernden Formen entstehen, *Chemie in unserer Zeit* **31**, pp. 58–65, 1997. His www.lowtem.hokudai.ac.jp/~frkw/index_e.html website gives more information. Cited on page 93.

94 de Gennes, Soft matter, *Reviews of Modern Physics* **63**, p. 645, 1992. Cited on page 93.

95 The fluid helium based quantum gyroscope was first used to measure the rotation of the Earth by the French research group O. Avenel & E. Varoquaux, Detection of the Earth rotation with a superfluid double-hole resonator, *Czechoslovak Journal of Physics (Supplement S6)* **48**, pp. 3319–3320, 1996. It was later repeated by the Berkeley group, K. Schwab, N. Bruckner & R. E. Packard, Detection of the Earth's rotation using superfluid phase coherence, *Nature* **386**, pp. 585–587, 1997. For an review, see O. Avenel, Yu. Mukharsky & E. Varoquaux, Superfluid gyrometers, *Journal of Low Temperature Physics* **135**, p. 745, 2004. Cited on page 93.

96 K. Schwab, E. A. Henriksen, J. M. Worlock & M. L. Roukes, Measurement of the quantum of thermal conductance, *Nature* **404**, p. 974, 2000. For the optical analog of the experiment, see the beautiful paper 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 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 quantised conductance of a point contact, *Physica B* **175**, pp. 149–152, 1991. For more details on this experiment, see the volume *Light, Charges and Brains* of the Motion Mountain series. Cited on page 93.

97 The present knowledge on how humans move is summarised by ... Cited on page 97.

98 See e.g. D. A. Bryant, Electron acceleration in the aurora, *Contemporary Physics* 35, pp. 165–179, 1994. Cited on page 97.

99 The discharges above clouds are described e.g. by ... Cited on page 97.

100 About Coulomb explosions, see ... Cited on page 97.

101 See e.g. L. O'C. Drury, Acceleration of cosmic rays, *Contemporary Physics* 35, pp. 231–242, 1994. Cited on page 97.

102 See Dirk Kreimer, New mathematical structures in renormalizable quantum field theories, arxiv.org/abs/hep-th/0211136 or *Annals of Physics* **303**, pp. 179–202, 2003, and the erratum ibid., **305**, p. 79, 2003. Cited on page 97.

103 S. Fray, C. Alvarez Diez, T. W. Hänsch & Martin Weitz, Atomic interferometer with amplitude gratings of light and its applications to atom based tests of the equivalence principle, *Physical Review Letters* **93**, p. 240404, 2004. Cited on page 99.

104 V. V. Nesvizhevsky & al., Quantum states of neutrons in the Earth's gravitational field, *Nature* **415**, p. 297, 17 January 2002. See also the longer paper V. V. Nesvizhevsky & al., Measurement of quantum states of neutrons in the Earth's gravitational field, *Physical Review D* **67**, p. 102002, 2003, preprint at arxiv.org/abs/hep-th/0306198. Cited on page 99.

105 H. Rauch, W. Treimer & U. Bonse, Test of a single crystal neutron interferometer, *Physics Letters A* **47**, pp. 369–371, 1974. Cited on page 101.

106 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 101.

107 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 101.

108 J. D. Bekenstein, Black holes and entropy, *Physical Review D* **7**, pp. 2333–2346, 1973. Cited on page 102.

109 R. Bousso, The holographic principle, *Review of Modern Physics* **74**, pp. 825–874, 2002, also available as arxiv.org/abs/hep-th/0203101. The paper is an excellent review; however, it has some argumentation errors, as explained on page 102. Cited on page 102.

110 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 103.

111 S. A. Fulling, Nonuniqueness of canonical field quantization in Riemannian spacetime, *Physical Review D* **7**, pp. 2850–2862, 1973, P. Davies/ Scalar particle production in Schwarzchild and Rindler metrics, *Journal of Physics A: General Physics* **8**, pp. 609–616, 1975, W. G. Unruh, Notes on black hole evaporation, *Physical Review D* **14**, pp. 870–892, 1976. Cited on page 103.

112 About the possibility to measure Fulling–Davies–Unruh radiation directly, 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. arxiv.org/abs/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 103.

113 E. T. Akhmedov & D. Singleton, On the relation between Unruh and Sokolov–Ternov effects, preprint at www.arxiv.org/abs/hep-ph/0610394; see also E. T. Akhmedov & D. Singleton, On the physical meaning of the Unruh effect, preprint at www.arxiv.

org/abs/0705.2525. Unfortunately, as Bell and Leinaas mention, the Sokolov–Ternov effect cannot be used to confirm the expression of the black hole entropy, because the radiation is not thermal, and so a temperature of the radiation is hard to define. Nevertheless, C. Schiller predicts (October 2009) that careful analysis of the spectrum could be used (1) to check the proportionality of entropy and area in black holes, and (2) to check the transformation of temperature in special relativity. Cited on page 104.

114 W. G. Unruh & R. M. Wald, Acceleration radiation and the generalised second law of thermodynamics, *Physical Review D* **25**, pp. 942–958 , 1982. Cited on page 104.

115 R. M. Wald, The thermodynamics of black holes, *Living Reviews of Relativity* 2001, www-livingreviews.org/lrr-2001-6. Cited on page 106.

116 For example, if neutrinos were massless, they would be emitted by black holes more frequently than photons. 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 106 and 111.

117 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, arxiv.org/abs/gr-qc/9712010. Cited on page 107.

118 Optical black holes ... Cited on page 107.

119 T. Damour & R. Ruffini, Quantum electrodynamical effects in Kerr–Newman geometries, *Physical Review Letters* **35**, pp. 463–466, 1975. Cited on page 109.

120 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 108.

121 An excellent 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 109.

122 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, arxiv.org/abs/astro-ph/0001425. Cited on page 110.

123 arxiv.org/abs/astro-ph/0301262, arxiv.org/abs/astro-ph/0303539, D.W. Fox & al., Early optical emission from the γ-ray burst of 4 October 2002, *Nature* **422**, pp. 284–286, 2003. Cited on page 109.

124 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 111.

125 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 111.

126 About the membrane description of black holes, see KIP S. THORNE, RICHARD H. PRICE & DOUGLAS A. MACDONALD, editors, *Black Holes: the Membrane Paradigm*, Yale University Press, 1986. Cited on page 111.

127 Page 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, which is based on his earlier papers, such as Information in black hole radiation, *Physical Review Letters* **71**, pp. 3743–3746, 1993. See also his preprint at arxiv.org/abs/hep-th/9305040. Cited on page 112.

128 See DON N. PAGE, Average entropy of a subsystem, *Physical Review Letters* **71**, pp. 1291–1294, 1993. The entropy formula of this 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. Cited on page 113.

129 R. LAFRANCE & R. C. MYERS, Gravity's rainbow: limits for the applicability of the equivalence principle, *Physical Review D* **51**, pp. 2584–2590, 1995, arxiv.org/abs/hep-th/9411018. Cited on page 113.

130 M. YU. KUCHIEV & V. V. FLAMBAUM, Scattering of scalar particles by a black hole, arxiv.org/abs/gr-qc/0312065. See also M. YU. KUCHIEV & V. V. FLAMBAUM, Reflection on event horizon and escape of particles from confinement inside black holes, arxiv.org/abs/gr-qc/0407077. Cited on page 114.

131 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 114.

132 I speculate that this version of the coincidences could be original; I have not found it in the literature. Cited on page 114.

133 STEVEN WEINBERG, *Gravitation and Cosmology*, Wiley, 1972. See equation 16.4.3 on page 619 and also page 620. Cited on page 114.

134 It could be that knot theory provides a relation between a local knot invariant, related to particles, and a global one. Cited on page 115.

135 This point was made repeatedly by STEVEN WEINBERG, namely in Derivation of gauge invariance and the equivalence principle from Lorentz invariance of the S-matrix , *Physics Letters* **9**, pp. 357–359, 1964, in Photons and gravitons in S-matrix theory: derivation of charge conservation and equality of gravitational and inertial mass, *Physical Review B* **135**, pp. 1049–1056, 1964, and in Photons and gravitons in perturbation theory: derivation of Maxwell's and Einstein's equations, *Physical Review B* **138**, pp. 988–1002, 1965. Cited on page 115.

136 E. JOOS, Why do we observe a classical space-time?, *Physics Letters A* **116**, pp. 6–8, 1986. Cited on page 116.

137 L. H. FORD & T. A. ROMAN, Quantum field theory constrains traversable wormhole geometries, arxiv.org/abs/gr-qc/9510071 or *Physical Review D* **53**, pp. 5496–5507, 1996. Cited on page 117.

138 B. S. KAY, M. RADZIKOWSKI & R. M. WALD, Quantum field theory on spacetimes with a compactly generated Cauchy horizon, arxiv.org/abs/gr-qc/9603012 or *Communications in Mathematical Physics 183* pp. 533–556, 1997. Cited on page 117.

139 M. J. PFENNING & L. H. FORD, The unphysical nature of 'warp drive', *Classical and Quantum Gravity* **14**, pp. 1743–1751, 1997. Cited on page 117.

140 A excellent technical introduction to nuclear physics is BOGDAN POVH, KLAUS RITH, CHRISTOPH SCHOLZ & FRANK ZETSCHE, *Teilchen und Kerne*, Springer, 5th edition, 1999. It is also available in English translation. Cited on page 119.

141 For MRI movies of the heart beat, search for "cardiac MRI" on the internet. See for example www.youtube.com/watch?v=58l6oFhfZU. Cited on page 120.

142 W.C.M. WEIJMAR SCHULTZ & al., Magnetic resonance imaging of male and female genitals during coitus and female sexual arousal, *British Medical Journal* **319**, pp. 1596–1600, December 18, 1999, available online as www.bmj.com/cgi/content/full/319/7225/1596. Cited on page 121.

143 M. CHANTELL, T. C. WEEKES, X. SARAZIN & M. URBAN, Antimatter and the moon, *Nature* **367**, p. 25, 1994. M. AMENOMORI & al., Cosmic ray shadow by the moon observed with the Tibet air shower array, Proceedings of the 23rd International Cosmic Ray Conference, Calgary **4**, pp. 351–354, 1993. M. URBAN & al., *Nuclear Physics, Proceedings Supplement* **14B**, pp. 223–236, 1990. Cited on page 130.

144 An excellent summary on radiometric dating is by R. WIENS, Radiometric dating – a christian perpective, www.asa3.org/ASA/resources/Wiens.html. The absurd title is due to the habit in many religious circles to put into question radiometric dating results. Putting apart the extremely few religious statements in the review, the content is well explained. Cited on page 135.

145 A good overview is given by A. N. HALLIDAY, Radioactivity, the discovery of time and the earliest history of the Earth, *Contemporary Physics* **38**, pp. 103–114, 1997. Cited on page 137.

146 J. DUDEK, A. GOD, N. SCHUNCK & M. MIKIEWICZ, Nuclear tetrahedral symmetry: possibly present throughout the periodic table, *Physical Review Letters* **88**, p. 252502, 24 June 2002. Cited on page 138.

147 A good introduction is R. CLARK & B. WODSWORTH, A new spin on nuclei, *Physics World* pp. 25–28, July 1998. Cited on page 139.

148 JOHN HORGAN, *The End of Science – Facing the Limits of Knowledge in the Twilight of the Scientific Age*, Broadway Books, 1997, chapter 3, note 1. Cited on page 142.

149 M. BRUNETTI, S. CECCHINI, M. GALLI, G. GIOVANNINI & A. PAGLIARIN, Gamma-ray bursts of atmospheric origin in the MeV energy range, *Geophysical Research Letters* **27**, p. 1599, 1 June 2000. Cited on page 143.

150 See the interview A conversation with Peter Mansfield, *Europhysics Letters* **37**, p. 26, 2006. Cited on page 146.

151 A book with nuclear explosion photographs is MICHAEL LIGHT, *100 Suns*, Jonathan Cape, 2003. Cited on page 145.

152 An older but still fascinating summary of solar physics is R. KIPPENHAHN, *Hundert Milliarden Sonnen*, Piper, 1980. It was a famous bestseller and is available also in English translation. Cited on page 148.

153 H. BETHE, On the formation of deuterons by proton combination, *Physical Review* **54**, pp. 862–862, 1938, and H. BETHE, Energy production in stars, *Physical Review* **55**, pp. 434–456, 1939. Cited on page 148.

154 M. NAUENBERG & V. F. WEISSKOPF, Why does the sun shine?, *American Journal of Physics* **46**, pp. 23–31, 1978. Cited on page 149.

155 The slowness of the speed of light inside stars is due to the frequent scattering of photons by the star matter. The most common estimate for the Sun is an escape time of 40 000 to 1 million years, but estimates between 17 000 years and 50 million years can be found in the literature. Cited on page 150.

156 See the freely downloadable book by JOHN WESSON, *The Science of JET - The Achievements of the Scientists and Engineers Who Worked on the Joint European Torus 1973-1999*, JET Joint

Undertaking, 2000, available at www.jet.edfa.org/documents/wesson/wesson.html. Cited on page 152.

157 J. D. Lawson, Some criteria for a power producing thermonuclear reactor, *Proceedings of the Physical Society, London* B **70**, pp. 6–10, 1957. The paper had been kept secret for two years. Cited on page 153.

158 R. A. Alpher, H. Bethe & G. Gamow, The Origin of Chemical Elements, *Physical Review* 73, p. 803-804, 1948. Cited on page 154.

159 Kendall, Friedman and Taylor received the 1990 Nobel Prize for Physics for a series of experiments they conducted in the years 1967 to 1973. The story is told in the three Nobel lectures R. E. Taylor, Deep inelastic scattering: the early years, *Review of Modern Physics* **63**, pp. 573–596, 1991, H. W. Kendall, Deep inelastic scattering: Experiments on the proton and the observation of scaling, *Review of Modern Physics* **63**, pp. 597–614, 1991, and J. I. Friedman, Deep inelastic scattering: Comparisons with the quark model, *Review of Modern Physics* **63**, pp. 615–620, 1991. Cited on page 157.

160 G. Zweig, An SU3 model for strong interaction symmetry and its breaking II, CERN Report No. 8419TH. 412, February 21, 1964. Cited on page 157.

161 About the strange genius of Gell-Mann, see the beautiful book by George Johnson, *Murray Gell-Mann and the Revolution in Twentieth-Century Physics*, Knopf, 1999. Cited on page 157.

162 The best introduction might be the wonderfully clear text by Donald H. Perkins, *Introduction to High Energy Physics*, Cambridge University Press, fourth edition, 2008. Also beautiful, with more emphasis on the history and more detail, is Kurt Gottfried & Victor F. Weisskopf, *Concepts of Particle Physics*, Clarendon Press, Oxford, 1984. Victor Weisskopf was one of the heroes of the field, both in theoretical research and in the management of CERN, the European organization for particle research. Cited on pages 159, 168, and 318.

163 The official reference for all particle data, worth a look for every physicist, is the massive collection of information compiled by the Particle Data Group, with the website pdg.web.cern.ch containing the most recent information. A printed review is published about every two years in one of the major journals on elementary particle physics. See for example C. Amsler & al., The Review of Particle Physics, *Physics Letters B* **667**, p. 1, 2008. For some measured properties of these particles, the official reference is the set of so-called CODATA values, found at physics.nist.gov/cuu/Constants/. The most recent list was published on paper as P. J. Mohr, B. N. Taylor & D. B. Newell, CODATA recommended values of the fundamental physical constants: 2006, *Reviews of Modern Physics* **80**, pp. 633–730, 2008. It is available also on the NIST site. Cited on pages 160, 169, 171, 172, 178, 182, 186, and 192.

164 H. Fritsch, M. Gell-Mann & H. Leutwyler, Advantages of the color octet picture, *Physics Letters B* 47, pp. 365–368, 1973. Cited on page 160.

165 Quantum chromodynamics can be explored in books of several levels. The first, the popular one, is described most clearly by one of its founders, in the longseller Harald Fritzsch, *Quarks – Urstoff unserer Welt*, Piper Verlag, 2006, or, in English language, *Quarks: the stuff of matter*, Penguin Books, 1983. In the second level, the field can be explored in texts on high energy physics, as those of Ref. 162. The third level contains books like Kerson Huang, Quarks, Leptons and Gauge Fields, World Scientific, 1992, or Felix J. Ynduráin, The Theory of Quark and Gluon Interactions, Springer Verlag, 1992, or Walter Greiner & Andreas Schäfer, Quantum Chromodynamics, Springer Verlag, 1995, or the modern and detailed text by Stephan Narison, *QCD as a Theory of Hadrons*, Cambridge University Press, 2004. As always, a student has to discover by himself or herself which text he or she likes most. Cited on pages 162, 167, and 169.

166 S. Dürr & al., Ab initio determination of the light hadron masses, *Science* **322**, pp. 1224–1227, 2008. Cited on page 165.

167 See for example C. Bernard & al., Light hadron spectrum with Kogut–Susskind quarks, *Nuclear Physics, Proceedings Supplement* 73, p. 198, 1999, and references therein. Cited on page 165.

168 R. Brandelik & al., Evidence for planar events in e^+e^- annihilation at high energies, *Physics Letters B* **86**, pp. 243–249, 1979. Cited on page 166.

169 For a pedagogical introduction to lattice QCD calculations, see R. Gupta, Introduction to Lattice QCD, preprint at arxiv.org/abs/hep-lat/9807028, or the clear introduction by Michael Creutz, *Quarks, Gluons and Lattices*, Cambridge University Press, 1983. Cited on page 168.

170 S. Bethke, Experimental tests of asymptotic freedom, *Progress in Particle and Nuclear Physics* 58, pp. 351–368, 2007, preprint at arxiv.org/abs/hep-ex/0606035. Cited on page 169.

171 F. Abe & al., Measurement of dijet angular distributions by the collider detector at Fermilab, *Physical Review Letters* 77, pp. 5336–5341, 1996. Cited on page 170.

172 The approximation of QCD with zero mass quarks is described by F. Wilczek, Getting its from bits, *Nature* **397**, pp. 303–306, 1999. It is also explained in F. Wilczek, Asymptotic freedom, *Nobel lecture* 2004. The proton's mass is set by the energy scale at which the strong coupling, coming from its value at Planck energy, becomes of order unity. Cited on pages 171 and 188.

173 A. J. Buchmann & E. M. Henley, Intrinsic quadrupole moment of the nucleon, *Physical Review C* **63**, p. 015202, 2000. Alfons Buchmann also predicts that the quadrupole moment of the other, strange $J = 1/2$ octet baryons is positive, and predicts a prolate structure for all of them (private communication). For the decuplet baryons, with $J = 3/2$, the quadrupole moment can often be measured spectroscopically, and is always negative. The four Δ baryons are thus predicted to have a negative intrinsic quadrupole moment and thus an oblate shape. This explained in A. J. Buchmann & E. M. Henley, Quadrupole moments of baryons, *Physical Review D* 65, p. 073017, 2002. For recent updates, see A. J. Buchmann, Charge form factors and nucleon shape, pp. 110–125, in the Shape of Hadrons Workshop Conference, Athens Greece, 27-29 April 2006, AIP Conference Proceedings 904, Eds. C.N. Papanicolas, Aron Bernstein. For updates on other baryons, see A. J. Buchmann, Structure of strange baryons, Hyp 2006, International Conference on Hypernuclear and Strange Particle Physics, Oct.10-14, Mainz, Germany, to be published in European Physics Journal A 2007. The topic is an active field of research; for example, results on magnetic octupole moments are expected soon. Cited on page 171.

174 S. Strauch & al., Polarization transfer in the ^{4}He (e,e'p) ^{3}H reaction up to $Q^2 = 2.6$(GeV/c)^2, *Physical Review Letters* **91**, p. 052301, 2003. Cited on page 171.

175 F. Close, Glueballs and hybrids: new states of matter, *Contemporary Physics* **38**, pp. 1–12, 1997. See also the next reference. Cited on page 173.

176 A tetraquark is thought to be the best explanation for the $f_0(980)$ resonance at 980 MeV. The original proposal of this explanation is due to R. L. Jaffe, Multiquark hadrons I: phenomenology of $Q^2\overline{Q}^2$ mesons, *Physical Review D* **15**, pp. 267–280, 1977, R. L. Jaffe, Multiquark hadrons II: methods, *Physical Review D* **15**, pp. 281–289, 1977, and R. L. Jaffe, *Physical Review D* $Q^2\overline{Q}^2$ resonances in the baryon-antibaryon system,**17**, pp. 1444–1458,

1978. For a clear and detailed modern summary, see the excellent review by E. Klempt & A. Zaitsev, Glueballs, hybrids, multiquarks: experimental facts versus QCD inspired concepts, *Physics Reports* **454**, pp. 1–202, 2007, preprint at arxiv.org/abs/0708.4016. See also F. Giacosa, Light scalars as tetraquarks, preprint at arxiv.org/abs/0711.3126, and V. Crede & C. A. Meyer, The experimental status of glueballs, preprint at arxiv.org/abs/0812.0600. Cited on page 173.

177 Pentaquarks were first predicted by Maxim Polyakov, Dmitri Diakonov, and Victor Petrov in 1997. Two experimental groups in 2003 claimed to confirm their existence, with a mass of 1540 MeV; see K. Hicks, An experimental review of the Θ^+ pentaquark, arxiv.org/abs/hep-ex/0412048. Results from 2005, however, seem to rule out that the 1540 MeV particle is a pentaquark. Cited on page 173.

178 See, for example, Ya. B. Zel'dovich & V. S. Popov, Electronic structure of superheavy atoms, *Soviet Physics Uspekhi* **17**, pp. 673–694, 2004 in the English translation. Cited on page 173.

179 J. Tran Thanh Van, ed., *CP violation in Particle Physics and Astrophysics*, Proc. Conf. Chateau de Bois, France, May 1989, Editions Frontières, 1990. Cited on page 178.

180 P.L. Anthony & al., Observation of parity nonconservation in Møller scattering, *Physical Review Letters* **92**, p. 181602, 2004. Cited on page 180.

181 M. A. Bouchiat & C. C. Bouchiat, Weak neutral currents in atomic physics, *Physics Letters B* **48**, pp. 111–114, 1974. U. Amaldi, A. Böhm, L. S. Durkin, P. Langacker, A. K. Mann, W. J. Marciano, A. Sirlin & H. H. Williams, Comprehensive analysis of data pertaining to the weak neutral current and the intermediate-vector-boson masses, *Physical Review D* **36**, pp. 1385–1407, 1987. Cited on page 180.

182 M. C. Noecker, B. P. Masterson & C. E. Wiemann, Precision measurement of parity nonconservation in atomic cesium: a low-energy test of electroweak theory, *Physical Review Letters* **61**, pp. 310–313, 1988. See also D.M. Meekhof & al., High-precision measurement of parity nonconserving optical rotation in atomic lead, *Physical Review Letters* **71**, pp. 3442–3445, 1993. Cited on page 180.

183 S. C. Bennet & C. E. Wiemann, Measurement of the 6S – 7S transition polarizability in atomic cesium and an improved test of the standard model, *Physical Review Letters* **82**, pp. 2484–2487, 1999. The group has also measured the spatial distribution of the weak charge, the so-called the *anapole moment*; see C.S. Wood & al., Measurement of parity nonconservation and an anapole moment in cesium, *Science* **275**, pp. 1759–1763, 1997. Cited on page 180.

184 C. Jarlskog, Commutator of the quark mass matrices in the standard electroweak model and a measure of maximal CP nonconservation, *Physical Review Letters* **55**, pp. 1039–1042, 1985. Cited on page 182.

185 K. Grotz & H. V. Klapdor, *Die schwache Wechselwirkung in Kern-, Teilchen- und Astrophysik*, Teubner Verlag, Stuttgart, 1989. Also available in English and in several other languages. Cited on page 186.

186 P. W. Higgs, Broken symmetries, massless particles and gauge fields, *Physics Letters* **12**, pp. 132–133, 1964. He then expanded the story in P. W. Higgs, Spontaneous symmetry breakdown without massless bosons, *Physical Review* **145**, p. 1156-1163, 1966. Higgs gives most credit to Anderson, instead of to himself; he also mentions Brout and Englert, Guralnik, Hagen, Kibble and 't Hooft. Cited on page 186.

187 D. Treille, Particle physics from the Earth and from the sky: Part II, *Europhysics News* **35**, no. 4, 2004. Cited on page 187.

188 Rumination is studied in P. Jordan & de Laer Kronig, in *Nature* **120**, p. 807, 1927. Cited on page 189.

189 K.W.D. Ledingham & al., Photonuclear physics when a multiterawatt laser pulse interacts with solid targets, *Physical Review Letters* **84**, pp. 899–902, 2000. K.W.D. Ledingham & al., Laser-driven photo-transmutation of Iodine-129 – a long lived nuclear waste product, *Journal of Physics D: Applied Physics* **36**, pp. L79–L82, 2003. R. P. Singhal, K. W. D. Ledingham & P. McKenna, Nuclear physics with ultra-intense lasers – present status and future prospects, *Recent Research Developments in Nuclear Physics* **1**, pp. 147–169, 2004. Cited on page 190.

190 The electron radius limit is deduced from the $g-2$ measurements, as explained in the Nobel Prize talk by Hans Dehmelt, Experiments with an isolated subatomic particle at rest, *Reviews of Modern Physics* **62**, pp. 525–530, 1990, or in Hans Dehmelt, Is the electron a composite particle?, *Hyperfine Interactions* **81**, pp. 1–3, 1993. Cited on page 194.

191 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 194.

192 For the bibliographic details of the latest print version of the *Review of Particle Physics*, see Appendix B. The online version can be found at pdg.web.cern.ch. The present status on grand unification can also be found in the respective section of the overview. Cited on page 196.

193 Slides of a very personal review talk by H. Georgi, The future of grand unification, can be found at www2.yukawa.kyoto-u.ac.jp/~yt100sym/files/yt100sym_georgi.pdf. A modern research approach is S. Raby, Grand unified theories, arxiv.org/abs/hep-ph/0608183. Cited on page 196.

194 H. Jeon & M. Longo, Search for magnetic monopoles trapped in matter, *Physical Review Letters* 75, pp. 1443–1447, 1995. Cited on page 197.

195 On proton decay rates, see the data of the Particle Data Group, at pdg.web.cern.ch. Cited on page 198.

196 U. Amaldi, W. de Boer & H. Fürstenau, Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP, *Physics Letters* **260**, pp. 447–455, 1991. This widely cited paper is the standard reference for this issue. An update is found in W. de Boer & C. Sander, Global electroweak fits and gauge coupling unification, *Physics Letters B 585* pp. 276–286, 2004, or preprint at arxiv.org/abs/hep-ph/0307049. The figure is taken with permission from the home page of Wim de Boer at www-ekp.physik.uni-karlsruhe.de/~deboer/html/Forschung/forschung.html. Cited on page 198.

197 Peter G. O. Freund, *Introduction to Supersymmetry*, Cambridge 1988. Julius Wess & Jonathan Bagger, *Supersymmetry and Supergravity*, Princeton University Press, 1992. This widely cited book contains a lot of mathematics but little physics. Cited on page 199.

198 P. C. Argyres, Dualities in supersymmetric field theories, *Nuclear Physics Proceedings Supplement* **61A**, pp. 149–157, 1998, preprint available at arxiv.org/abs/hep-th/9705076. Cited on page 201.

199 Michael Stone editor, *Bosonization*, World Scientific, 1994. R. Rajaraman, *Solitons and Instantons*, North Holland, 1987. Cited on page 203.

200 In 1997, the smallest human-made flying object was the helicopter built by a group of the Institut für Mikrotechnik in Mainz, in Germany. A picture is available at their web page, to be found at www.imm-mainz.de/English/billboard/f_hubi.html. The helicopter is 24 mm long, weighs 400 mg and flies (though not freely) using two built-in electric motors driving two rotors, running at between 40 000 and 100 000 revolutions per minute. See also the helicopter from Stanford University at www-rpl.stanford.edu/RPL/htmls/mesoscopic/mesicopter/mesicopter.html, with an explanation of its battery problems. Cited on page 205.

201 Henk Tennekes, *De wetten van de vliegkunst – over stijgen, dalen, vliegen en zweven*, Aramith Uitgevers, 1992. This clear and interesting text is also available in English. Cited on page 205.

202 The most recent computational models of lift still describe only two-dimensional wing motion, e.g. Z. J. Wang, Two dimensional mechanism for insect hovering, *Physical Review Letters 85* pp. 2216–2219, 2000. A first example of a mechanical bird has been constructed by Wolfgang Send; it can be studied on the www.aniprop.de website. See also W. Send, Physik des Fliegens, *Physikalische Blätter* 57, pp. 51–58, June 2001. Cited on page 206.

203 R. B. Srygley & A. L. R. Thomas, Unconventional lift-generating mechanisms in free-flying butterflies, *Nature* **420**, pp. 660–664, 2002. Cited on page 206.

204 The book by John Brackenbury, *Insects in Flight*, 1992. is a wonderful introduction into the biomechanics of insects, combining interesting science and beautiful photographs. Cited on page 208.

205 The simulation of insect flight using enlarged wing models flapping in oil instead of air is described for example in www.dickinson.caltech.edu/research_robofly.html. The higher viscosity of oil allows to achieve the same Reynolds number with larger sizes and lower frequencies than in air. Cited on page 208.

206 E. Purcell, Life at low Reynolds number, *American Journal of Physics* **45**, p. 3, 1977. Cited on page 209.

207 Most bacteria are flattened, ellipsoidal sacks kept in shape by the membrane enclosing the cytoplasma. But there are exceptions; in salt water, quadratic and triangular bacteria have been found. More is told in the corresponding section in the interesting book by Bernard Dixon, *Power Unseen – How Microbes Rule the World*, W.H. Freeman, New York, 1994. Cited on page 209.

208 S. Pitnick, G. Spicer & T. A. Markow, How long is a giant sperm?, *Nature* 375, p. 109, 1995. Cited on page 210.

209 M. Kawamura, A. Sugamoto & S. Nojiri, Swimming of microorganisms viewed from string and membrane theories, *Modern Journal of Physics Letters A* **9**, pp. 1159–1174, 1994. Also available as arxiv.org/abs/hep-th/9312200. Cited on page 210.

210 W. Nutsch & U. Rüffer, Die Orientierung freibeweglicher Organismen zum Licht, dargestellt am Beispiel des Flagellaten Chlamydomonas reinhardtii, *Naturwissenschaften* **81**, pp. 164–174, 1994. Cited on page 210.

211 They are also called prokaryote flagella. See for example S. C. Schuster & S. Khan, The bacterial flagellar motor, *Annual Review of Biophysics and Biomolecular Structure* **23**, pp. 509–539, 1994, or S. R. Caplan & M. Kara-Ivanov, The bacterial flagellar motor, *International Review of Cytology* **147**, pp. 97–164, 1993. See also the information on the topic that can be found on the website www.id.ucsb.edu:16080/fscf/library/origins/graphics-captions/flagellum.html. Cited on page 210.

212 For an overview of the construction and the motion of coli bacteria, see H. C. Berg, Motile behavior of bacteria, *Physics Today* 53, pp. 24–29, January 2000. Cited on page 210.

213 This is from the book by David Dusenbery, *Life at a Small Scale*, Scientific American Library, 1996. Cited on page 210.

214 F. Wilczek & A. Zee, Appearance of gauge structures in simple dynamical systems, *Physical Review Letters* **52**, pp. 2111–2114, 1984, A. Shapere & F. Wilczek, Self-propulsion at low Reynold number, *Physical Review Letters* **58**, pp. 2051–2054, 1987, A. Shapere & F. Wilczek, Gauge kinematics of deformable bodies, *American Journal of Physics* **57**, pp. 514–518, 1989, A. Shapere & F. Wilczek, Geometry of self-propulsion at low Reynolds number, *Journal of Fluid Mechanics* **198**, pp. 557–585, 1989, A. Shapere & F. Wilczek, Efficiencies of self-propulsion at low Reynolds number, *Journal of Fluid Mechanics* **198**, pp. 587–599, 1989. See also R. Montgomery, Gauge theory of the falling cat, *Field Institute Communications* **1**, pp. 75–111, 1993. Cited on page 211.

215 E. Putterman & O. Raz, The square cat, *American Journal of Physics* **76**, pp. 1040–1045, 2008. Cited on page 211.

216 J. Wisdom, Swimming in spacetime: motion by cyclic changes in body shape, *Science* **299**, pp. 1865–1869, 21st of March, 2003. A similar effect was discovered later on by E. Guéron & R. A. Mosna, The relativistic glider, *Physical Review D* **75**, p. 081501(R), 2007, preprint at arxiv.org/abs/gr-qc/0612131. Cited on page 214.

217 S. Smale, A classification of immersions of the two-sphere, *Transactions of the American Mathematical Society* **90**, pp. 281–290, 1958. Cited on page 215.

218 G.K. Francis & B. Morin, Arnold Shapiro's Eversion of the Sphere, *Mathematical Intelligencer* pp. 200–203, 1979. See also the unique manual for drawing manifolds by George Francis, *The Topological Picturebook*, Springer Verlag, 1987. It also contains a chapter on sphere eversion. Cited on page 215.

219 B. Morin & J.-P. Petit, Le retournement de la sphere, *Pour la Science* **15**, pp. 34–41, 1979. See also the clear article by A. Phillips, Turning a surface inside out, *Scientific American* pp. 112–120, May 1966. Cited on page 215.

220 S. Levy, D. Maxwell & T. Munzner, *Making Waves – a Guide to the Ideas Behind Outside In*, Peters, 1995. Cited on page 216.

221 George K. Batchelor, *An Introduction to Fluid Mechanics*, Cambridge University Press, 1967, and H. Hashimoto, A soliton on a vortex filament, *Journal of Fluid Mechanics* **51**, pp. 477–485, 1972. A summary is found in H. Zhou, On the motion of slender vortex filaments, *Physics of Fluids* **9**, p. 970-981, 1997. Cited on pages 217 and 218.

222 V. P. Dmitriyev, Helical waves on a vortex filament, *American Journal of Physics* **73**, pp. 563–565, 2005, and V. P. Dmitriyev, Mechanical analogy for the wave-particle: helix on a vortex filament, arxiv.org/abs/quant-ph/0012008. Cited on page 219.

223 T. Jacobson, Thermodynamics of spacetime: the Einstein equation of state, *Physical Review Letters* **75**, pp. 1260–1263, 1995, or arxiv.org/abs/gr-qc/9504004. Cited on page 220.

224 J. Frenkel & T. Kontorowa, Über die Theorie der plastischen Verformung, *Physikalische Zeitschrift der Sowietunion* **13**, pp. 1–10, 1938. F. C. Frank, On the equations of motion of crystal dislocations, *Proceedings of the Physical Society A* **62**, pp. 131–134, 1949, J. Eshelby, Uniformly moving dislocations, *Proceedings of the Physical Society A* **62**, pp. 307–314, 1949. See also G. Leibfried & H. Dietze, *Zeitschrift für Physik* **126**, p. 790, 1949. A general introduction can be found in A. Seeger & P. Schiller, Kinks in dislocations lines and their effects in internal friction in crystals, *Physical Acoustics* **3A**, W. P. Mason, ed., Academic Press, 1966. See also the textbooks by Frank R.N. Nabarro, *Theory of Crystal Dislocations*, Oxford University Press, 1967, or J. P. Hirth & J. Lothe, *Theory of Dislocations*, McGraw Hills Book Company, 1968. Cited on page 222.

225 Enthousiastic introductions into the theoretical aspects of polymers are Alexander Yu. Grosberg & Alexei R. Khokhlov, *Statistical Physics of Macromolecules*, AIP, 1994, and Pierre-Gilles de Gennes, *Scaling Concepts in Polymer Physics*, Cornell University Press, 1979. Cited on page 223.

226 The master of combining research and enjoyment in mathematics is Louis Kauffman. An example is his art is the beautiful text Louis H. Kauffman, *Knots and Physics*, World Scientific, third edition, 2001. It gives a clear introduction to the mathematics of knots and their applications. Cited on page 224.

227 A good introduction to knot tabulation is the paper by J. Hoste, M. Thistlethwaite & J. Weeks, The first 1,701,936 knots, *The Mathematical Intelligencer* **20**, pp. 33–47, 1998. Cited on page 224.

228 I. Stewart, *Game, Set and Math*, Penguin Books, 1989, pp. 58–67. Cited on page 225.

229 P. Pieranski, S. Przybyl & A. Stasiak, Tight open knots, arxiv.org/abs/physics/0103016. Cited on page 226.

230 T. Ashton, J. Cantarella, M. Piatek & E. Rawdon, Self-contact sets for 50 tightly knotted and linked tubes, arxiv.org/abs/math/0508248. Cited on pages 226 and 227.

231 J. Cantarella, J. H. G. Fu, R. Kusner, J. M. Sullivan & N. C. Wrinkle, Criticality for the Gehring link problem, *Geometry and Topology* **10**, pp. 2055–2116, 2006, preprint at arxiv.org/abs/math/0402212. Cited on page 227.

232 W. R. Taylor, A deeply knotted protein structure and how it might fold, *Nature* **406**, pp. 916–919, 2000. Cited on page 228.

233 Alexei Sossinsky, *Nœuds – histoire d'une théorie mathématique*, Editions du Seuil, 1999. D. Jensen, Le poisson noué, *Pour la science*, dossier hors série, pp. 14–15, April 1997. Cited on page 228.

234 D. M. Raymer & D. E. Smith, Spontaneous knotting of an agitated string, *Proceedings of the National Academy of Sciences (USA)* **104**, pp. 16432–16437, 2007, or www.pnas.org/cgi/doi/10.1073/pnas.0611320104. This work won the humorous Ignobel Prize in Physics in 2008; seeimprobable.com/ig. Cited on page 229.

235 A. C. Hirshfeld, Knots and physics: Old wine in new bottles, *American Journal of Physics* **66**, pp. 1060–1066, 1998. Cited on page 229.

236 For some modern knot research, see P. Holdin, R. B. Kusner & A. Stasiak, Quantization of energy and writhe in self-repelling knots, *New Journal of Physics* **4**, pp. 20.1–20.11, 2002. Cited on page 230.

237 H. R. Pruppacher & J. D. Klett, *Microphysics of Clouds and Precipitation*, Reidel, 1978, pp. 316–319. Falling drops are flattened and look like a pill, due to the interplay between surface tension and air flow. See also U. Thiele, Weine nicht, wenn der Regen zerfällt, *Physik Journal* **8**, pp. 16–17, 2009. Cited on page 230.

238 J. J. Socha, Becoming airborne without legs: the kinematics of take-off in a flying snake, Chrysopelea paradisi, *Journal of Experimental Biology* **209**, pp. 3358–3369, 2006, J. J. Socha, T. O'Dempsey & M. LaBarbera, A three-dimensional kinematic analysis of gliding in a flying snake, Chrysopelea paradisi, *Journal of Experimental Biology* **208**, pp. 1817–1833, 2005, J. J. Socha & M. LaBarbera, Effects of size and behavior on aerial performance of two species of flying snakes (Chrysopelea), *Journal of Experimental Biology* **208**, pp. 1835–1847, 2005. A full literature list on flying snakes can be found on the website www.flyingsnake.org. Cited on page 231.

239 An informative account of the world of psychokinesis and the paranormal is given by the famous professional magician James Randi, *Flim-flam!*, Prometheus Books, Buffalo 1987, as well as in several of his other books. See also the www.randi.org website. Cited on page 237.

240 James Clerk Maxwell, *Scientific Papers*, 2, p. 244, October 1871. Cited on page 242.

241 A good introduction is C. J. Hogan, Why the universe is just so, *Reviews of Modern Physics* 72, pp. 1149–1161, 2000. Most of the material of Table 24 is from the mighty book by John D. Barrow & Frank J. Tipler, *The Anthropic Cosmological Principle*, Oxford University Press, 1986. Discarding unrealistic options is also an interesting pastime. See for example the reasons why life can only be carbon-based, as explained in the essay by I. Asimov, The one and only, in his book *The Tragedy of the Moon*, Doubleday, Garden City, New York, 1973. Cited on page 242.

242 L. Smolin, The fate of black hole singularities and the parameters of the standard models of particle physics and cosmology, arxiv.org/abs/gr-qc/9404011. Cited on page 244.

243 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 246.

244 Le Système International d'Unités, Bureau International des Poids et Mesures, Pavillon de Breteuil, Parc de Saint Cloud, 92310 Sèvres, France. All new developments concerning SI units are published in the journal *Metrologia*, edited by the same body. Showing the slow pace of an old institution, the BIPM launched a website only in 1998; it is now reachable at www.bipm.fr. See also the www.utc.fr/~tthomass/Themes/Unites/index.html website; this includes the biographies of people who gave their names to various units. The site of its British equivalent, www.npl.co.uk/npl/reference, is much better; it provides many details as well as the English-language version of the SI unit definitions. Cited on page 247.

245 The bible in the field of time measurement is the two-volume work by J. Vanier & C. Audoin, *The Quantum Physics of Atomic Frequency Standards*, Adam Hilge, 1989. A popular account is Tony Jones, *Splitting the Second*, Institute of Physics Publishing, 2000.

The site opdaf1.obspm.fr/www/lexique.html gives a glossary of terms used in the field. For precision *length* measurements, the tools of choice are special lasers, such as mode-locked lasers and frequency combs. There is a huge literature on these topics. Equally large is the literature on precision *electric current* measurements; there is a race going on for the best way to do this: counting charges or measuring magnetic forces. The issue is still open. On *mass* and atomic mass measurements, see page 64. On high-precision *temperature* measurements, see page 425. Cited on page 248.

246 The unofficial prefixes were first proposed in the 1990s by Jeff K. Aronson of University of Oxford, and might come into general usage in the future. Cited on page 249.

247 For more details on electromagnetic unit systems, see the standard text by John David Jackson, *Classical Electrodynamics*, 3rd edition, Wiley, 1998. Cited on page 252.

248 D.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 252.

249 P. J. Hakonen, R. T. Vuorinen & J. E. Martikainen, Nuclear antiferromagnetism in rhodium metal at positive and negative nanokelvin temperatures, *Physical Review Letters* 70, pp. 2818–2821, 1993. See also his article in *Scientific American*, January 1994. Cited on page 253.

250 G. Charpak & R. L. Garwin, The DARI, *Europhysics News* 33, pp. 14–17, January/February 2002. Cited on page 253.

251 A. Zeilinger, The Planck stroll, *American Journal of Physics* 58, p. 103, 1990. Can you find another similar example? Cited on page 253.

Challenge 197 e

252 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. Cited on page 253.

253 The most precise clock built in 2004, a caesium fountain clock, had a precision of one part in 10^{15}. Higher precision has been predicted to be possible soon, among others by M. Takamoto, F.-L. Hong, R. Higashi & H. Katori, An optical lattice clock, *Nature* 435, pp. 321–324, 2005. Cited on page 253.

254 J. Bergquist, ed., *Proceedings of the Fifth Symposium on Frequency Standards and Metrology*, World Scientific, 1997. Cited on page 253.

255 See the information on $D_s^{\pm}$ mesons from the particle data group at pdg.web.cern.ch/pdg. Cited on page 254.

256 About the long life of tantalum 180, see D. Belic & al., Photoactivation of $^{180}Ta^m$ 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 254.

257 The various concepts are even the topic of a separate international standard, ISO 5725, with the title *Accuracy and precision of measurement methods and results*. A good introduction is John R. Taylor, *An Introduction to Error Analysis: the Study of Uncertainties in Physical Measurements*, 2nd edition, University Science Books, Sausalito, 1997. Cited on page 254.

258 P. J. Mohr & B. N. Taylor, CODATA recommended values of the fundamental physical constants: 1998, *Reviews of Modern Physics* 59, p. 351, 2000. This is the set of constants resulting from an international adjustment and recommended for international use by the Committee on Data for Science and Technology (CODATA), a body in the International Council of Scientific Unions, which brings together the International Union of Pure and Applied Physics (IUPAP), the International Union of Pure and Applied Chemistry (IUPAC) and other organizations. The website of IUPAC is www.iupac.org. Cited on pages 254 and 255.

259 Some of the stories can be found in the text by N. W. Wise, *The Values of Precision*, Princeton University Press, 1994. The field of high-precision measurements, from which the results on these pages stem, is a world on its own. A beautiful introduction to it is J. D. Fairbanks, B. S. Deaver, C. W. Everitt & P. F. Michaelson, eds., *Near Zero: Frontiers of Physics*, Freeman, 1988. Cited on page 255.

260 The details are given in the well-known astronomical reference, Kenneth Seidelmann, *Explanatory Supplement to the Astronomical Almanac*, 1992. Cited on page 259.

261 For information about the number π, and about some other mathematical constants, the website oldweb.cecm.sfu.ca/pi/pi.html provides the most extensive information and references. It also has a link to the many other sites on the topic, including the overview at mathworld.wolfram.com/Pi.html. Simple formulae for π are

$$\pi + 3 = \sum_{n=1}^{\infty} \frac{n\, 2^n}{\binom{2n}{n}} \qquad (155)$$

or the beautiful formula discovered in 1996 by Bailey, Borwein and Plouffe

$$\pi = \sum_{n=0}^{\infty} \frac{1}{16^n} \left(\frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6} \right) . \qquad (156)$$

The mentioned site also explains the newly discovered methods for calculating specific binary digits of π without having to calculate all the preceding ones. The known digits of π pass all tests of randomness, as the mathworld.wolfram.com/PiDigits.html website explains. However, this property, called *normality*, has never been proven; it is the biggest open question about π. It is possible that the theory of chaotic dynamics will lead to a solution of this puzzle in the coming years.

Another method to calculate π and other constants was discovered and published by D. V. Chudnovsky & G. V. Chudnovsky, The computation of classical constants, *Proceedings of the National Academy of Sciences (USA)* **86**, pp. 8178–8182, 1989. The Chudnowsky brothers have built a supercomputer in Gregory's apartment for about 70 000 euros, and for many years held the record for calculating the largest number of digits of π. They have battled for decades with Kanada Yasumasa, who held the record in 2000, calculated on an industrial supercomputer. However, the record number of (consecutive) digits in 2010 was calculated in 123 days on a simple desktop PC by Fabrice Bellard, using a Chudnovsky formula. Bellard calculated over 2.7 million million digits, as told on bellard.org. New formulae to calculate π are still occasionally discovered.

For the calculation of Euler's constant γ see also D. W. DeTemple, A quicker convergence to Euler's constant, *The Mathematical Intelligencer*, pp. 468–470, May 1993. Cited on page 260.

262 The proton charge radius was determined by measuring the frequency of light emitted by hydrogen atoms to high precision by T. 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 261.

263 For a full list of isotopes, see R. B. Firestone, *Table of Isotopes, Eighth Edition, 1999 Update*, with CD-ROM, John Wiley & Sons, 1999. For a list of isotopes on the web, see the Corean website by J. Chang, atom.kaeri.re.kr. For a list of precise isotope masses, see the csnwww.in2p3.fr website. Cited on pages 262, 265, and 275.

264 The ground state of bismuth 209 was thought to be stable until early 2003. It was then discovered that it was radioactive, though with a record lifetime, as reported by de Marcillac, N. Coron, G. Dambier, J. Leblanc & J. -P. Moalic, Experimental detection of α-particles from the radioactive decay of natural bismuth, *Nature* **422**, pp. 876–878, 2003. By coincidence, the excited state 83 MeV above the ground state of the same bismuth 209 nucleus is the shortest known radioactive nuclear state. Cited on page 262.

265 For information on the long life of tantalum 180, see D. Belic & al., Photoactivation of $^{180}Ta^{m}$ 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 262.

266 Stephen J. Gould, *The Panda's thumb*, W.W. Norton & Co., 1980. This is one of several interesting and informative books on evolutionary biology by the best writer in the field. Cited on page 263.

267 F. Marques & al., Detection of neutron clusters, *Physical Review C* **65**, p. 044006, 2002. Opposite results have been obtained by B. M. Sherrill & C. A. Bertulani, Proton-tetraneutron elastic scattering, *Physical Review C* **69**, p. 027601, 2004, and D.V. Aleksandrov & al., Search for resonances in the three- and four-neutron systems in the 7Li(7Li, 11C)3n and 7Li(7Li, 10C)4n reactions, *JETP Letters* **81**, p. 43, 2005. Cited on page 263.

268 For a good review, see the article by P. T. Greenland, Antimatter, *Contemporary Physics* **38**, pp. 181–203, 1997. Cited on page 263.

269 Almost everything known about each element and its chemistry can be found in the encyclopaedic Gmelin, *Handbuch der anorganischen Chemie*, published from 1817 onwards. There are over 500 volumes, now all published in English under the title *Handbook of Inor-*

ganic and Organometallic Chemistry, with at least one volume dedicated to each chemical element. On the same topic, an incredibly expensive book with an equally bad layout is Per Enhag, *Encyclopedia of the Elements*, Wiley–VCH, 2004. Cited on page 265.

270 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 website. For an isotope mass list, see the csnwww.in2p3.fr website. Cited on pages 265 and 275.

271 The metallic, covalent and Van der Waals radii are from Nathaniel W. Alcock, *Bonding and Structure*, Ellis Horwood, 1999. This text also explains in detail how the radii are defined and measured. Cited on page 275.

272 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 was written before the superstring revolution, so that the topic is missing from the otherwise excellent presentation. Cited on page 284.

273 For more puzzles, see the excellent book James Tanton, *Solve This – Math Activities for Studends and Clubs*, Mathematical Association of America, 2001. Cited on page 284.

274 For an introduction to topology, see for example Mikio Nakahara, *Geometry, Topology and Physics*, IOP Publishing, 1990. Cited on page 285.

275 An introduction to the classification theorem is R. Solomon, On finite simple groups and their classification, *Notices of the AMS* 42, pp. 231–239, 1995, also available on the web as www.ams.org/notices/199502/solomon.ps Cited on page 289.

276 A pedagogical explanation is given by C. G. Wohl, Scientist as detective: Luis Alvarez and the pyramid burial chambers, the JFK assassination, and the end of the dinosaurs, *American Journal of Physics* 75, pp. 968–977, 2007. Cited on page 299.

CREDITS

Many people who have kept their gift of curiosity alive have helped to make this project come true. Most of all, Saverio Pascazio has been – present or not – a constant reference for this project. Fernand Mayné, Anna Koolen, Ata Masafumi, Roberto Crespi, Serge Pahaut, Luca Bombelli, Herman Elswijk, Marcel Krijn, Marc de Jong, Martin van der Mark, Kim Jalink, my parents Peter and Isabella Schiller, Mike van Wijk, Renate Georgi, Paul Tegelaar, Barbara and Edgar Augel, M. Jamil, Ron Murdock, Carol Pritchard, Richard Hoffman, Stephan Schiller and, most of all, my wife Britta have all provided valuable advice and encouragement.

Many people have helped with the project and the collection of material. Most useful was the help of Mikael Johansson, Bruno Barberi Gnecco, Lothar Beyer, the numerous improvements by Bert Sierra, the detailed suggestions by Claudio Farinati, the many improvements by Eric Sheldon, the detailed suggestions by Andrew Young, the continuous help and advice of Jonatan Kelu, the corrections of Elmar Bartel, and in particular the extensive, passionate and conscientious help of Adrian Kubala.

Important material was provided by Bert Peeters, Anna Wierzbicka, William Beaty, Jim Carr, John Merrit, John Baez, Frank DiFilippo, Jonathan Scott, Jon Thaler, Luca Bombelli, Douglas Singleton, George McQuarry, Tilman Hausherr, Brian Oberquell, Peer Zalm, Martin van der Mark, Vladimir Surdin, Julia Simon, Antonio Fermani, Don Page, Stephen Haley, Peter Mayr, Allan Hayes, Norbert Dragon, Igor Ivanov, Doug Renselle, Wim de Muynck, Steve Carlip, Tom Bruce, Ryan Budney, Gary Ruben, Chris Hillman, Olivier Glassey, Jochen Greiner, squark, Martin Hardcastle, Mark Biggar, Pavel Kuzin, Douglas Brebner, Luciano Lombardi, Franco Bagnoli, Lukas Fabian Moser, Dejan Corovic, Paul Vannoni, John Haber, Saverio Pascazio, Klaus Finkenzeller, Leo Volin, Jeff Aronson, Roggie Boone, Lawrence Tuppen, Quentin David Jones, Arnaldo Uguzzoni, Frans van Nieuwpoort, Alan Mahoney, Britta Schiller, Petr Danecek, Ingo Thies, Vitaliy Solomatin, Carl Offner, Nuno Proença, Elena Colazingari, Paula Henderson, Daniel Darre, Wolfgang Rankl, John Heumann, Joseph Kiss, Martha Weiss, Antonio González, Antonio Martos, André Slabber, Ferdinand Bautista, Zoltán Gácsi, Pat Furrie, Michael Reppisch, Enrico Pasi, Thomas Köppe, Martin Rivas, Herman Beeksma, Tom Helmond, John Brandes, Vlad Tarko, Nadia Murillo, Ciprian Dobra, Romano Perini, Harald van Lintel, Andrea Conti, François Belfort, Dirk Van de Moortel, Heinrich Neumaier, Jarosław Królikowski, John Dahlman, Fathi Namouni, Paul Townsend, Sergei Emelin, Freeman Dyson, S.R. Madhu Rao, David Parks, Jürgen Janek, Daniel Huber, Alfons Buchmann, William Purves, Pietro Redondi, Douglas Singleton, Emil Akhmedov Damoon Saghian, plus a number of people who wanted to remain unnamed.

I also thank the lawmakers and the taxpayers in Germany, who, in contrast to most other countries in the world, allow residents to use the local university libraries.

The software tools were refined with extensive help on fonts and typesetting by Michael Zedler and Achim Blumensath and with the repeated and valuable support of Donald Arseneau; help came also from Ulrike Fischer, Piet van Oostrum, Gerben Wierda, Klaus Böhncke, Craig Upright, Herbert Voss, Andrew Trevorrow, Danie Els, Heiko Oberdiek, Sebastian Rahtz, Don Story, Vincent Darley, Johan Linde, Joseph Hertzlinger, Rick Zaccone, John Warkentin, Ulrich Diez, Uwe

Siart, Will Robertson, Joseph Wright Enrico Gregorio, Rolf Niepraschk, Alexander Grahn and Paul Townsend.

I also thank the lawmakers and the taxpayers in Germany, who, in contrast to most other countries in the world, allow residents to use the local university libraries.

The typesetting and book design is due to the professional consulting of Ulrich Dirr. The typography was improved with the help of Johannes Küster. The design of the book and its website owe also much to the suggestions and support of my wife Britta.

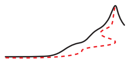
Since May 2007, the electronic edition and distribution of the Motion Mountain text is generously supported by the Klaus Tschira Foundation.

Film credits

The animation of the actin–myosin system on page 20 is copyright and courtesy by San Diego State University, Jeff Sale and Roger Sabbadini. It can be found on the website www.sci.sdsu.edu/movies/actin_myosin.html. The film on page 80, showing single electrons moving through liquid helium, is copyright and courtesy of Humphrey Maris; it can be found on his website physics.brown.edu/physics/researchpages/cme/bubble. The film of the solar flare on page 155 was taken by NASA's TRACE satellite, whose website is trace.lmsal.com. The film of a flying snake on page 231 is copyright and courtesy by Jake Socha. It can be found on his website www.flyingsnake.org.

Image credits

The mountain photograph on the front cover is courtesy and copyright by Dave Thompson and found on his website www.daveontrek.co.uk. The photograph of a soap bubble on page 14 is courtesy and copyright of Jason Tozer for Creative Review/Sony; its making is described on creativereview.co.uk/crblog. The picture of the orchid on page 19 is courtesy and copyright of Elmar Bartel. The sea urchin photograph on page 21 was taken by Kristina Yu, and is copyright and courtesy of the Exploratorium, found at www.exploratorium.edu. The illustrations of DNA on page 49 are copyright and courtesy by David Deerfield and dedicated to his memory. They are taken from his website www.psc.edu/~deerfiel/Nucleic_Acid-SciVis.html. The images of DNA molecules on page 50 are copyright and courtesy by Hans-Werner Fink and used with permission of Wiley VCH. The mineral photographs on page 54 onwards are copyright and courtesy of Rob Lavinsky at irocks.com, and taken from his large and beautiful collection there and at www.mindat.org/photo-49529.html. The corundum photograph on page 55, the malachite photograph on page 147 and the magnetite photograph on page page 149 are copyright and courtesy of Stephan Wolfsried and taken from his beautiful collection of microcrystals at www.mindat.org/user-1664.html. The photograph of Paraiba tourmalines on page page 55 is copyright and courtesy of Manfred Fuchs and found on his website at www.tourmalinavitalis.de. The synthetic crystals photographs on page 56 and page 56 are copyright of Northrop Grumman and courtesy of Scott Griffin. The alexandrite photographs on page 56 are copyright and courtesy of John Veevaert, Trinity Mineral Co., trinityminerals.com. The photographs of the silicon crystal growing machines and of the resulting silicon crystals are copyright of PVATePla at www.PVATePla.com, and are used by kind permission of Gert Fisahn. The shark teeth photograph on page 59 is courtesy and copyright of Peter Doe and found at www.flickr.com/photos/peteredin. The climate graph on page 66 is copyright and courtesy of Dieter Lüthi. The collection of images of single atom sheets on page 68 is taken from Ref. 53, copyright of the National Academy of Sciences, and are courtesy of the authors and of Tiffany Millerd. The rhenium photographs on page 92 are copyright and courtesy of Hans-Werner Fink, APS and AIP and reprinted with permission from the cited references. The pictures of snow flakes on page 93 are courtesy and copyright of Furukawa Yoshinori. The picture of neutron interferometers on page 100 is courtesy and copyright of Helmut Rauch and Erwin Seidl, and found at the website of the Atominstitut in Vienna, www.ati.ac.at. The map with the gamma ray bursts on page 108 is courtesy of NASA and taken from the BATSE website at www.batse.msfc.nasa.gov/batse. The MRI images of the head and the spine on page 120 are courtesy and copyright of Joseph Hornak and taken from his website www.cis.rit.edu/htbooks/mri. The MRI image on page 122 of humans making love is courtesy and copyright of Willibrord Weijmar Schultz.IndexWeijmar Schultz, Willibrord The nuclide chart on page 127 is used with permission of Alejandro Sonzogni of the National Nuclear Data Centre, using data extracted from the NuDat database. The electroscope photograph on page 128 is courtesy and copyright of Harald Chmela and taken from his website www.hcrs.at. The photograph of an aurora on page 132 is courtesy and copyright of Jan Curtis and taken from his website climate.gi.alaska.edu/Curtis/curtis.html. The photograph of the Geigerpod on page 131 is courtesy and copyright of Joseph Reinhardt, and found on his website flickr.com/photos/javamoose. The figure of the Erta Ale volcano on page 137 is courtesy and copyright of Marco Fulle and part of the wonderful photograph collection at www.stromboli.net. The drawing of the JET reactor on page 152 is courtesy and copyright of EFDA-JET. The drawing of the strong coupling constant on page page 169 is courtesy and copyright of Siegfried Bethke. The photograph of the watch on page 177 is courtesy and copyright of Traser, found at www.traser.com. The unification graph on page 198 is courtesy and copyright of Wim de Boer and taken from his home page at www-ekp.physik.uni-karlsruhe.de/~deboer. The picture of the butterfly in a wind tunnel on page 205 is courtesy and copyright of Robert Srygley and Adrian Thomas. The picture of the vulture on page 206 is courtesy and copyright of S.L. Brown from the website SLBrownPhoto.com. The picture of the hummingbird on page 206 is courtesy and copyright of the Pennsylvania Game Comission and Joe Kosack and taken from the website www.pgc.state.pa.us. The picture of the dragonfly on page 206 is courtesy and copyright of nobodythere and found on his website xavier.favre2.free.fr/coppermine. The pictures of feathers insects on page 208 are material used with kind permission of HortNET, a product of The Horticulture and Food Research Institute of New Zealand. The picture of the scallop on page 209 is courtesy and copyright of Dave Colwell. The picture of cliff diver Artem Silchenko on page 212 is copyright of the World High Diving Federation, found at the web site www.whdf.com, and courtesy of Frederic Weill. The picture of the eversion of the sphere on page 215 is courtesy and copyright of John Sullivan and found on his *Optiverse* website on new.math.uiuc.edu/optiverse. The knot and link diagrams on pages 225 and 227 are courtesy and copyright of Robert Scharein and taken from his website on www.knotplot.com. The clasp images on page 227 is courtesy and copyright of Jason Cantarella. The photograph of a hagfish on page 228 is courtesy and copyright of Christine Ortlepp; it is also found on the web page www.lemonodor.com/scruz-2003-01-24/17m.shtml. The photograph of a waterspout on page 217 is copyright and courtesy of Zé Nogueira and found on his website at www.flickr.com/photos/zenog. The photograph of a vortex lattice on page 217 is copyright of Andre Schirotzek, courtesy of Wolfgang Ketterle, and found on their research website at cua.mit.edu/ketterle_group. The drawing of a mixed dislocation on page 221 has been produced especially for this text and is copyright and courtesy of Ulrich Kolberg. The poster of the elements on page 265 is copyright and courtesy by Theodore Gray and found on his website www.theodoregray.com, where paper prints can be ordered. The historical portraits of physicists in the text do not have copyright, except where mentioned. The photograph on the back cover, of a basilisk running over water, is courtesy and copyright by the Belgian group TERRA vzw and found on their website www.terra.vzw.org. All drawings are copyright by Christoph Schiller. If you suspect that your copyright is not correctly given or obtained, this has not done on purpose; please contact me in this case.



NAME INDEX

Page numbers in *italic* typeface refer to pages where the person is presented in more detail.

L

LAMOREAUX

T

SUBJECT INDEX

Page numbers in *italic* typeface refer to pages where the keyword is defined or presented in detail. The subject index thus acts as a glossary.

B
BECQUEREL

F

G

G
GELL-MANN

H

I

J

N
NUCLIDES

Q
QED

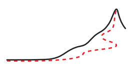
T

T

THORIUM

W

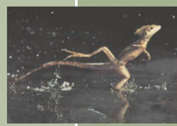
Wikipedia



MOTION MOUNTAIN

The Adventure of Physics – Vol. V

Pleasure, Technology and Stars



Why do change and motion exist?
How does a rainbow form?
What is the most fantastic voyage possible?
Is 'empty space' really empty?
How can one levitate things?
At what distance between two points does it become impossible to find room for a third one in between?
What does 'quantum' mean?
Which problems in physics are unsolved?

Answering these and other questions on motion, this series gives an entertaining and mind-twisting introduction into modern physics – one that is surprising and challenging on every page.

Starting from everyday life, the adventure provides an overview of the recent results in mechanics, thermodynamics, electrodynamics, relativity, quantum theory, quantum gravity and unification. It is written for undergraduate students and for anybody interested in physics.

Christoph Schiller, PhD Université Libre de Bruxelles, is a physicist with more than 25 years of experience in the presentation of physical topics.

ISBN 978-3-00-021946-7
9 783000 219467