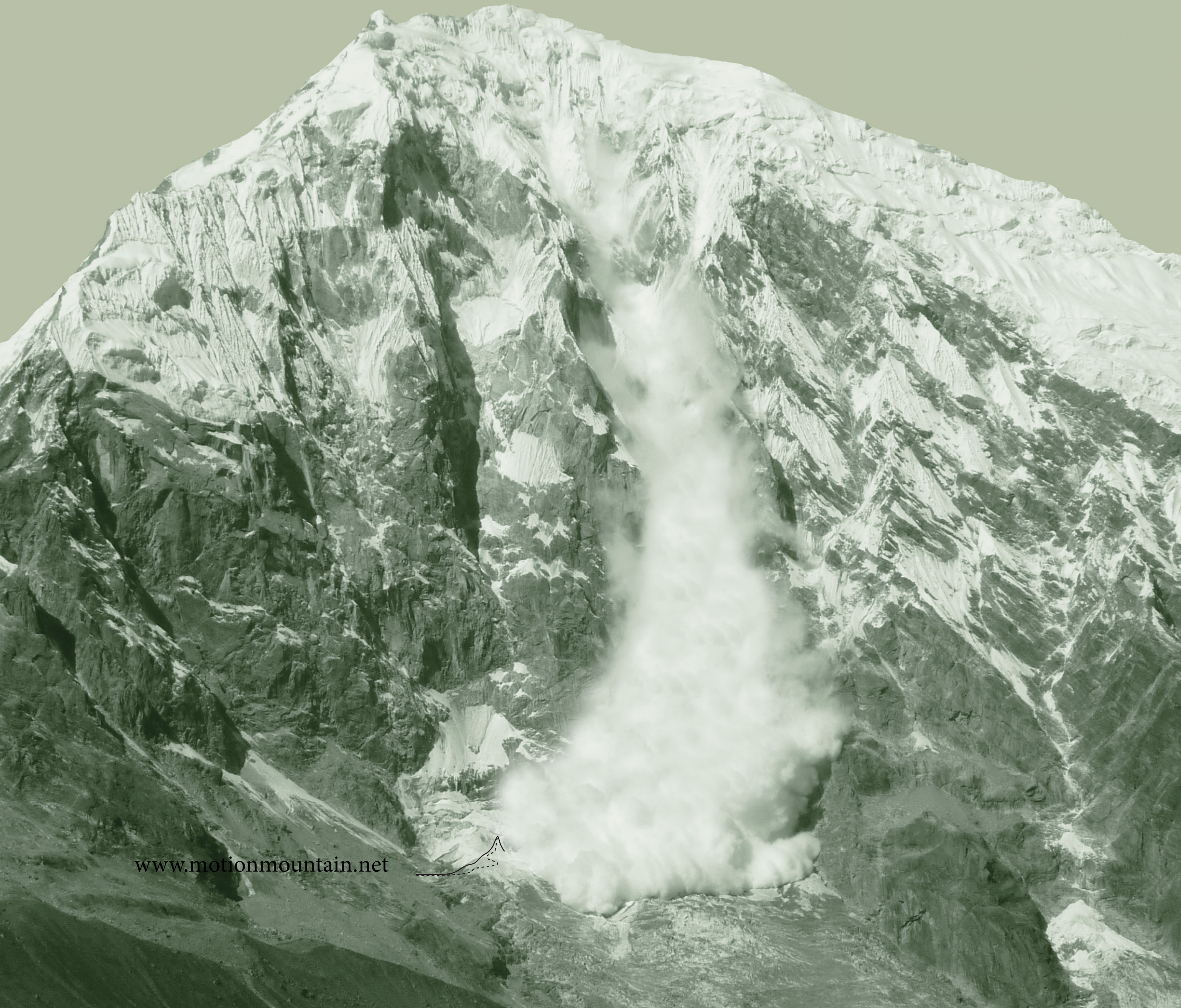
Christoph Schiller

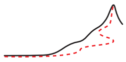
MOTION MOUNTAIN

THE ADVENTURE OF PHYSICS – VOL.V

MOTION INSIDE MATTER –
PLEASURE, TECHNOLOGY AND STARS

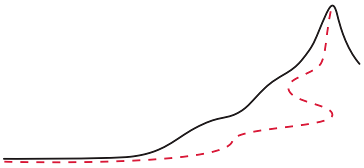


www.motionmountain.net



Christoph Schiller

Motion Mountain



The Adventure of Physics
Volume V

Motion Inside Matter – Pleasure, Technology and Stars

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

Editio trigesima prima.

Thirty-first edition.

To Britta, Esther and Justus Aaron

τῷ ἐμοὶ δαὶμονι

Die Menschen stärken, die Sachen klären.

PREFACE

> “Primum movere, deinde docere.*
>
> Antiquity”

This book series is for anybody who is curious about motion in nature. How do things, people, animals, images and empty space move? The answer leads to many adventures, and this volume presents those about motion inside everyday matter, inside people and animals, and inside stars and nuclei.

Motion *inside bodies* – dead or alive – is tiny: thus it is described by quantum theory. Quantum theory describes all motion with the quantum of action $\hbar$, the smallest change observed in nature. Building on this basic idea, the text first shows how to describe life, death and pleasure. Then, the text explains the observations of chemistry, materials science, astrophysics and particle physics. In the structure of physics, these topics correspond to the three ‘quantum’ points in Figure 1. The story of motion found inside living cells, inside the coldest gases and throughout the hottest stars is told here 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 on pedagogy, translates as: ‘To fortify people, to clarify things.’ Clarifying things – and adhering only to the truth – requires courage, as changing the habits of thought produces fear, often hidden by anger. But by overcoming our fears we grow in strength. And we experience intense and beautiful emotions. All great adventures in life allow this, and exploring motion is one of them. Enjoy it.

Christoph Schiller

---

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

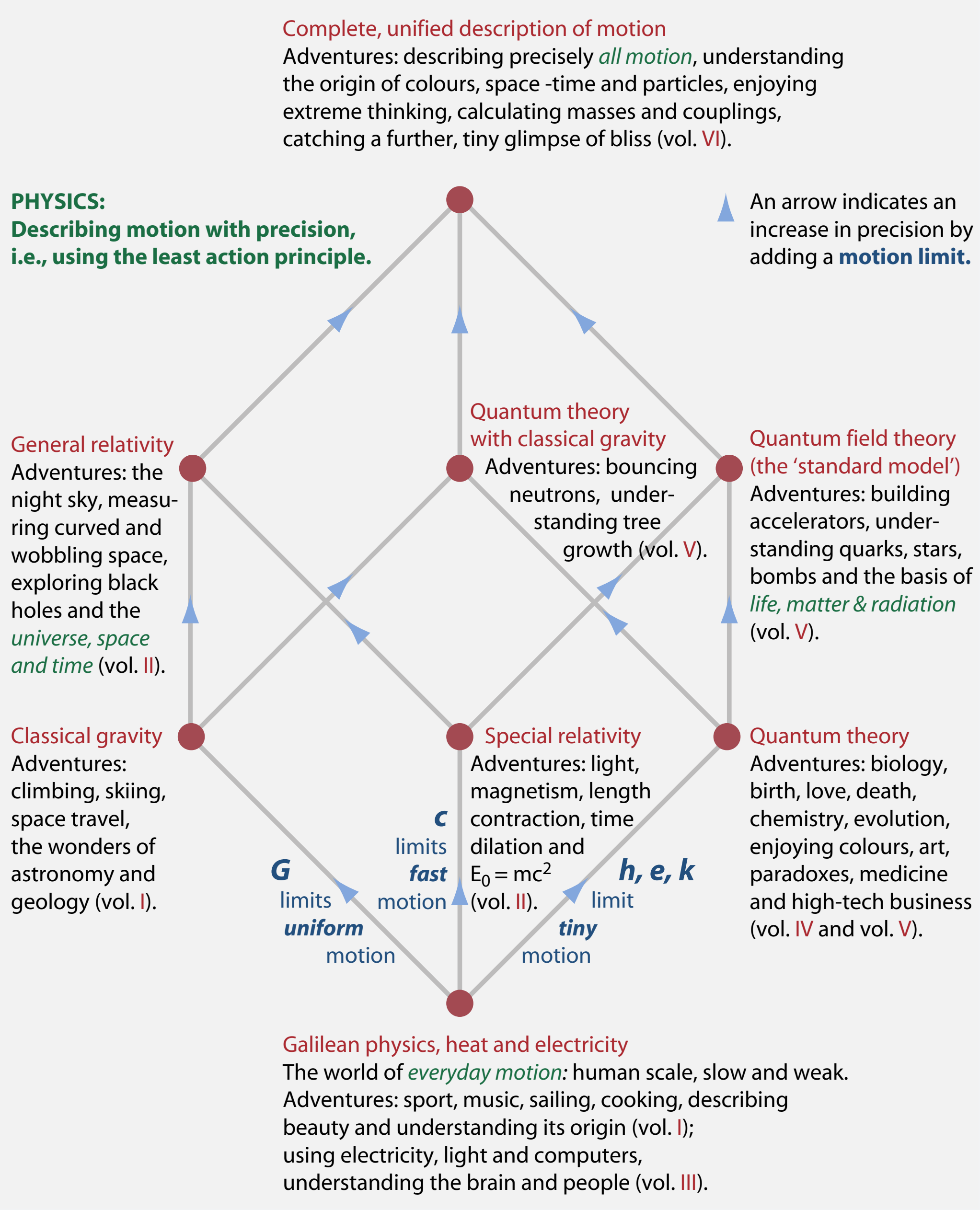


FIGURE 1 A complete map of physics, the science of motion, as first proposed by Matvei Bronshtein (b. 1907 Vinnytsia, d. 1938 Leningrad). The Bronshtein cube starts at the bottom with everyday motion, and shows the connections to the fields of modern physics. Each connection increases the precision of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

USING THIS BOOK

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

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

ADVICE FOR LEARNERS

Learning allows us to discover what kind of person we can be. Learning widens knowledge, improves intelligence and provides a sense of achievement. Therefore, learning from a book, especially one about nature, should be efficient and enjoyable. Avoid bad learning methods like the plague! Do not use a marker, a pen or a pencil to highlight or underline text on paper. It is a waste of time, provides false comfort and makes the text unreadable. And do not learn from a screen. In particular, never, ever, learn from the internet, from videos, from games or from a smartphone. Most of the internet, almost all videos and all games are poisons and drugs for the brain. Smartphones are dispensers of drugs that make people addicted and prevent learning. Nobody putting marks on paper or looking at a screen is learning efficiently or is enjoying doing so.

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

ADVICE FOR TEACHERS

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

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

FEEDBACK

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

Challenge 1 s

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

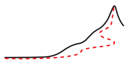
Also help on the specific points listed on the www.motionmountain.net/help.html web page is welcome. All feedback will be used to improve the next edition. You are welcome to send feedback by mail or by sending in a pdf with added yellow notes, to provide illustrations or photographs, or to contribute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

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

SUPPORT

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

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



CONTENTS

MOTION INSIDE MATTER – PLEASURE, TECHNOLOGY AND STARS

In our quest to understand how things move
as a result of a smallest change value in nature, we discover
how pleasure appears,
why the floor does not fall but keeps on carrying us,
that interactions are exchanges of radiation particles,
that matter is not permanent,
how quantum effects increase human wealth and health,
why empty space pulls mirrors together,
why the stars shine,
where the atoms inside us come from,
how quantum particles make up the world,
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 have explored quantum effects in the previous volume, let us now have some serious fun with *applied* quantum physics. The quantum of action $\hbar$ has significant consequences for medicine, biology, chemistry, materials 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 *motion inside matter*.

Inside matter, we observe, above all, tiny motions of quantum particles.[***] Therefore the understanding and the precise description of matter requires *quantum physics*. In the following, we will only explore a cross-section, but it will be worth it. We start our exploration of motion inside matter with three special forms that are of special importance to us: life, reproduction and death. We mentioned at the start of quantum physics that none of these forms of motion can be described by classical physics. Indeed, life, reproduction and death are quantum effects. In addition, every perception, every sense, and thus every kind of pleasure is a quantum effect. The same is true for all our actions. Let us find out why.

Vol. IV, page 15

FROM QUANTUM PHYSICS TO BIOLOGICAL MACHINES AND MINIATURIZATION

We know that all of quantum theory can be resumed in one sentence:

▹ In nature, action or change below $\hbar = 1.1 \cdot 10^{-34}\,\mathrm{Js}$ is not observed.

In the following, we want to understand how this observation explains life, pleasure and death. An important consequence of the quantum of action is well-known.

▹ If it moves, it is made of quantons, or quantum particles.

[**] 'I am a man and nothing human is alien to me.' Terence is Publius Terentius Afer (b. *c.* 190 Carthago, d. 159 BCE Greece), 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.

Starting size: the dot on the letter i – final size:

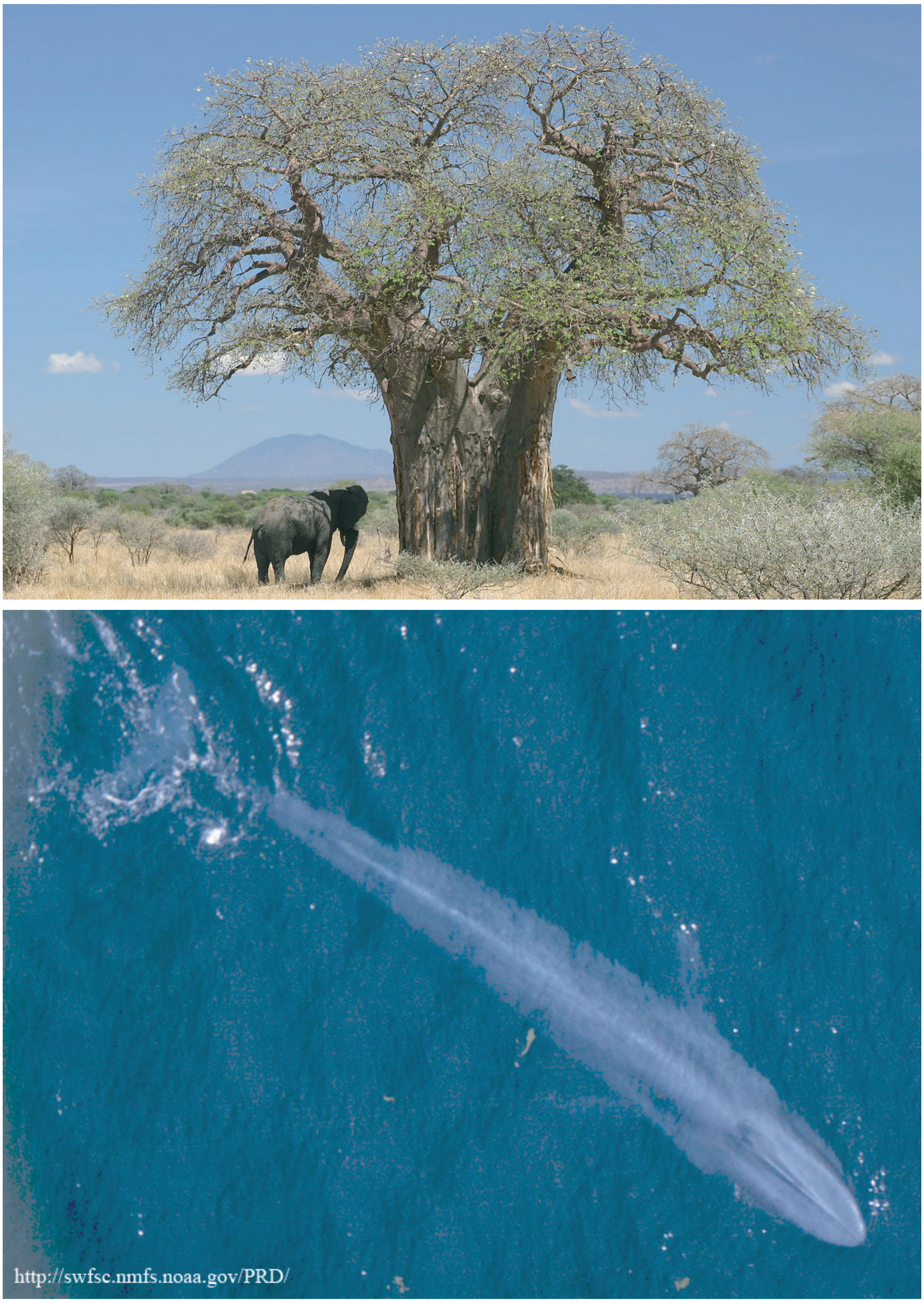


FIGURE 2 Metabolic growth can lead from single cells, about 0.1 mm in diameter, to living beings of 25 m in size, such as the baobab or the blue whale (© Ferdinand Reus, NOAA).

Step by step we will discover how these statements are reflected in the behaviour of living beings. But what are living beings?

Living beings are physical systems that show metabolism, information processing, information exchange, reproduction and motion. All these properties can be condensed in a single statement:

▷ A *living being* is a collection of machines that is able to self-reproduce.

By *self*-reproduction, we mean that a system uses its own metabolism to reproduce. There are examples of objects which reproduce and which nobody would call living. Can you find some examples? To avoid misunderstandings, whenever we say 'reproduction' in the following, we always mean 'self-reproduction'.

Challenge 2 s

Before we explore the definition of living beings in more detail, we stress that self-reproduction is simplified if the system is miniaturized. Therefore, most living beings are extremely small machines for the tasks they perform. This is especially clear when living beings are compared to human-made machines. The smallness of living beings is often astonishing, because the design and construction of human-made machines has considerably fewer requirements.

1. Human-made machines do not need to be able to reproduce; as a result, they can be made of many parts and can include rotating macroscopic parts. This is in contrast to living beings, who are all made of a single piece of matter, and cannot use wheels, propellers, gearwheels or even screws.
2. Human made machines can make use of metals, ceramics, poisonous compounds and many other materials that living beings cannot use.
3. Human machines do not need to self-assemble and grow; in contrast, living beings always need to carry a built-in chemical factory with them.
4. Human machines can be assembled and can operate at various temperatures, in strong contrast to living beings.

Despite these extreme engineering restrictions, 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 and lost. 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 thousand times larger in volume.
- Motors in living beings are several 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 achieves the movement changes of a grasshopper, a fly or a tadpole.
- 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.

Vol. III, page 268

In particular, already the navigation systems built by nature are far smaller than anything built by human technology.
- 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.
- 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. The efficiency has three main aspects. First of all, 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 electric, 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. Secondly, *all parts in living systems have more than one function.* Challenge 4 e For example, bones provide structure and produce blood; fingernails are tools and shed chemical waste. Living systems use many such optimizations. Challenge 5 e Last but not least, living machines are *well* miniaturized because they make efficient use of quantum effects. Indeed, every single function in living beings relies on the quantum of action. And every such function is extremely well miniaturized. We explore a few important cases.

REPRODUCTION

> “Finding a mate is life's biggest prize.”
> The view of biologists.

All the astonishing complexity of life is geared towards reproduction. *Reproduction*, more precisely, *self-reproduction*, is the ability of an object to build other objects similar to itself. Vol. IV, page 122 Quantum theory told us that it is only possible to build a *similar* object, since an *exact* copy would contradict the quantum of action. But this limitation is not a disadvantage: an imperfect copy is required for life; indeed, a similar, thus imperfect copy is essential for biological evolution, and thus for change and specialization.

Reproduction is characterized by random changes, called *mutations*, that distinguish one generation from 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 heredity are quantum effects.

Reproduction requires *growth*, and growth needs metabolism. *Metabolism* is a chemical process, and thus a quantum process, to harness energy, harness materials, realize growth, heal injuries and realize reproduction. Page 58

Since reproduction requires an increase in mass, as shown in Figure 2, all 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 thus deduce that reproduction and growth implies that matter is made of smallest entities. If matter were not made of smallest entities, there would be no way to realize

FIGURE 3 A quantum machine (© Elmar Bartel).

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 and secure its stability over time. The large number of particles is necessary to protect the information from the influences of the outside world. Vol. I, page 404

Our own construction plan is stored in DNA molecules in the nucleus and the mitochondria of each of the millions of cells inside our body. We will explore some details below. The plan is thus indeed stored and secured with the help of *many* particles. Therefore, reproduction is first of all a transfer of parent's DNA to the next generation. The transfer is an example of motion. It turns out that this and all other examples of motion in our bodies occur in the same way, namely with the help of molecular machines. Page 62

QUANTUM MACHINES

Living beings move. In order to reproduce, living beings must be able to move in self-directed ways.

▷ A system able to perform self-directed motion is called a *machine*.

All self-reproducing beings, such as the one of Figure 3, are thus machines. Even machines that do not grow still need fuel, and thus need a metabolism. All machines, living or not, are based on quantum effects.

How do living machines work? From a fundamental physics point of view, we need only a few sections of our walk so far to describe them: we need QED and sometimes universal gravity. Simply stated, life is an electromagnetic process taking place in weak gravity.* But the details of this statement are tricky and interesting.

---

* In fact, also the nuclear interactions play some role for life: cosmic radiation is one source for random

TABLE 1 Motion and molecular machines found in living beings.

Motion type	Examples	Involved motors
Growth	collective molecular processes in cell growth, cell shape change, cell motility	linear molecular motors, ion pumps
	gene turn-on and turn-off	linear molecular motors
	ageing	linear molecular motors
Construction	material transport (polysaccharides, lipids, proteins, nucleic acids, others)	muscles, linear molecular motors
	forces and interactions between biomolecules and cells	pumps in cell membranes
Functioning	metabolism (respiration, digestion)	muscles, ATP synthase, ion pumps
	muscle working	linear molecular motors, 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
	memory: synapse growth	linear molecular motors
	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	cilial motors
Sensing	eye	chemical pumps, ion pumps
	ear	hair motion sensors, ion pumps, rotary molecular motors
	smell	ion pumps
	touch	ion pumps
Reproduction	information storage and retrieval	linear and rotational molecular motors inside nuclei
	cell division, organelle motion	linear molecular motors, polymerase
	sperm motion	linear molecular motors
	courting, using brain and muscles	linear molecular motors, ion pumps

We can say that *living beings are systems that move against their environment faster than molecules do.* Ref. 2 Observation shows that living systems move faster the bigger they are. Observation also shows that living beings achieve this speed by making use of a huge number of tiny machines, often made of one or only a few molecules, that work together. These machines realize the numerous processes that are part of life.

An overview of processes taking place in living beings is given in Table 1. Above all, the table shows that the processes are due to *molecular machines*.

▷ A living being is a collection of a huge number of specialized molecular machines.

Molecular machines are among the most fascinating devices found in nature. Table 1 also shows that nature only needs a few such devices to realize all the motion types used by humans and by all other living beings: *molecular pumps* and *molecular motors*. Given the long time that living systems have been around, these devices are extremely efficient. They are found in every cell, including those of Figure 5. The specialized molecular machines in living beings are ion pumps, chemical pumps and rotational and linear molecular motors. Ion and chemical pumps are found *in* membranes and transport matter Ref. 3 *across* membranes. Rotational and linear motors move structures *against* membranes. Even though there is still a lot to be learned about molecular machines, the little that is known is already 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.

▷ *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.

▷ Each shape-changing molecule is a molecular motor.

There are three basic classes of molecular motors in nature: linear motors, rotational motors and pumps.

---

mutations, which are so important in evolution. Plant growers often use radioactive sources to increase Ref. 1 mutation rates. Radioactivity can also terminate life or be of use in medicine.

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).

Summing up, the nuclear interactions occasionally 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.

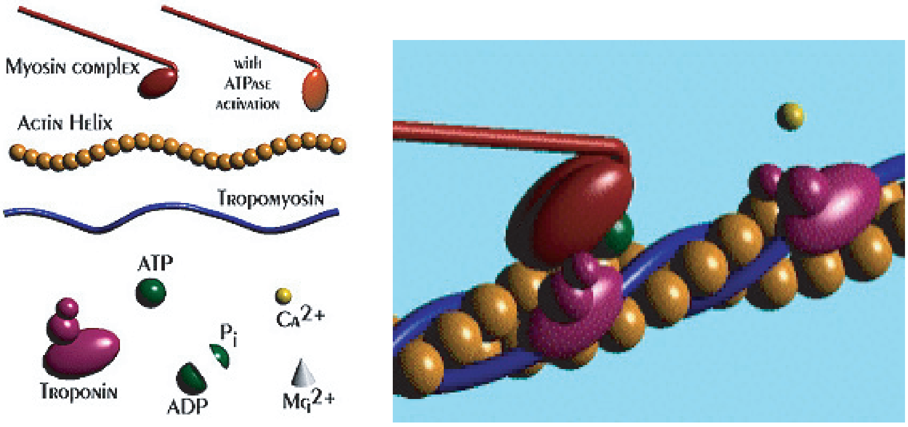


FIGURE 4 Left: myosin and actin are the two protein molecules that realize the most important linear molecular motor in living beings, including the motion in muscles. Right: the resulting motion step is 5.5 nm long; it has been slowed down by about a factor of ten (image and QuickTime film © San Diego State University, Jeff Sale and Roger Sabbadini).

1. *Linear molecular motors* are at the basis of muscle motion; an example is given in Figure 4. Other linear motors separate genes during cell division. Linear motors also move organelles inside cells and displace cells through the body during embryo growth, when wounds heal, and in all other cases of cell motility. Also assembler molecules, for example those that replicate DNA, can be seen as linear motors.

Challenge 6 s

A typical molecular motor consumes around 100 to 1000 ATP molecules per second, thus about 10 to 100 aW. The numbers are small; there are more astonishing if we take into account that 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 molecular motor is. At our scale, this would correspond to a car that drives, all the time, through an ongoing storm and earthquake.

Vol. I, page 92
Ref. 4
Ref. 5
Page 25

2. We encountered *rotational motors* already earlier on; nature uses them to rotate the cilia of many bacteria as well as sperm tails. Researchers have also discovered that evolution produced molecular motors which turn around DNA helices like a motorized bolt would turn around a screw. Such motors are attached at the end of some viruses and insert the DNA into virus bodies when they are being built by infected cells, or extract the DNA from the virus after it has infected a cell. The most important rotational motor, and 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 equally essential to life. They pump chemicals, such as ions or specific molecules, into every cell or out of it, using energy. They do so 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 health problems, for example for the water loss in cholera infection.

Ref. 6

In the following, we explore a few specific molecular motors found in cells. How molecules produce movement in linear motors was uncovered during the 1980s. The results

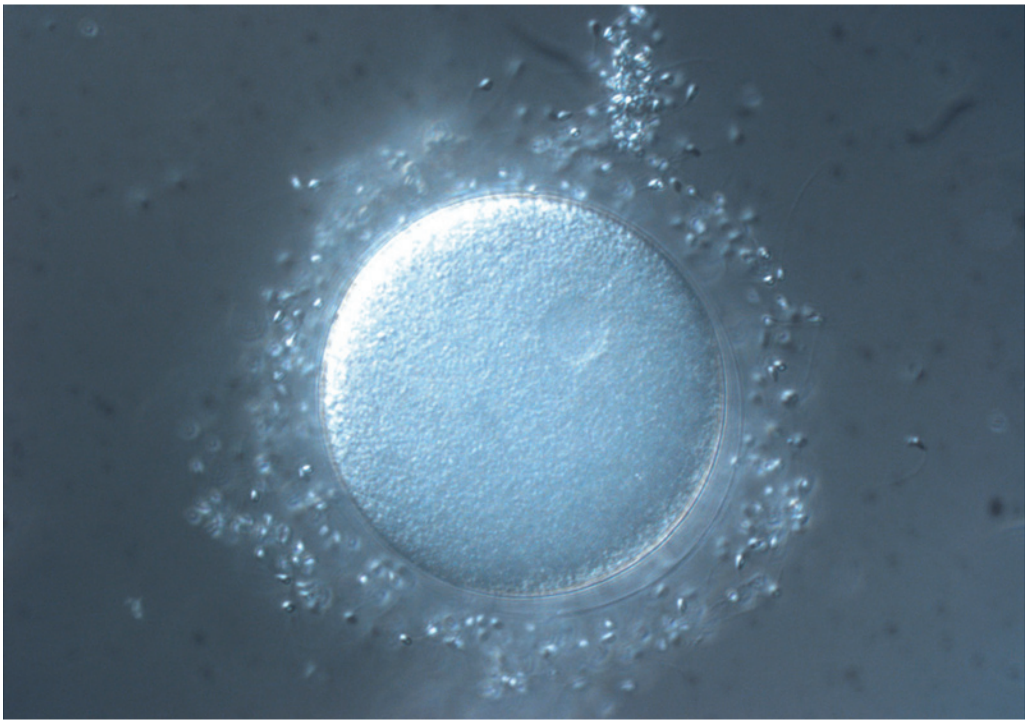


FIGURE 5 A sea urchin egg surrounded by sperm, or molecular motors in action: molecular motors make sperm move, make fecundation happen, and make cell division occur (photo by Kristina Yu, © Exploratorium www.exploratorium.edu).

started a wave of research on all other molecular motors found in nature. The research showed that molecular motors differ from most everyday motors: molecular motors do *not* involve temperature gradients, 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 4. The motion step size has been measured, with the help of some beautiful experiments, to always be an integer 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) and releases the energy contained in the chemical bond. The force generated is about 3 to 4 pN; the steps can be repeated several times a second. Muscle motion is the result of thousand of millions of such elementary steps taking place in concert. Ref. 6

Why does this molecular motor work? The molecular motor is so small that the noise due to the Brownian motion of the molecules of the liquid around it is extremely intense. Indeed, the transformation of disordered molecular motion into ordered macroscopic

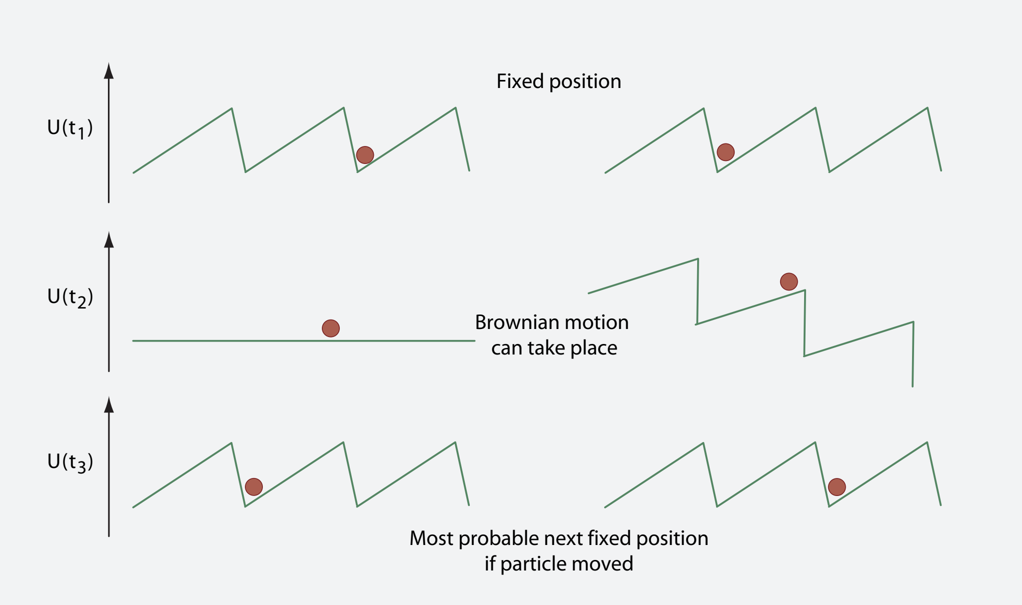


FIGURE 6 Two types of Brownian motors: switching potential (left) and tilting potential (right).

motion is one of the great wonders of nature.

Evolution is smart: with three tricks it takes advantage of Brownian motion and transforms it into macroscopic motion. (Molecular motors are therefore also called *Brownian motors*.) 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 of the ratchet, together with an energy input to make it happen. The two most important realizations are shown in Figure 6. Molecular motors thus work away from thermal equilibrium. The third trick is to take a large number of these molecular motors and to add their effects. Ref. 7

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 4 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 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. Nature then takes millions of these ratchets together: that is how our muscles work.

Engineering and evolution took different choices. A moped contains one motor. An expensive car contains about 100 motors. A human contains at least 10^{16} motors. Challenge 7 e

---

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

FIGURE 7 A classical ratchet, here of the piezoelectric kind, moves like a linear molecular motor (© PiezoMotor).

Another well-studied linear molecular motor is the kinesin–microtubule system that carries organelles from one place to the other within a cell. Like 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. These applications are pursued in various experimental research programmes.

Also *classical* ratchets exist. One example is found in every mechanical clock; also many ballpoint pens contain one. Another example of ratchet with asymmetric of mechanical steps uses the Leidenfrost effect to rapidly move liquid droplets, as shown in the video www.thisiscolossal.com/2014/03/water-maze. A further example is shown in Figure 7; indeed, many piezoelectric actuators work as ratchets and the internet is full of videos that show how they work. Piezoelectric ratchets, also called *ultrasound motors*, are found in precision stages for probe motion and inside certain automatic zoom objectives in expensive photographic cameras. Also many atomic force microscopes and scanning electron microscopes use ratchet actuators. Vol. I, page 377

Molecular motors are essential for the growth and the working of nerves and the brain. A nerve contains large numbers of dozens of molecular motors types from all three main families: dynein, myosin and kinesin. These motors transport chemicals, called ‘cargos’, along axons and all have loading and unloading mechanisms at their ends. They are necessary to realize the growth of nerves, for example from the spine to the tip of the toes. Other motors control the growth of synapses, and thus ensure that we have long-term memory. Malfunctioning molecular motors are responsible for Alzheimer disease, Huntington disease, multiple sclerosis, certain cancers and many other diseases due either to genetic defects or to environmental poisons. Ref. 8

In short, without molecular motors, we could neither move nor think.

A ROTATIONAL MOLECULAR MOTOR: ATP SYNTHASE

In cells, the usual fuel for most chemical reactions is adenosine triphosphate, or ATP. In plants, most ATP is produced on the membranes of cell organelles called *chloroplasts*, and in animal cells, in the so-called *mitochondria*. These are the power plants in most cells. ATP also powers most bacteria. It turns out that ATP is synthesized by a protein

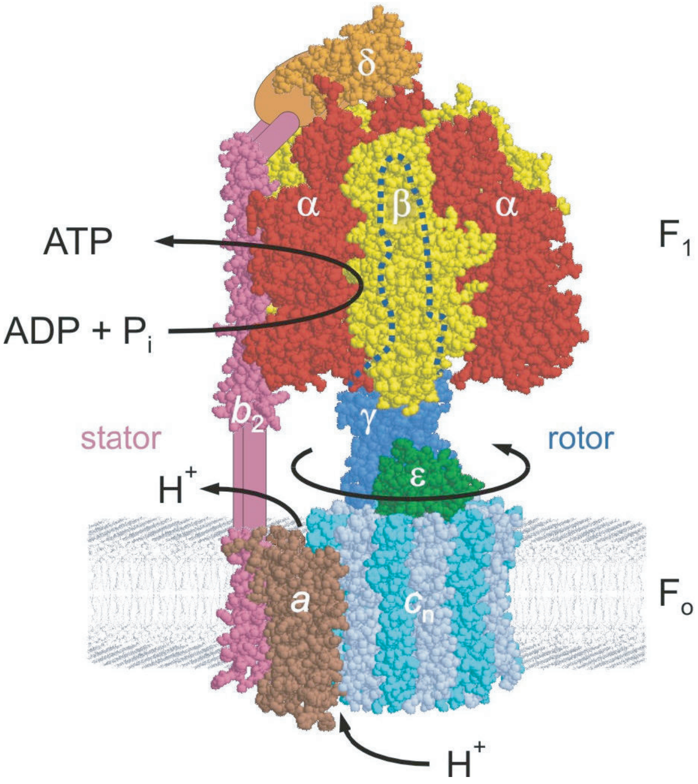


FIGURE 8 The structure of ATP synthase (© Joachim Weber).

located in membranes. The protein is itself powered by protons, H^+, which form the basic fuel of the human body, whereas ATP is the high-level fuel. For example, most other molecular motors are powered by ATP. ATP releases its energy by being changed into to adenosine diphosphate, or ADP. The importance of ATP is simple to illustrate: every human synthesizes, during a typical day, an amount of ATP that is roughly equal to his or her body mass.

The protein that synthesizes ATP is simply called *ATP synthase*. In fact, ATP synthase differs slightly from organism to organism; however, the differences are so small that they can be neglected in most cases. (An important variation are those pumps where Na^+ ions replace protons.) Even though ATP synthase is a highly complex protein, its function it easy to describe: it works like a *paddle wheel* that is powered by a proton gradient across the membrane. Figure 8 gives an illustration of the structure and the process. The research that led to these discoveries was rewarded with the 1997 Nobel Prize in Chemistry. Ref. 9

In fact, ATP synthase also works in the reverse: if there is a large ATP gradient, it pumps protons out of the cell. In short, ATP synthase is a rotational motor and molecular pump at the same time. It resembles the electric starter motor, powered by the battery, found in the older cars; in those cars, during driving, the electric motor worked as a dynamo charging the battery. (The internet also contains animations of the rotation of ATP synthase.) ATP

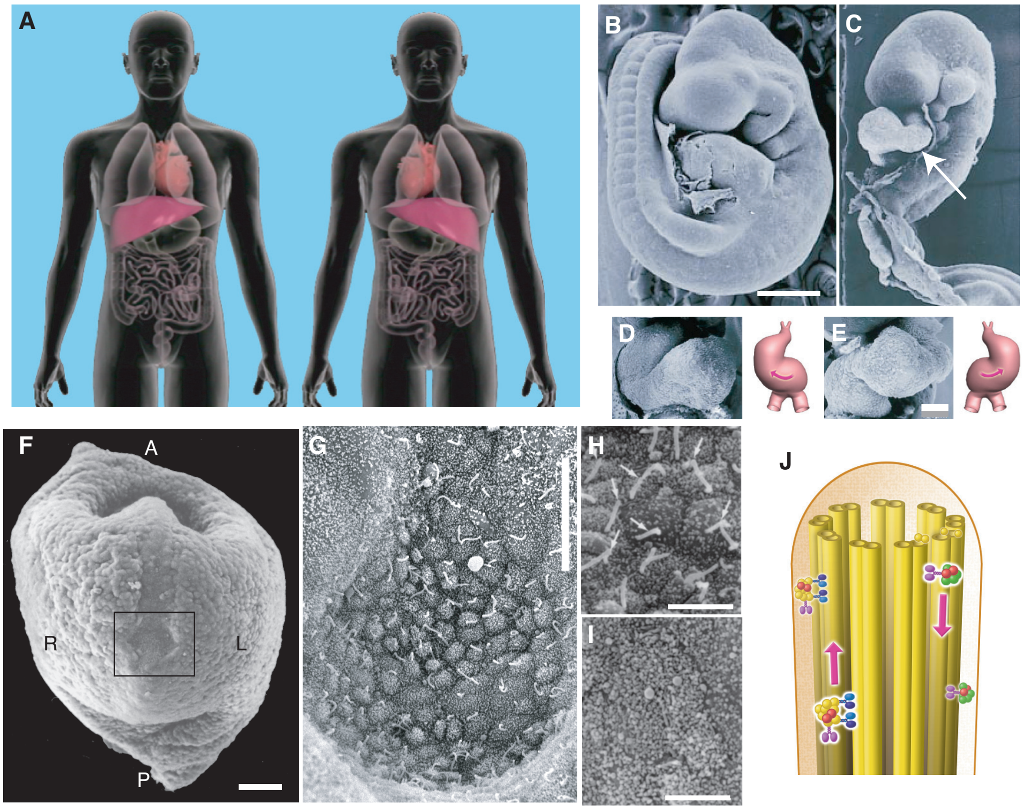


FIGURE 9 A: The asymmetric arrangement of internal organs in the human body: Normal arrangement, situs solitus, as common in most humans, and the mirrored arrangement, situs inversus. Images B to E are scanning electron micrographs of mouse embryos. B: Healthy embryos at this stage already show a right-sided tail. C: In contrast, mutant embryos with defective cilial motors remain unturned, and the heart loop is inverted, as shown by the arrow. D: Higher-magnification images and schematic representations of a normal heart loop. E: Similar image showing an inverted loop in a mutant embryo. Images F to I are scanning electron micrographs of a mouse node. F: A low-magnification view of a 7.5 day-old mouse embryo observed from the ventral side, with the black rectangle indicating the node. The orientation is indicated with the letters A for anterior, P for posterior, L for left and R for right; the scale bar is 100 μm. G: A higher-magnification image of the mouse node; the scale bar is 20 μm. H: A still higher-magnification view of healthy nodal cilia, indicated by arrows, and of the nodal pit cells; the scale bar is 5 μm. I: The nodal pit cells of mutant embryos lacking cilia. J: Illustration of the molecular transport inside a healthy flagellum (© Hirokawa Nobutaka).

synthase has been studied in great detail. For example, it is known that it produces three ATP molecules per rotation, that it produces a torque of around $20kT/2\pi$, where kT is the kinetic energy of a molecule at temperature T. There are at least 10^{16} such motors in an adult body. The ATP synthase paddle wheel is one of the central building blocks of life.

Rotational motors and parity breaking

Why is our heart on the left side and our liver on the right side? The answer of this old question is known only since a few years. The left-right asymmetry, or chirality, of human bodies must be connected to the chirality of the molecules that make up life. In all living

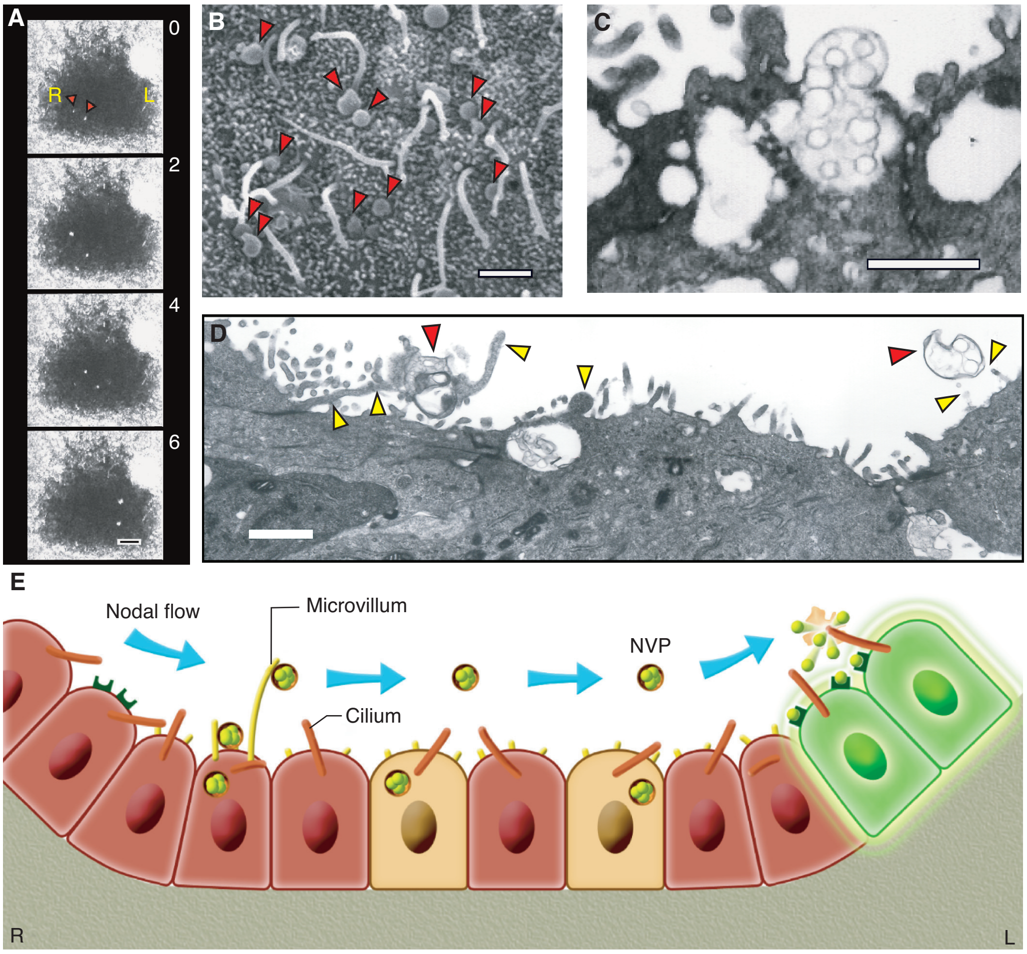


FIGURE 10 A: Optical microscope images of flowing nodal vesicular parcels (NVPs). L and R indicate the orientation. The NVPs, indicated by arrowheads, are transported to the left side by the nodal flow. The scale bar is 10 μm. B: A scanning electron micrograph of the ventral surface of nodal pit cells. The red arrowheads indicate NVP precursors. The scale bar is 2 μm. C and D: Transmission electron micrographs of nodal pit cells. The scale bar is 1 μm. E: A schematic illustration of NVP flow induced by the cilia. The NVPs are released from dynamic microvilli, are transported to the left side by the nodal flow due to the cilia, and finally are fragmented with the aid of cilia at the left periphery of the node. The green halos indicate high calcium concentration – a sign of cell activation that subsequently starts organ formation (© Hirokawa Nobutaka).

beings, sugars, proteins and DNS/DNA are chiral molecules, and in all living beings, only one of the two molecular mirror types is actually used. But how does nature translate the chirality of molecules into the chirality of a body? The answer was deduced only recently by Hirokawa Nobutaka and his team; and surprisingly, rotational molecular motors are the key to the puzzle. Ref. 10

The position of the internal organs is fixed during the early development of the embryo. At an early stage, a central part of the embryo, the so-called *node*, is covered with rotational cilia, i.e., rotating little hairs. They are shown in Figure 9. In fact, all vertebrates have a node at some stage of embryo development. The nodal cilia are powered by a molecular motor; they all rotate in the *same* (clockwise) direction about ten times per second. The rotation direction is a consequence of the chirality of the molecules that are contained in the motors. However, since the cilia are inclined with respect to the surface – towards the tail end of the embryo – the rotating cilia effectively move the fluid above the node towards the left body side of the embryo. The fluid of this newly discovered *nodal flow* contains substances – nodal vesicular parcels – that accumulate on the left side of the node and subsequently trigger processes that determine the position of the heart. If the cilia do not rotate due to a genetic defect, or if the flow is reversed by external means, the heart and other organs get misplaced. This connection also explains all the other consequences of such genetic defects or interventions.

In other words, through the rotation of the cilia and the mentioned mechanism, the chirality of molecules is mapped to the chirality of the whole vertebrate organism. It might even be that similar processes occur also elsewhere in nature, for example in the development of the brain asymmetry. This is still a field of intense research. In summary, molecular motors are truly central to our well-being and life.

Curiosities and fun challenges about biology

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

With modern microscopic methods it is possible to film, in all three dimensions, the evolution of the eye of a fruit fly until it walks away as a larva. Such a film allows to follow every single cell that occurs during the 20 hours: the film shows how cells move around during development and show every single cell division. Watch this amazing film, taken at the EMBL in Heidelberg, at youtu.be/MefTPoeVQ3w.

∗ ∗

Biological evolution can be summarized in three principles:

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.

The last principle is often called the ‘survival of the fittest’. As a result of these three principles, with each generation, *species and living beings can change*. The result of accumulated generational change is called *biological evolution*. In particular, these three principles explain the change from unicellular to multicellular life, from fish to land animals, and from animals to people.

Quantum effects are fundamental in all three principles of evolution. Of course, life and metabolism are quantum effects. The differences mentioned in the first principle are due to quantum physics: perfect copies of macroscopic systems are impossible. The second principle mentions competition; that is a kind of measurement, which, as we saw, is only possible due to the existence of a quantum of action. The third principle mentions

Vol. IV, page 122

Vol. IV, page 20

---

* ‘A *fur* is a skin that has changed beast.’

reproduction: that is again a quantum effect, based on the copying of genes, which are quantum structures. In short, biological evolution is a process due to the quantum of action. Page 18

* *

Challenge 8 d 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 9 s

It is well known that the first ever 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.

Many properties of a mammal are not determined by genes, but by the environment of the pregnancy, in particular by the womb and the birth experience. Womb influences include fur patches, skin fold shapes, but also character traits. The influence of the birth experience on the character is well-known and has been studied by many psychologists.

* *

Discuss the following argument: If nature were classical instead of quantum, there would not be just *two* sexes – nor 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 10 s

* *

Here is a well-known 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 11 r

* *

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.

* *

When the ciliary motors that clear the nose are overwhelmed and cannot work any more, they send a distress signal. When enough such signals are sent, the human body triggers the sneezing reaction. The sneeze is a reaction to blocked molecular motors.

* *

The growth of human embryos is one of the wonders of the world. The website embryo.soad.umich.edu provides extensive data, photos, animations and magnetic resonance images on the growth process.

* *

Challenge 12 s Do birds have a navel?

* *

All animals with the possibility of regenerating themselves from a small piece, such as *Planaria*, reproduce asexually, by dividing. All animals that reproduce sexually are unable to regenerating the whole animal from a small part.

* *

Challenge 13 s All animals that move with limbs are left-right symmetric. Why?

* *

Many molecules found in living beings, such as sugar, have mirror molecules. However, in all living beings only one of the two sorts is found. Life is intrinsically asymmetric. How can this be? Challenge 14 s

* *

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 in 46, in other words, about 2.2 %? Challenge 15 s

* *

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 survival time was estimated to lie 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 the laboratory. A similar result has been recently claimed by the discovery of another bacterium in a North-American salt deposit in the Salado formation. Ref. 11 Ref. 12

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 of the original sample. So the question of the longest survival time of bacteria is still open. Ref. 13

* *

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 Earth. The bacteria were still alive. They had survived for three years without food, water or air. Life can be resilient indeed. This widely quoted story is so unbelievable that it was checked again in 2011. The conclusion: the story is false; the bacteria were added by mistake in the laboratory after the return of the camera. Ref. 15

* *

TABLE 2 Approximate numbers of living species.

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

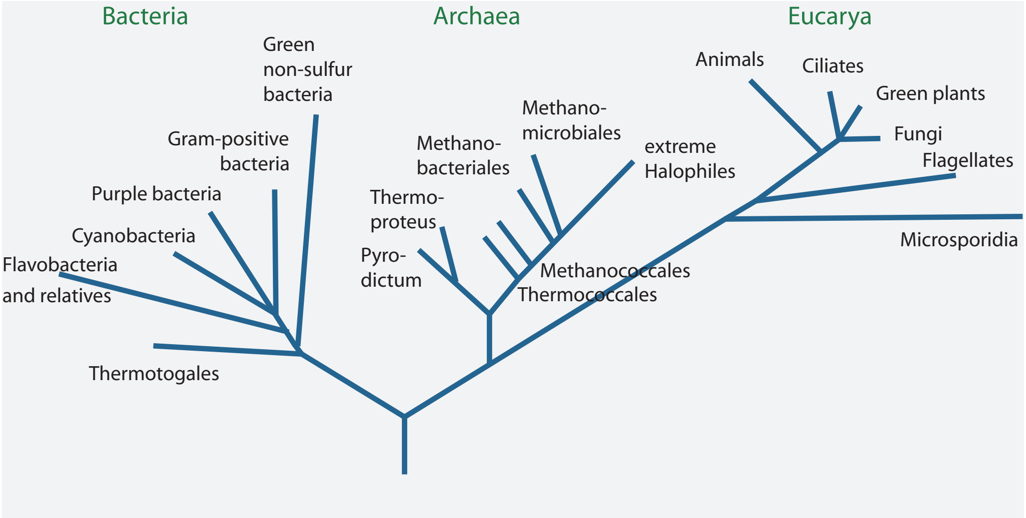


FIGURE 11 A modern version of the evolutionary tree.

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 magnitude of the task. This wealth of material can be summarized in one graph, shown in Figure 11. Newer research seems to suggest some slight changes to the picture. So far however, there still is only a single root to the tree. Ref. 16

* *

Muscles produce motion through electrical stimulation. Can technical systems do the same? Candidate are appearing: 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.

* *

Life is not a clearly defined concept. The definition used above, the ability to self-reproduce, has its limits. Can it be applied to old animals, to a hand cut off by mistake, to sperm, to ovules or to the first embryonal stages of a mammal? The definition of life also gives problems when trying to apply it to single cells. Can you find a better definition? Is the definition of living beings as 'what is made of cells' useful? Challenge 16 e

* *

Every example of growth is a type of motion. Some examples are 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 1000 species have been isolated so far.

Bacteria are essential for our life: they help us to digest and they defend us against illnesses due to dangerous bacteria. Ref. 17 In fact, the number of bacteria in a human body is estimated to be $3.8(2.0) * 10^{13}$, more than 99 % of which are in the gut. The number of cells in a adult, average human body is estimated to be $3.0(0.3) * 10^{13}$ – of which 70 to 85 % are red blood cells. In short, a human body contains more bacteria than own cells! Nevertheless, the combined mass of all bacteria in a human body is estimated to be only around 0.2 kg, because gut bacteria are much smaller than human cells.

Of the around 100 groups of bacteria in nature, the human body mainly contains species from four of them: actinobacteria, bacteroidetes, firmicutes and proteobacteria. They play a role in obesity, malnutrition, heart disease, diabetes, multiple sclerosis, autism and many other conditions. These connections are an important domain of present research.

* *

How do trees grow? When a tree – biologically speaking, 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 that tree. Just check this observation with the trees in your garden. Challenge 17 e

* *

Challenge 18 d Mammals have a narrow operating temperature. In contrast to machines, humans function only if the internal temperature is within a narrow range. Why? And does this requirement also apply to extraterrestrials – provided they exist?

* *

Challenge 19 r How did the first cell arise? This important question is still open. As a possible step towards the answer, 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 the origins of cells is one of the great open riddles of biology – though the answer will not be of much use.

* *

Challenge 20 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, so far, all statements that claim to have detected an example were *lies*. Not mistakes, but actual lies. The fantasy of extraterrestrial life poses an interesting challenge to everybody: Why would an extraterrestrial being be of interest to you? If you can answer, realize the motivation in some other way, *now*, without waiting. If you cannot answer, do something else. Challenge 21 e

* *

What could *holistic medicine* mean to a scientist, i.e., avoiding nonsense and false 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;
- *psychological* 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*.*

---

* ‘The physician helps, but nature heals.’

* *

Life is, above all, beautiful. For example, the book by CLAIRE NOUVIAN, *The Deep: The Extraordinary Creatures of the Abyss*, presented at www.press.uchicago.edu/books/nouvian/index.html allows one to savour the beauty of life deep in the ocean.

* *

What are the effects of *environmental pollution* on life? Answering this question is an intense field 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. An example is France, where the lack of bees posed a threat to the wine industry.
- 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 partly filled with concrete because its high radioactivity killed anybody that walked along it for an hour. Page 195
- 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.
- Even arctic and antarctic animals have livers full of human-produced chemical poisons.
- Burning fossil fuels raises 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.

* *

Some researchers prefer to define living beings as *self-reproducing* systems, others prefer to define them as *metabolic* systems. Among the latter, Mike Russell and Eric Smith propose the following definition of life: '*The purpose of life is to hydrogenate carbon dioxide.*' In other terms, the aim of life is to realize the reaction

$$CO_2 + 4\,H_2 \rightarrow CH_4 + 2\,H_2O\ . \qquad (1)$$

This beautifully dry description is worth pondering – and numerous researchers are indeed exploring the consequences of this view.

* *

Not only is death a quantum process, also aging is one. Research in the details of this vast field is ongoing. A beautiful example is the loss of leaves in autumn. The loss is triggered by ethene, a simple gas. You can trigger leave loss yourself, for example by putting cut apples – a strong source of ethene – together with a rose branch in a plastic bag: the roses will lose their leaves.

∗ ∗

The reanimation of somebody whose heart and breathing stopped is an useful movement sequence, called *cardiopulmonary resuscitation*. Do learn it.

∗ ∗

New research has shown that motion is important to staying healthy. In particular, it is important to do sports, but it is even more important to reduce the time of being seated. People who sit many hours per day have increased risk to get diabetes, breast cancer, white mass reduction in the brain, dementia, and various other diseases. Research into the dangerous effects of sitting is still in its infancy. For example, research has shown that sitting in front of a tv, in front of a PC or in a car for many hours a day *cannot* be compensated by doing sport.

∗ ∗

To stay fit and to ensure that you feel fit, enjoy life and enjoy the books by Mark Verstegen.

∗ ∗

Are there living beings that contain metal parts? Or is every metal object automatically not part of a living being? Astonishingly, there are exceptions. Enjoy the search. Challenge 22 e

∗ ∗

Which species of living being is the most successful, if we measure success as the species' biomass? This simple question has no known answer. Among animals, cattle (*Bos taurus*), humans (*Homo sapiens*) and Antarctic krill (*Euphasia superba*) have similar biomass values, but it is not clear whether these are the highest values. No good data seem to exist for plants – except for crops. It is almost sure that several species of bacteria, such as from the marine genus *Prochlorococcus* or some other bacteria species found in soil, and several species of fungi achieve much higher biomass values. But no reliable overview is available.

∗ ∗

Trees move in many interesting ways. For example, trees fight with their neighbours over space and access to light and nutrients. This occurs with most vehemence if the neighbour is of another species. Most trees do not like to be touched by other trees – but there are exceptions, such as beeches. For example, when beeches fight with oaks, after a few years, the oak is left with little space and light, and the beech has taken over most of it. But trees also help neighbours, for example in case of sickness, by providing nutrients and water. Many more fascinating stories about trees – including the way the communicate via aiborne chemical signals such as ethylene (ethene) – are told by Peter Wohlleben, *Das geheime Leben der Bäume: Was sie fühlen, wie sie kommun-*

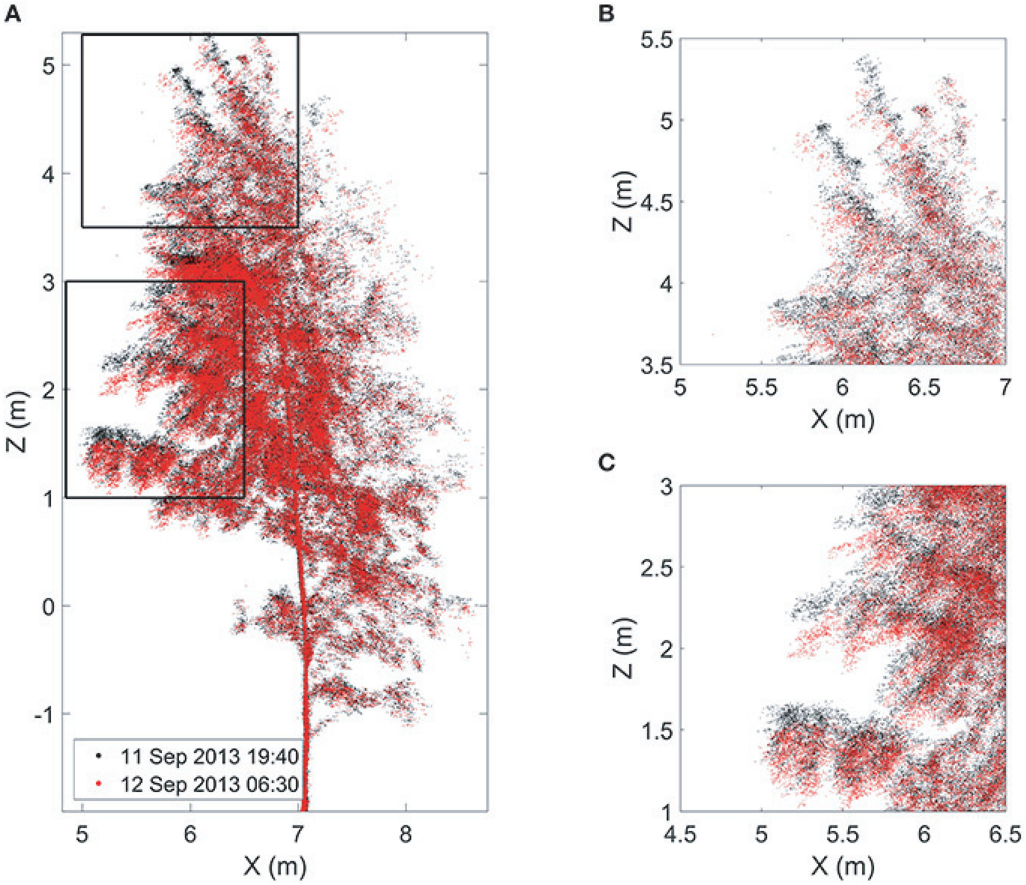


FIGURE 12 Branch and leaf position of a birch at the end of the day (black) and in the early morning (red), with magnified sections shown on the right (© Eetu Puttonen et al.).

izieren – die Entdeckung einer verborgenen Welt, Ludwig Verlag, 2015.

* *

Ref. 18 Trees sleep at night and get up in the morning. The observation is known since centuries; a beautiful measurement of the effect, using a laser scanner, was performed by Eetu Puttonen and his group. Their results, shown in Figure 12, show that the height of a typical birch branch and its leaves in the early morning is up to 10 cm lower than during the day. The measurements also show that the trees move most in the early morning, when they wake up. The origin of these effects seems to be the difference of water intake during day and night.

THE PHYSICS OF PLEASURE

> “What is mind but motion in the intellectual sphere?”
>
> Oscar Wilde, *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 built 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 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.

Research into pleasure and biological sensors is a fascinating field that is still evolving; here we can only have a quick tour of the present knowledge.

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 pressures vary 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 (for 20 Hz) to 17 mm (for 20 kHz). In this range, the ear, with its 16 000 to 20 000 hair cells and 30 000 cochlear neurons, is able to distinguish at least 1500 pitches. But the ear is also able to distinguish nearby frequencies, such as 400 and 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 $1'$, or 0.29 mrad, and the capacity to consciously detect down to 60 *incident* photons in 0.15 s, or 4 *absorbed* photons in the same time interval.

Each 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.

Vol. III, page 199 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 Ref. 19 simple test shown in Figure 13. The sensitivity difference between the colour-detecting cones and the colour-blind rods is the reason that at night all cats are grey.

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 eye also contains about 1 million retinal ganglion cells. All signals from the eye are transmitted through 1 million optical nerve fibres to a brain region, the virtual cortex, that contains over 500 million cells.

Human *touch sensors* are distributed over the skin, with a surface density which varies

FIGURE 13 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.

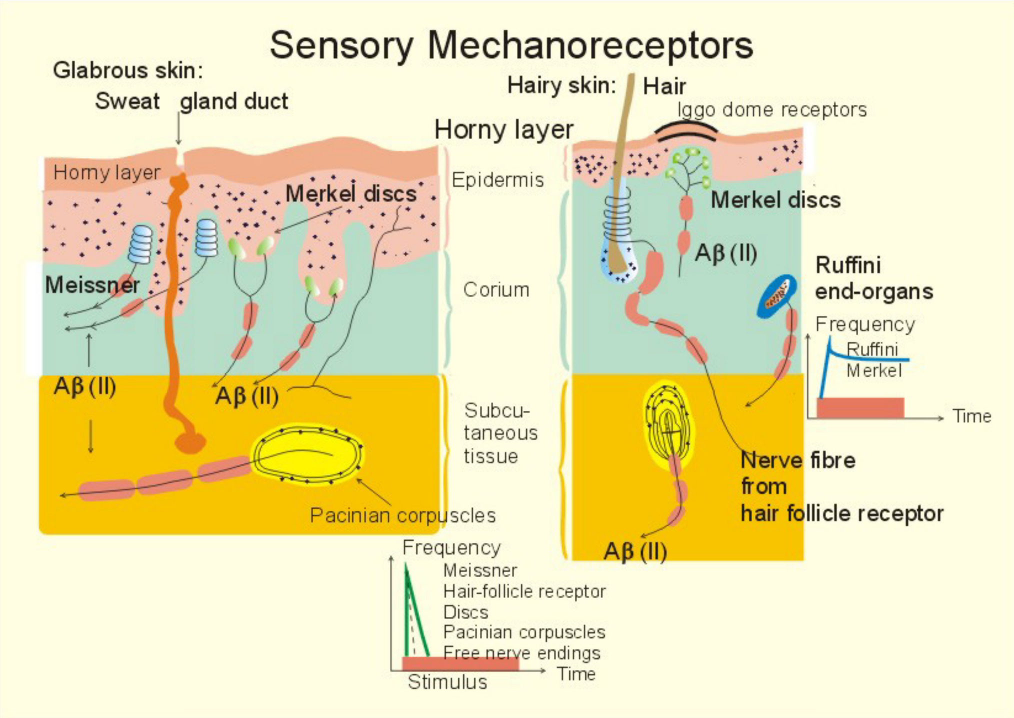


FIGURE 14 The five sensors of touch in humans, from the most to the least common ones: Meissner’s corpuscles, Merkel cells, Ruffini corpuscles, Pacinian corpuscles and hair receptors.

from one region to the other. The density is lowest on the back and highest in the face and on the tongue. The hand has about 17 000 tactile receptors, most of them at the finger tips. There are separate sensors for light touch (Meissner’s corpuscles) and pressure (Merkel cells), for deformation (Ruffini corpuscles), for vibration (Pacinian corpuscles), and for tickling (unmyelinated fibers); there are additional separate sensors for heat, for coldness,* and for pain. Some of the sensors, whose general appearance is shown in Fig-

---

* There are *four* sensors for heat; one is triggered above 27°C, one above 31°C, one above 42°C, and one

ure 14, 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 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.

The *taste sensor* 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. Ref. 20 The tongue, palate and cheeks have about 10 000 taste buds, 90 % of which are on the tongue. Each taste bud has between 50 and 150 receptors; their diameter is around 10 μm.

In ancient Greece, 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 food industry. Research is also ongoing to find substances to block taste receptors; one aim is to reduce the bitterness of medicines or of food.

The *nose* has about 350 different smell receptors and a total of about 40 million receptor cells. (Dogs have 25 times more.) Through the possible combinations it is estimated that the nose can detect about 10 000 different smells.** Together with the six 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 checking wine or cheese during their production, thus making their inventor extremely rich. Challenge 24 ny At the moment, humans, with all their technology at their disposal, 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.

---

above 52°C. The sensor for temperatures above 42°C, TRPV1, is also triggered by capsaicin, the sharp chemical in chilli peppers.

There seems to be only *one* sensor for coldness, the ion channel TRPM8, triggered between 8 and 26°C. It is also triggered by menthol, a chemical contained in mojito and mint. Coldness neurons, i.e., neurons with TRPM8 at their tips, can be seen with special techniques using fluorescence and are known to arrive into the teeth; they provide the sensation you get at the dentist when he applies his compressed air test.

[*] 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 23 s

[**] Linda Buck and Richard Axel received the 2004 Nobel Prize in Physiology or Medicine for their unravelling of the working of the sense of smell.

Vol. III, page 32 Other animals feature additional types of sensors. Sharks can *feel electrical fields*. Many snakes have *sensors for infrared light*, such as the pit viper or vampire bats. These sensors are used to locate prey or food sources. Some beetles, such as *Melanophila acuminata*, can also detect infrared; they use this sense to locate the wildfires they need to make their eggs hatch. Also other insects have such organs. 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 *hear ultrasound* up to 100 kHz and more. Whales and elephants can detect and localize *infrasound* signals. Vol. I, page 325

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 possible. Studying physics is one way to do this. Ref. 21

> “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, *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 the brain. Vol. I, page 315 In the brain itself, the arriving signals are *classified* and *stored*, sometimes for a short time, sometimes for a long time. Most storage mechanisms take place in the structure and the connection strength between brain cells, the *synapses*, as we have seen. Vol. III, page 265 The process remaining to understand is the classification, a process we usually call *thinking*. 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 or understanding in terms of signal motions.

---

* 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. The author heard him making the above statement in 1982, during one of his lectures.

** This is not in contrast with the fact that a few *whale* species have brains with a 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.

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 motion.

Living clocks

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

Vol. I, page 44 We have given an overview of living clocks already at the beginning of our adventure. They are common in bacteria, plants and animals. And as Table 3 shows, without biological clocks, neither life nor pleasure would 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 Ref. 22 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.

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, Ref. 23 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 Ref. 24 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.

The clocks with the longest cycle in the human body control *ageing*. One of the more famous ageing clock limits the number of divisions that a cell can undergo. Indeed, the number of cell divisions is finite for most cell types of the human body and typically lies between 50 and 200. (An exception are reproductory cells – we would not exist if they

---

* '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 nucleus (SCN), circadian hormone concentration, temperature, etc.; leads to jet lag	90 ks
	skin clock	circadian
	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 (photoperiodism); 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

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. 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 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 interesting, complex, and unclear.

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. Ref. 25 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.

Even though not all the clock mechanisms in humans are known, biological clocks share a property with all human-built and all non-living 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.

When we explored general relativity we found out that purely gravitational clocks do not exist, Vol. II, page 282 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.

Ref. 26 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, as a mathematical theorem. Challenge 25 ny

Take a mechanical pendulum clock. In all such clocks the weight has to stop when the chain end is reached. More generally, all clocks have to stop when the battery or the energy source is empty. In other words, in all real clocks the Hamiltonian is bounded

---

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

from below. And the above theorem from quantum theory then states that such a clock cannot really work.

In short, quantum theory shows that exact *clocks do not exist in nature*. Quantum theory states that any clock can only be *approximate*. Time cannot be measured exactly; time can only be measured approximately. 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?

High precision implies high sensitivity to fluctuations. Now, all clocks have an oscillator inside, e.g., a motor, that makes them work. A high precision clock thus needs a high precision oscillator. In all clocks, the position of this oscillator is read out and shown on the dial. Now, the quantum of action implies that even the most precise clock oscillator has a position indeterminacy. The precision of any clock is thus limited.

Worse, like any quantum system, any clock oscillator even has a small, but finite probability to stop or to run backwards for a while. You can check this conclusion 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. The clock 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 26 e

In summary, 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. For example, this is regularly done in atomic optical clocks.

THE PRECISION OF CLOCKS

Given the limitations due to quantum theory, what is the ultimate accuracy τ of a clock? To start with, the indeterminacy relation provides the limit on the mass of a clock. The clock mass M must obey Challenge 27 ny

$$M > \frac{\hbar}{c^2 \tau} \qquad (2)$$

Challenge 28 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. 27

$$M > \frac{\hbar}{c^2 \tau} \frac{T}{\tau} \qquad (3)$$

where T is the time to be measured. You might check that this condition directly requires that any clock must be *macroscopic*. Challenge 29 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

$$l > \sqrt{\frac{\hbar t}{m}} \ . \qquad (4)$$

Ref. 28 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 (4) 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 of a single atom. In fact, such oscillations promise to be the most precise human built clocks. But the oscillator is only one part of any clock, as explained above. Ref. 29 Page 42

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

$$\frac{\delta\nu}{\nu} = \sqrt{\frac{\Delta E}{E_{\text{tot}}}} \qquad (5)$$

where $\Delta E = \hbar/T$ is the energy indeterminacy stemming from the finite measuring time T and $E_{\text{tot}} = NE_{\text{bind}}$ 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 become a central issue. Vol. VI, page 65

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

Nature limits predictions in four ways:

1. We have seen that quantum theory, through the uncertainty relations, limits the precision of measurements, and of clocks and time measurements in particular. Thus,

the quantum of action makes it hard to determine initial states to full precision – even for a single particle.

Vol. I, page 239

2. We have seen that high numbers of particles make it difficult to predict the future due to the often statistical nature of their initial conditions.
3. We have found in our adventure that predictions of the future are made difficult by non-linearities and by the divergence from nearby initial conditions. Vol. I, page 424
4. We have seen that a non-trivial space-time topology can limit predictability. For example, we will discover that black hole and horizons can limit predictability due to their one-way inclusion of energy, mass and signals. Vol. II, page 110
5. We will find out in the last pat of our adventure that quantum gravity effects even make a precise definition of time and space impossible. Vol. VI, page 40

Measurements and practical predictability are thus limited. The central reason for this limitation is the quantum of action. But if the quantum of action makes perfect clocks impossible, is determinism still the correct description of nature? And does time exist after all? The answer is clear: 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 practice, in all quantum systems both determinism and the concept of time remain applicable.

This conclusion is valid even though theory says otherwise. Our ability to enjoy the pleasures due to the flow of time remains intact.

However, when extremely large momentum flows or extremely large dimensions need to be taken into account, quantum theory cannot be applied alone; in those cases, general relativity needs to be taken into account. The fascinating effects that occur in those situations will be explored in detail later on. Vol. VI, page 57

DECAY AND THE GOLDEN RULE

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

All pleasure only makes sense in the face of death. And death is a form of decay. Decay is any spontaneous change. Like the wave aspect of matter, decay is a process with no classical counterpart. Of course, any decay – including the emission of light by a lamp, the triggering of a camera sensor, radioactivity or the ageing of humans – can be observed classically; however, the *origin* of decay is a pure quantum effect.

In any decay of unstable systems or particles, the decoherence of superpositions of macroscopically distinct states plays an important role. Indeed, experiments confirm that the prediction of decay for a specific system, like a scattering of a particle, is only possible on *average*, for a large number of particles or systems, and never for a single one. These observations confirm the quantum origin of decay. 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, Vol. IV, page 143 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.

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

$$\dot{N} = -\frac{N}{\tau} \quad \text{where} \quad \frac{1}{\tau} = \frac{2\pi}{\hbar} \left| \langle \psi_{\text{initial}} | H_{\text{int}} | \psi_{\text{final}} \rangle \right|^2 . \qquad (6)$$

This result for $\dot{N} = \mathrm{d}N/\mathrm{d}t$ 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

Challenge 30 e

$$N(t) = N_0\, \mathrm{e}^{-t/\tau} . \qquad (7)$$

Decay is thus predicted to follow an exponential behaviour, 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 almost a century, all experiments confirmed that quantum 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. 30

Ref. 31

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

Challenge 31 s

THE PRESENT IN QUANTUM THEORY

> “Utere temporibus.**”
>
> Ovidius

---

* Originally, the golden rule is a statement from the christian bible, (Matthew 7,12) namely the precept ‘Do to others whatever you would like them to do to you.’

** ‘Use the occasions.’ *Tristia* 4, 3, 83

Many sages advise to enjoy the present. As shown by perception research, what humans call 'present' has a duration of between 20 and 70 milliseconds. This result on the biological present leads us to ask whether the *physical* present might have a duration as well. Ref. 32

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 the flight of an arrow, he assumed that this is possible. However, quantum theory changes the situation.

Can we really say that a moving system is at a given spot at a given time? In order to find an answer through experiment, 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. All such investigations confirm: Whenever we reduce shutter times as much as possible, observations become unsharp. The lack of sharpness is due to the quantum of action. 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.

In summary, the indeterminacy relation and the smallest action value prevent that moving objects are at a fixed position at a given time. Zeno's discussion was based on an extrapolation of classical concepts 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 span.

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, physical *present* is always an average over a non-vanishing interval of time. In nature, the present has a finite duration. To give a rough value that guides our thought, in most situations the length of the present will be less than a yoctosecond, so that it can usually be neglected.

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 *only approximately* the change of position with time.

Why then can we observe and 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 in practice, quantum evolution

---

* Also the discussion of the quantum Zeno effect, below, does not change the conclusions of this section. Page 50

is applicable in all cases. As long as energy and time have no limits, all problems are avoided, and *motion is a time sequence of quantum states*.

In summary, we can observe motion because for any known observation energy we can find a still higher energy and a still longer averaging time that can be used by the measurement instruments to define space and time with higher precision than for the system under observation. In the final part of our mountain ascent, we will discover that there is a maximum energy in nature, so that we will need to change our description in those situations. However, this energy value is so huge that it does not bother us at all at the present point of our exploration.

Vol. VI, page 46

Rest and the quantum Zeno effect

The quantum of action implies that there is no rest in nature. Rest is thus 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. But there is another apparent case of rest in quantum theory, the *quantum Zeno effect*. Usually, observation *changes* the state of a system. However, for certain systems, observation can have the opposite effect, and *fix* a system.

Quantum mechanics predicts that an unstable particle can be 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 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. In the meantime, other groups have confirmed the measurements. Thus, quantum theory has been confirmed also in this surprising aspect. Ref. 33

The quantum Zeno effect is also connected to the deviations from exponential decay – due to the golden rule – that are predicted by quantum theory. Indeed, quantum theory predicts that every decay is exponential only for intermediate times, and quadratic for short times and polynomial for extremely long times. These issues are research topics to this day.

Ref. 34 In a fascinating twist, in 2002, 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.

In summary, the quantum Zeno effect does not contradict the statement that there is no rest in nature; in situations showing the effect, there is an non-negligible interaction between the system and its environment. The details of the interaction are important: in certain cases, frequent observation can actually accelerate the decay or evolution. Quantum physics still remains a rich source of fascinating effects. Ref. 35

CONSCIOUSNESS – A RESULT OF THE QUANTUM OF ACTION

In the pleasures of life, consciousness plays an essential role.

Vol. III, page 339

▷ *Consciousness* is our ability to observe what is going on in our mind.

This activity, like any type of change, can itself be observed and studied. Though it is hard and probably impossible to do so by introspection, we can study consciousness in others. 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 located on Earth 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.

Modern imaging machines 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 imaging machines are based on magnetic resonance, as described below. Another, more questionable imaging technique, positron tomography, works by letting people swallow radioactive sugar. Both techniques confirm 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. Page 162

Today, we are thus in the same situation as material scientists were a century ago: they knew that matter is made of charged particles, but they could not say *how* matter is built up. Similarly, we know today 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 neurology. This is one of the great challenges of twenty-first century science.

Can you add a few arguments to the ones given here, showing that consciousness is a physical process? Can you show in particular that not only the consciousness of others, but also your own consciousness is a quantum process? Can you show, in addition, that despite being a quantum process, coherence plays no essential role in consciousness? Challenge 32 s

In short, our consciousness is a consequence of the matter that makes us up. Consciousness and pleasure depend on matter, its interactions and the quantum of action.

WHY CAN WE OBSERVE MOTION? – AGAIN

Studying nature can be one of the most intense pleasures of life. All pleasures are based on our ability to observe or detect motion. And our human condition is central to this ability. In particular, in our adventure so far we found the following connections: We

experience motion

- only because we are of finite size, and in particular, because we are large compared to our quantum mechanical wavelength (so that we do not experience wave effects in everyday life),
- only because we are large compared to a black hole of our same mass (so that we have useful interactions with our environment),
- only because we are made of a large but finite number of atoms (to produce memory and enable observations),
- only because we have a limited memory (so that we can clear it),
- only because we have a finite but moderate temperature (finite so that we have a lifetime, not zero so that we can be working machines),
- only because we are a mixture of liquids and solids (enabling us to move and thus to experiment),
- only because we are approximately electrically neutral (thus avoiding that our sensors get swamped),
- only because our brain forces us to approximate space and time by continuous entities (otherwise we would not form these concepts),
- only because our brain cannot avoid describing nature as made of different parts (otherwise we would not be able to talk or think),
- only because our ancestors reproduced,
- only because we are animals (and thus have a brain),
- only because life evolved here on Earth,
- only because we live in a relatively quiet region of our galaxy (which allowed evolution), and
- only because the human species evolved long after the big bang (when the conditions were more friendly to life).

If any of these conditions – and many others – were not fulfilled, we would not observe motion; we would have no fun studying physics. In fact, we can also say: if any of these conditions were not fulfilled, motion would not exist. In many ways motion is thus an illusion, as Zeno of Elea had claimed a long time ago. Of course, motion is a *inevitable* illusion, one that is shared by many other animals and machines. To say the least, the observation and the concept of motion is a result of the properties and limitations of the human condition. A complete description of motion and nature must take this connection into account. Before we attempt that in the last volume of this adventure, we explore a few additional details.

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 33 s

* *

The sense of smell is quite complex. For example, the substance that smells most badly to humans is *skatole*, also called, with his other name, *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. Skatole does not smell bad to all animals; in contrast to humans, flies are attracted by its smell. Skatole is also produced by some plants for this reason.

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

* *

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 energies as small as about 10 μJ. Is one of the two systems relativistic? Challenge 34 s

* *

The human construction plan is stored in the DNA. The DNA is structured into 20 000 genes, which make up about 2 % of the DNA, and 98 % non-coding DNA, once called ‘junk DNA’. Humans have as many genes as worms; plants have many more. Only around 2010 it became definitely clear, through the international ‘Encode’ project, what the additional 98 % of the DNA do: they switch genes on and off. They form the *administration* of the genes and work mostly by binding to specific proteins.

Research suggests that most genetic defects, and thus of genetic diseases, are not due to errors in genes, but in errors of the control switches. All this is an ongoing research field.

* *

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 35 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. 37

∗ ∗

The importance of ion channels in the human body can not be overstressed. Ion channels malfunctions are responsible for many infections, for certain types of diabetes and for many effects of poisons. But above all, ion channels, and electricity in general, are essential for life. Ref. 38

∗ ∗

Our body contains many systems that avoid unpleasant outcomes. For example, it was discovered in 2006 that saliva contains a strong pain killer, much stronger than morphine; it is now called *opiorphin*. It prevents that small bruises inside the mouth disturb us too much. Opiorphine also acts as an antidepressant. Future research has to show whether food addiction is related to this chemical.

∗ ∗

It is still unknown why people – and other mammals – *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, shown in Figure 16, 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 developed 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 of the pleasure system are possible. One, shown in Figure 15, is the ‘neurochemical mobile’ model of the brain. Ref. 39 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.

---

[*] Neurotransmitters come in many types. They can be grouped into *mono- and diamines* – such as dopamine, serotonin, histamine, adrenaline – *acetylcholine*, *amino acids* – such as glycine, GABA and glutamate – *polypeptides* – such as oxytocin, vasopressin, gastrin, the opiods, the neuropetides etc. – *gases* – such as NO and CO – and a number of molecules that do not fit into the previous classes – such as anandamide or NAAG.

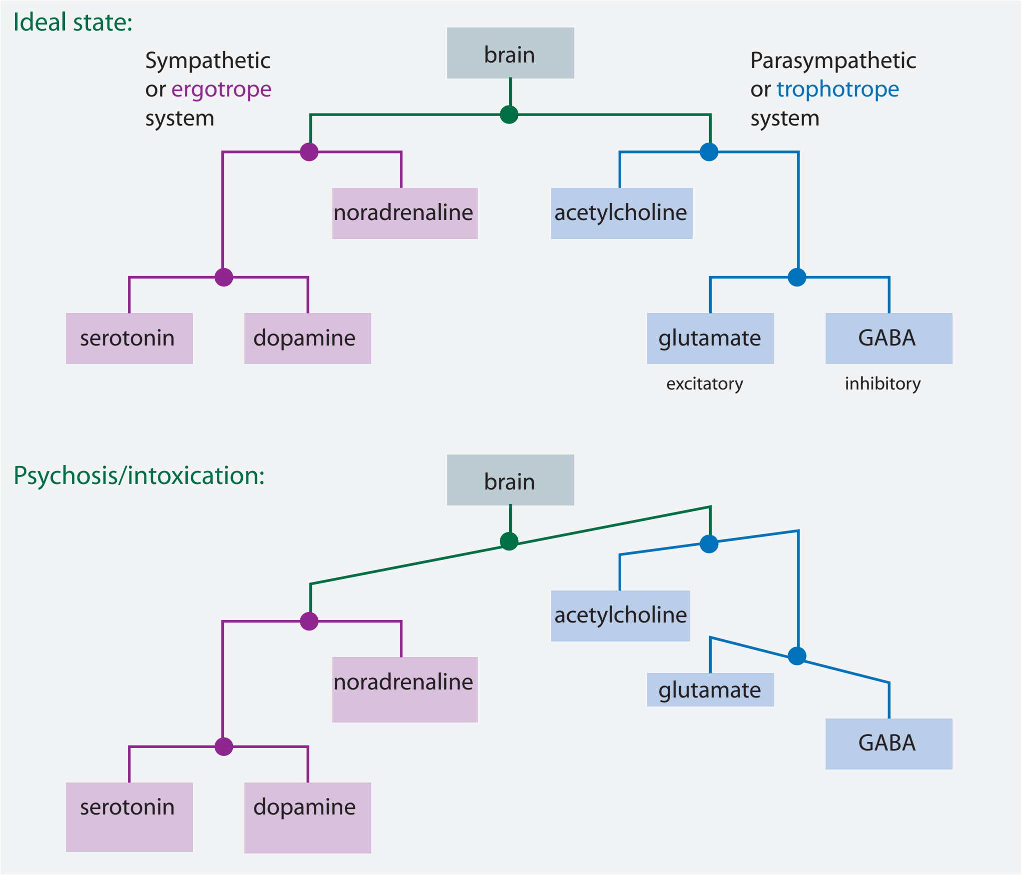


FIGURE 15 The 'neurochemical mobile' model of well-being, with one of the way it can get out of balance.

* *

Ref. 40 The pleasure system in the brain 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 love is found in many animal species. Romantic love is a kind of positive addiction, and works like cocaine. In short, in life, we can all choose between addiction and love.

* *

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 36 s

* *

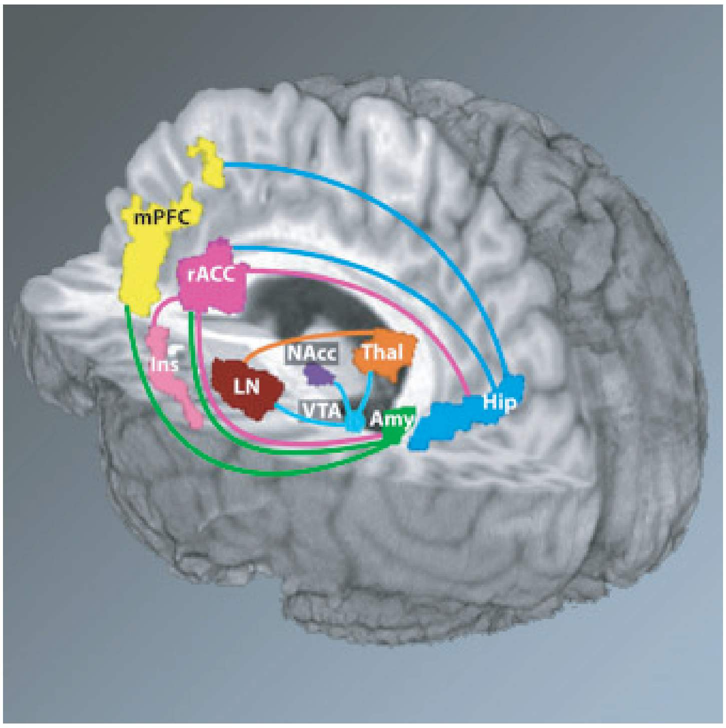


FIGURE 16 The location of the ventral tegmental area (VTA) and of the nucleus accumbens (NAcc) in the brain. The other regions involved in pleasure and addiction are Amy, the amygdala, Hip, the hippocampus, Thal, the thalamus, rACC, the rostral anterior cingulate cortex, mPFC, the medial prefrontal cortex, Ins, the insula, and LN, the lentiform nucleus (courtesy NIH).

We all know the smell that appears in the open field at the start of a summer rainfall, or the smell of fresh earth. It is due to a substance called *geosmin*, a bicyclic alcohol that is produced by bacteria in the soil. The bacteria produce it when it rains. For reasons not fully understood, the human nose is especially sensitive to the smell of geosmin: we are able to smell it at concentrations below 10^{-10}.

* *

Also plants have sensors. Plants can sense light, touch, gravity, chemicals, as well as electric fields and currents. Many plants grow differently when touched; roots sense and conduct electric signals and then grow accordingly; obviously, plants grow against gravity and towards light. Above all, plants are known to be able to distinguish many different chemicals in the air, above all, ethene (ethylene), which is also an important plant hormone.

* *

Many plants have built-in sensors and clocks that measure the length of the day. For example, *spinach* does not grow in the tropics, because in order to flower, spinach must sense for at least fourteen days in a row that the day is at least 14 hours long – and this never happens in the tropics. This property of plants, called *photoperiodism*, was discovered in 1920 by Wightman Garner and Harry Allard while walking across tobacco fields. They discovered – and proved experimentally – that tobacco plants and soya plants only flower when the length of the day gets sufficiently short, thus around September. Garner and Allard found that plants could be divided into species that flower when days are short – such as chrysanthemum or coffee – others that flower when days are long – such as carnation or clover – and still others that do not care about the length of the day at all – such as roses or tomatoes. The measurement precision for the length of the day is around 10 min. The day length sensor itself was discovered only much later; it is located in the leaves of the plant, is called the *phytochrome system* and is based on specialised proteins. The proteins are able to measure the ratio between bright red and dark red light and control the moment of flower opening in such plants.

Summary on biology and pleasure

To ensure the successful reproduction of living beings, evolution has pursued miniaturization as much as possible. Molecular motors, including molecular pumps, are the smallest motors known so far; they work as quantum ratchets. Molecular motors are found in huge numbers in every living cell. In short:

▷ Every human consists of trillions of machines.

To increase pleasure and avoid pain, evolution has also supplied the human body also with numerous sensors, sensor mechanisms, and a pleasure system deep inside the brain. In short, nature has invented *pleasure* as a guide for behaviour. Neurologists have thus proven what Epicurus said 23 centuries ago and Sigmund Freud repeated one century ago:

▷ Pleasure controls human life.*

All biological pleasure sensors and pleasure systems are based on quantum motion, in particular on chemistry and materials science. We therefore explore both fields in the following.

---

* But 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 honourably and justly without living pleasantly.’ This is one of his *Principal Doctrines*.

Chapter 2

CHANGING THE WORLD WITH QUANTUM EFFECTS

The discovery of quantum effects 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, materials science, electronics and medicine would not have been possible without quantum effects. Many other improvements of our everyday life are due to quantum physics, and many are still expected. In the following, we give a short overview of this vast field.

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. In fact, 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. 41

To understand molecules and everyday matter, the first step is to understand atoms. The early quantum theorists, lead by Niels Bohr, dedicated their life to understanding their atoms and their detailed structure. The main result of their efforts is what you learn in secondary school: in atoms with more than one electron, the various electron clouds form spherical layers around the nucleus. The electron layers can be grouped into groups of related clouds that are called *shells*. For electrons outside the last fully occupied 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. The relation is shown in Figure 17.

Quantum theory shows that the first atomic shell has room for two electrons, the second for 8, the third for 18, and the general n-th shell for $2n^2$ electrons. The (neutral)

[**] ‘Beer makes stupid.’

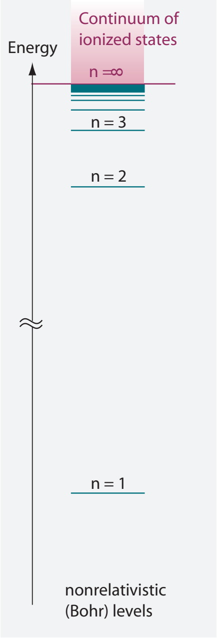


FIGURE 17 The principal quantum numbers in hydrogen.

atom with one electron is hydrogen, the atom with two electrons is called helium. Every chemical element has a specific number of electrons (and the same number of protons, as we will see). A way to picture this connection is shown in Figure 18. It is called the Ref. 42 *periodic table of the elements*. The standard way to show the table is found on page 345 and, more vividly, in Figure 19. (For a periodic table with a video about each element, see www.periodicvideos.com.)

Experiments show that different atoms that share the *same* number of electrons in their outermost shell show *similar* chemical behaviour. Chemists know that the chemical behaviour of an element is decided by the ability of its atoms to from bonds. For example, the elements with one electron in their outer 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.

Atomic 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 20. This distance is due to an energy minimum: a lower distance would lead to electrostatic repulsion

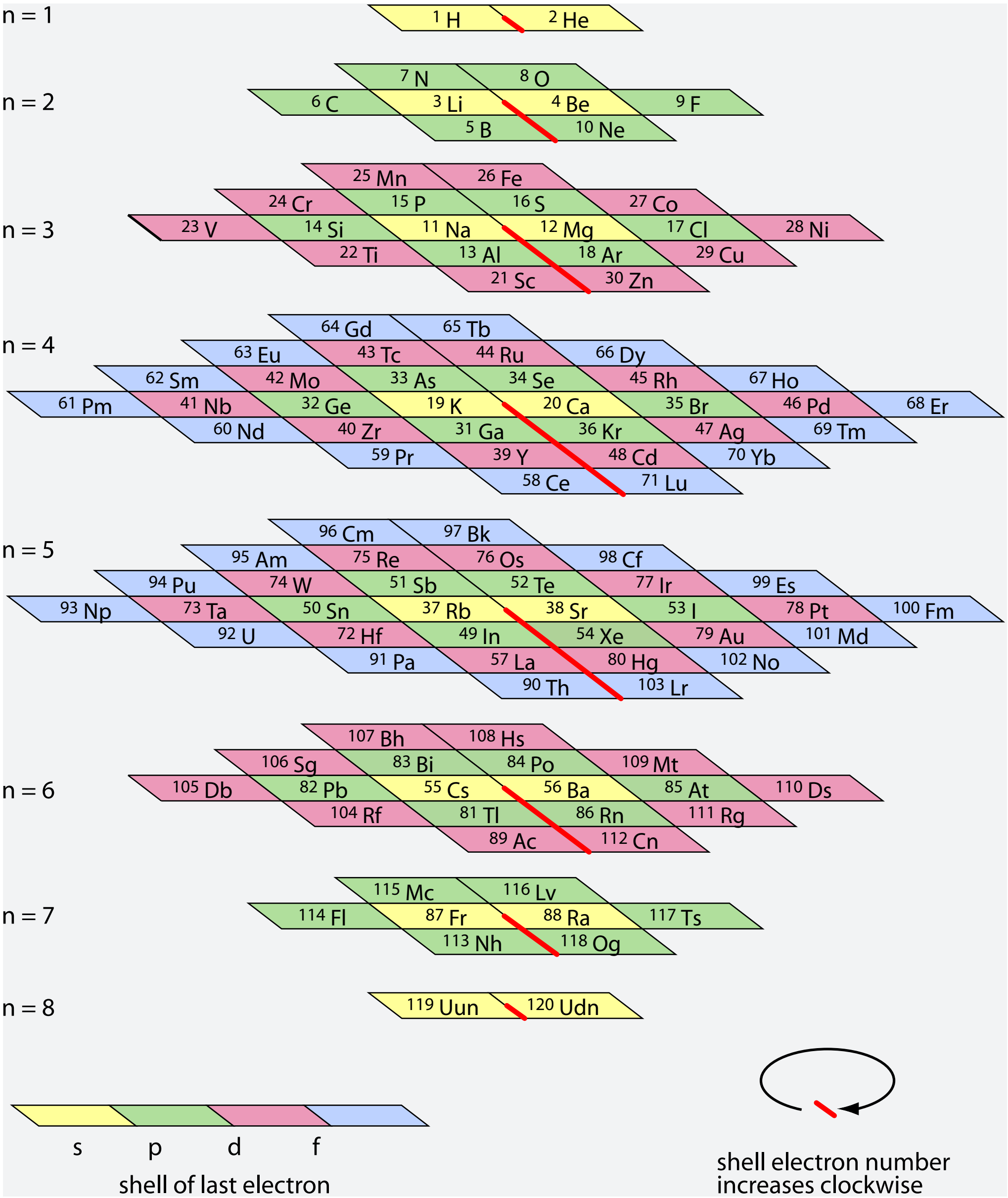


FIGURE 18 An unusual form of the periodic table of the elements.

between the atomic cores, a higher distance would increase the electron cloud energy.

Many atoms can bind to more than one neighbours. In this case, energy minimization also leads to specific bond angles, as shown in Figure 21. Maybe you remember those funny pictures of school chemistry about orbitals and dangling bonds. Such dangling bonds can now be measured and observed. Several groups were able to image them using scanning force or scanning tunnelling microscopes, as shown in Figure 22.

Ref. 43

Ref. 44

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,

Challenge 37 e

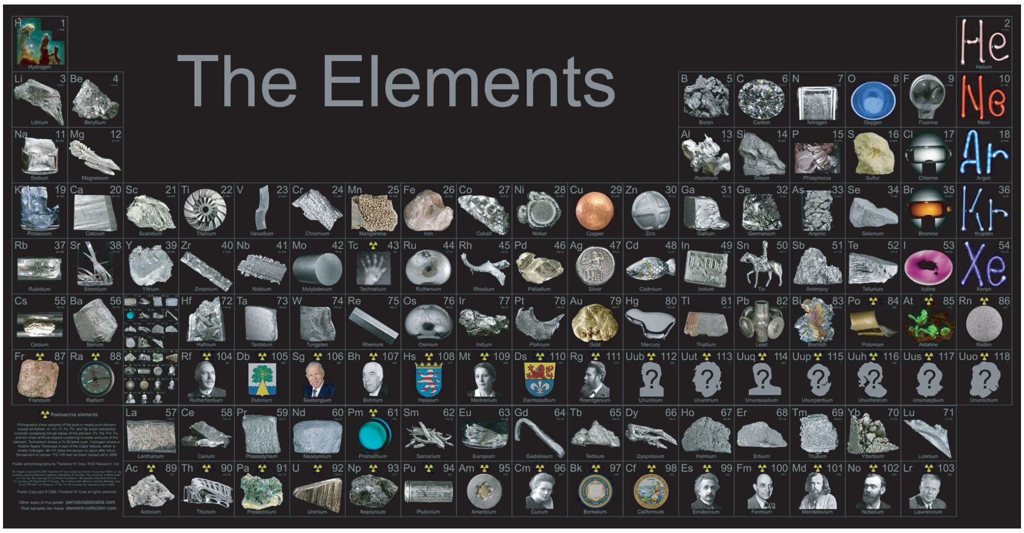


FIGURE 19 A modern periodic table of the elements (© Theodore Gray, for sale at www.theodoregray.com).

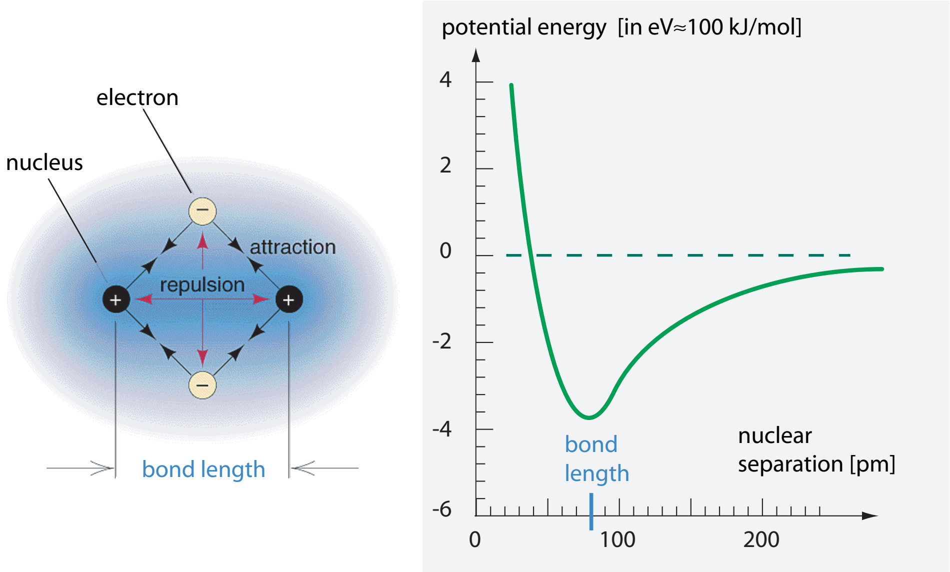


FIGURE 20 The forming of a chemical bond between two atoms, and the related energy minimum (left hand image © chemistry4gcms2011.wikispaces.com).

the H–O–H angle in water molecules is 107°.

Atoms can also be connected by *multiple* bonds. Double bonds appear in carbon dioxide, or CO_2, which is therefore often written as $O = C = O$, triple bonds appear in carbon monoxide, CO, which is often written as $C \equiv O$. Both double and triple bonds are common in organic compounds. (In addition, the well-known hexagonal benzene ring molecule C_6H_6, like many other compounds, has a one-and-a-half-fold bond.) Higher bonds are rare but do exist; quadruple bonds occur among transition metal atoms such

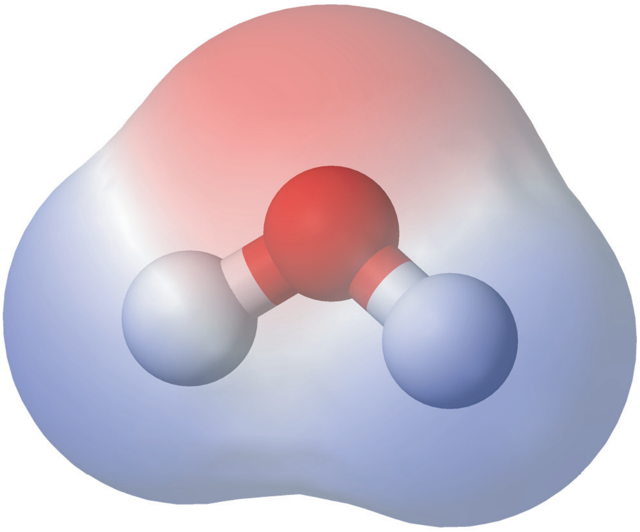


FIGURE 21 An artistic illustration of chemical bond angles when several atoms are involved: in a water molecule, with its charge distribution due to its covalent bonds, blue colour at the two ends indicates positive charge, and the red colour in upper vertex indicates negative charge. The central drawing shows a typical structural rendering of the water molecule (© Benjah-bmm27).

as rhenium or tungsten. Research also confirmed that the uranium U_2 molecule, among others, has a quintuple bond, and that the tungsten W_2 molecule has a hextuple bond. Ref. 45

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 physician Friedrich Miescher (b. 1844 Basel, d. 1895 Davos) in white blood cells. He also found it in cell nuclei, and thus called the substance 'Nuklein'. 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, as he died before they were established). They changed the name to 'nucleic acid'.

DNA is, as shown in Figure 23, 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 molecules; an example is shown in Figure 24. Ref. 46

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

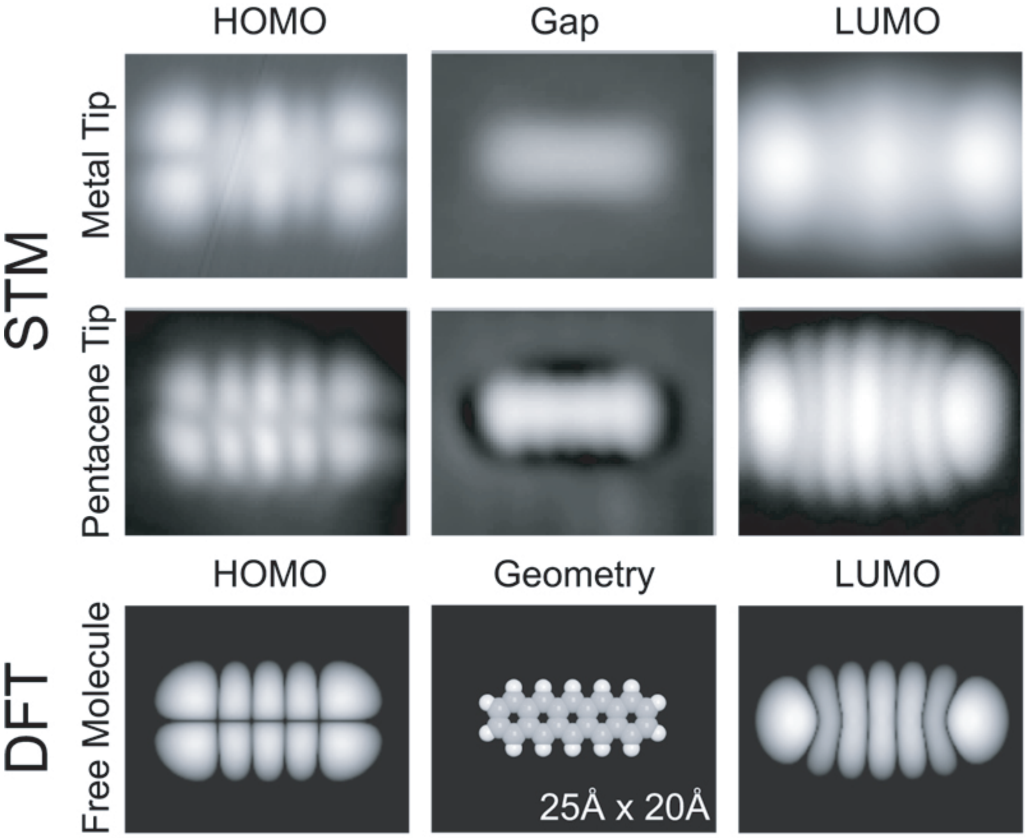


FIGURE 22 Top two rows: measured chemical bonds in the pentacene molecule, using different techniques; bottom row: textbook calculations and illustrations of the same experiment (© IBM).

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 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.

To experience the wonders of DNA, have a look at the animations of DNA copying and of other molecular processes at the unique website www.wehi.edu.au/education/wehitv.

CURIOSITIES AND FUN CHALLENGES ABOUT CHEMISTRY

Among the fascinating topics of chemistry are the studies of substances that influence humans: toxicology explores *poisons*, pharmacology explores *medicines* (pharmaceutical drugs) and endocrinology explores *hormones*.

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

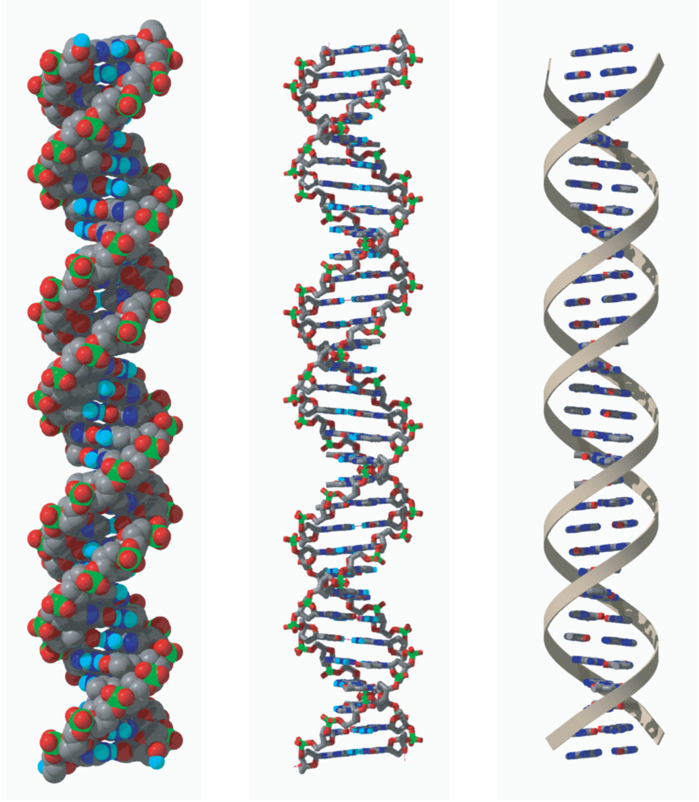


FIGURE 23 Several ways to picture B-DNA, all in false colours (© David Deerfield).

polonium 210 (kills in doses as low as 5 ng, far less than a spec of dust). Most countries have publicly accessible poison databases; see for example www.gsbl.de.

Challenge 38 e

Can you imagine why ‘toxicology’, the science of poisons, actually means ‘bow science’ in Greek? In fact, not all poisons are chemical. Paraffin and oil for lamps, for example, regularly kill children who taste it because some 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. Paraffin should never be present in homes with children.

In the 1990s, the biologist Binie Ver Lipps discovered a substance, a simple polypeptide, that helps against venoms of snakes and other poisonous animals. The medical industry worldwide refuses to sell the substance – it could save many lives – because it is too cheap.

* *

Whether a substance is a poison depends on the animal ingesting it. Chocolate is poison for dogs, but not for children. Poisonous mushrooms are edible for snails; bitten mushrooms are thus not a sign of edibility.

* *

Hormones are internal signalling substances produced by the human body. Chemically,

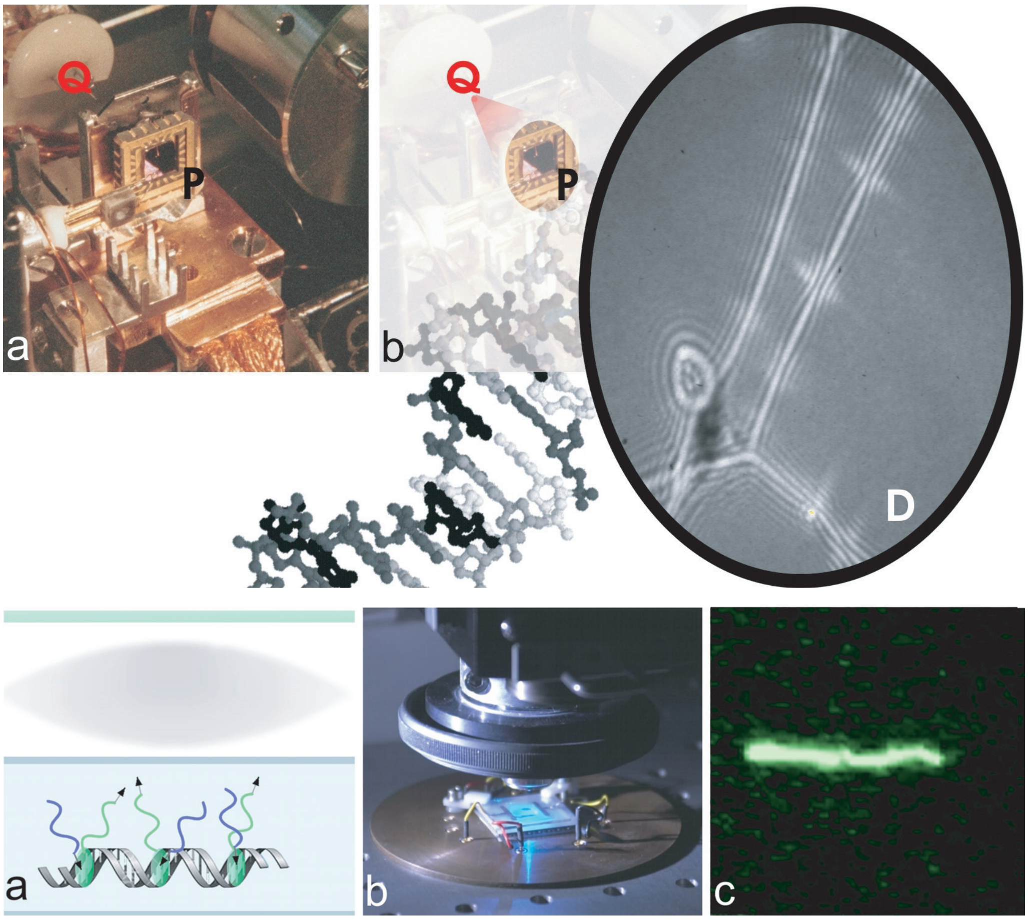


FIGURE 24 Two ways to image single DNA molecules: by holography with electrons emitted from atomically sharp tips (top) and by fluorescence microscopy, with a commercial optical microscope (bottom) (© Hans-Werner Fink/Wiley VCH).

they can be peptides, lipids or monoamines. Hormones induce mood swings, organize the fight, flight or freeze responses, stimulate growth, start puberty, control cell death and ageing, activate or inhibit the immune system, regulate the reproductive cycle and activate thirst, hunger and sexual arousal.

* *

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 39 s

* *

Why do *organic* materials, i.e., materials that contain several carbon atoms, usually burn at much lower temperature than *inorganic* materials, such as aluminium or magnesium? Challenge 40 ny

* *

A cube of sugar does not burn. However, if you put some cigarette ash on top of it, it burns. Why? Challenge 41 ny

* *

Sugars are essential for life. One of the simplest sugars is *glucose*, also called *dextrose* or *grape sugar*. Glucose is a so-called *monosaccharide*, in contrast to cane sugar, which is a disaccharide, or starch, which is a polysaccharide.

The digestion of glucose and the burning, or combustion, of glucose follow the same chemical reaction:

$$C_6H_{12}O_6 + 6\,O_2 \rightarrow 6\,CO_2 + 6\,H_2O + 2808\,\mathrm{kJ} \qquad (8)$$

This is the simplest and main reaction that fuels the muscle and brain activities in our body. The reaction is the reason we have to eat even if we do not grow in size any more. The required oxygen, O_2, is the reason that we breathe in, and the resulting carbon dioxide, CO_2, is the reason that we breathe out. Life, in contrast to fire, is thus able to 'burn' sugar at 37°. That is one of the great wonders of nature. Inside cells, the energy gained from the digestion of sugars is converted into adenosinetriphosphate (ATP) and then converted into motion of molecules.

* *

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.

* *

Two atoms can form bound states by a number of effects that are weaker than electron bonds. A famous one is the bound state formed by two sodium atoms at a distance of around 60 Bohr radii, thus much larger than usual bond distances. The bond appears due to the continuous exchange of a photon between the two atoms. Ref. 47

* *

What happens if you take the white powder potassium iodide – KJ – and the white powder lead nitrate – $Pb(NO_3)_2$ – and mix them with a masher? (This needs to be done with proper protection and supervision.) Challenge 42 s

* *

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.

* *

How is the concentration of *ozone*, with the chemical composition O_3, maintained in the high atmosphere? It took many years of research to show that the coolants used in refrigerators, the so-called fluoro-chloro-hydrocarbons or FCHCs, slowly destroyed this important layer. The reduction of ozone has increased the rate of skin cancer all across the world. By forbidding the most dangerous refrigerator coolants all over the world, it is hoped that the ozone concentration can recover. The first results are encouraging. In 1995, Paul Crutzen, Mario Molina and Sherwood Rowland received the Nobel Price for Chemistry for the research that led to these results and policy changes.

* *

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. In 2013, Jens Pahnke found that an extract of *Hypericum* has positive effects on the cognition and memory of Alzheimer patients. The extract is already available as prescription-free medication, for other uses, with the name LAIF900.

* *

Ref. 48

Cyanoacrylate is a fascinating substance. It is the main ingredient of instant glue, the glue that starts to harden after a few seconds of exposure to moisture. The smell of evaporating cyanoacrylate glue is strong and known to everybody who has used this kind of adhesive. The vapour also has another use: they make finger prints visible. You can try this at home!

Challenge 43 e

* *

Fireworks fascinate many. A great challenge of firework technology is to produce forest green and greenish blue colours. Producers are still seeking to solve the problem. For more information about fireworks, see the cc.oulu.fi/~kempmp website.

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 136

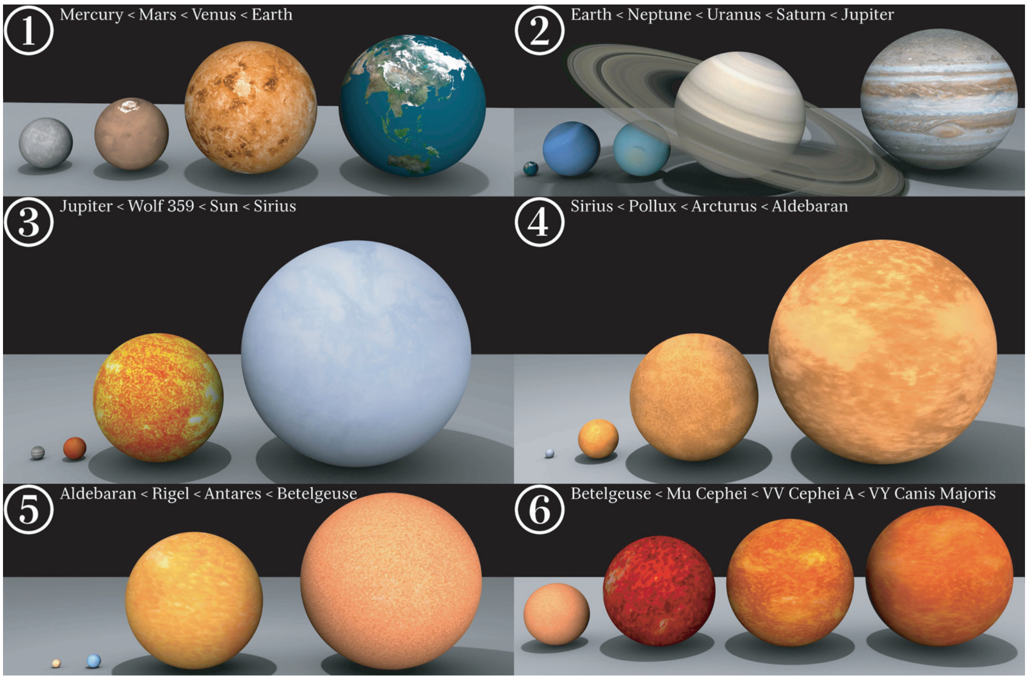


FIGURE 25 A comparison of star sizes (© Dave Jarvis).

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. Some people have spent their whole life to understand why such other floors, namely *surfaces of stars*, do not fall, or when they do, how it happens.

The floors and the sizes of all astronomic objects are due to quantum effects. Figure 25 illustrates the range of sizes that are found in astronomic objects. In each object, a quantum effect leads to an internal pressure which fixes the floor, and thus the size of the object.

In solid or liquid planets, the size is given by the incompressibility of condensed matter, which in turn is due to Pauli's exclusion principle. The effective internal pressure of condensed matter is often called the *Pauli pressure*. In gaseous planets, such as Jupiter, and in *usual stars*, such as in the Sun, the gas pressure takes the role that the incompressibility of solids and liquids has for smaller planets. The gas pressure is due to the heat stored in them; the heat is usually released by internal nuclear reactions.

Light pressure does play a role in determining the size of *red giants*, such as Betelgeuse; but for average stars, light pressure is negligible.

Other quantum effects appear in *dense* stars. Whenever light pressure, gas pressure and the electronic Pauli pressure cannot keep atoms from interpenetrating, 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 186 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 Vol. II, page 262 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 the conjectured *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 the age of the stone, within an error of a few per cent, simply by looking at it. The full story behind this astonishing ability forms a large part of geology, but the general lines should also be known to every physicist.

Generally speaking, the mass density of the Earth decreases from the centre towards the surface. The *upper mantle*, below the solid crust of the Earth, is mostly composed of *peridotite*, a dense, obviously igneous rock with a density of around 3.3 g/cm^3. The *oceanic crust*, with a thickness between 5 and 10 km, is mainly composed of igneous rocks such as basalt, diabase and gabbro. These rocks are somewhat less dense, around 3 g/cm^3, and are typically 200 million years old. The *continental crust* has a depth of 30 to 50 km, consists of lighter rocks, around 2.7 g/cm^3, such as granite. The age of the continental crust varies strongly; on average it is 2000 million years old, with a range from extremely young rocks to some older than 4300 million years. The continental crust contains most of the incompatible elements.

Every stone arrives in our hand through the *rock cycle(s)*. The main 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*, such as sandstone. (Sedimentary rocks can also form from biogenic base materials.) 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 main rock cycle takes around 110 to 170 million years. For this reason, rocks that are older than this age are less common on Earth. Any stone that we collect during a walk is the product of erosion of one of the rock types. A geologist can usually tell, just by looking at the stone, the type of rock it belongs to; if he sees or knows the original environment, he can also give the age and often tell the story of the formation, without any laboratory.

In the course of millions of years, minerals float upwards from the mantle or are pushed down the crust, they are transformed under heat and pressure, they dissolve or precipitate, and they get enriched in certain locations. These captivating stories about minerals are explored in detail by geologists. Geologists can tell where to find beaches with green sand (made of olivine); they can tell how contact between sedimentary limestone with molten igneous rocks leads to marble, ruby and other gemstones, and under Ref. 49 which precise conditions; they can also tell from small crystals of quartz that enclose coesite that in earlier times the rock has been under extremely high pressure – either

FIGURE 26 Igneous rocks (top three rows): gabbro, andesite, permatite, basalt, pumice, porphyry, obsidian, granite, tuff; sedimentary rocks (centre): clay, limestone, sandstone; and (below) two specimen of a metamorphic rock: marble (© Siim Sepp at www.sandatlas.org, Wikimedia).

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, a few %	clastic	shale, siltstone, sandstone
		biogenic	limestone, chalk, dolostone
		precipitate	halite, gypsum
Metamorphic rocks (metamorphites)	transformed by heat and pressure, a few %	foliated	slate, schist, gneiss (Himalayas)
		non-foliated (grandoblastic or hornfelsic)	marble, skarn, quartzite
Meteorites	from the solar system	rock meteorites	
		iron meteorites	

because it once was at a depth of the order of 70 km, or because of an asteroid impact, or because of an atomic bomb explosion.

From the point of view of materials science, rocks are mixtures of minerals. Even though more than 5000 minerals are known, only about 200 form rocks. These rock-forming minerals can be grouped in a few general types. The main group are the silica-based rocks. They contain SiO_4 tetrahedra and form around 92 % of all rocks. The remaining 8 % of rocks are of different composition, such as carbonates or oxides. Table 5 gives more details. The table covers the minerals found on the Earth's crust. However, the most common mineral in absolute is *Bridgmanite*, a form of $MgSiO_3$. About one third of the Earth is made of Bridgmanite, a silicate perovskite; it is formed in the lower mantle, at temperatures of around 1800°C and pressures above 24 GPa. The mineral never appears on the Earth's crust. Recent research suggests that some forms of Bridgmanite may even have contributed, when it rose to the surface by convection, to the oxygen in the atmosphere. Ref. 50

From the point of view of chemistry, rocks are even more uniform. 99 % of all rocks are made of only nine elements. Table 6 shows the details.

Almost all minerals are crystals. *Crystals* are solids with a regular arrangement of atoms and are a fascinating topic by themselves.

TABLE 5 The mineralogic composition of rocks and stones in the Earth’s crust.

Group	Mineral	Volume fraction
Inosilicates	*single chain* silicates: pyroxenes, e.g., diopside	11(2) %
	double chain silicates: amphiboles/hornblende, e.g., tremolite	5(1) %
Phyllosilicates	*sheet* silicates:	10 (2) %
	clays, e.g., kaolinite, talc	5(1) %
	mica-based minerals, e.g., biotite, muscovite	5(1) %
Tectosilicates	*volume* silicates:	65(5) %
	quartz, tridymite, cristobalite, coesite	11(1) %
	the plagioclase feldspar series, e.g., albite	43(4) %
	the alkali feldspars, e.g., orthoclase	14(2) %
Other silicates	with *isolated*, *double* or *cyclic* silica groups, e.g., olivine, beryl and garnets, or amorphous silicates, e.g., opal	3(1) %
Oxide-based rocks	e.g., magnetite, hematite, bauxite	5(1) %
Carbonate-based rocks	e.g., calcite, dolomite	
Sulfate-based rocks	e.g., gypsum, anhydrite	
Halide-based rocks	e.g., rock salt or halite, fluorite	
Other rocks		
	phosphates, e.g., apatite	1(0.5) %
	sulfides, e.g., pyrite	
	native metals, e.g., gold	
	borates, e.g.,	
	and many others.	

CRYSTAL FORMATION

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.

TABLE 6 The chemical composition of rocks and stones in the Earth's crust.

Element	Volume fraction
Oxygen	46.7(1.0) %
Silicon	27.6(0.6) %
Aluminium	8.1(0.1) %
Iron	5(1) %
Calcium	4.3(0.7) %
Sodium	2.5(0.2) %
Potassium	2.0(0.5) %
Magnesium	2.5(0.4) %
Titanium	0.5(0.1) %
Other elements	0.8(0.8) %

Are regular crystal lattices *energetically optimal*? This simple question leads to a wealth of problems. We might start with the much simpler question whether a regular dense packing of spheres is the most *dense* packing possible. Its density is $\pi/\sqrt{18}$, i.e., a bit over 74 %. Even though this was conjectured to be the maximum possible value already in 1609 by Johannes Kepler, the statement was proven only in 1998 by Tom Hales. The proof is difficult because in small volumes it is possible to pack spheres up to almost 78 %. To show that over large volumes the lower value is correct is a tricky business.

Challenge 44 s

Ref. 51

Next, does a regular crystal of solid spheres, in which the spheres do *not* touch, really have the *highest possible entropy*? This simple problem has been the subject of research only from the 1990s onwards. 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 reaching the entropy values of a solid.

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. 52

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. 53

For an exact study of crystal energy, the interactions between all atoms have to be

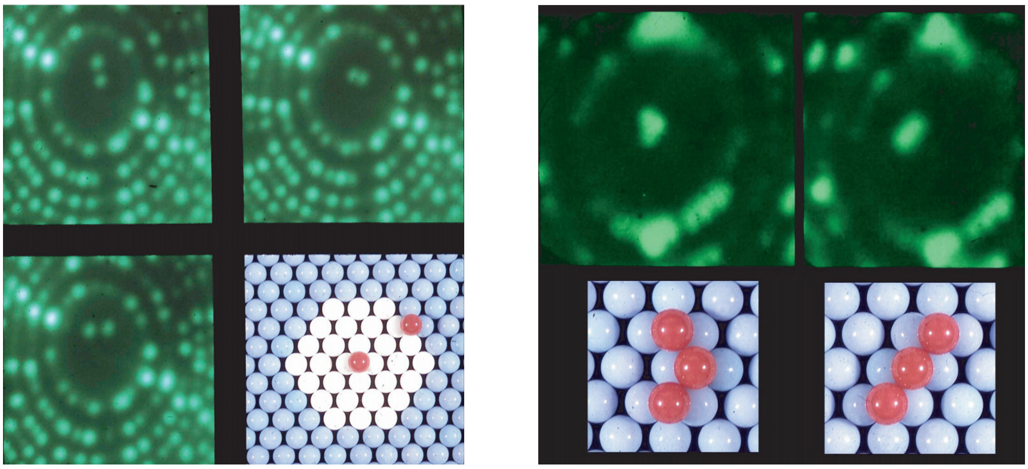


FIGURE 27 On tungsten tips, rhenium atoms, visible at the centre of the images, do not form dimers (left) but do form trimers (right) (© Hans-Werner Fink/APS, from Ref. 53).

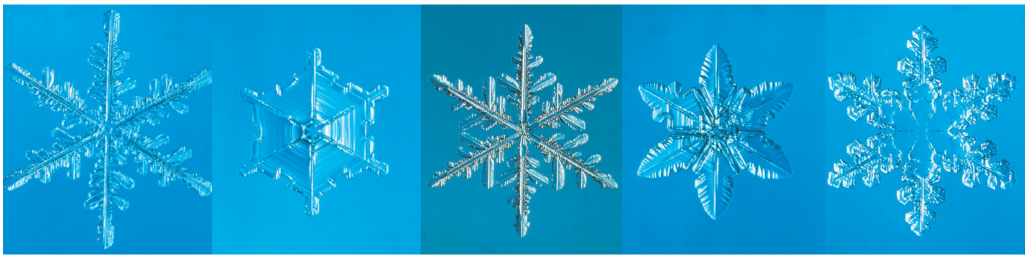


FIGURE 28 Some snow flakes (© Furukawa Yoshinori).

included. The simplest question about crystal energy is to determine whether a regular array of alternatively charged spheres has lower energy than some irregular collection. Already such simple questions are still topic of research; the answer is still open.

The previous topics concerned bulk crystals. The next topic is the 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 45 s Ref. 54

However, not always the slowest growing planes win out during crystal growth. Figure 28 shows some well-known exceptions: snow flakes. Explaining the shapes of snow flakes is possible today. Furukawa Yoshinori is one of the experts in the field, heading a dedicated research team. Ref. 55 These explanations also settle the question of symmetry: why are crystals often symmetric, instead of asymmetric? This is a topic of self-organization, as mentioned already in the section of classical physics. Vol. I, page 415 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 growth processes. But as usual, the details are still a topic of research.

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 *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: 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. Page 69

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 only the *slowest* growing facets are found in crystals? In the following, a few important crystals are presented. Challenge 46 e


FIGURE 29 Quartz found at St. Gotthard, Switzerland, picture size 12 cm (© Rob Lavinsky).


FIGURE 30 Citrine found on Magaliesberg, South Africa, crystal height 9 cm (© Rob Lavinsky).


FIGURE 31 Amethystine and orange quartz found in the Orange River, Namibia, picture size 6 cm (© Rob Lavinsky).

* *

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 sometimes be larger than humans. 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. A film of an oscillating clock quartz is found in 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.

Vol. I, page 291

FIGURE 32 Corundum found in Laacher See, Germany, picture size 4 mm (© Stephan Wolfsried).

FIGURE 33 Ruby found in Jagdalak, Afghanistan, picture height 2 cm (© Rob Lavinsky).

FIGURE 34 Sapphire found in Ratnapura, Sri Lanka, size 1.6 cm (© Rob Lavinsky).

∗ ∗

Corundum, *ruby* and *sapphire* are crystalline variations of alumina, or Al_2O_3. Corundum is pure and colourless crystalline alumina, 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 the second-hardest material known, just after diamond, and is therefore used as scratch-resistant 'glass' in watches and, since a short time, in mobile phones. 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. Modern synthetic single crystals of corundum can weigh 30 kg and more. Also alumina ceramics, which can be white or even transparent, are important in industrial and medical systems.

∗ ∗

Tourmaline is a frequently found mineral and can be red, green, blue, orange, yellow,

FIGURE 35 Left: raw crystals, or boules, of synthetic corundum, picture size c. 50 cm. Right: a modern, 115 kg corundum single crystal, size c. 50 cm (© Morion Company, GT Advanced).

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 natural and untreated, more expensive than diamonds.

FIGURE 36 Natural bicoloured tourmaline found in Paprok, Afghanistan, picture size 9 cm (© Rob Lavinsky).

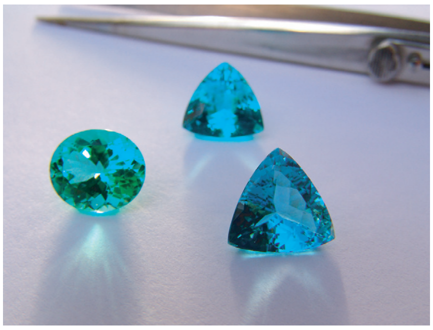


FIGURE 37 Cut Paraiba tourmaline from Brazil, picture size 3 cm (© Manfred Fuchs).

* *

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

FIGURE 39 Green demantoid, a garnet owing its colour to chromium doping, found in Tubussis, Namibia, picture size 5 cm (© Rob Lavinsky).

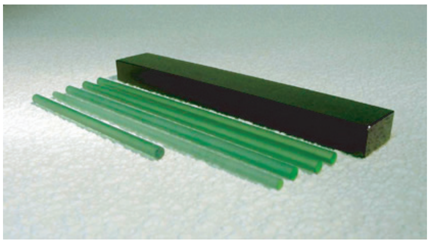


FIGURE 40 Synthetic Cr,Tm,Ho:YAG, a doped yttrium aluminium garnet, picture size 25 cm (© Northrop Grumman).

* *

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 41. 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 41 Alexandrite found in the Setubal river, Brazil, crystal height 1.4 cm, illuminated with daylight (left) and with incandescent light (right) (© Trinity Mineral).

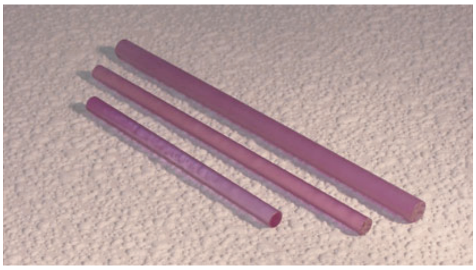


FIGURE 42 Synthetic alexandrite, picture size 20 cm (© Northrop Grumman).

* *

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

* *

FIGURE 43 Perovskite found in Hillesheim, Germany. Picture width 3 mm (© Stephan Wolfsried).

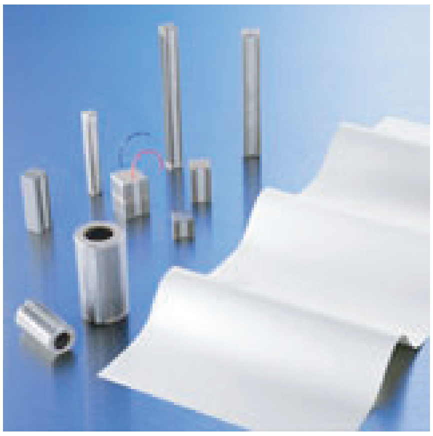


FIGURE 44 Synthetic PZT, or lead zirconium titanate, is a perovskite used in numerous products. Picture width 20 cm (© Ceramtec).

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

∗ ∗

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. 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 $Ca_{10}(PO_4)_6(OH)_2$, pos-

FIGURE 45 Natural diamond from Saha republic, Russia, picture size 4 cm (© Rob Lavinsky).

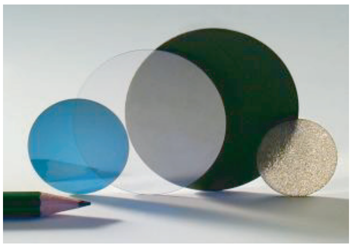


FIGURE 46 Synthetic diamond, picture size 20 cm (© Diamond Materials GmbH).

FIGURE 47 Ophthalmic diamond knife, picture size 1 cm (© Diamatrix Ltd.).

sesses hexagonal crystal structure, is hard (more than steel) but relatively brittle. It occurs as mineral in sedimentary rocks (see Figure 49), 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.

* *

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.

* *

In 2009, Luca Bindi of the Museum of Natural History in Florence, Italy, made headlines

FIGURE 48 A silicon crystal growing machine and two resulting crystals, with a length of c. 2 m (© www.pvatepla.com).

Ref. 56 across the world with his discovery of the first natural *quasicrystal*. Quasicrystals are materials that show non-crystallographic symmetries. Until 2009, only synthetic materials 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

FIGURE 49 Hydroxylapatite found in Oxsoykollen, Snarum, Norway, length 65 mm (© Aksel Österlöf).

FIGURE 50 The main and the reserve teeth on the jaw bone of a shark, all covered in hydroxylapatite, picture size 15 cm (© Peter Doe).

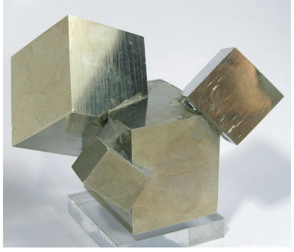


FIGURE 51 Pyrite, found in Navajún, Spain, picture width 5.7 cm (© Rob Lavinsky).

FIGURE 52 Silver from Colquechaca, Bolivia, picture width 2.5 cm (© Rob Lavinsky).

FIGURE 53 Synthetic copper single crystal, picture width 30 cm (© Lachlan Cranswick).

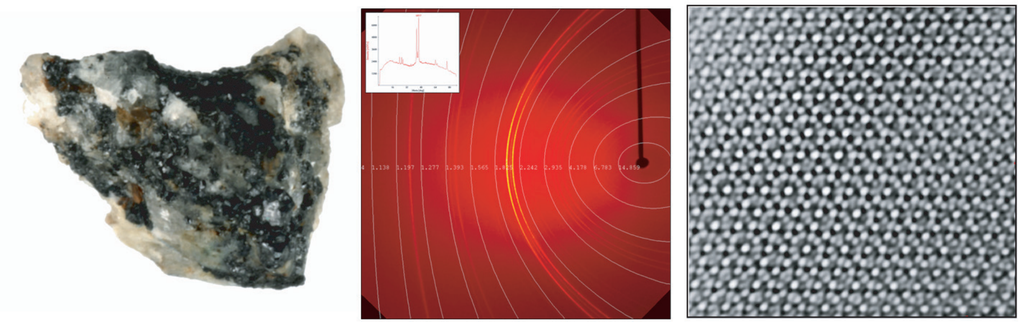


FIGURE 54 The specimen, found in the Koryak Mountains in Russia, is part of a triassic mineral, about 220 million years old; the black material is mostly khatyrkite ($CuAl_2$) and cupalite ($CuAl_2$) but also contains quasicrystal grains with composition $Al_{63}Cu_{24}Fe_{13}$ that have fivefold symmetry, as clearly shown in the X-ray diffraction pattern and in the transmission electron image. (© Luca Bindi).

various places on the internet, including www.mindat.org or www.mineralienatlas.de. To explore the world of crystal shapes, see the www.smorf.nl website. Around 40 new minerals are discovered each year. Searching for minerals and collecting them is a fascinating pastime.

HOW CAN WE LOOK THROUGH MATTER?

Vol. IV, page 28

The quantum of action tells us that all obstacles have only finite potential heights. The quantum of action implies that matter is penetrable. That leads to a question: Is it possible to look through solid matter? For example, can we see what is hidden inside a mountain? To achieve this, we need a signal which fulfils two conditions: the signal must be able to *penetrate* the mountain, and it must be scattered in a *material-dependent* way. Indeed, such signals exist, and various techniques use them. Table 7 gives an overview of the possibilities.

TABLE 7 Signals penetrating mountains and other matter.

SIGNAL	PENETRATION DEPTH IN STONE	ACHIEVED RESOLUTION	MATERIAL DEPENDENCE	USE
Fluid signals				
Diffusion of gases, such as helium	*c.* 5 km	*c.* 100 m	medium	exploring vacuum systems and tube systems
Diffusion of water or liquid chemicals	*c.* 5 km	*c.* 100 m	medium	mapping hydrosystems
Sound signals				
Infrasound and earthquakes	100 000 km	100 km	high	mapping of Earth crust and mantle
Sound, explosions, short 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, sonar, echo systems
Acousto-optic or photoacoustic imaging		1 mm	medium	blood stream imaging, mouse imaging
Electromagnetic signals				
Static magnetic field variations			medium	cable search, cable fault localization, search for structures and metal inside soil, rocks and the 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 resonance imaging, research into solar interior

TABLE 7 (Continued) Signals penetrating mountains and other matter.

SIGNAL	PENETRATION DEPTH IN STONE	ACHIEVED RESOLUTION	MATERIAL DEPENDENCE	USE
Ultra-wide band radio	10 cm	1 mm	sufficient	searching for wires and tubes in walls, breast cancer detection
THz and mm waves	below 1 mm	1 mm		see through clothes, envelopes and teeth Ref. 57
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, including breast tumour screening
X-rays	a few metres	5 µm	high	medicine, material analysis, airports, food production check
γ-rays	a few metres	1 mm	high	medicine
Matter particle signals				
Neutrons from a reactor	up to *c.* 1 m	1 mm	medium	tomography of metal structures, e.g., archaeologic statues or car engines
Muons created by cosmic radiation or technical sources	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	brain tomography
Electrons	up to *c.* 1 µm	10 nm	small	transmission electron microscopes
Neutrino beams	light years	none	very weak	studies of Sun
Radioactivity mapping	1 mm to 1 m			airport security checks
Gravitation				
Variation of g		50 m	low	oil and ore search

We see that many signals are able to penetrate a mountain, and even more signals are able to penetrate other condensed matter. To distinguish different materials, or to distinguish solids from liquids and from air, sound and radio waves are the best choice. In addition, any useful method requires a large number of signal sources and of signal receptors, and thus a large amount of money. Will there ever be a simple method that allows looking into mountains as precisely as X-rays allow looking into human bodies? For example, is it possible to map the interior of the pyramids? A motion expert should be able to give a definite answer. Challenge 47 s

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 162

Various modern imaging techniques, such as X-rays, ultrasound imaging and several

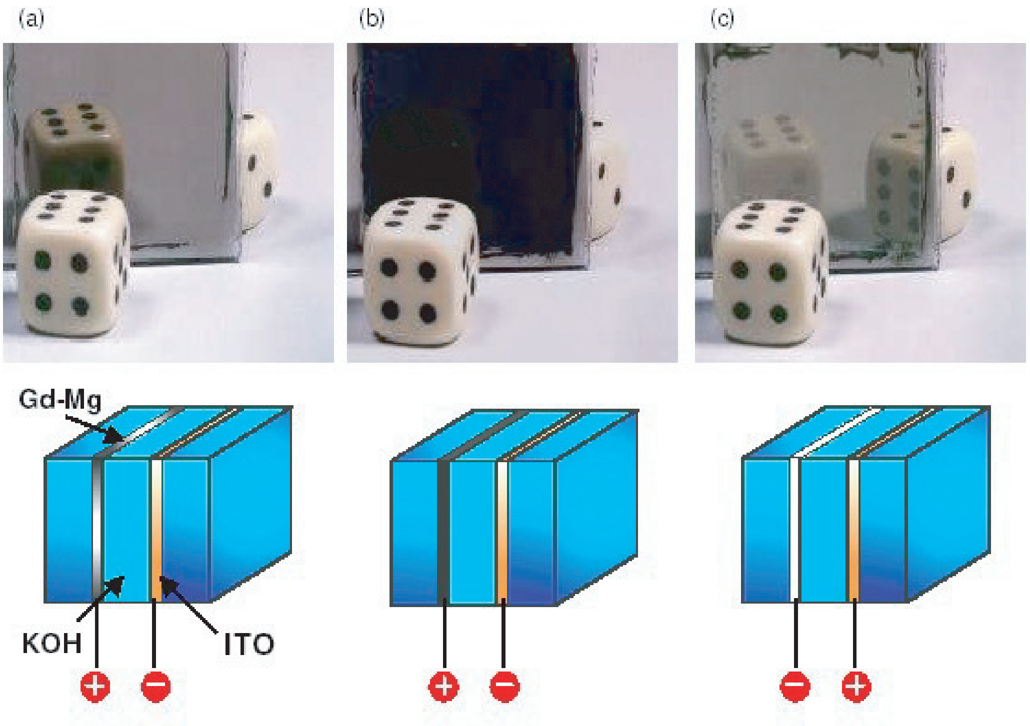


FIGURE 55 A switchable Mg-Gd mirror (© Ronald Griessen).

Vol. I, page 313
Vol. I, page 313

future ones, are useful in medicine. As mentioned before, the use of ultrasound imaging for prenatal diagnostics of embryos is not recommended. Studies have found that ultrasound produces extremely high levels of audible sound to the embryo, especially when the ultrasound is repeatedly switched on and off, and that babies react negatively to this loud noise.

Looking into the ground is important for another reason. It can help in *locating land mines*. Detecting land mines, especially metal-free mines, buried in the ground is a big technological challenge that is still unsolved. Many technologies have been tested: X-ray backscatter devices working at 350 to 450 keV, ground-penetrating radar and ultra-wideband radar, infrared detection, thermal or fast neutron bombardment and analysis, acoustic and sonar detection, electric impedance tomography, radio-frequency bombardment, nuclear quadrupole resonance, millimetre waves, visual detection, ion mobility spectrometers, using dogs, using rats, and explosive vapour detection with dedicated sensors. (And of course, for metallic mines, magnetometers and metal detectors are used.) But so far, the is still no solution in sight. Can you find one? If you do, get in touch with www.gichd.org.

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, yttrium hydride or YH_3, that can be switched from mirror mode to transparent mode using an electrical

Ref. 58 signal. A number of other materials were also discovered. An example of the effect for Mg-Gd layers is shown in Figure 55.

Switchable mirrors might seem 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. For example, it is often claimed that objects can be made invisible by covering them with metamaterials. The impossibility of this aim has been already shown earlier on. But we now can say more. Vol. III, page 169

Nature shows us how to realize invisibility. An object is invisible if it has no surface, no absorption and small size. In short, invisible objects are either small clouds or composed of them. Most atoms and molecules are examples. Homogeneous non-absorbing gases also realize these conditions. That is the reason that air is (usually) invisible. When air is not homogeneous, it can be visible, e.g. above hot surfaces.

In contrast to gases, solids or liquids do have surfaces. Surfaces are usually visible, even if the body is transparent, because the refractive index changes there. For example, quartz can be made so transparent that one can look through 1 000 km of it; pure quartz is thus *more* transparent than usual air. Still, objects made of pure quartz are visible to the eye, due to 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, because 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 likely. In summary, quantum theory shows that only the dead can be invisible. Quantum theory has a reassuring side: we already found that quantum theory forbids ghosts; we now find that it also forbids any invisible beings. Vol. IV, page 136

What moves inside matter?

All matter properties are due to the motion of the components of matter. Therefore, we can correctly argue that understanding the motion of electrons and nuclei implies understanding all properties of matter. Sometimes, however, it is more practical to explore the motion of collections of electrons or nuclei as a whole. Here is a selection of such collective motions. Collective motions that appear to behave like single particles are called *quasiparticles*.

In crystalline solids, sound waves can be described as the motion of *phonons*. For example, transverse phonons and longitudinal phonons describe many processes in semiconductors, in solid state lasers and in ultrasound systems. Phonons approximately behave as bosons.

Page 298 In metals, the motion of crystal defects, the so-called *dislocations* and *disclinations*, is central to understand their hardening and their breaking.

Also in metals, the charge waves of the conductive electron plasma, can be seen as composed of so-called *plasmons*. Plasmons are also important in the behaviour of high-speed electronics.

In magnetic materials, the motion of spin orientation is often best described with the help of magnons. Understanding the motion of magnons and that of magnetic domain walls is useful to understand the magnetic properties of magnetic material, e.g., in permanent magnets, magnetic storage devices, or electric motors. Magnons behave approximately like bosons.

In semiconductors and insulators, the motion of *conduction electrons* and *electron holes*, is central for the description and design of most electronic devices. They behave as fermions with spin 1/2, elementary electric charge, and a mass that depends on the material, on the specific conduction band and on the specific direction of motion. The bound system of a conduction electron and a hole is called an *exciton*. It can have spin 0 or spin 1.

In polar materials, the motion of light through the material is often best described in terms of *polaritons*, i.e., the coupled motion of photons and dipole carrying material excitations. Polaritons are approximate bosons.

In dielectric crystals, such as in many inorganic ionic crystals, the motion of an electron is often best described in terms of *polarons*, the coupled motion of the electron with the coupled polarization region that surrounds it. Polarons are fermions.

In fluids, the motion of vortices is central in understanding turbulence or air hoses. Especially in superfluids, vortex motion is quantized in terms of *rotons* which determines flow properties. Also in fluids, bubble motion is often useful to describe mixing processes.

In superconductors, not only the motion of *Cooper pairs*, but also the motion of *magnetic flux tubes* determines the temperature behaviour. Especially in thin and flat superconductors – so-called 'two-dimensional' systems – such tubes have particle-like properties.

In all condensed matter systems, the motion of *surface states* – such as surface magnons, surface phonons, surface plasmons, surface vortices – has also to be taken into account.

Many other, more exotic *quasiparticles* exist in matter. Each quasiparticle in itself is an important research field where quantum physics and material science come together. To clarify the concepts, we mention that a *soliton* is not, in general, a quasiparticle. 'Soliton' is a mathematical concept; it applies to macroscopic waves with only one crest that remain unaltered after collisions. Many domain walls can be seen as solitons. But quasiparticles are concepts that describe physical observations similar to quantum particles. Vol. I, page 316

In summary, all the mentioned examples of collective motion inside matter, both macroscopic and quantized, are of importance in electronics, photonics, engineering and medical applications. Many are quantized and their motion can be studied like the motion of real quantum particles.

CURIOSITIES AND FUN CHALLENGES ABOUT MATERIALS SCIENCE

What is the maximum height of a mountain? This question is of course of interest to all climbers. Many effects limit the height. The most important is that under heavy pressure, solids become liquid. For example, on Earth this happens for a mountain with a height of about 27 km. This is quite a bit more than the highest mountain known, which is

Ref. 59

Challenge 48 ny

the volcano Mauna Kea in Hawaii, whose top is about 9.45 km above the base. On Mars gravity is weaker, so that mountains can be higher. Indeed the highest mountain on Mars, Olympus mons, is 80 km high. Can you find a few other effects limiting mountain height? Challenge 49 s

* *

Do you want to become rich? Just invent something that can be produced in the factory, is cheap and can fully substitute duck feathers in bed covers, sleeping bags or in badminton shuttlecocks. Challenge 50 r 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.

* *

Materials differ in density, in elasticity, in strength, stiffness, toughness, melting temperature, heat insulation, electric resistivity, and many other parameters. To get an overview, the so-called *Ashby charts* are most useful, of which Figure 56 shows an example. The race to find materials that are lighter and stiffer than wood, in particular balsa wood, is still ongoing.

* *

How much does the Eiffel tower change in height over a year due to thermal expansion and contraction? Challenge 51 s

* *

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 stiffness. For a picture of dislocations, see below. Page 298

* *

Challenge 52 e What is the difference between solids, liquids and gases?

* *

One subject of materials science is the way a solid object breaks. The main distinction is between *brittle* fraction and *ductile* fraction. In brittle fraction, as in a breaking of a glass pane, the resulting edges are sharp and irregular; in ductile fraction, as occurs in hot glass, the edges are rounded and regular. The two fraction types also differ in their mechanisms, i.e., in the motion of the involved defects and atoms. This difference is important: when a car accident occurs at night, looking at the shapes of the fraction surface of the tungsten wire inside the car lamps with a microscope, it is easy to decide whether the car lamps were on or off at the time of the accident.

* *

Material science can also help to make erased information visible again. Many laborat-

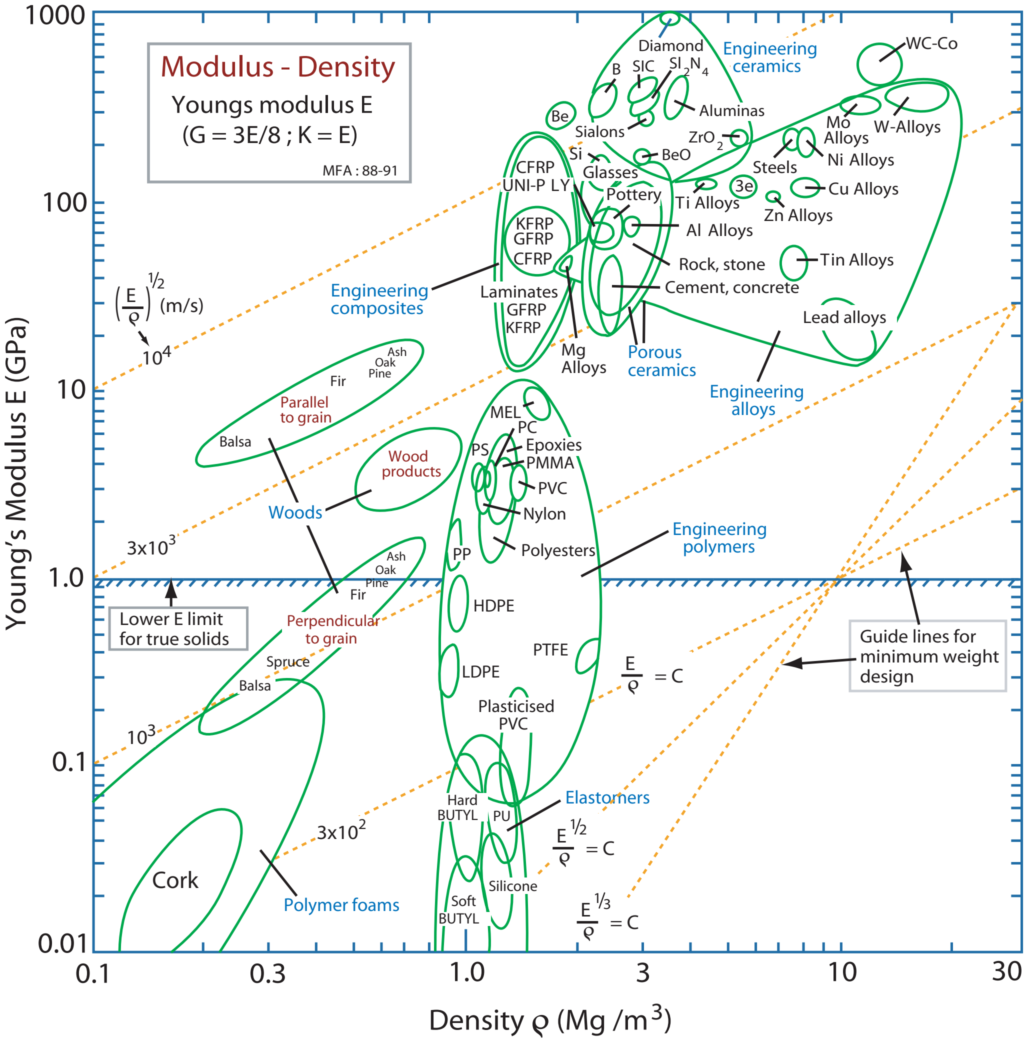


FIGURE 56 An overview of the elastic modulus and the density of materials. For structures that need to be light and stiff, a high ratio E / ρ^3 is required; the graph shows that wood is well optimized for this task. (© Carol Livermore/Michael Ashby).

ories are now able to recover data from erased magnetic hard disks. Other laboratories can make erased serial numbers in cars bodies visible again, either by heating the metal part or by using magnetic microscopy.

* *

Challenge 53 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

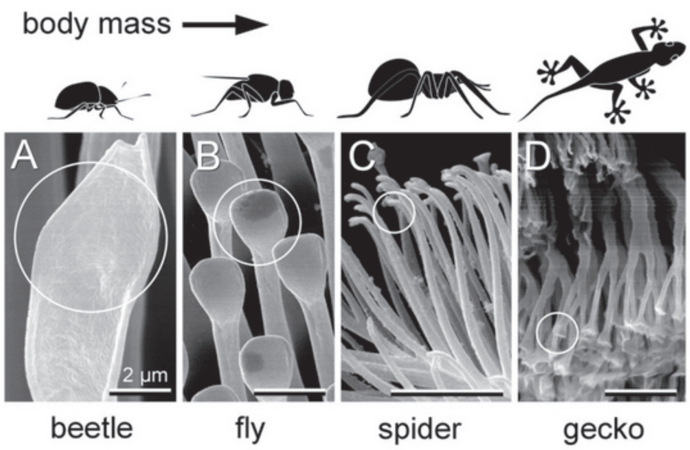


FIGURE 57 Insects and geckos stick to glass and other surfaces using the van der Waals force at the ends of a high number of spatulae (© Max Planck Gesellschaft).

Challenge 54 s walls can never be tightly shut. Can you provide the argument?

* *

In 1936, Henry Eyring proposed that the shear viscosity of a liquid η obeys

$$\eta \geqslant \rho\hbar \, , \qquad (9)$$

Challenge 55 ny where ρ is the density of the fluid. Is the lower limit valid?

* *

Heat can flow. Like for all flows, quantum theory predicts that heat transport quantized. This implied that thermal conductance is quantized. And indeed, in the year 2000, experiments have confirmed the prediction. Can you guess the smallest unit of thermal conductance? Ref. 60 Challenge 56 s

* *

Ref. 61 Vol. I, page 91 Robert Full has shown that van der Waals forces are responsible for the way that geckos walk on walls and ceilings. (A picture is found in Figure 56.) 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 or setae, 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 of these 500 million sticking points, 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 adhesion forces are so high that detaching the foot requires a special technique. The internet has slow-motion videos showing how geckos perform the feat, in each step they take.

Hairy feet as adhesion method are also used by jumping spiders (*Salticidae*). For example, *Evarcha arcuata* have hairs at their feet which are covered by hundred of thou-

Ref. 62 sands of setulae. Again, the van der Waals force in each setula helps the spider to stick on surfaces. Also many insects use small hairs for the same aim. Figure 57 shows a Ref. 63 comparison. Researchers have shown that the hairs – or setae – are finer the more massive the animal is. Eduard Arzt likes to explain that small flies and beetles have simple, spherical setae with a diameter of a few micrometers whereas the considerably bigger and heavier geckos have branched nanohairs with diameters of 200 nm.

Researchers have copied the hairy adhesion mechanism for the first time in 2003, using microlithography on polyimide, and they hope to make durable sticky materials – without using any glue – in the future.

* *

One of the most fascinating materials in nature are bones. Bones are light, stiff, and can Ref. 64 heal after fractures. If you are interested in composite materials, read more about bones: their structure, shown in Figure 58, and their material properties are fascinating and complex, and so are their healing and growth mechanisms. All these aspects are still subject of research.

* *

A cereal stalk has a height-to-width ratio of about 300. No human-built tower or mast Challenge 57 s achieves this. Why?

* *

Millimetre waves or terahertz waves are emitted by all bodies at room temperature. Modern camera systems allow producing images with them. In this way, it is possible to see through clothes, as shown by Figure 59. (Caution is needed; there is the widespread suspicion that the image is a fake produced to receive more development funding.) 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.

* *

Does the melting point of water depend on the magnetic field? This surprising claim was Ref. 65 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. A few are shown in Figure 60. Not only can plasmas 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.

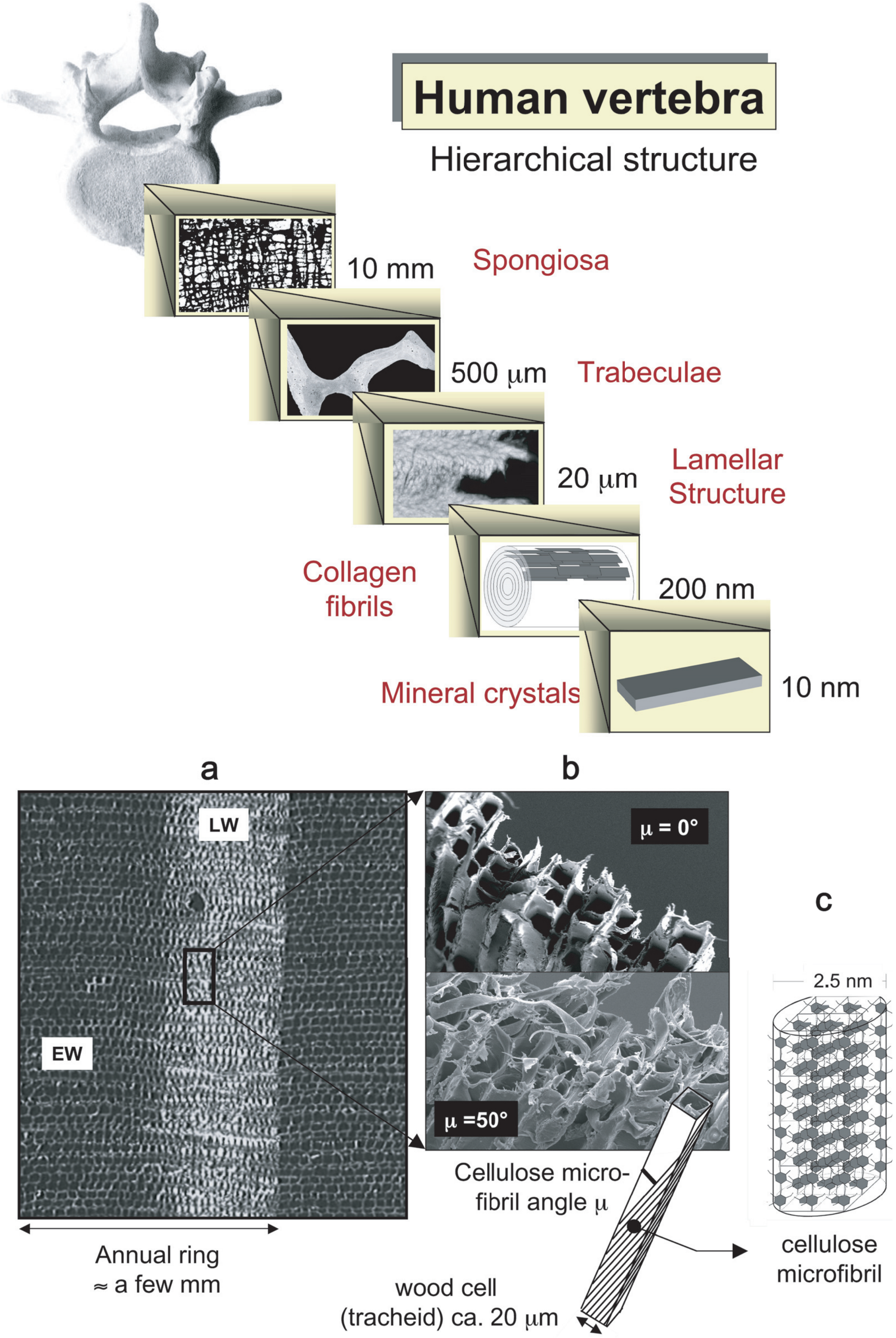


FIGURE 58 The structure of bones, shown for a human vertebra (© Peter Fratzl and Physik Journal).

FIGURE 59 An alleged image acquired with terahertz waves. Can you explain why it is a fake? (© Jefferson Lab)

* *

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 61. (In 2016, the value was already 403 ppm. 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. On average, the Earth has cooled over the past 10 million years; since a few thousand years, the temperature has, however, slowly risen; together with the fast rise during the last decades the temperature is now at the same level as 3 million years ago. How do the decade global warming trend, the thousand year warming trend and the million year cooling trend interact? This is a topic of intense research.

Ref. 66
Challenge 58 s
Ref. 67

* *

Making crystals can make one rich. The first man who did so, the Frenchman Auguste Verneuil (b. 1856 Dunkerque, d. 1913 Paris), sold rubies grown in his laboratory for many years without telling anybody. Many companies now produce synthetic gems with machines that are kept secret. An example is given in Figure 62.

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. 68

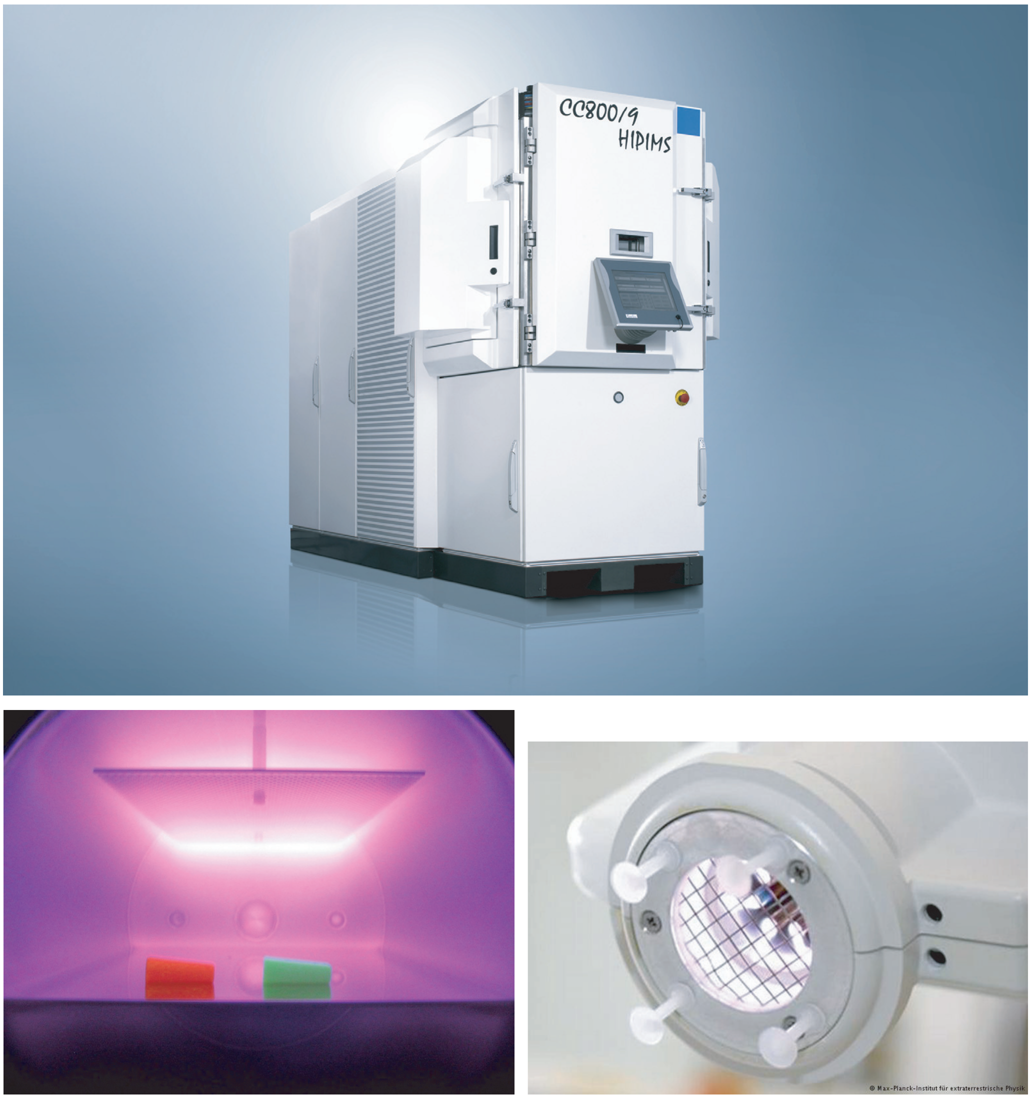


FIGURE 60 Some plasma machines: a machine for coating metal parts, a machine for cleaning polymer parts and a device for healing wounds (© www.cemecon.de, www.diener.de, Max Planck Gesellschaft).

* *

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 59 s How can a small plant pierce through tarmac?

* *

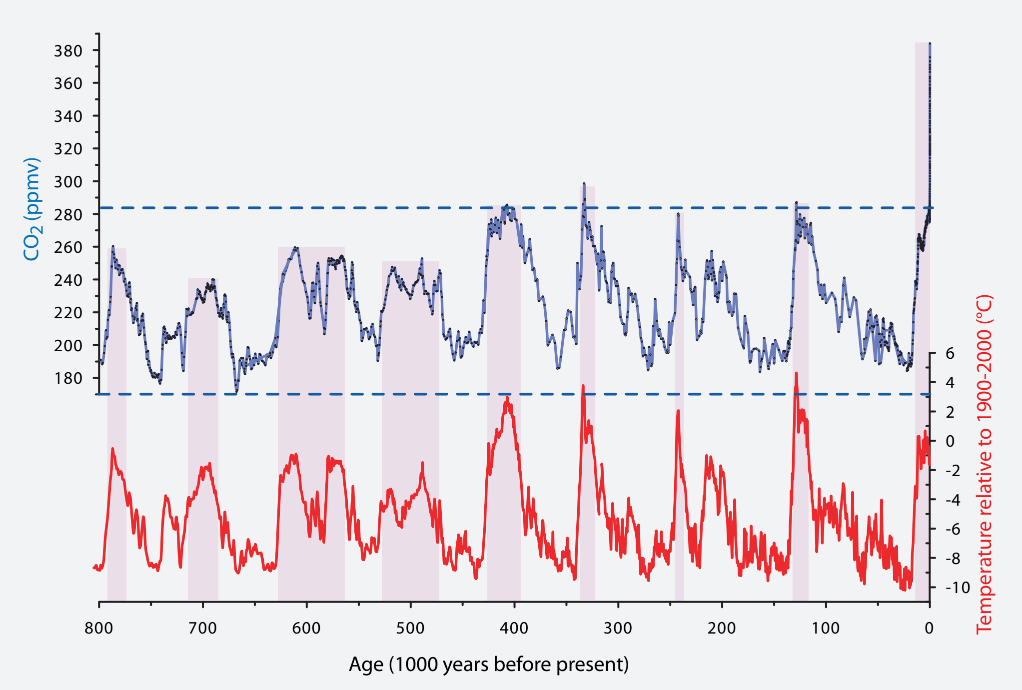


FIGURE 61 The concentration of CO_2 and the change of average atmospheric temperature in the past 0.8 million years (© Dieter Lüthi).

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 Dierking. Ref. 69

* *

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 (b. 1855 Great Falls, d. 1938 Cambridge), who discovered it in 1879, during his PhD. The effect, shown in Figure 63, 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

$$V = \frac{IB}{ned} , \qquad (10)$$

where n is the electron number density, e the electron charge, and d is the thickness of the probe, as shown in Figure 63. Deducing the equation is a secondary school exercise. Challenge 60 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

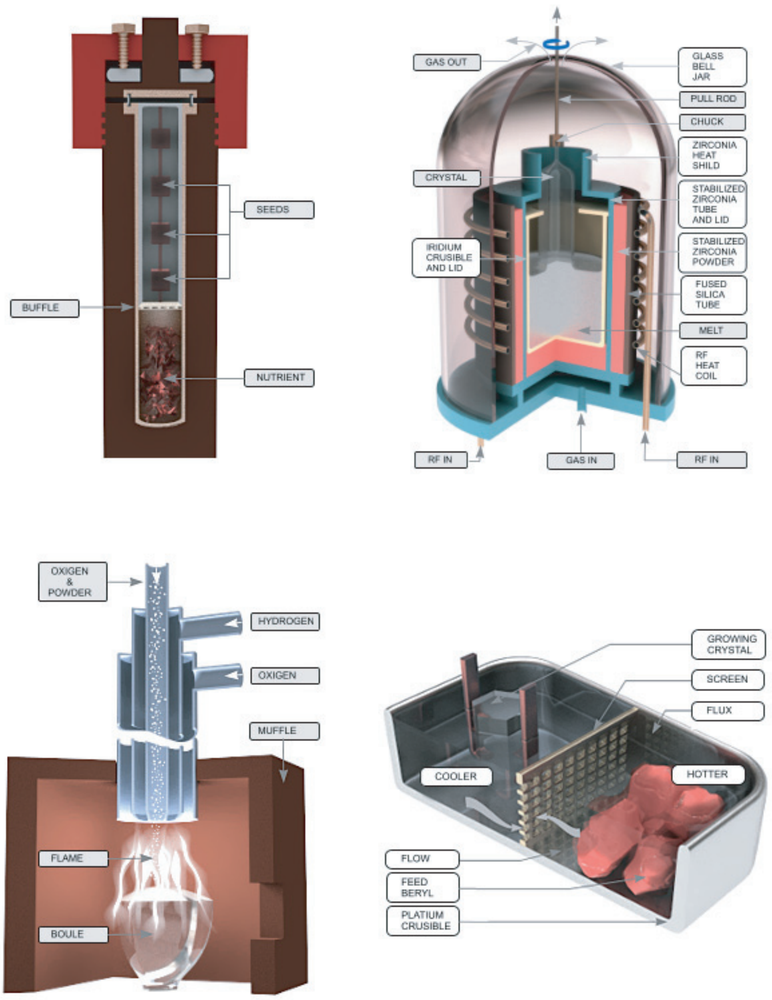


FIGURE 62 Four crystal and synthetic gemstone growing methods. Top: the hydrothermal technique, used to grow emeralds, quartz, rock crystal and amethyst, and Czochralski's pulling technique, used for growing ruby, sapphire, spinel, yttrium-aluminium-garnet, gadolinium-gallium-garnet and alexandrite. Bottom: Verneuil's flame fusion technique, used for growing corundum, sapphire, ruby and spinel boules, and the flux process, used for chrysoberyl (© Ivan Golota).

charge carriers; indeed, for metal strips the voltage polarity is opposite to the one shown in the figure. Challenge 61 e

Many variations of the Hall effect have been studied. For example, the *quantum* Hall effect and the fractional quantum Hall effect will be explored below. Page 107

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. 70

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

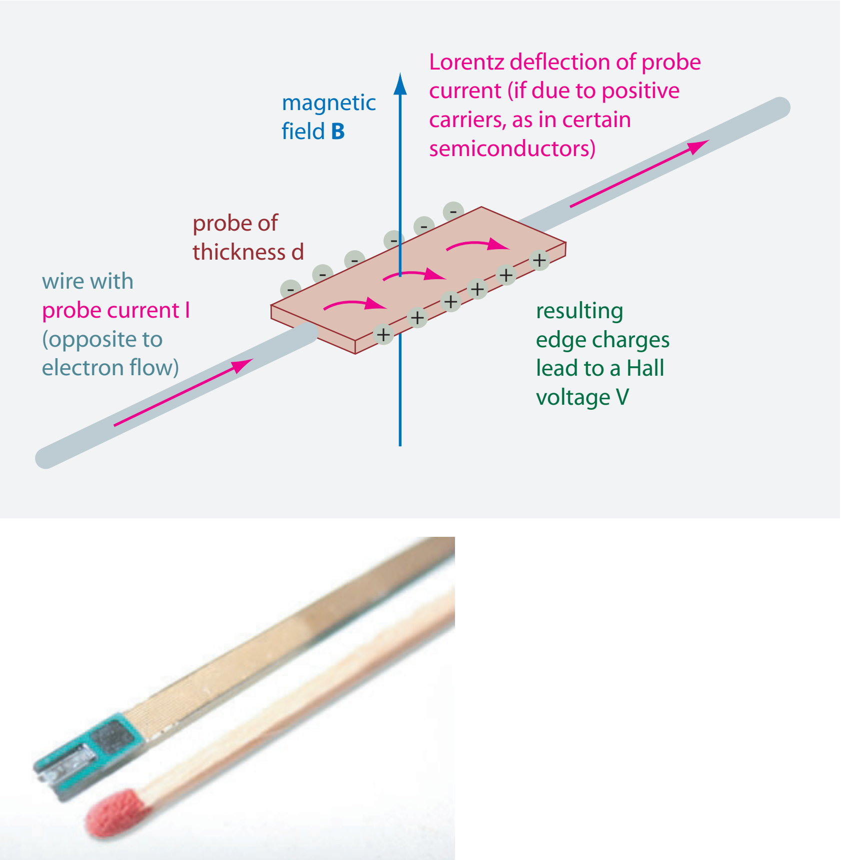


FIGURE 63 Top: the (classical) Hall effect. Bottom: a modern miniature Hall probe using the effect to measure magnetic fields (© Metrolab).

Ref. 71 the two sides. They called this the *phonon Hall effect*.

* *

Do magnetic fields influence the crystallization of calcium carbonate in water? This issue Ref. 72 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. 73 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

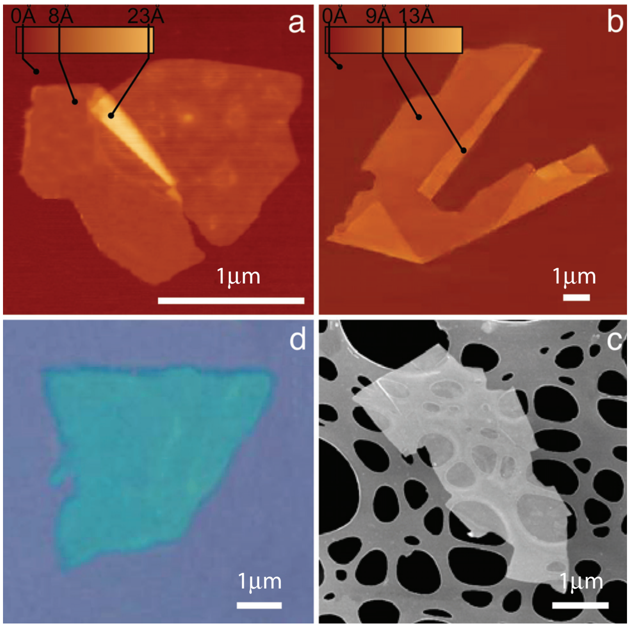


FIGURE 64 Single atom sheets, mapped by atomic force microscopy, of a: $NbSe_2$, b: of graphite or graphene, d: a single atom sheet of MoS_2 imaged by optical microscopy, and c: a single atom sheet of $Bi_2Sr_2CaCu_2O_x$ on a holey carbon film imaged by scanning electron microscopy (from Ref. 73, © 2005 National Academy of Sciences).

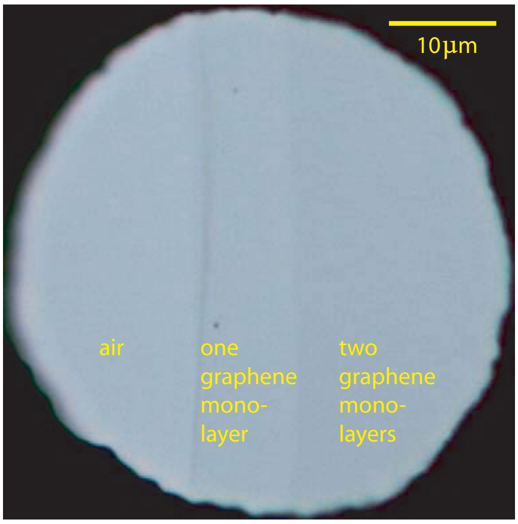


FIGURE 65 A microscope photograph shows the absorption of a single and of a double layer of graphene – and thus provides a way to see the fine structure constant. (© Andre Geim).

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 the future, they might even have applications in high-performance batteries.

* *

A monolayer of graphene has an astonishing optical property. Its optical absorption over the full optical spectrum is $\pi\alpha$, where α is the fine structure constant. (The exact expres-

FIGURE 66 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).

sion for the absorption is $A = 1 - (1 + \pi\alpha/2)^{-2}$.) The expression for the absorption is the consequence of the electric conductivity $G = e^2/4\hbar$ for every monolayer of graphene. The numeric value of the absorption is about 2.3 %. This value is visible by the naked eye, as shown in Figure 65. Graphene thus yields a way to 'see' the fine structure constant. Ref. 74

* *

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.

* *

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 66 shows. Making hydrophobic surfaces is an important part of modern materials 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 40

* *

Sometimes research produces bizarre materials. An example are the so-called *aerogels*, highly porous solids, shown in Figure 67. 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.

* *

Where do the minerals in the Amazonian rainforest come from? The Amazonas river

FIGURE 67 A piece of aerogel, a solid that is so porous that it is translucent (courtesy NASA).

washes many nutrient minerals into the Atlantic Ocean. How does the rainforest get its minerals back? It was a long search until it became clear that the largest supply of minerals is airborne: from the Sahara. Winds blow dust from the Sahara desert to the Amazonas basin, across the Atlantic Ocean. It is estimated that 40 million tons of dust are moved from the Sahara to the Amazonas rainforest every year.

* *

Some materials undergo almost unbelievable transformations. What is the final state of *moss*? Large amounts of moss often become peat (turf). Old turf becomes lignite, or brown coal. Old lignite becomes black coal (bituminous coal). Old black coal can become *diamond*. In short, diamonds can be the final state of moss.

* *

In materials science, there is a dream: to make a material that is harder than diamond. It is not clear whether this dream can be realized. The coming years will tell.

QUANTUM TECHNOLOGY

> I were better to be eaten to death with a rust than to be scoured to nothing with perpetual motion.
>
> William Shakespeare *King Henry IV*.

Quantum effects do not appear only in microscopic systems or in material properties. Also *applied* quantum effects are important in modern life: technologies such as transistors, lasers, superconductivity and other effects and systems have deeply affected our civilisation.

Transistors

Transistors are found in almost all devices that improve health, as well as in almost all devices for telecommunication. A *transistor*, shown in Figure 68 is a device that allows controlling a large electric current with the help of a small one; therefore it can play

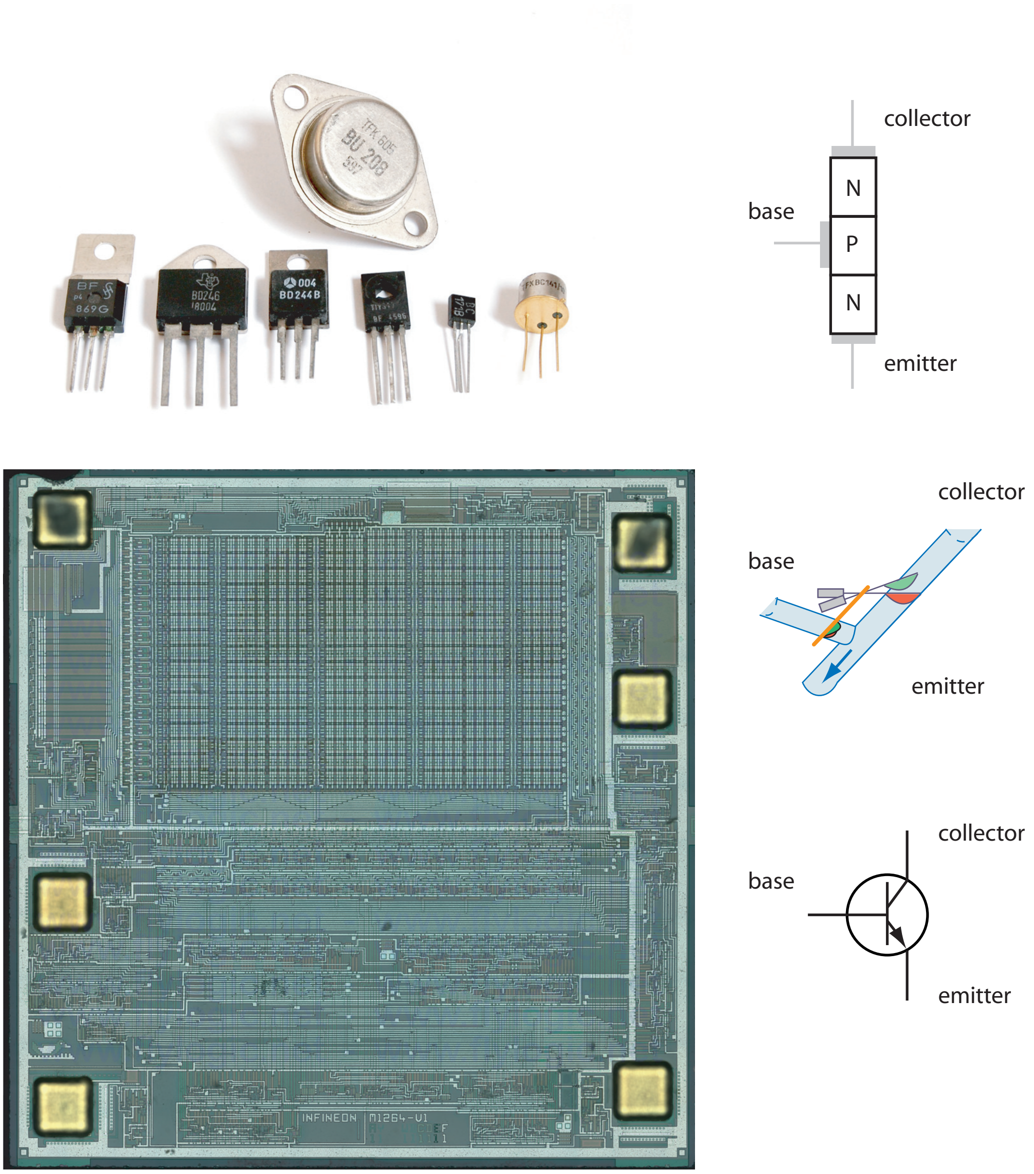


FIGURE 68 Top: examples of packaged single transistors. Right: the basic semiconductor structure, the equivalent water structure, and the technical drawing of an NPN transistor. Bottom: a typical integrated circuit for smart cards incorporating a large number of transistors. (© Benedikt Seidl, blog.ioactive.com)

the role of an electrically controlled switch or of an amplifier. Transistors are made from silicon and can be as small as a 2 by 2 μm and as large as 10 by 10 cm. Transistors are used to control the signals in pacemakers for the heart and the current of electric train engines. Amplifying transistors are central to the transmitter in mobile phones and switching transistors are central to computers and their displays.

Transistors are (almost exclusively) based on *semiconductors*, i.e., on materials where the electrons that are responsible for electric conductivity are *almost* free. The devices

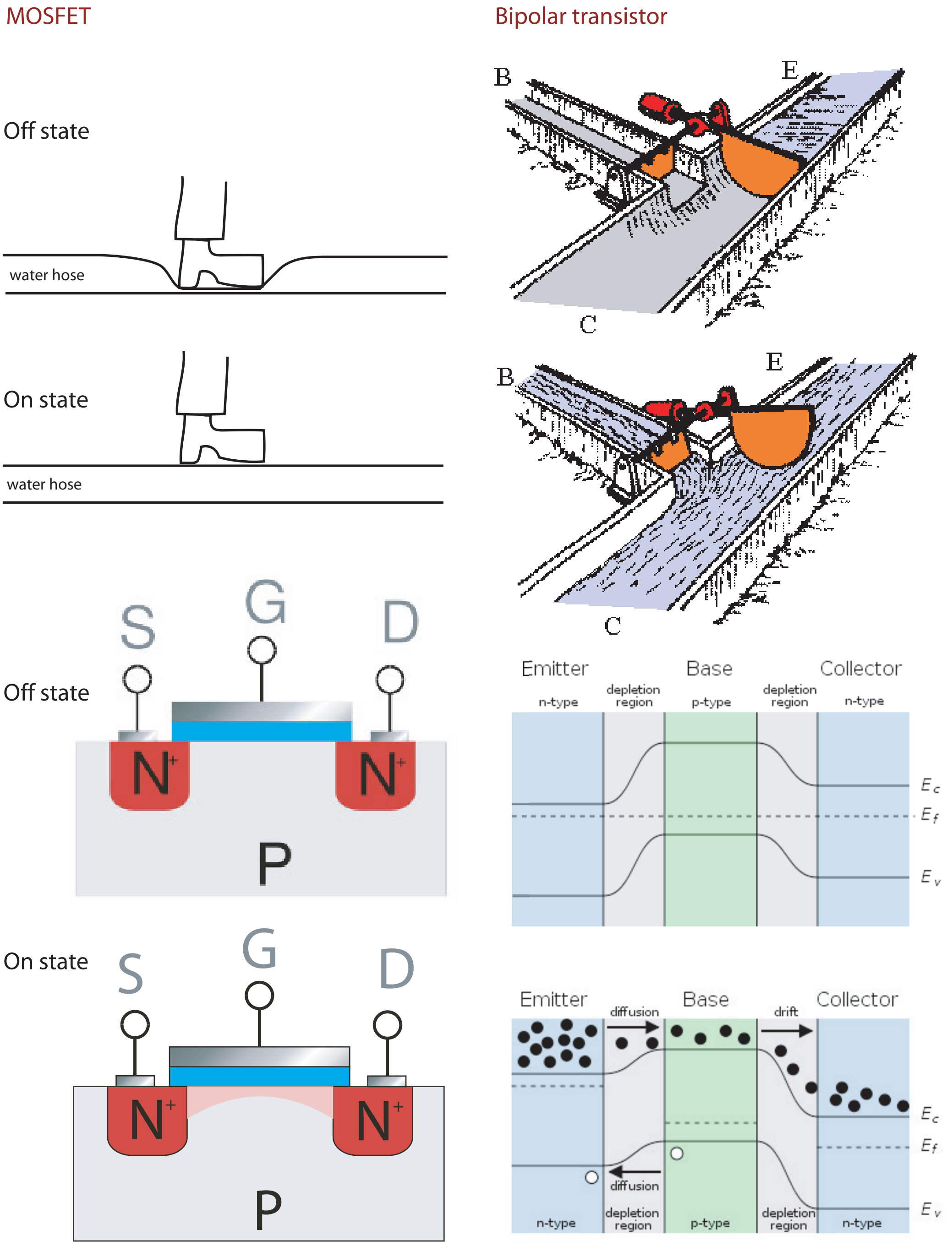


FIGURE 69 The working and construction of a metal-oxide silicon field effect transistor (left) and of a bipolar transistor (right). The 'off' and 'on' states are shown (© Leifi Physik, Wikimedia).

are built in such a way that applying an electric signal changes the conductivity. Every explanation of a transistor makes use of potentials and tunnelling; transistors are applied quantum devices.

The transistor is just one of a family of semiconductor devices that includes the field-effect transistor (FET), the metal-oxide-silicon field-effect transistor (MOSFET), the junction gate field-effect transistor (JFET), the insulated gate bipolar transistor (IGBT) and the unijunction transistor (UJT), but also the memristors, diode, the PIN diode, the Zener diode, the avalanche diode, the light-emitting diode (LED), the photodiode, the photovoltaic cell, the diac, the triac, the thyristor and finally, the integrated circuit (IC). These are important in industrial applications: the semiconductor industry has at least 300 thousand million Euro sales every year (2010 value) and employ millions of people across the world.

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 indeed 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. To ensure this, low temperature is usually needed. Despite a large amount of data, it took over 40 years to reach a full understanding of superconductivity. Ref. 75 This happened in 1957, when Bardeen, Cooper and Schrieffer published their results. At low temperatures, electron behaviour in certain materials is dominated by an *attractive* interaction that makes them form pairs. These so-called *Cooper pairs* are effective bosons. And bosons can all be in the same state, and can thus effectively move without friction.

In superconductivity, the attractive interaction between electrons is due to the deformation of the lattice. At low temperature, 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, these 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. As a result, they received the Nobel Prize in 1972.*

Another type of motion without friction is *superfluidity*. In 1937, Pyotr Kapitsa had understood that usual liquid helium, i.e., ^{4}He, below a transition observed at the temperature of 2.17 K, is a *superfluid*: the liquid effectively moves without friction through

---

* For John Bardeen (b. 1908 Madison, d. 1991 Boston), 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.

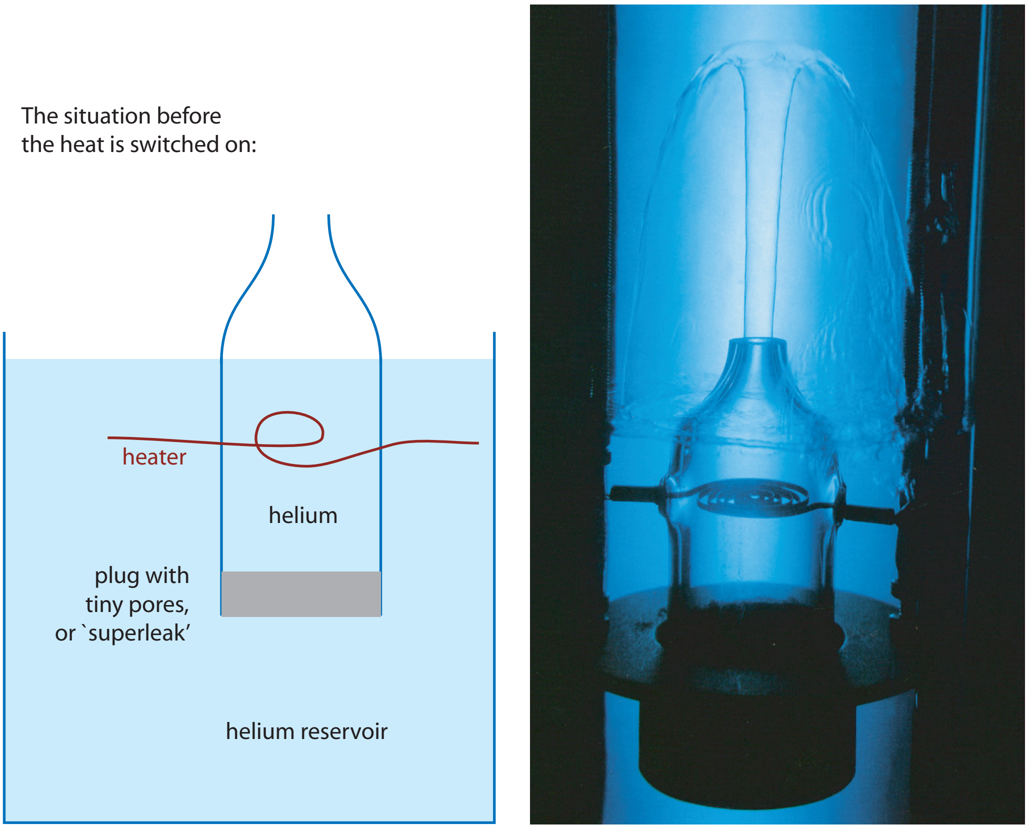


FIGURE 70 The superfluidity of helium 4 can be used to produce the fountain effect above a disc with very small pores, through which superfluid helium can pass, but normal fluid cannot. Superfluid helium 4 has a large thermal conductivity and flows towards a heated region trying to cool it down again, whereas the normal liquid cannot return back through the pores. This thermomechanical effect leads to the fountain (© Pacific Institute of Theoretical Physics).

devices, tubes, etc. More precisely, liquid helium remains a mixture of a superfluid component and a normal component; only the superfluid component moves without friction. Superfluid helium is even able, after an initial kick, to flow over obstacles, such as glass walls, or to flow out of bottles. A well-known effect of superfluidity is shown in Figure 70. Superfluidity occurs because the ^{4}He atom is a boson. Therefore no pairing is necessary for it to move without friction. This research earned Kapitsa a Nobel Prize in 1978.

In 1972, Richardson, Lee and Osheroff found that even ^{3}He is superfluid, provided that the temperature is 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. They received the Nobel Prize in 1996 for this discovery.

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

---

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

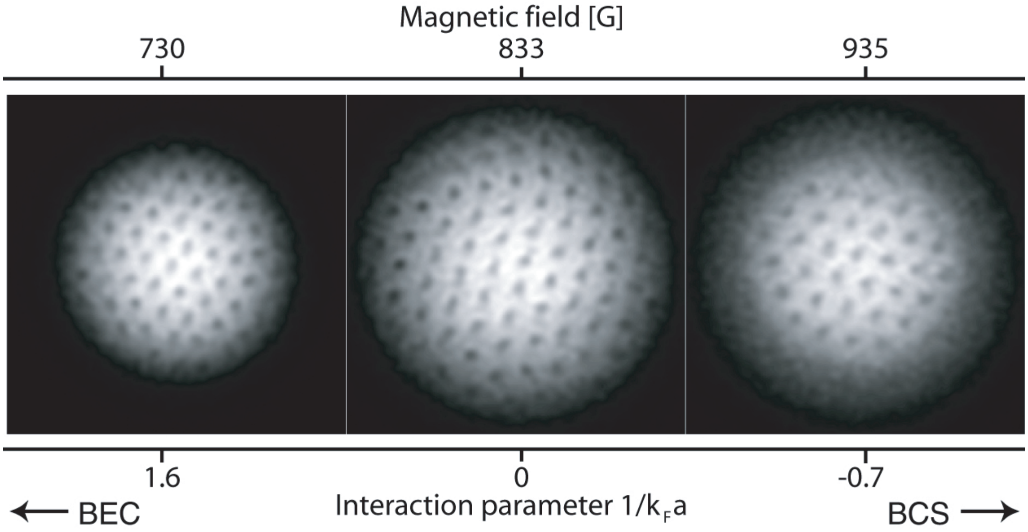


FIGURE 71 A vortex lattice in cold lithium gas, showing their quantized structure (© Andre Schirotzek).

^{3}He atoms, the basic mechanism that leads to superfluidity, turns out to be the spin-spin interaction.

Superfluidity has also been observed in a number of *gases*, though at much lower temperatures. Studying the behaviour of gases at lowest temperatures has become popular in recent years. 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 such states 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. All these condensates show superfluidity.

Superfluids are fascinating substances. Vortices also exist in them. But in superfluids, be they gases or liquids, 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 $h/2m_{^3\mathrm{He}}$. (This is also the case in superconductors.) Vortices in superfluids have quantized angular momentum. An effect of the quantization can be seen in Figure 71. In superfluids, these quantized vortices flow forever! Like in ordinary fluids, also in superfluids one can distinguish between laminar and turbulent flow. The transition between the two regimes is mediated by the behaviour the *vortices* in the fluid. Present research is studying how these vortices behave and how they induce the transition. Ref. 76

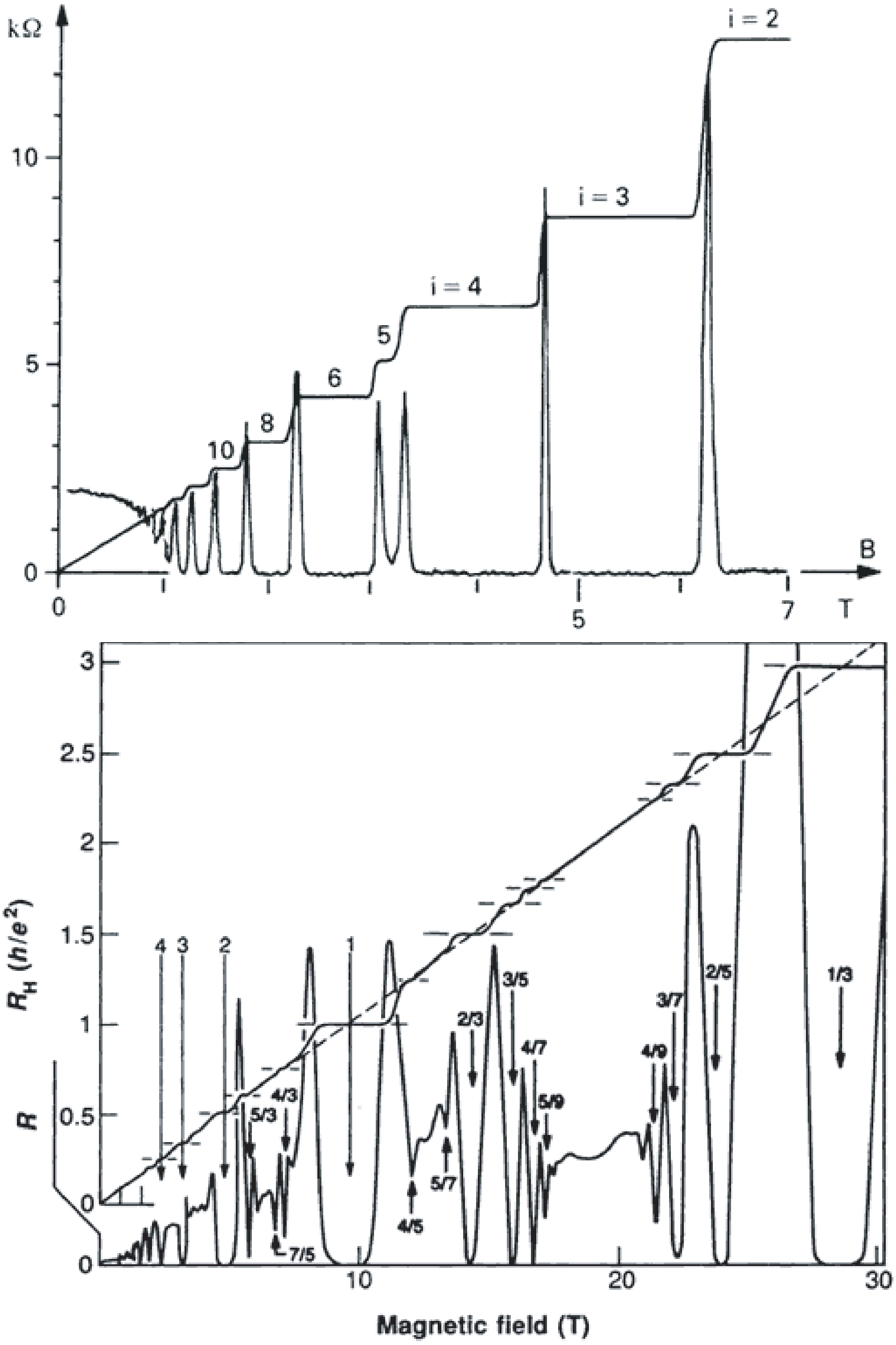


FIGURE 72 The quantum Hall effect (above) and the fractional quantum Hall effect (below): each graph yielded a Nobel prize. The graphs show how the Hall resistance and the Ohmic resistance vary with the applied magnetic field at very low temperature. The step height is quantized in integer or simple fractions of $h/e^2 = 25.812\,807\,557(18)\,\mathrm{k}\Omega$. Quantum Hall experiments allow the most precise determination known to date of this constant of nature.

THE FRACTIONAL QUANTUM HALL EFFECT

The fractional quantum Hall effect is one of the most intriguing discoveries of materials science, and possibly, of physics as a whole. The effect concerns the flow of electrons in Ref. 77 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 Ref. 78 prediction was indeed verified in 1997.

Page 97 We encountered the (classical) Hall effect above. The story continues with the discovery by Klaus von Klitzing of the *quantum* Hall effect. In 1980, Klitzing and his Ref. 79 collaborators found that in two-dimensional systems at low temperatures – about 1 K – the electrical conductance S, also called the Hall conductance, is quantized in multiples of the quantum of conductance

$$S = n\,\frac{e^2}{h}\ . \qquad (11)$$

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

$$R = \frac{1}{n}\frac{h}{e^2} = \frac{1}{n}\,25,812\,807\,557(18)\,\mathrm{k}\Omega\ . \qquad (12)$$

The values are independent of material, temperature, or magnetic field. They are constants of nature. Von Klitzing received the Nobel Prize in 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, in 1982, it was found that in extremely strong magnetic fields and at extremely low temperatures, the conductance could vary in steps *one third* that Ref. 80 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 Ref. 77 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.

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 quasiparticles might well be the most complex and fascinating aspects of physics, but exploring them would lead us too far from the aim of our adventure.

Ref. 81 In 2007, a new chapter in the story was opened by Andre Geim and his team, and a

---

2003.

TABLE 8 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

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.* Page 98

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 a mechanism similar to 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.

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 8.

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.

LASERS AND OTHER SPIN-ONE VECTOR BOSON LAUNCHERS

Photons are vector bosons; a lamp is thus a vector boson launcher. All existing 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 *recombination lamps* generate (cold) light through recombination of charges in semiconductors or li- Ref. 82

---

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

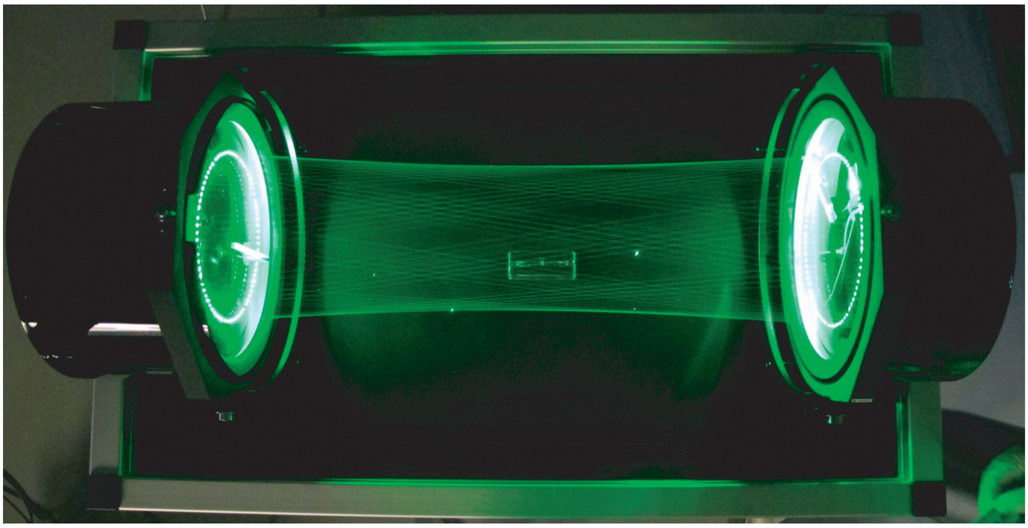


FIGURE 73 The beauty of lasers: the fine mesh created by a green laser delay line (© Laser Zentrum Hannover).

quids. The latter are the only lamp types found in living systems. The other main sources of light are *lasers*. All light sources are based on quantum effects, but for lasers the connection is especially obvious. The following table gives an overview of the main types and their uses.

TABLE 9 A selection of lamps and lasers.

Lamp type, application	Wave-length	Bright-ness or power	Cost	Life-time
Incandescent lamps				
Oil lamps, candles, for illumination	white	up to 500 lm	1 cent/lm	5 h
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	up to 10^{44} W	free	up to thousands of millions of years
Gas discharge lamps				
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

TABLE 9 A selection of lamps and lasers (continued).

Lamp type, application	Wave-length	Bright-ness or power	Cost	Life-time
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	up to 10^{20} W	free	up to thousands of millions of years
Recombination lamps				
Foxfire in forests, e.g. due to *Armillaria mellea*, *Neonothopanus gardneri* or other bioluminescent fungi	green	just visible	free	years
Firefly, to attract mates	green-yellow		free	*c.* 10 h
Large deep sea squid, *Taningia danae*, producing light flashes, to confuse prey	red	*c.* 1 W	free	years
Deep-sea fish, such as *angler fish*, to attract prey or find mates	white	*c.* 1 μW	free	years
Deep-sea medusae, to produce attention so that predators of predators are attracted	blue and all other colours		free	years
Light emitting diodes, for measurement, illumination and communication	red, green, blue, UV	up to 150 lm/W, up to 5 W	10 cent/lm	15k to 100 kh
Synchrotron radiation sources				
Electron synchroton source	X-rays to radio waves	pulsed	many MEuro	years
Maybe some stars	broad spectrum		free	thousands of years
Ideal white lamp or laser	visible	*c.* 300 lm/W	0	∞
Ideal coloured lamp or laser	green	683 lm/W	0	∞
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, now obsolete	several blue and green lines	up to 100 W	10 kEuro	

TABLE 9 A selection of lamps and lasers (continued).

Lamp type, application	Wave-length	Brightness or power	Cost	Life-time
Krypton laser, for pumping and laser shows, now obsolete	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 hobbyists	337.1 nm	pulsed up to 1 MW	down to a few hundred Euro	limited by metal electrode lifetime
Water vapour laser, for research, now obsolete	many lines between 7 and 220 µm, often 118 µm	CW 0.5 W, pulsed much higher	a few kEuro	
CO_2 laser, for cutting, welding, glass welding and surgery	10.6 µm	CW up to 100 kW, pulsed up to 10 TW	*c.* 100 Euro/W	1500 h
Excimer laser, for lithography in silicon chip manufacturing, eye surgery, laser pumping, psoriasis treatment, laser deposition	193 nm (ArF), 248 nm (KrF), 308 nm (XeCl), 353 nm (XeF)	100 W	10 to 500 kEuro	years
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	10 kEuro	1 khour
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, all obsolete	1.3 to 4.2 µm	up to MW in CW mode	over 10 MEuro	unknown
Liquid dye lasers				

TABLE 9 A selection of lamps and lasers (continued).

Lamp type, application	Wavelength	Brightness or power	Cost	Lifetime
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, diluted marmelade and many other liquids work as laser material	IR, visible	usually mW	1 kEuro	a few minutes
Solid state lasers				
Ruby laser (obsolete), for holography and tattoo removal	694 nm		1 kEuro	
Nd:YAG (neodymium:yttrium aluminium granate) laser, for material processing, surgery, pumping, range finding, velocimetry, also used with doubled frequency (532 nm), with tripled frequency (355 nm) and with quadrupled frequency (266 nm), also used as slab laser	1064 nm	CW 10 kW, pulsed 300 MW	50 to 500 kEuro	1000 h
Er:YAG laser, for dermatology	2940 nm			
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 fibre 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	

TABLE 9 A selection of lamps and lasers (continued).

Lamp type, application	Wave-length	Bright-ness or power	Cost	Life-time
Semiconductor lasers				
GaN laser diode, for optical recording	355 to 500 nm, depending on doping	up to 150 mW	a few Euro to 5 kEuro	*c.* 10 000 h
AlGaAs laser diode, for optical recording, pointers, data communication, laser fences, bar code readers (normal or vertical cavity)	620 to 900 nm, depending on doping	up to 1 W	below 1 Euro to 100 Euro	*c.* 10 000 h
InGaAsP laser diode, for fiberoptic communication, laser pumping, material processing, medical 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	up to 4 W	*c.* 10 kEuro	*c.* 1 000 h
Hybrid silicon lasers, for research	IR	nW	0.1 MEuro	
Free electron lasers				
Used for materials 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				

FROM LAMPS TO LASERS

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 the most friendly to the environment? Challenge 62 s

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 whenever a lamp is sufficiently turned up, it becomes a laser. 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). Most materials can be used to make lasers for fun, including water, beer and vodka.

Lasers produce radiation in the range from microwaves and extreme ultraviolet. They have the special property of emitting *coherent* light, usually in a collimated beam. Therefore lasers achieve much higher light intensities than lamps, allowing their use as tools. In modern lasers, the coherence length, i.e., the length over which interference can be observed, can be thousands of kilometres. Such high quality light is used e.g. in gravitational wave detectors.

People have become pretty good at building lasers. Lasers are used to cut metal sheets up to 10 cm thickness, others are used instead of knives in surgery, others increase surface hardness of metals or clean stones from car exhaust pollution. Other lasers drill holes in teeth, measure distances, image biological tissue or grab living cells.

Some materials amplify light so much that end mirrors are not necessary. This is the case for nitrogen lasers, in which nitrogen, or simply air, is used to produce a UV beam. Even a laser made of a single atom (and two mirrors) has been built; in this example, only Ref. 83 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.

THE THREE LIGHTBULB SCAMS

In the 1990s, all major light bulb producers in the world were fined large sums because they had agreed to keep the lifetimes of light bulbs constant. It is no technical problem to make light bulbs that last 2000 hours; however, the producers agreed not to increase the lifetime above 700 hours, thus effectively making every lightbulb three times as expensive as it should. This was the first world-wide light bulb scam.

Despite the fines, the crooks in the light bulb industry did not give up. In 2012, a large German light bulb maker explained in its advertising that its new light sources were much longer living than its conventional light bulbs, which, they explained on their ads, lasted only 500 hours. In other words, not only did the fines not help, the light bulb industry even *reduced* the lifetimes of the light bulbs from the 1990s to 2012. This was the second light bulb scam.

Parallel to the second scam, in the years around 2000, the light bulb industry started lobbying politics with the false statement that light bulbs were expensive and would waste energy. As a result of the false data provided from the other two scams, light bulbs were forbidden in Europe, with the result that consumers in Europe are now forced to buy other, much more expensive means of illumination. On top of this, many of these more expensive light sources are bad for the eyes. Indeed, flickering mercury or flickering LED lamps, together with their reduced colour spectrum, force the human visual system in overload mode, a situation that does not occur with the constantly glowing light bulbs. In other words, with this third scam, the light bulb industry increased their profits even more, while ruining the health of consumers at the same time. One day, maybe, parliaments will be less corrupt and more sensible. The situation will then again improve.

Applications of lasers

As shown in Figure 74, lasers can be used to make beautiful parts – including good violins and personalized bicycle parts – via sintering of polymer or metal powders. Lasers are used in rapid prototyping machines and to build architectural models. Lasers can cut paper, metal, plastics and flesh.

Lasers are used to read out data from compact discs (CDs) and digital versatile discs (DVDs), are used in the production of silicon integrated circuits and for the transport telephone signals through optical fibres. In our adventure, we already encountered lasers that work as loudspeakers. Vol. I, page 408 Important advances in recent years came from the applications of *femtosecond laser* pulses. Femtosecond pulses generate high-temperature plasmas in the materials they propagate through; this happens even in air, if the pulses are focused. The effect has been used to create luminous three-dimensional displays floating in mid-air, as shown in Figure 74, or in liquids. Such short pulses can also be used to cut material without heating it, for example to cut bones in skull operations. The lack of heating is so complete that femtosecond lasers can be used to engrave matches and even dynamite without triggering a reaction. Femtosecond lasers have been used to make high resolution holograms of human heads within a single flash. Recently such lasers have been used to guide lightning along a predetermined path; they seem promising candidates for laser ligtning rods. Ref. 84 A curious demonstration application of femtosecond lasers is the storage of information in fingernails (up to 5 Mbit for a few months), in a way not unlike that used in recordable compact discs (CD-R). Ref. 85

Lasers are used in ophthalmology, with a technique called *optical coherence tomography*, to diagnose eye and heart illnesses. Around 2025, there will finally be laser-based breast screening devices that use laser light to search for cancer without any danger to the patient. The race to produce the first working system is already ongoing since the 1990s. Additional medical laser applications will appear in the coming years.

Lasers have been used in recent demonstrations, together with image processing software, to kill mosquitos in flight; other lasers are burning weeds while the laser is moved over a field of crops. One day, such combined laser and vision systems will be used to evaporate falling rain drops one by one; as soon as the first such *laser umbrella* will be available, it will be presented here. The feat should be possible before the year 2022. Ref. 86

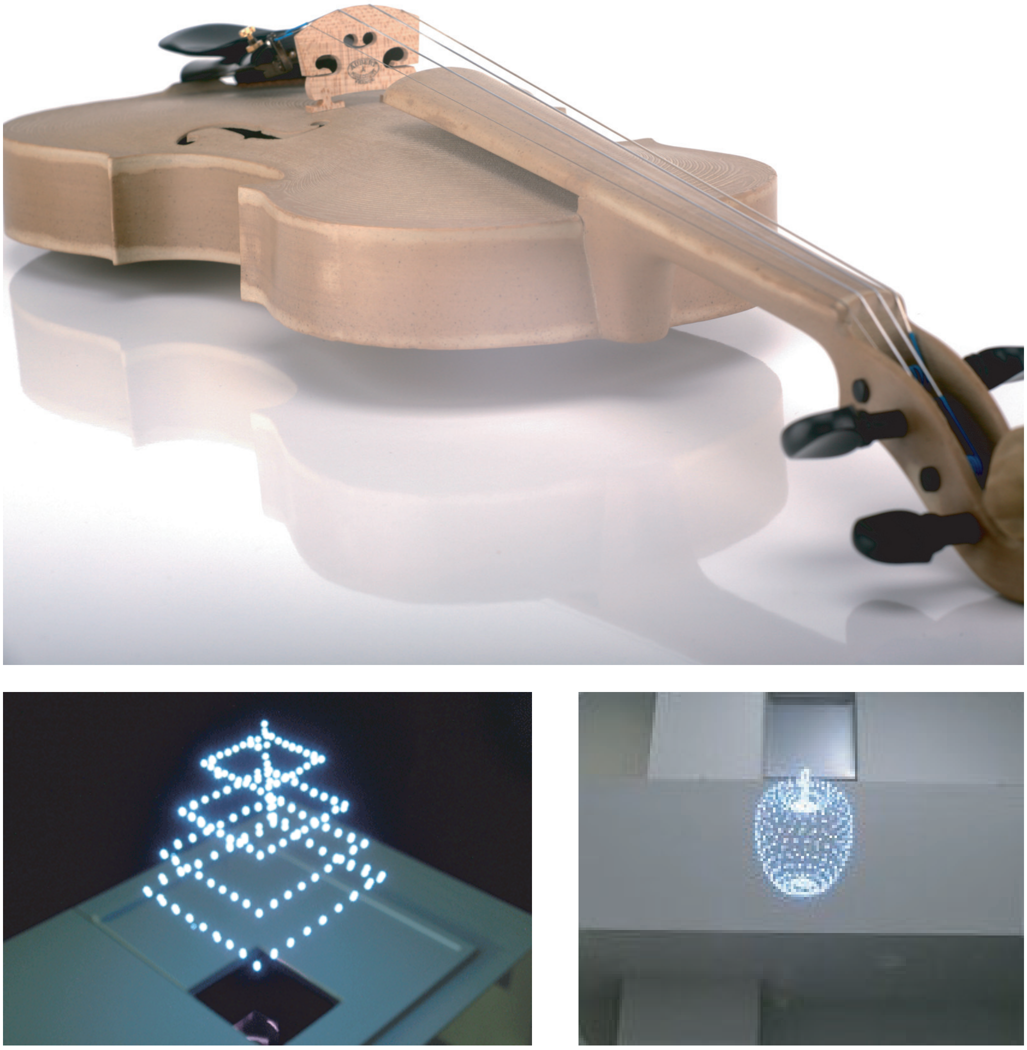


FIGURE 74 Some laser applications. Top: a violin, with excellent sound quality, made of a single piece of polymer (except for the chords and the black parts) through laser sintering of PEEK by EOS from Krailling, in Germany. Bottom: a display floating in mid-air produced with a galvanometer scanner and a fast focus shifter (© Franz Aichinger, Burton).

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.

* *

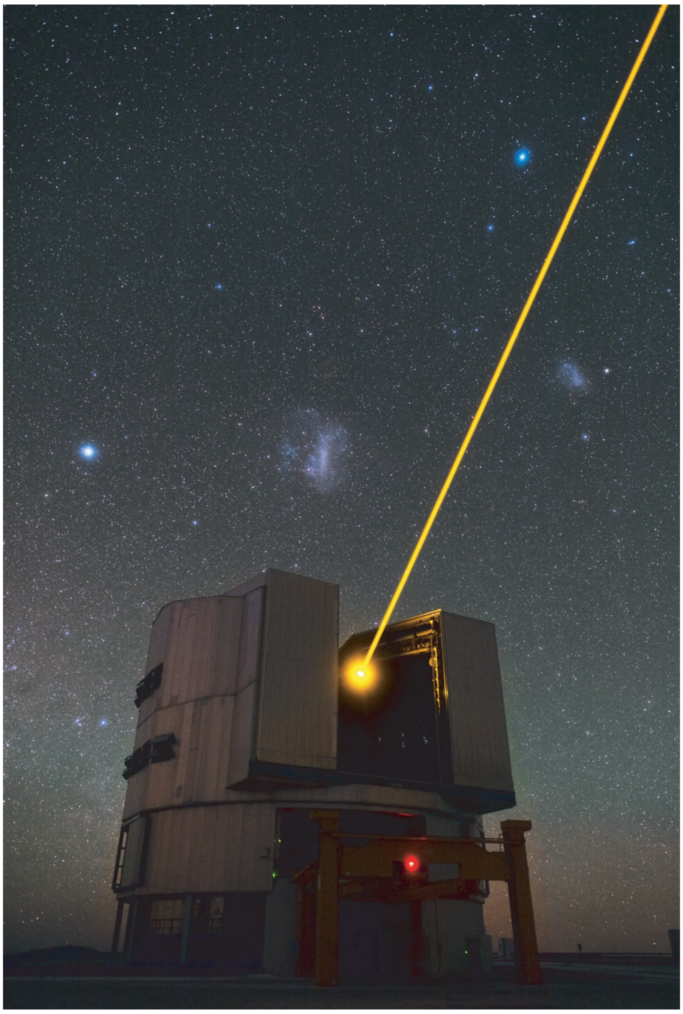


FIGURE 75 The most expensive laser pointer: a yellow 10 W laser that is frequency-stabilized at the wavelength of sodium lamps allows astronomers to improve the image quality of terrestrial telescopes. By exciting sodium atoms found at a height of 80 to 90 km, the laser provides an artificial guide star that is used to compensate for atmospheric turbulence, using the adaptive optics built into the telescope. (© ESO/Babak Tafreshi).

In 2007, Humphrey Maris and his student Wei Guo performed an astonishing experiment: they filmed *single electrons* with a video camera. Actually the truth is a bit more complicated, but it is not a lie to summarize it in this way. Ref. 87

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

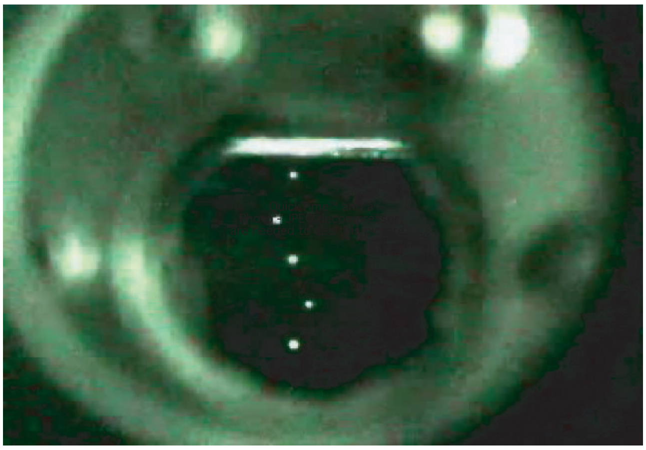


FIGURE 76 How to image single electrons with a video camera: isolated electrons surrounded by bubbles that explode in liquid helium under negative pressure produce white spots (mpg film © Humphrey Maris).

shone light through the helium. When the pressure became negative enough they saw the explosions of the bubbles. Figure 76 shows the video. The experiment is one of the highlights of experimental physics in the last decade.

* *

Is it possible to make A4-size flexible colour displays for an affordable price and with print-like quality? Challenge 63 d

* *

Will there ever be rechargeable batteries with an energy content per mass that is comparable to diesel oil? How long will it take, from 2014 onwards, until the last company producing electric cars powered by batteries stops production?

* *

How many companies promising free energy, engineers promising cars powered by water, politicians promising fusion energy or quacks promising food additives or sugar pills that cure cancer will we see every year?

* *

Challenge 64 r Will there ever be room-temperature superconductivity?

* *

Challenge 65 r Will there ever be desktop laser engravers for 1000 euro?

* *

Challenge 66 s Will there ever be teleportation of everyday objects?

* *

One process that quantum physics does *not* allow is telepathy. An unnamed space agency

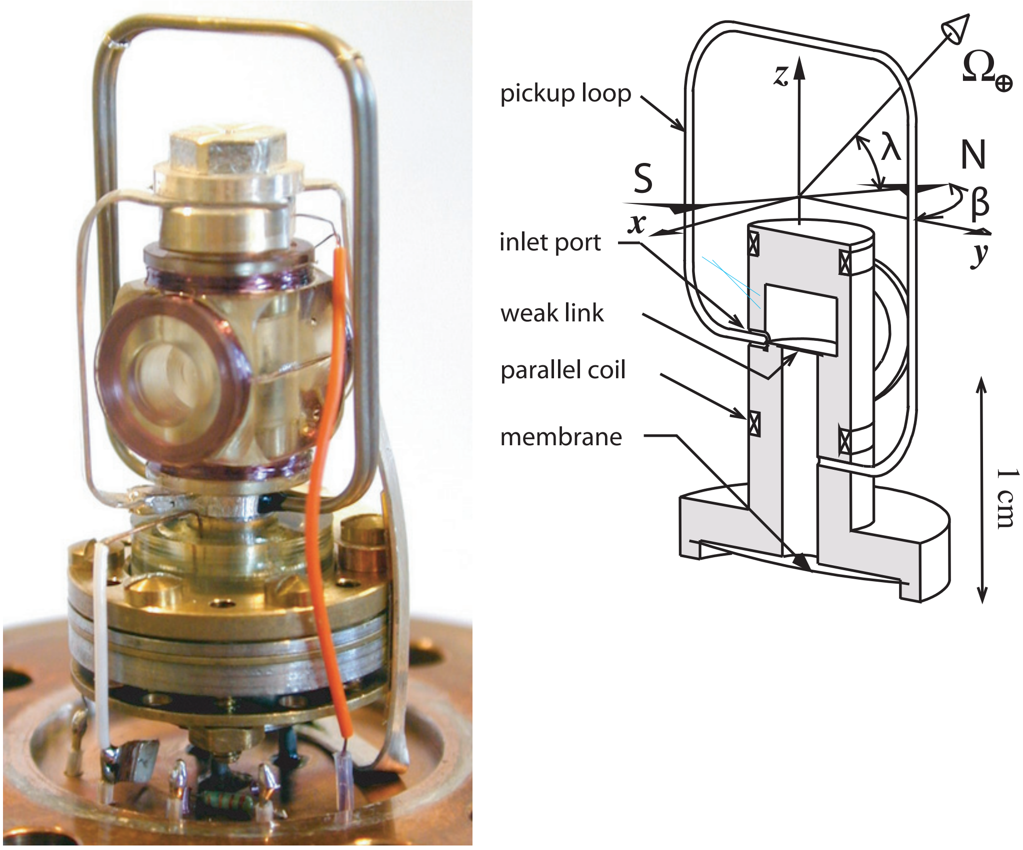


FIGURE 77 The interior of a gyroscope that uses superfluid helium (© Eric Varoquaux).

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 Ref. 88 that telepathy was useless. (This not a joke.) It is unclear why the 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 67 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 68 d

* *

Challenge 69 r Will there ever be radio-controlled flying toys in the size of insects?

* *

Ref. 107 By shining an invisible and harmless laser onto cars driving by, it is now possible to

FIGURE 78 These colours were produced on steel using just an infrared laser shining on it. (© Trotec Laser at www.troteclaser.com)

detect whether the persons inside have drunk alcohol. Will this method ever become widespread? Challenge 70 s

* *

Ref. 108 In 1997, Eric Varoquaux and his 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 helium can behave like a Foucault pendulum. With a clever arrangement, shown in Figure 77, they were able to measure the rotation of the helium in the ring using phonon signals, and, finally, to detect the rotation of the Earth.

* *

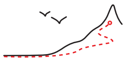
Lasers are quantum devices that can be used for many applications. Figure 78 shows a way to produce colours on steel by scanning a focused infrared laser beam over the surface. Why and how do the colours appear? Challenge 71 e

Summary on changing the world with quantum effects

Atoms form bonds. Quantum effects thus produce molecules, gases, liquids and solids, as well as all effects and properties of all materials. In the past, quantum effects have been used to develop numerous materials with desired properties, such as new steel types, new carbon fibre composites, new colourants, new magnetic materials and new polymers.

Quantum effects have been used to develop modern electronics, lasers, light detectors, data storage devices, superconducting magnets, new measurement systems and new production machines. Magnetic resonance imaging, computers, polymers, telecommunication and the internet resulted from applying quantum effects to technology.

Quantum effects will continue to be used to design new materials and systems: new nanoparticles to deliver drugs inside the body, new polymers, new crystals, new environmentally friendly production processes and new medical devices, among others.



CHAPTER 3

QUANTUM ELECTRODYNAMICS – THE ORIGIN OF VIRTUAL REALITY

The central concept that quantum field theory adds to 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 usual, *real* particles. Virtual particles can move faster than light and can move backward in time. Despite these strange properties, they have many observable effects. We explore the most spectacular ones.

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. The situation is illustrated in Figure 79. 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

$$E = mgh\,\alpha^2/2 \tag{13}$$

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

---

** Sipko Boersma published a paper in which he gave his reading of shipping manuals, advising captains to
Ref. 89 let the ships be pulled apart using a well-manned rowing boat. This reading has been put into question by
Ref. 90 subsequent research, however.

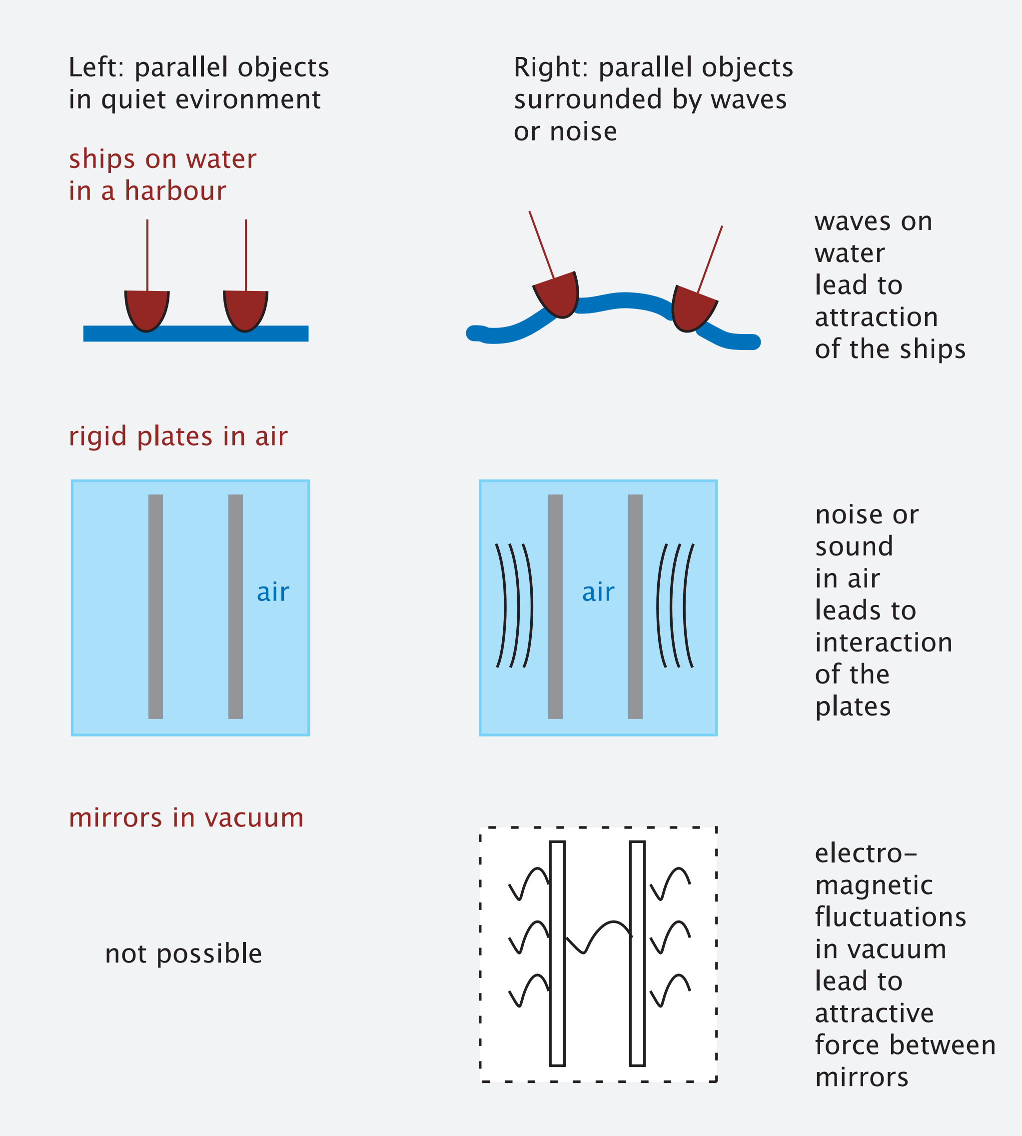


FIGURE 79 The analogy between ships in a harbour, metal plates in air and metal mirrors in vacuum.

$e = 2.718$. If the quality factor Q of an oscillating ship and its oscillation period T are given, the radiated power W is

$$W = 2\pi \frac{E}{QT} . \tag{14}$$

Vol. III, page 120 We saw above that radiation force (radiation pressure times area) is W/c, where c is the wave propagation velocity. For gravity water waves in deep water, we have the well-known relation Vol. III, page 295

$$c = \frac{gT}{2\pi} . \tag{15}$$

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

$$ma = m2\pi^2 \frac{h\alpha^2}{QT^2} . \tag{16}$$

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. 91

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 in a cathode ray tube – or today, in a plasma display – of a television, the phosphor, 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. 92, Ref. 93

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 Aat distance d is given by

$$\frac{ma}{A} = \frac{\pi^3}{120} \frac{\hbar c}{d^4} . \tag{17}$$

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. 94 Ref. 95 Ref. 96 Ref. 97

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 72 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. 98

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 that spontaneous emission is emission stimulated by the zero point fluctuations.

Ref. 99 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 73 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 Ref. 90 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 an analogy however, it remains valid.

THE LAMB SHIFT

In the old days, it was common that a person receives the Nobel Prize in Physics for observing the colour of a lamp – if the observation was sufficiently careful. In 1947, Willis Lamb (b. 1913 Los Angeles, d. 2008 Tucson) performed such a careful measurement of the spectrum of hydrogen. He found that the $2S_{1/2}$ energy level in atomic hydrogen lies slightly above the $2P_{1/2}$ level. This observation is in contrast to the calculation performed Vol. IV, page 188 earlier on, where the two levels are predicted to have the same energy. In contrast, the measured energy difference is 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?

The reason for contrast between calculation and observation is an approximation performed in the relativistic calculation of the hydrogen levels that took over twenty years to clarify. There are two equivalent explanations. One explanation is to say that the relativistic 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 second explanation is in line with the modern vocabulary of quantum electrodynamics. Quantum electrodynamics, or QED, is the (perturbative) approach to solve the coupled Dirac and Maxwell equations.

In short, Lamb discovered the first effect due to virtual particles. In fact, Lamb used microwaves for his experiments; only in the 1970 it became possible to see the Lamb shift with optical means. For this and similar feats Arthur Schawlow received the Nobel Prize in Physics in 1981.

The QED Lagrangian and its symmetries

In simplified 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:

$$\begin{aligned} \mathcal{L}_{\text{QED}} = \ & \overline{\psi}(i\hbar c \not{\partial} - c^2 m_k)\psi && \Big\} \ \text{the matter term} \\ & - \frac{1}{4\mu_0} F_{\mu\nu}F^{\mu\nu} && \Big\} \ \text{the electromagnetic field term} \\ & + e\hbar c A_\mu \overline{\psi}\gamma^\mu\psi \ . && \Big\} \ \text{the electromagnetic interaction term} \end{aligned} \qquad (18)$$

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 nuclei must be added; we will explore it below.)

Page 162

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. In fact, the Lagrangian also describes the motion of muons, tau leptons and all other charged particles. 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*, the electromagnetic coupling 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 developed independently by Julian Schwinger, Freeman Dyson, Richard Feynman and Tomonaga Shin'ichiro, four among the smartest physicists ever. *

The QED Lagrangian density contains the strength of the electromagnetic interaction in the form of the fine structure constant $\alpha = e^2/(4\pi\varepsilon_0\hbar c) \approx 1/137.036(1)$. This number is part of the Lagrangian; no explanation for its value is given, and the explanation was still unknown in the year 2016. It is one of the hardest puzzles of physics. 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 requirements were clearly spelled out, as we will discover in the rest of this volume.

Vol. I, page 435

Challenge 74 e

The Lagrangian density retains all symmetries that we know from classical physics. 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 relation.

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

---

* Tomonaga Shin'ichiro (b. 1906 Tokio, d. 1979 Tokio) developed quantum electrodynamics and won the 1965 Nobel Prize in 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

$$\frac{E}{V} = \frac{4\pi h}{c^3} \int_0^\infty \nu^3 \mathrm{d}\nu \,. \qquad (19)$$

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

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*

Vol. VI, page 40

$$l_{\mathrm{Pl}} = \sqrt{\hbar G/c^3} \approx 1.6 \cdot 10^{-35}\ \mathrm{m} \,. \qquad (20)$$

Vol. VI, page 40

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 value. 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 every sound source or for every wall speed. For example, experiments show that walls do not produce echoes if the wall or the sound source moves faster than sound. Walls do not produce echoes even if the sound source moves with them, if both objects move faster than sound. On the other hand, light mirrors *always* produce an image, whatever the involved speed of the light source or the mirror may be. These observations confirm that the speed of light is the same for all observers: it is *invariant* and a limit speed. (Can you detail the argument?) In contrast, the speed of sound in air depends on the observer; it is *not* invariant.

Challenge 75 s

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 of mirrors and of their properties are sufficient to derive special relativity.

Vol. II, page 22

---

* 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.

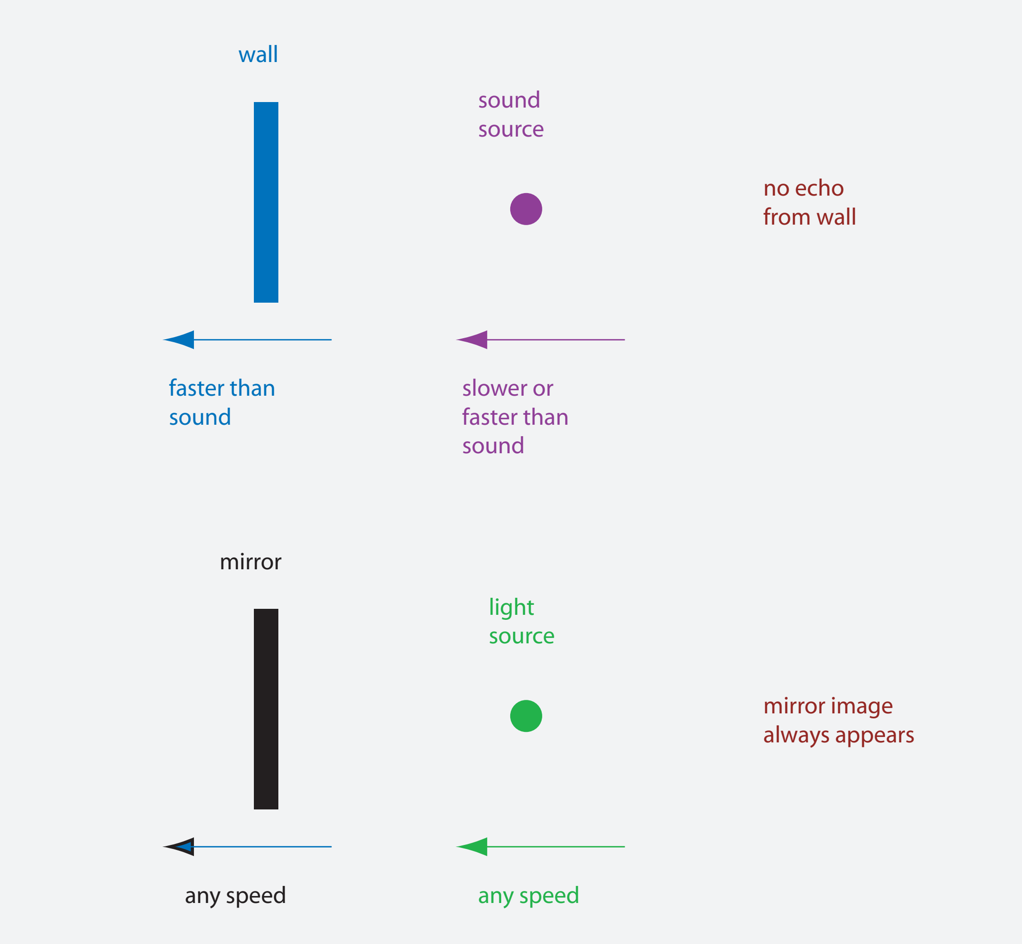


FIGURE 80 A fast wall does not produce an echo; a fast mirror does.

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 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 been measured directly, up to now. We will explore the issue in more detail below. (Can you explain why accelerated mirrors emit light, but not matter?)

Page 146
Challenge 76 s

Photons hitting photons

Usually, light can cross light undisturbed: interference is the proof and the result of this basic property of light. But there is an exception. When virtual particles are taken into account, light beams can 'bang' onto each other – though only slightly. 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 other words, the bang is extremely weak. When two light beams cross, most photons will pass undisturbed. The cross-section A is approximately

$$A \approx \frac{973}{10\,125\pi}\,\alpha^4 \left(\frac{\hbar}{m_\mathrm{e}c}\right)^2 \left(\frac{\hbar\omega}{m_\mathrm{e}c^2}\right)^6 \tag{21}$$

for the everyday case that the energy $\hbar\omega$ of the photon is much smaller than the rest energy $m_\mathrm{e}c^2$ 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 will ever be observable 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, this prediction was also confirmed experimentally. Ref. 100

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 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 Challenge 77 e

$$\gamma_{29.2\mathrm{GeV}} + n\,\gamma_\mathrm{green} \rightarrow \mathrm{e}^+ + \mathrm{e}^- \tag{22}$$

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. 101 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, as described above. The damping depends on the shape and thus also on the direction of the body's motion.

Vol. I, page 391 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. 102 of the vacuum should produce a mean square displacement

$$\langle d^2 \rangle = \frac{\hbar}{m} t \qquad (23)$$

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 78 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 243 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 79 s its total mass. Can you confirm this?

QED is a *perturbative* description. This means, that any predicted result R is found as a Taylor series of powers of a small parameter:

$$R = R_0 + R_1\alpha + R_2\alpha^2 + R_3\alpha^3 + R_4\alpha^4 + ... \qquad (24)$$

In QED, the small parameter is the fine structure constant $\alpha = 1/137.036(1)$. With the help of the perturbation series, the exact result R is approximated more and more precisely.

Now, 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

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \lim_{\Delta t \to 0} \Delta x \frac{1}{\Delta t} , \qquad (25)$$

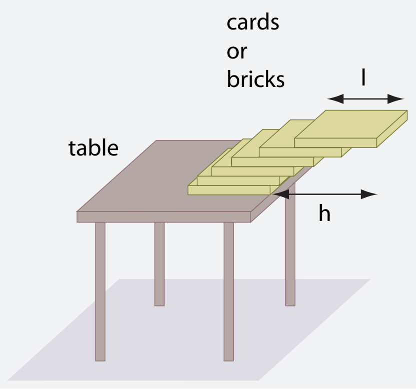


FIGURE 81 What is the maximum possible value of h/l?

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. 103

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 37

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.

* *

A famous riddle, illustrated in Figure 81, asks 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 nor 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 80 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? In short, the fine structure constant limits the number of chemical elements in nature. 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. 104 Challenge 81 s Page 153

* *

Stripping 91 of the 92 electrons off an uranium atom allows researchers to check with high precision whether the innermost electron still is described by QED. The electric field near the uranium nucleus, 1 EV/m, is the highest achievable in the laboratory; the field value 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; but even for these extremely high fields, the value matches the calculation. Ref. 105

* *

Is there a *critical magnetic field* in nature, like there is a critical electric field, limited by spontaneous pair production? Challenge 82 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. Ref. 106

In radioactive decay, the slowness record is held by ^{209}Bi, with over 10^{19} years of half-life. Page 342

* *

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 83 s

* *

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 84 s Ref. 109

* *

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 85 d

* *

Challenge 86 s Can you explain why mud is not clear?

* *

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. 201

* *

Challenge 87 s Can the universe ever have been smaller than its own Compton wavelength?

* *

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.

* *

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 you to check your understanding and to compare it with that of others.

Imagine that you are on a perfectly frictionless surface and that you want to move to its border. How many methods can you find to achieve this? Any method, so tiny its effect may be, is allowed.

Classical physics provided quite a number of methods. We saw that for rotating ourselves, we just need to turn our arm above the head. For translation motion, throwing a shoe or inhaling vertically and exhaling horizontally are the simplest possibilities. Can you list at least six additional methods, maybe some making use of the location of the surface on Earth? What would you do in space? Challenge 88 s

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 89 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 90 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 91 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 92 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 93 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? Challenge 94 s Can you find at least two other methods to move yourself using quantum gravity effects? Can you find one from string theory?

If you want points for the test, the marking is simple. For students, every working method gives one point. Eight points is ok, twelve points is good, sixteen points is very good, and twenty points or more is excellent. For graduated physicists, the point is given only when a back-of-the-envelope estimate for the ensuing momentum or acceleration is provided.

A SUMMARY OF QUANTUM ELECTRODYNAMICS

The shortest possible summary of quantum electrodynamics is the following:

▷ Everyday matter is made of charged elementary particles that interact through photon exchange in the way described by Figure 82.

No additional information is necessary. In a bit more detail, quantum electrodynamics starts with *elementary particles* – characterized by their mass, spin, charge, and parities – 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 electromagnetic *decay* is described as the interaction with the virtual photons of the vacuum.

▷ The Feynman diagram of Figure 82 provides an *exact* description of *all* electromagnetic phenomena and processes.

No contradiction between observation and calculation are known. In particular, the Feynman diagram is equivalent to the QED Lagrangian of equation (18). Page 126 Because QED is a perturbative theory, the Feynman diagram directly describes the first order effects;

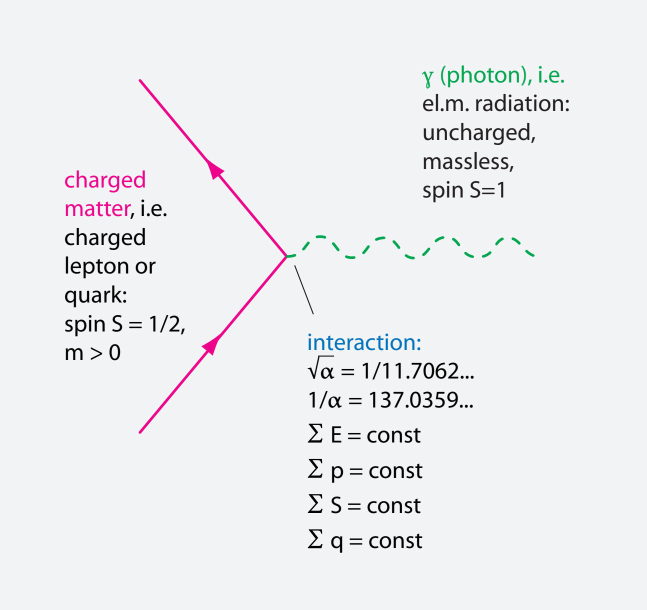


FIGURE 82 The basis of QED; more precisely, the fundamental diagram of QED as a perturbation theory in space-time.

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 82. 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.

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

$$d \approx \lambda_C = \frac{h}{m\,c} \;. \tag{26}$$

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:

$$d \approx \lambda_{dB} = \frac{h}{m\,v} \;. \tag{27}$$

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 electromagnetic radiation in daily life. All everyday objects and all images are described, 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 due to electromagnetism without any discrepancy from experiment, that does not mean that we understand every detail of every example of such motion. For example, nobody has described the motion of an animal with QED yet.* In fact, there is beautiful and fascinating research 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. 111 Ref. 112

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. 113

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. 114

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 82 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. Ref. 115

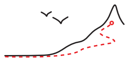
---

* 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. 110

FIGURE 83 The rainbow can only be explained fully if the fine structure constant α can be calculated (© ed g2s, Christophe Afonso).

If we want to be very strict, we need to add that we do *not* fully understand any colour, because we still do not know the origin of the fine structure constant. In particular, the fine structure constant determines the refractive index of water, and thus the formation of a rainbow, as pictured in Figure 83. Ref. 116

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. In our adventure however, our quest is the description of the *fundamentals* of motion. And so far, we have not achieved it. In particular, we still need to understand motion in the realm of atomic nuclei and the effect of the quantum of action in the domain of gravitation. We start with the latter topic.



Chapter 4

QUANTUM MECHANICS WITH GRAVITATION – FIRST STEPS

Gravitation is a weak effect. Indeed, 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 all in order to complete our mountain ascent.

Gravity acts on quantum systems: 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 and neutrons, as we will see. And the quantum of action plays an important role in the behaviour of black holes. We explore these topics now.

Falling atoms

In 2004 it finally became possible to repeat Galileo's leaning tower experiment with single atoms instead of steel balls. This is not an easy experiment because even the smallest effects disturb the motion. 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. 117

The experiment was difficult to perform, but the result is not surprising, because all falling everyday objects are made of atoms. Indeed, Galileo himself had predicted that atoms fall like stones, because parts of a body have to fall with the same acceleration as the complete body. But what is the precise effect of gravity on wave functions? This question is best explored with the help of neutrons.

Vol. I, page 202

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, as long as friction can be neglected. The general layout of the experiment is shown in Figure 84. How does a quantum particle behave in the same setting?

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

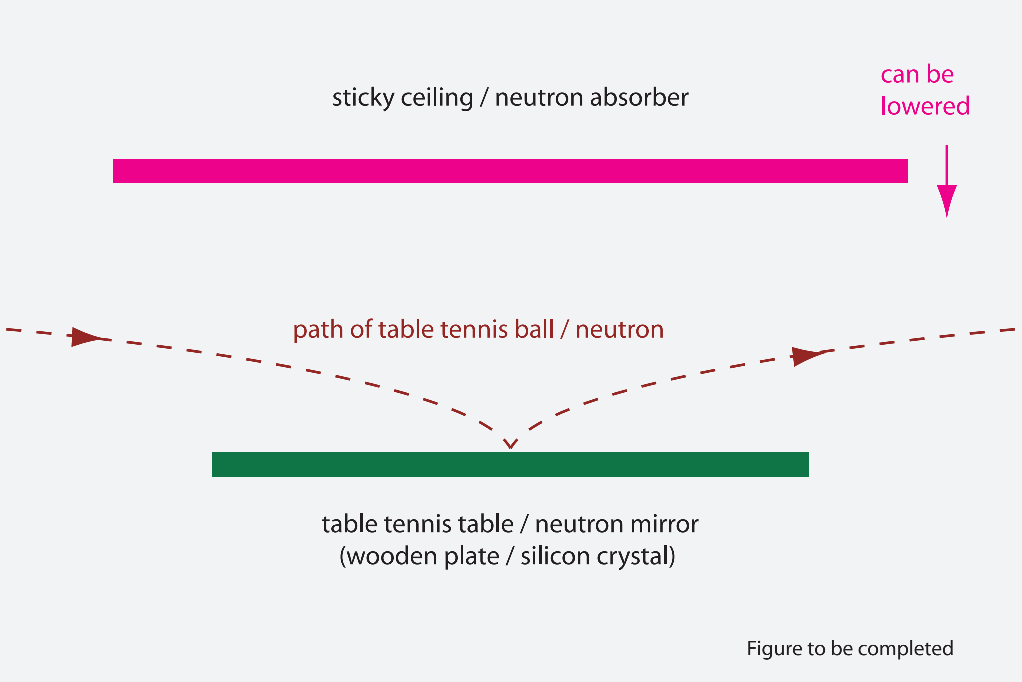


FIGURE 84 Table tennis and neutrons.

Challenge 95 e bounce of energy E_n and duration t_n,

$$n\hbar \sim E_n t_n \sim \frac{E_n^{3/2}}{g m^{1/2}} \, . \tag{28}$$

In other words, only discrete bounce heights, distinguished by the number n, are possible in the quantum case. The discreteness leads to an expected probability density that changes with height in discrete steps, as shown in Figure 84.

The best way to realize the experiment with quantum particles is to produce an intense beam of neutral particles, because neutral particles are not affected by the stray electromagnetic fields that are present in every laboratory. Neutrons are ideal, as they are produced in large quantities by nuclear reactors. The experiment was first performed in 2002, by Hartmut Abele and his group, after years of preparations. Ref. 118 Using several clever tricks, they managed to slow down neutrons from a nuclear reactor to the incredibly small value of 8 m/s, comparable to the speed of a table tennis ball. (The equivalent temperature of these ultracold neutrons is 1 mK, or 100 neV.) They then directed the neutrons onto a neutron mirror made of polished glass – the analogue of the table tennis table – and observed the neutrons bouncing back up. To detect the bouncing, they lowered an absorber – the equivalent of a sticky ceiling – towards the table tennis table, i.e., towards the neutron mirror, and measured how many neutrons still reached the other end of the table. (Both the absorber and the mirror were about 20 cm in length.)

Why did the experiment take so many years of work? 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,

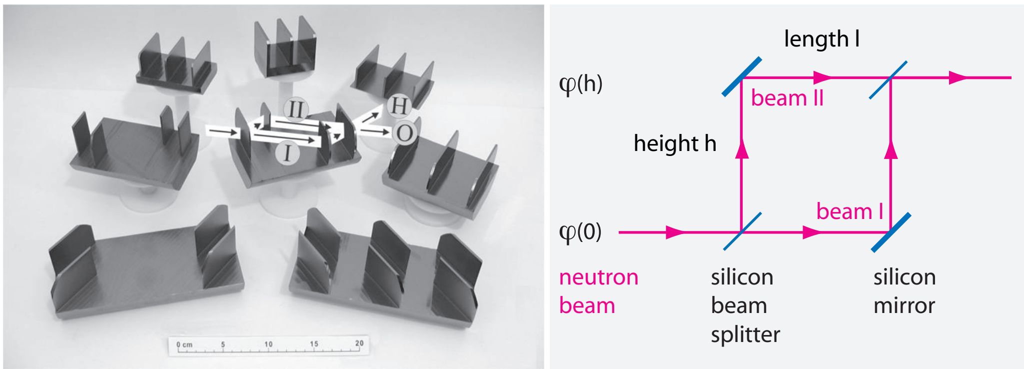


FIGURE 85 The weakness of gravitation. A neutron interferometer made of a silicon single crystal (with the two neutron beams I and II) can be used to detect the effects of gravitation on the phase of wave functions (photo © Helmut Rauch and Erwin Seidl).

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. Despite these small energy values, the team managed to measure the first few discrete energy levels. The results confirmed the prediction of the Schrödinger equation, with the gravitational potential included, to the achievable measurement precision.

In short, gravity influences wave functions. In particular, gravity changes the phase of 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 by Helmut Rauch and his team. Rauch had developed neutron interferometers based on single silicon crystals, shown in Figure 85, in which a neutron beam – again from a nuclear reactor – is split into two beams and the two beams are then recombined and brought to interference.

Challenge 96 s

Ref. 119

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 two different locations. The experiment is shown schematically on the right of Figure 85. The neutron beam is split; the two beams are deflected upwards, one directly, one a few centimetres further on, and then recombined.

Ref. 120

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

Challenge 97 ny

$$\Delta\varphi = \frac{mghl}{\hbar v} , \qquad (29)$$

where l is the horizontal distance between the two climbs and v and m are the speed and mass of the neutrons. All experiments – together with several others of similar simple elegance – have confirmed the prediction by quantum theory within experimental errors.

Ref. 121

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 in the potential, the Schrödinger and Dirac equations can thus be used, for example, to describe the growth and the processes inside trees. Trees can mostly be described with quantum electrodynamics in weak gravity.

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 gravitational Bohr radius of Challenge 98 ny

$$r_{\text{gr.B.}} = \frac{\hbar^2}{G m_e^2 m_p} = 1.1 \cdot 10^{29}\ \text{m} \qquad (30)$$

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 components. (Are you able to confirm this?) Challenge 99 e

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.

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?) Challenge 100 ny 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. Vol. II, page 179 Therefore, the existence of normal atoms does not require a quantum theory of gravity.

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. Ref. 145 Predictably, the calculated results do not reproduce experiments.

* *

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 Ref. 146 Planck length, the cosmic horizon R_0, and the number of baryons N_b:

$$(N_b)^3 \approx \left(\frac{R_0}{l_{Pl}}\right)^4 = \left(\frac{t_0}{t_{Pl}}\right)^4 \approx 10^{244} \qquad (31)$$

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

$$m_p^3 \approx \frac{\hbar^2}{Gct_0} . \qquad (33)$$

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 (31) or (33) express some long-sought relation between local and global topo-Ref. 148 logical 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.

* *

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 Challenge 101 s for this to happen?

In summary, 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.

Ref. 147 Vol. VI, page 104 * 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

$$m_0^2/R_0^2 \approx m_{Pl}^2/R_{Pl}^2 = c^4/G^2 . \qquad (32)$$

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.

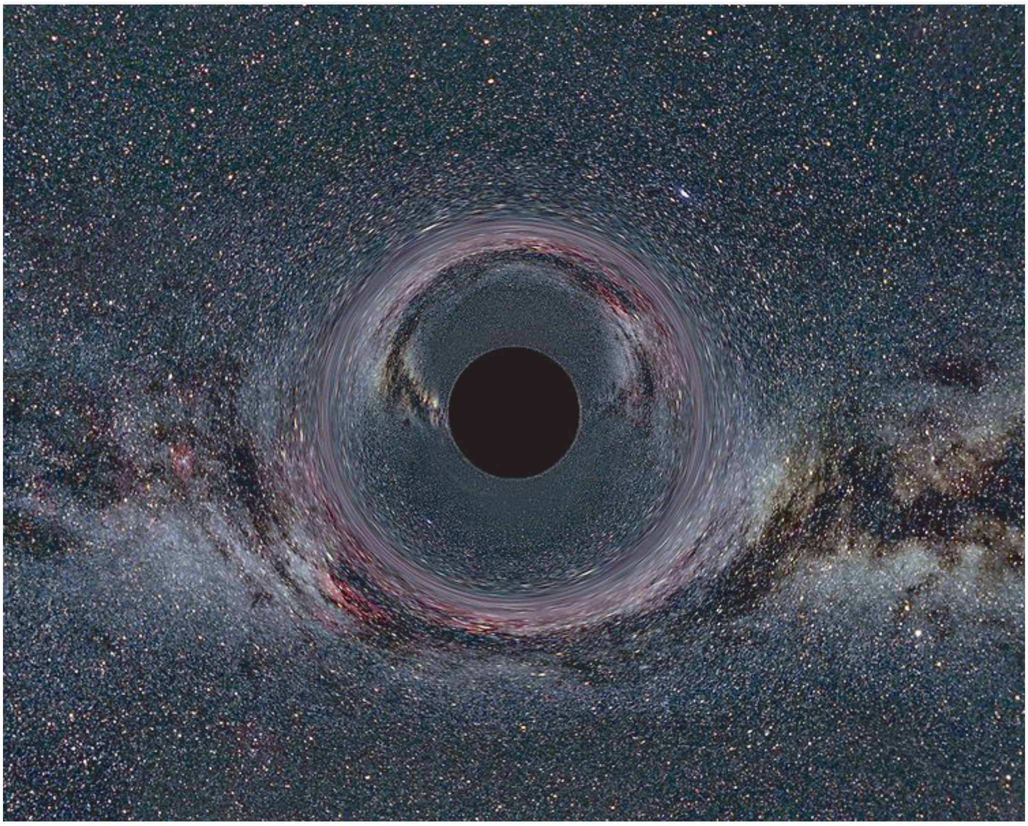


FIGURE 86 A simplified simulated image – not a photograph – of how a black hole of ten solar masses, with Schwarzschild radius of 30 km, seen from a constant distance of 600 km, will distort an image of the Milky Way in the background (image © Ute Kraus at www.tempolimit-lichtgeschwindigkeit.de).

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 also found that the Boltzmann constant k is the smallest entropy value found in nature. Vol. I, page 256

What is the influence of gravitation on entropy, and on 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. 122 Bekenstein was investigating the consequences gravity has for quantum physics. He

---

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

found that the entropy S of an object of energy E and size L is bound by

$$S \leqslant EL\frac{k\pi}{\hbar c} \qquad (34)$$

for all physical systems, where k is the Boltzmann constant. In particular, he deduced that (nonrotating) *black holes* saturate the bound. We recall that black holes are the densest systems for a given mass. They occur when matter collapses completely. Figure 86 shows an artist's impression. Vol. II, page 262

Challenge 102 s Bekenstein found that black holes have an entropy given by

$$S = A\,\frac{kc^3}{4G\hbar} = M^2\,\frac{4\pi kG}{\hbar c} \qquad (35)$$

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 thermodynamics. In fact, *black holes are the most disordered systems known.*[*]

As an interesting note, the maximum entropy also implies an upper memory limit for memory chips. Can you find out how? Challenge 103 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 somehow due to the diffeomorphism invariance at this boundary. 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. Ref. 124 Challenge 104 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.

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: 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*. All Ref. 125

---

[*] 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 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. Ref. 123

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

The radiation has a spectrum corresponding to the temperature

$$T = \frac{\hbar}{2\pi k c} a \,, \qquad (36)$$

where a is the magnitude 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 precise way. But if it will, it will be a beautiful experimental result. Ref. 126

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 (36). 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 65

Surprisingly, at the end of the twentieth century it became clear that the Fulling–Davies–Unruh effect possibly had already been observed *before* it was predicted! The Fulling–Davies–Unruh effect turned out be related to a well-established observation: the so-called *Sokolov–Ternov effect*. Ref. 127 In 1963, the Russian physicist Igor Ternov, together with Arsenji Sokolov, had used the Dirac equation to predict 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 Nuclear Physics in 1971, and then confirmed by experiments in Orsay, in Stanford and in Hamburg. Nowadays, the effect is used routinely 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! Ref. 127 The equivalence is somewhat surprising. 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 another case of a theoretical discovery that was made much later than necessary. In 2006 however, this equivalence was put into question again. The issue is not closed. Ref. 127

BLACK HOLES AREN'T BLACK

In 1973 and 1974, Jacob Bekenstein, and independently, Stephen Hawking, famous for the intensity 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

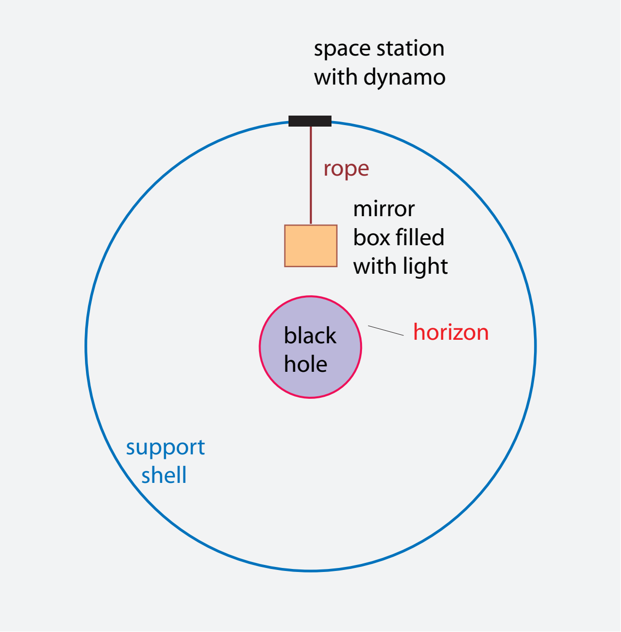


FIGURE 87 A thought experiment allowing you to deduce the existence of black hole radiation.

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 thought experiment was published by William Unruh and Robert Wald, showing that the whole result could have been deduced already 50 years earlier! Ref. 128

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 87. 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 generating energy. Obviously, when the box reaches the black hole horizon, the heat radiation is red-shifted to infinite wavelength. At that point, the full amount of energy originally contained in the heat radiation has been provided to the space station. We can now do the following: we can open the box on the horizon, let drop out whatever is still inside, and wind the empty and massless box back up again. As a result, we have completely converted heat radiation into mechanical energy. Nothing else has changed: the black hole has the same mass as beforehand.

But the lack of change 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 $g_{\text{surf}} = c^4/4GM$. The energy E gained by lowering the box is

$$E = c^2 m - m g_{\text{surf}} \frac{d}{2} = c^2 m \left(1 - \frac{dc^2}{8GM}\right) . \tag{37}$$

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

$$\eta = \frac{E}{c^2 m} = 1 - \frac{T_{\text{BH}}}{T} , \tag{38}$$

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

$$T_{\text{BH}} = \frac{\hbar c^3}{8\pi kGM} = \frac{\hbar c}{4\pi k}\frac{1}{R} = \frac{\hbar}{2\pi kc}\, g_{\text{surf}} \quad \text{with} \quad g_{\text{surf}} = \frac{c^4}{4GM} , \tag{39}$$

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 – much too weak to be detectable.

Challenge 105 ny
Ref. 129

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.*

TABLE 10 The principles of thermodynamics and those of horizon mechanics.

PRINCIPLE	THERMODYNAMICS	HORIZONS
Zeroth principle	the temperature T is the same across a body at equilibrium	the surface gravity a is the same across the horizon
First principle	energy is conserved: $\mathrm{d}E = T\mathrm{d}S - p\mathrm{d}V + \mu\mathrm{d}N$	energy is conserved: $\mathrm{d}(c^2 m) = \frac{ac^2}{8\pi G}\mathrm{d}A + \Omega\mathrm{d}J + \Phi\mathrm{d}q$
Second principle	entropy never decreases: $\mathrm{d}S \geqslant 0$	surface area never decreases: $\mathrm{d}A \geqslant 0$ (except for black hole radiation)
Third principle	$T = 0$ cannot be achieved	$a = 0$ cannot be achieved

All thermodynamic principles are valid for black holes, of course. A summary of the meaning of each thermodynamic principle in the case of black holes is given in Table 10.

Black hole radiation is thus so weak that we must speak of an academic effect! It leads Challenge 106 ny to a luminosity that increases with decreasing mass or size as

$$L \sim \frac{1}{M^2} \sim \frac{1}{R^2} \quad \text{or} \quad L = nA\sigma T^4 = \frac{n}{15 \cdot 2^{11}\pi}\frac{c^6\hbar}{G^2 M^2} \qquad (40)$$

Vol. III, page 239 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 – Ref. 130 the only case of practical importance – we have $n = 2$.

Black holes thus shine, and the more the smaller they are. For example, a solar-mass black holes emits less than 0.1 xW. This is a genuine *quantum* effect, since classically, black holes, as the name says, cannot emit any light at all. 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 treating isolated black holes, surrounded only by pure vacuum.

THE LIFETIME OF BLACK HOLES

Due to the emitted radiation, black holes gradually lose mass. Therefore their theoretical Challenge 107 ny lifetime is *finite*. A calculation shows that the lifetime is given by

$$t = M^3 \frac{20\,480\,\pi G^2}{\hbar c^4} \approx M^3\ 3.4 \cdot 10^{-16}\ \mathrm{s/kg^3} \qquad (41)$$

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, Ref. 131 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. 132

BLACK HOLES ARE ALL OVER THE PLACE

Around the year 2000, astronomers amassed a large body of evidence that showed something surprising: there 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. Also quasars, active galactic nuclei and gamma-ray bursters seem to be due to supermassive black holes at the centre of galaxies. The masses of these black holes are typically higher than a million solar masses.

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 a hypothetical cloud 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 disappear 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 the future. Vol. II, page 174

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

FASCINATING GAMMA-RAY BURSTS

Nuclear explosions produce flashes of γ rays, or gamma rays. In the 1960s, several countries thought that detecting γ 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 satellites into the sky to check for such flashes, they discovered something surprising. They observed about two γ 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

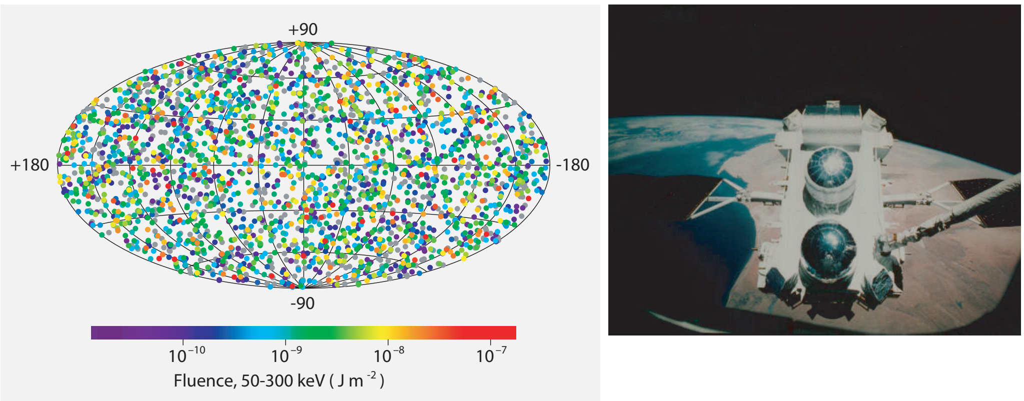


FIGURE 88 The location and energy of the 2704 γ ray bursts observed in the sky between 1991 and 2000 by the BATSE experiment on board of the Compton Gamma Ray Observatory, a large satellite deployed by the space shuttle after over 20 years of planning and construction. The Milky Way is located around the horizontal line running from +180 to −180 (NASA).

them in five minutes: the flashes, today called *gamma-ray bursts*, were coming from Ref. 134 outer 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 scientists, the Compton Gamma Ray Observatory, confirmed that the bursts were *extragalactic* in origin, as proven by the map of Figure 88.

Measurements of gamma-ray bursts 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, lasting many hours, sometimes even days. In 1997, afterglow was discovered also in the optical domain. The satellite also allowed researchers to find the corresponding X-ray, optical and radio sources for each burst. These measurements in turn allowed determining the distance of the burst sources; red-Ref. 135 shifts between 0.0085 and 4.5 were measured. In 1999 it finally became possible to detect the *optical* bursts corresponding to gamma-ray bursts.*

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 Ref. 137 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 Challenge 108 s universe taken together! Put differently, it is the same amount of energy that is released when converting several solar masses into radiation within a few seconds.

In fact, the measured luminosity of long bursts is near the theoretical maximum lu-Vol. II, page 108 minosity a body can have. This limit is given by

$$L < L_{\mathrm{Pl}} = \frac{c^5}{4G} = 0.9 \cdot 10^{52}\,\mathrm{W}\,, \tag{42}$$

---

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

Challenge 109 e as you might want to check yourself. In short, the sources of gamma ray bursts are the biggest bombs found in the universe. The are 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 Ref. 135 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.

Indeed, already 1975, a powerful radiation emission mechanism was predicted for dying Ref. 133 *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 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 Ref. 136 localized so far. All these exciting studies are still ongoing.

Understanding long gamma-ray bursts is one of the most fascinating open questions in astrophysics. The relation to black holes is generally accepted. But 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 Vol. II, page 271 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.

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

$$\rho = \frac{1}{M^2} \frac{3c^6}{32\pi G^3} \tag{43}$$

Challenge 110 e 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.

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

$$g_{\text{surf}} = \frac{1}{M} \frac{c^4}{4G} = \frac{c^2}{2R} \tag{44}$$

Challenge 111 e which is still $15\,\text{km/s}^2$ for an air density black hole.

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

$$\frac{1}{T} = \left.\frac{\partial S}{\partial E}\right|_\rho = \left.\frac{\partial S}{\partial (c^2 M)}\right|_\rho \tag{45}$$

All other thermal properties can be deduced by the standard relations from thermostatics.

In particular, black holes are the systems in nature with the largest possible entropy. Challenge 112 e Can you confirm this statement?

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 Challenge 113 ny 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 Challenge 114 s 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 Ref. 138 heat dE/dT is negative.

Black holes, like any object, oscillate when slightly perturbed. These vibrations have Ref. 139 also been studied; their frequency is proportional to the mass of the black hole.

Nonrotating black holes have no magnetic field, as was established already in the 1960s Ref. 130 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 Ref. 140 the horizon is described as a membrane, even though this model is not always applicable. In any case, we can study and describe isolated macroscopic black holes like any other macroscopic material body. The topic is not closed.

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 (41) 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., we can imagine the black hole being 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 115 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. 141 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 116 ny

$$S_m = \frac{1-m}{2n} + \sum_{k=n+1}^{mn} \frac{1}{k} \tag{46}$$

which is approximately given by

$$S_m = \ln m - \frac{m}{2n} \quad \text{for} \quad m \gg 1 \, . \tag{47}$$

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 (47) 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 (47) 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. Ref. 142 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.

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. Vol. II, page 275

* *

Apart from the questions of entropy, we can look for methods to cheat around conservation of energy, angular momentum, or charge. Challenge 117 ny But every thought experiment comes to the same conclusions. No cheats are possible. Every reasoning confirms that 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.

* *

A black hole transforms matter into antimatter with a certain efficiency. Indeed, a black hole formed by collapsing matter also radiates antimatter. Thus one might look for departures from particle number conservation. Are you able to find an example? Challenge 118 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. 143

* *

If lightweight boxes made of mirrors can float in radiation, one might deduce a strange consequence: such a box could 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 that this situation is impossible because of 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 119 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. 144

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 search was not successful. Indeed, modern searches take another direction, as will be explained in the last part of our adventure. But 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 of gravitation 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. 149

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

$$\alpha_{G1} = \frac{G}{\hbar c} = 2.2 \cdot 10^{-15}\,\mathrm{kg}^{-2} \quad \text{or by} \quad \alpha_{G2} = \frac{Gmm}{\hbar c} = \left(\frac{m}{m_{\mathrm{Pl}}}\right)^2 = \left(\frac{E}{E_{\mathrm{Pl}}}\right)^2 . \tag{48}$$

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 $\alpha_{\mathrm{em}} = 1/137.036$.

If all this is correct, *virtual* field quanta would also have to exist, to explain static gravitational fields. However, up to this day, the so-called *graviton* has not yet been detected, and there is in fact little hope that it ever will. On the experimental side, nobody knows yet how to build a graviton detector. Just try! Challenge 120 s 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

$$\Delta g \approx 2\frac{{l_{\mathrm{Pl}}}^2}{l^2} , \tag{49}$$

Challenge 121 ny where $l_{\mathrm{Pl}} = \sqrt{\hbar G/c^3}$ is the Planck length. Can you deduce the result? Quantum theory thus shows that like the momentum or the position of a particle, also the metric tensor g is a *fuzzy* observable.

But that is not all. Quantum theory is based on the principle that actions below $\hbar$ cannot be observed. This implies that the observable values for the metric g in a region of size L are bound by

$$g \geqslant \frac{2\hbar G}{c^3} \frac{1}{L^2} . \tag{50}$$

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 59

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. 150 The discussion is simplified for the simplest case of all, namely the superposition, in a vacuum region of size l, of a homogeneous gravitational field with value g and one with value g'. Vol. IV, page 148 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 calculation yields a decay time given by Challenge 122 ny

$$t_{\mathrm{d}} = \left(\frac{2kT}{\pi m}\right)^{3/2} \frac{n l^4}{(g-g')^2} , \tag{51}$$

where n is the particle number density, kT their kinetic energy and m their mass. Inserting typical numbers, we find that the variations in gravitational field strength are *extremely* small. Challenge 123 e 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. 151 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. 152 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. 153 In 1997 Pfenning and Ford showed that *warp drive* situations, which are also imagin-

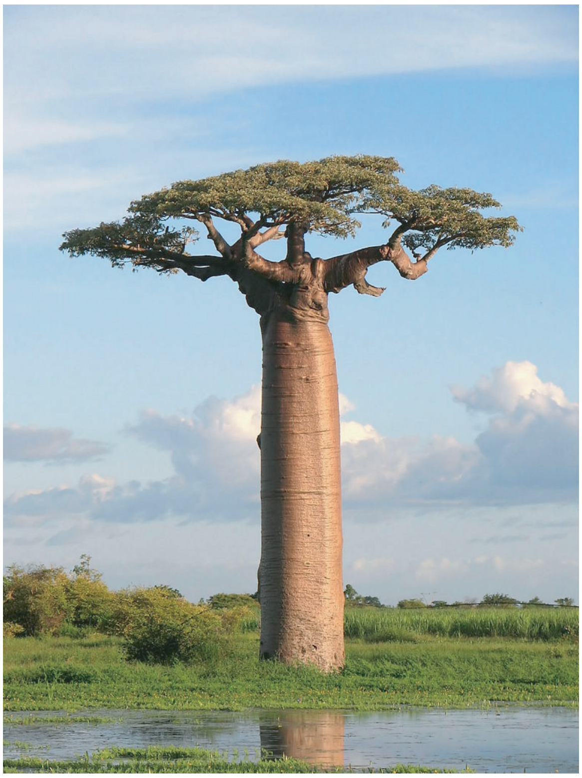


FIGURE 89 Every tree, such as this beautiful Madagascar baobab (*Adansonia grandidieri*), shows that nature, in contrast to physicists, is able to combine quantum theory and gravity (© Bernard Gagnon).

able 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 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.

Vol. IV, page 115

SUMMARY ON QUANTUM THEORY AND GRAVITY

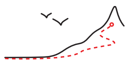
Every tree tells us: everyday, *weak* gravitational fields can be included in quantum theory. *Weak* gravitational fields have measurable and predictable effects on wave functions: quantum particles fall and their phases change in gravitational fields. Conversely, the inclusion of quantum effects into general relativity leads to space-time foam, space-time superpositions and gravitons. The inclusion of quantum effects into gravity prevents the existence of wormholes, time-like curves and negative energy regions.

The inclusion of *strong* gravitational fields into quantum theory works for practical situations but leads to problems with the particle concept. Conversely, the inclusion of quantum effects into situations with high space-time curvature leads to problems with the concept of space-time.

In summary, the combination of quantum theory and gravitation leads to problems with both the particle concept and the space-time concept. The combination of quantum theory and general relativity puts into question the foundations of the description of nature that we used so far. As shown in Figure 89, nature is smarter than we are.

In fact, up to now we hid a simple fact: conceptually, quantum theory and general relativity *contradict* each other. This contradiction was one of the reasons 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, but only *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 motion inside matter, and in particular the motion of and in nuclei.

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 open issues are left. But in its golden past, researchers produced new ways for medical doctors to dramatically improve the healing rate of patients. Ref. 154 Researchers also discovered why stars shine, how powerful bombs work, and how cosmic evolution produced the atoms we are made of. We will explore these topics now. A fascinating spin-off of nuclear physics, high energy particle physics, will keep us busy later on.

> “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 imaging technique allows one to image the interior of human bodies with a high resolution and with no damage or danger to the patient, in strong contrast to X-ray imaging. The technique is based on moving atomic nuclei. Though the machines are still expensive – costing up to several million euro – there is hope that they will become cheaper in the future. Such an MRI machine, shown in Figure 90, 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 91. The images allow detecting problems in bones, in the spine, in the beating heart and in general tissue. Many people owe their life and health to these machines; in many cases the machines allow making precise diagnoses and thus choosing the appropriate treatment for patients.

In MRI machines, a radio transmitter emits radio waves that are absorbed because hydrogen nuclei – protons – 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

$$E = \hbar\omega = \hbar\gamma B \tag{52}$$

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. The absorption of the radio wave is a pure quantum effect, as shown by the appearance of the quantum of action $\hbar$. Using some cleverly ap-

FIGURE 90 A commercial MRI machine (© Royal Philips Electronics).

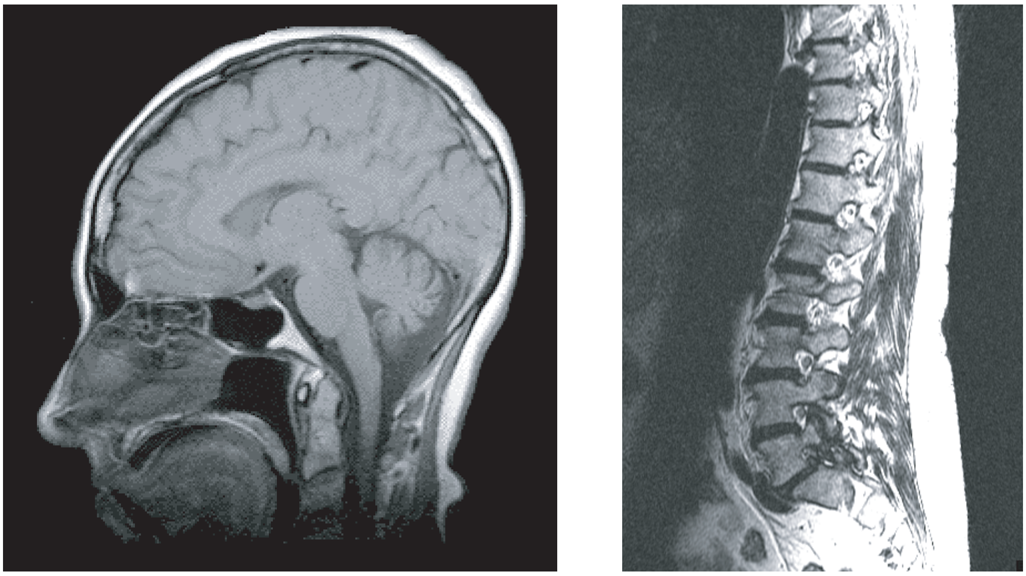


FIGURE 91 Sagittal images of the head and the spine (used with permission from Joseph P. Hornak, *The Basics of MRI*, www.cis.rit.edu/htbooks/mri, Copyright 2003).

plied magnetic fields, typically with a strength between 0.3 and 7 T for commercial and up to 21 T for experimental machines, the absorption for each volume element can be measured 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. When the intensity of the absorption is plotted as grey scale, an image is formed that retraces the different chemical compositions. Figure 91 shows two examples. Using additional tricks, modern machines can picture blood flow in the heart or air flow

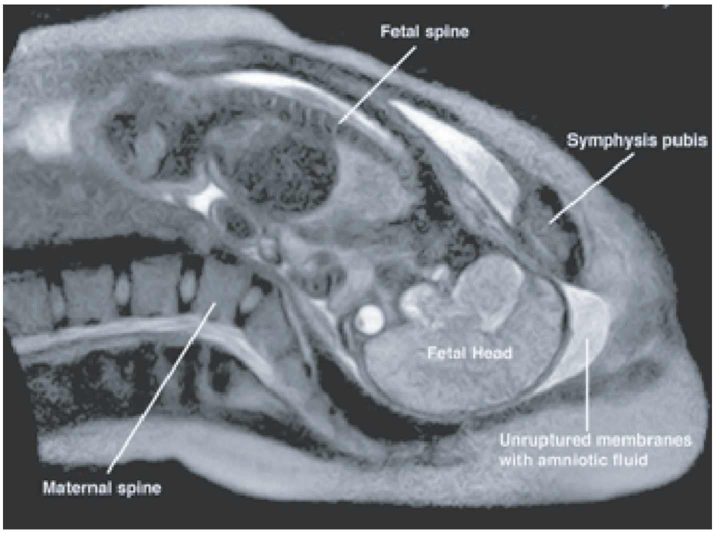


FIGURE 92 An image of the first magnetic resonance video of a human birth (© C. Bamberg).

Ref. 155 in 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.*

Magnetic resonance imaging can even image the great wonders of nature. The first video of a human birth, taken in 2010 and published in 2012, is shown in Figure 92. MRI scans are loud, but otherwise harmless for unborn children. The first scan of a married couple making love has been taken by Willibrord Weijmar Schultz and his group in 1999. It is shown on page 399.

Ref. 156

Ref. 157

Every magnetic resonance image thus proves that

▷ Many – but not all – atoms have nuclei that spin.

Like any other object, nuclei have size, shape, colour, composition and interactions. Let us explore them.

THE SIZE OF NUCLEI AND THE DISCOVERY OF RADIOACTIVITY

The magnetic resonance signal shows that hydrogen nuclei spin with high speed. Thus they must be small. Indeed, the g-factor of protons, defined using the magnetic moment μ, their mass m and charge e, is found to be

$$g = \mu 4 \frac{m}{e\hbar} \approx 5.6 \,. \qquad (53)$$

Vol. IV, page 107 This is a small value. Using the expression that relates the g-factor and the radius of a composite object, we deduce that the radius of the proton is about 0.9 fm; this value

---

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

FIGURE 93 Henri Becquerel (1852–1908)

FIGURE 94 Marie Curie (1867–1934)

is confirmed by many experiments and other measurement methods. Protons are thus about 30 000 times smaller than hydrogen atoms, whose radius is about 30 pm. The proton is the smallest of all nuclei; the largest known nuclei have radii about 7 times as large.

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 Röntgen 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, Gustave Bémont named the property of these minerals *radioactivity*.

Radioactive rays are also emitted from many elements other than uranium. This radiation can be 'seen': it can be detected by the tiny flashes of light that appear when the rays hit a scintillation screen. The light flashes are tiny even at a distance of several metres from the source; thus the rays must be emitted from point-like sources. In short, radio-

---

* Henri Becquerel (b. 1852 Paris, d. 1908 Le Croisic), important physicist; his primary research 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 Becquerel. For his discovery of radioactivity he received the 1903 Nobel Prize in Physics; he shared it with the Curies.

TABLE 11 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

activity has to be emitted from single atoms. Thus radioactivity confirmed unambiguously that atoms do exist. In fact, radioactivity even allows *counting* atoms: in a diluted radioactive substance, the flashes can be counted, either with the help of a photographic film or with a photon counting system. Vol. IV, page 42

The intensity of radioactivity cannot be influenced by magnetic or electric fields; and 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 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. The types of radiation are listed in Table 11. In the meantime, all

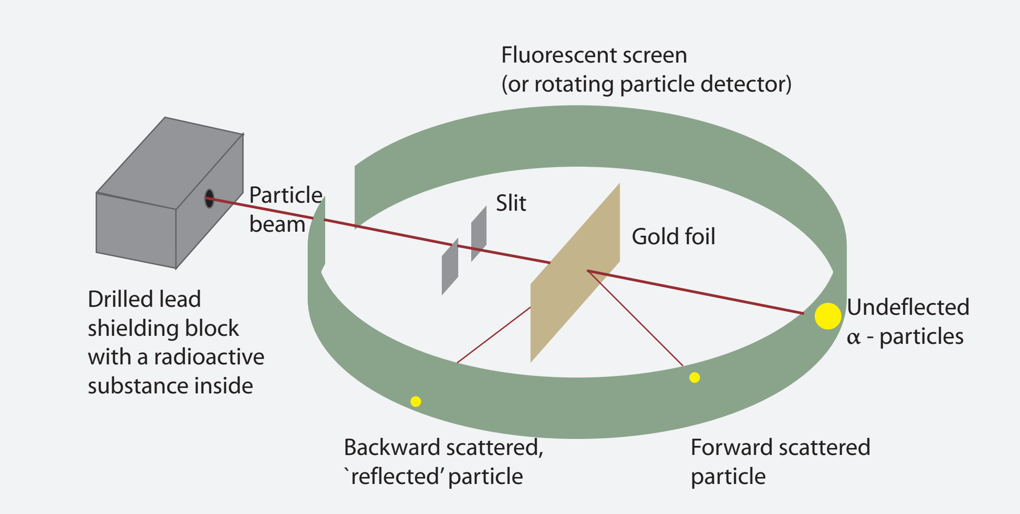


FIGURE 95 The schematics of the Rutherford–Geiger scattering experiment. The gold foil is about a square centimetre in size.

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 smart experiments. He asked his collaborator Hans Geiger to take an emitter of α radiation – a type of radioactivity 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 α rays would end up. The experiment is shown in Figure 95. 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. Both cloud chambers and bubble chambers thus allow *seeing* radioactivity, as shown in the examples of Figure 96.

The radiation detectors around the thin gold foil give a consistent, but strange result: most α particles pass through the 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 to show this?) Rutherford and Geiger deduced from their scattering experiment that first of all, the atoms in the metal foil are mainly transparent. Only transparency of atoms explains why most α particles pass the

Challenge 124 s

---

* Ernest Rutherford (b. 1871 Brightwater, d. 1937 Cambridge), important physicist. He emigrated to Britain and became professor at the University of Manchester. He coined the terms α particle, β particle, proton and neutron. A gifted experimentalist, he discovered that radioactivity transmutes the elements, explained the nature of α rays, discovered the nucleus, measured its size and performed the first nuclear reactions. Ironically, in 1908 he received the Nobel Prize in 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.

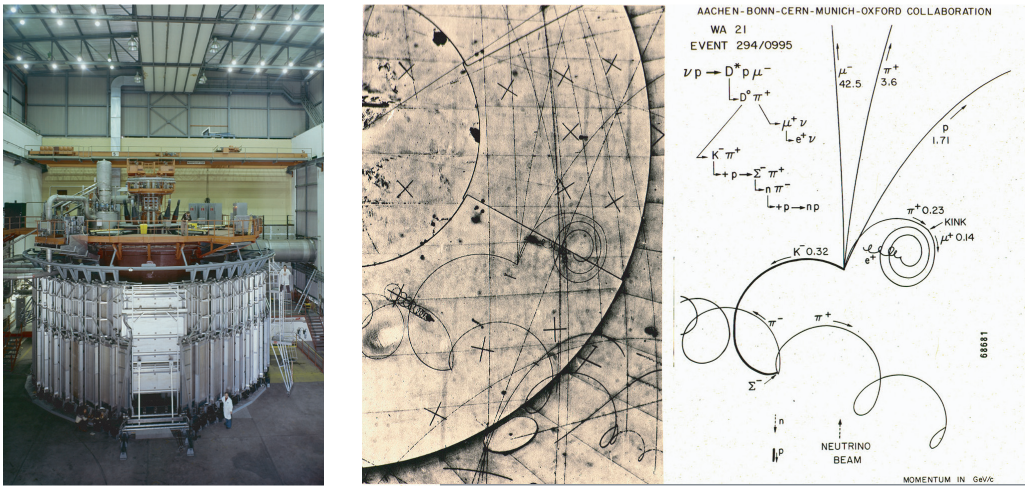
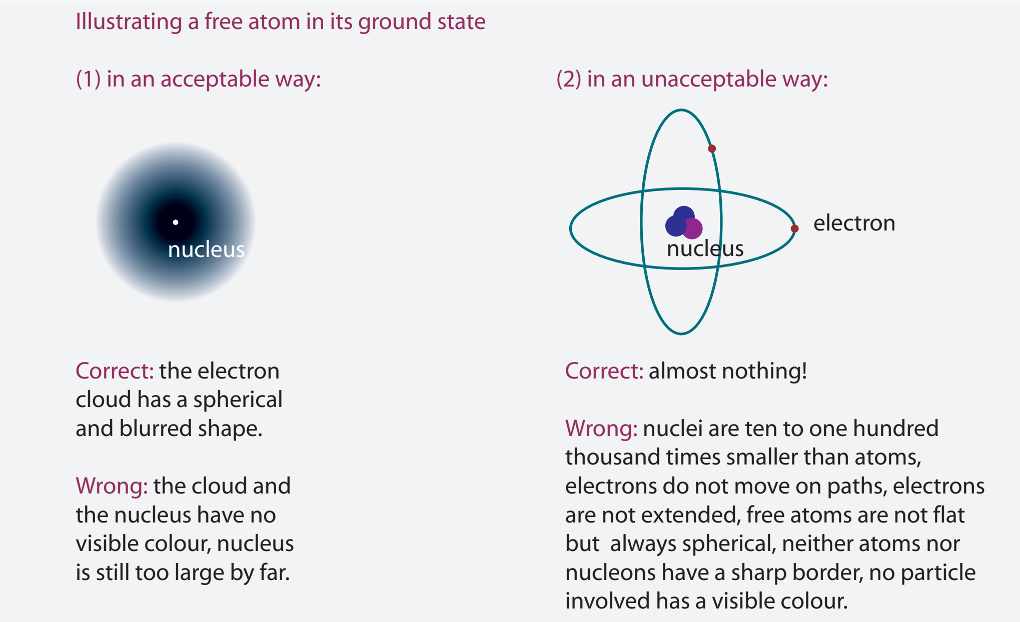


FIGURE 96 The ‘Big European Bubble Chamber’ – the biggest bubble chamber ever built – and an example of tracks of relativistic particles it produced, with the momentum values deduced from the photograph (© CERN).



Correct: the electron cloud has a spherical and blurred shape.

Wrong: the cloud and the nucleus have no visible colour, nucleus is still too large by far.

Correct: almost nothing!

Wrong: nuclei are ten to one hundred thousand times smaller than atoms, electrons do not move on paths, electrons are not extended, free atoms are not flat but always spherical, neither atoms nor nucleons have a sharp border, no particle involved has a visible colour.

FIGURE 97 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.

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*

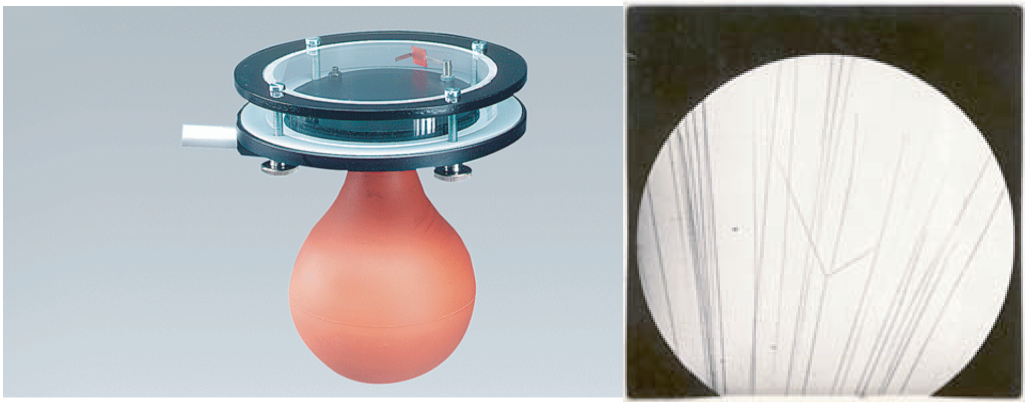


FIGURE 98 Left: a modern Wilson cloud chamber, diameter c. 100 mm. Right: one of the first pictures of α rays taken with a cloud chamber in the 1920s by Patrick Blackett, showing also a collision with an atom in the chamber (© Wiemann Lehrmittel, Royal Society)

of the reflecting entity and to estimate its *mass*. (This calculation is now a standard exercise in the study of physics at universities.) He found that the reflecting entity contains practically all of the mass of the atom in a diameter of a few fm. Rutherford named this concentrated mass 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. Thus

▷ Atoms resemble 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. In solids, the candy floss, i.e., the electron cloud, keeps the neighbouring nuclei at constant distance. Figure 97 shows the more and less correct ways to picture an atom. Candy floss explains the Rutherford–Geiger experiment: for the tiny and massive α particles however, the candy floss is essentially empty space, so that they 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 $5.5 \cdot 10^{17}$ 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 oil tanker. (Can you confirm this?) Challenge 125 e

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

NUCLEI ARE COMPOSED

Magnetic resonance images also show that nuclei are *composed*. Indeed, images can also be taken using heavier nuclei instead of hydrogen, such as certain fluorine or oxygen

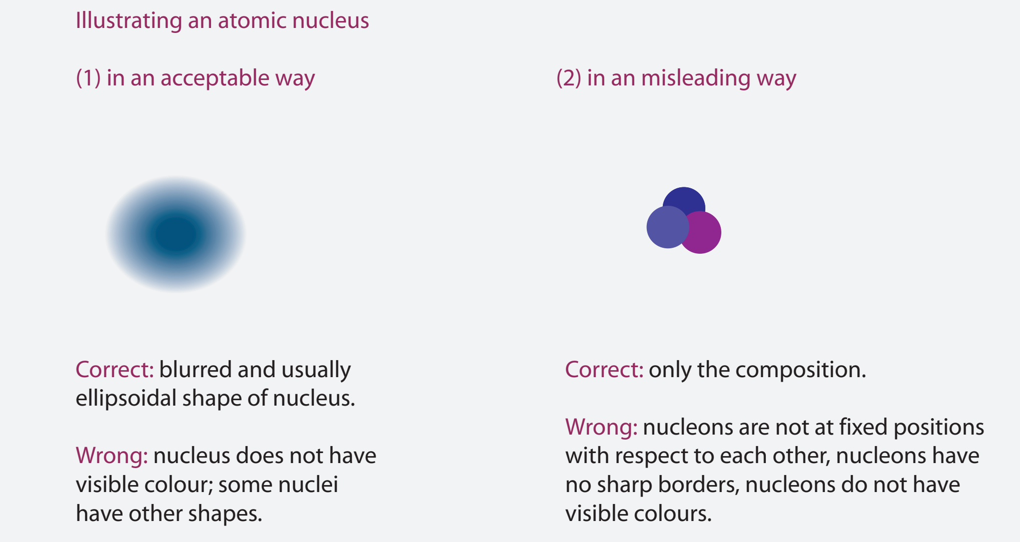


FIGURE 99 A reasonably realistic (left) and a misleading illustration of a nucleus (right) as is regularly found in school books. *Nuclei are spherical nucleon clouds.*

nuclei. Also the g-factors of these nuclei depart from the value 2 characteristic of point particles; the more massive the nuclei are, the bigger the departure. Therefore, all nuclei have a finite size. The size of nuclei can actually be measured; the Rutherford–Geiger experiment and many other scattering experiments allow to do so. The measured values confirm the values predicted by the g-factor. In short, both the values of the g-factor and the non-vanishing sizes show that *nuclei are composed.*

Vol. IV, page 107

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 discovered, 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 about 1836 times as heavy. More details on the proton are listed in Table 12.

However, the charge multiples and the mass multiples for the heavier 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 confirmed an idea formulated by Werner Heisenberg: all nuclei heavier than hydrogen nuclei are made of positively charged *protons* and of neutral *neutrons*. Neutrons are particles a tiny bit more massive than protons (the difference is less than a part in 700, as shown in Table 12), 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

TABLE 12 The properties of the nucleons: proton and neutron (source: pdg.web.cern.ch).

PROPERTY	PROTON	NEUTRON
Mass	$1.672\,621\,777(74) \cdot 10^{-27}$ kg 0.150 327 7484(66) nJ 938, 272 046(21) MeV 1.007 276 466 812(90) u 1836.152 6675(39)$\cdot\, m_e$	$1.674\,927\,351(74) \cdot 10^{-27}$ kg 0.150 534 9631(66) nJ 939, 565 379(21) MeV 1.008 664 916 00(43) u 1838.683 6605(11)$\cdot\, m_e$
Spin	1/2	1/2
P parity	+1	+1
Antiparticle	antiproton $\overline{\mathrm{p}}$	antineutron $\overline{\mathrm{n}}$
Quark content	uud	udd
Electric charge	1 e	0
Charge radius	0.88(1) fm	0.12(1) fm^2
Electric dipole moment	$< 5.4 \cdot 10^{-26}$ 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	$1.410\,606\,743(33) \cdot 10^{-26}$ J/T	$-0.966\,236\,47(23) \cdot 10^{-26}$ J/T
g-factor	5.585 694 713(46) $2.792\,847\,356\,(23) \cdot \mu_N$	−3.826 085 45(90) $-1.913\,042\,72(45) \cdot \mu_N$
Gyromagnetic ratio	0.267 522 2005(63) 1/nsT	
Magnetic polarizability	$1.9(5) \cdot 10^{-4}\,\mathrm{fm}^3$	$3.7(2.0) \cdot 10^{-4}\,\mathrm{fm}^3$
Mean life (free particle)	$> 2.1 \cdot 10^{29}$ a	880.1(1.1) s
Shape (quadrupole moment)	oblate	oblate
Excited states	more than ten	more than ten

neutral, neutrons do not leave tracks in clouds chambers and are more difficult to detect than protons, charged hadrons or charged leptons. For this reason, the neutron was discovered later than several other, more exotic 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 885.7(8) s. Nevertheless, we all know that inside most nuclei we are made of, neutrons do not decay for millions of years, because the decay products do not lead to a state of lower energy. (Why not?)

Challenge 126 s

Magnetic resonance images also show that most elements have different types of atoms. These elements have atoms with the same number of protons, but with different numbers of neutrons. One says that these elements have several *isotopes*.[*] This result also explains why some elements radiate with a mixture of different decay times. Though chemically isotopes are (almost) indistinguishable, they can differ strongly in their nuclear properties. Some elements, such as tin, caesium, or polonium, have over thirty isotopes each. Together, the 118 known elements have over 2000 isotopes. They are shown in Figure 100. (Isotopes without electrons, i.e., specific nuclei with a given number of neutrons and protons, are called *nuclides*.)

Because nuclei are so extremely dense despite containing 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 α rays. The additional force is called the *strong nuclear interaction*. We shall study it in detail shortly.

Page 219

The strong nuclear interaction binds protons and neutrons in the nucleus. It is essential to recall that inside a nucleus, the protons and neutrons – they are often collectively called *nucleons* – move in a similar way to the electrons moving in atoms. Figure 99 illustrates this. The motion of protons and neutrons inside nuclei allows us to understand the shape, the spin and the magnetic moment of nuclei.

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 the α particles emitted by many radioactive substance are helium nuclei. Like all nuclei, α 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). 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

Vol. III, page 23

---

[*] 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.

[**] Viktor Franz Heß, (b. 1883 Waldstein, d. 1964 Mount Vernon), nuclear physicist, received the Nobel Prize in 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.

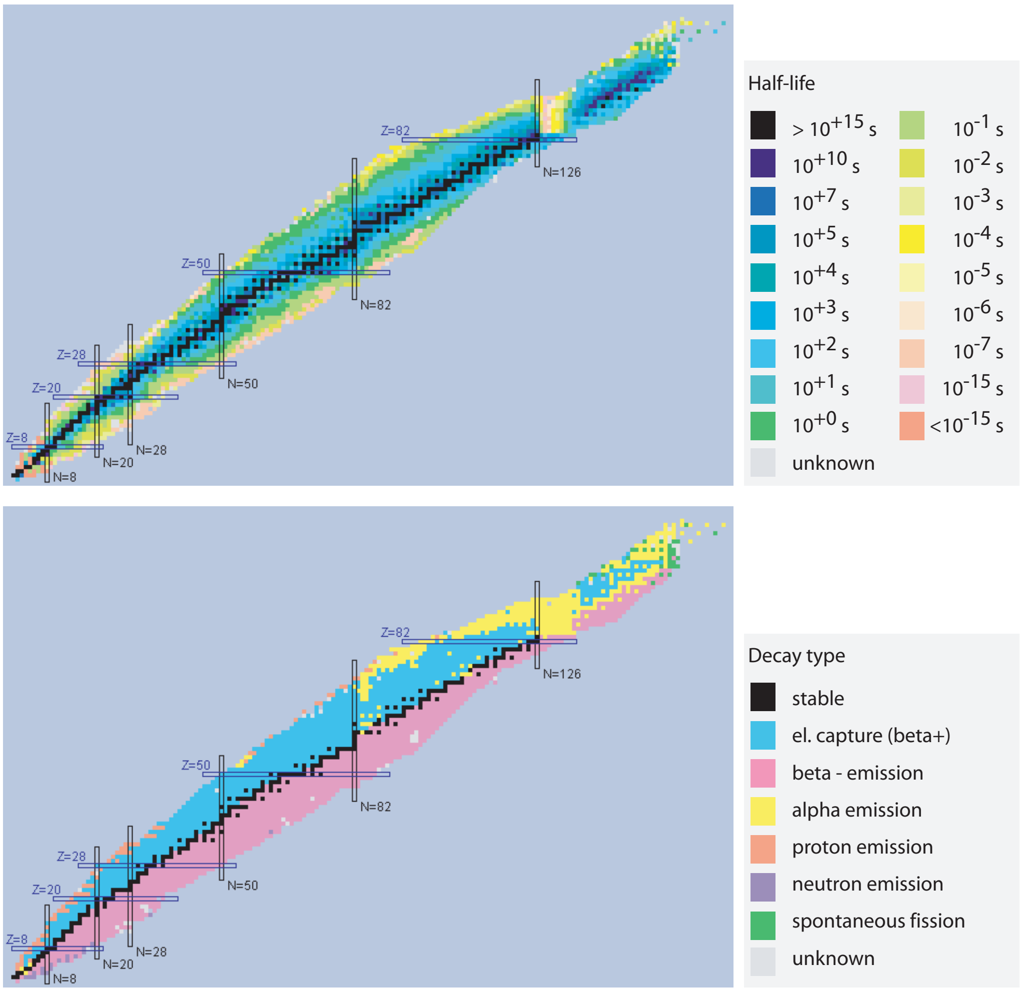


FIGURE 100 All known nuclides with their lifetimes (above) and main decay modes (below). The data are from www.nndc.bnl.gov/nudat2.

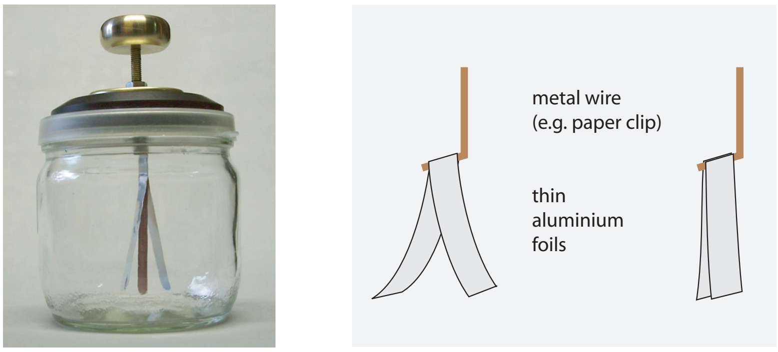


FIGURE 101 An electroscope (or electrometer) (© Harald Chmela) and its charged (middle) and uncharged state (right).

charged, the strips repel each other and move apart, as shown in Figure 101. (You can

FIGURE 102 Viktor Heß (1883–1964)

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 127 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 was several times faster 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*. One also speaks of *cosmic radiation*. During the last few centuries, many people have drunk from a glass and eaten chocolate covered by aluminium foil; but only Heß combined these activities with such careful observation and deduction that he earned a Nobel Prize.*

Today, the most common 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 by watching the spark chamber 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. In short, with a Geiger counter one cannot see ions or particles, but one can hear them. Later on, with the advances in electronics, ionized

---

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

TABLE 13 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
α 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	counters, films	1 mm of any material
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/ $m^2 s$	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., $n + {}^{16}\mathrm{O} \rightarrow \mathrm{p} + {}^{14}\mathrm{C}$	counters	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.				

atoms or particles could be counted.

Finding the right gas mixture for a Geiger–Müller counter 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 percentage of alcohol to the gas in the chamber. Nasty rumours maintained that this was discovered when another assistant tried,

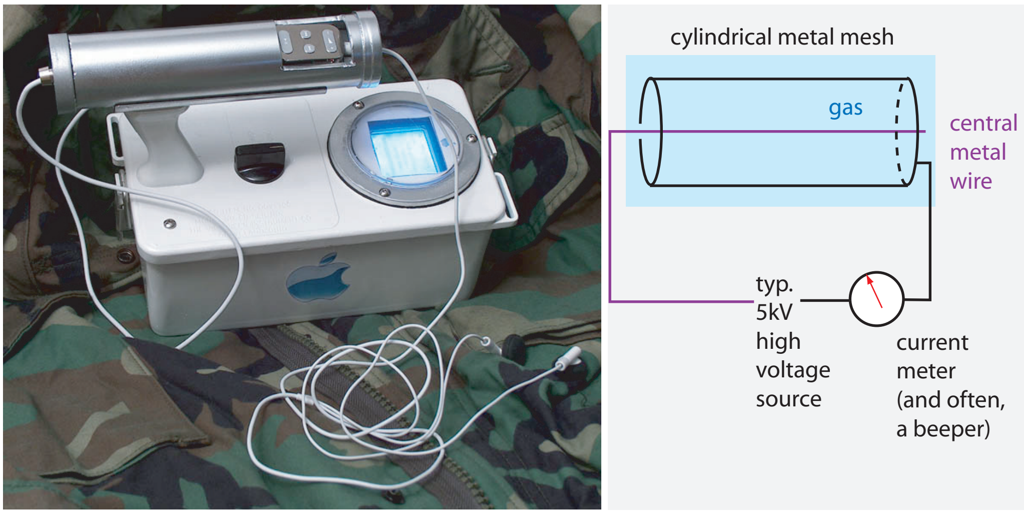


FIGURE 103 A Geiger–Müller counter with the detachable detection tube, the connection cable to the counter electronics, and, for this model, the built-in music player (© Joseph Reinhardt).

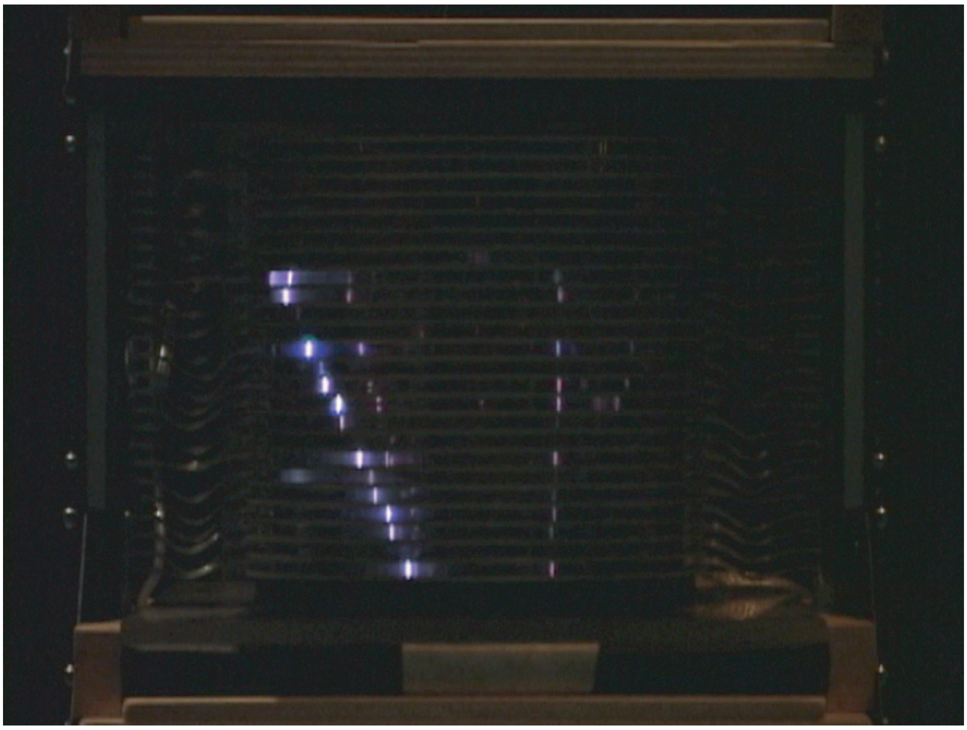


FIGURE 104 A modern spark chamber showing the cosmic rays that constantly arrive on Earth (QuickTime film © Wolfgang Rueckner).

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

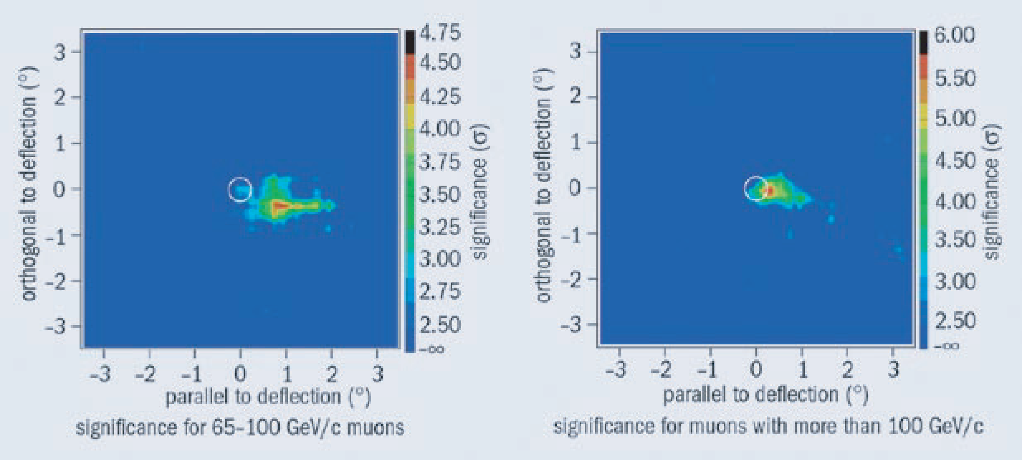


FIGURE 105 The cosmic ray moon shadow, observed with the L3 detector at CERN. The shadow is shifted with respect of the position of the moon, indicated by a white circle, because the Earth's magnetic field deflects the charged particles making up cosmic rays (© CERN Courier).

smallest versions fit in mobile phones and inside wrist watches. An example is shown in Figure 103.

If you can ever watch a working spark chamber, do so. The one in the CERN entrance hall is about 0.5 m^3 in size. A few times per minute, you can see the pink sparks showing the traces of cosmic rays. The rays appear in groups, called *showers*. And they hit us all the time.

Various particle detectors also allow measuring the energy of particles. The particle energy in cosmic rays spans a range between 10^3 eV and at least 10^{20} eV; the latter is the same energy as a tennis ball after serve, but for a single ion. This is a huge range in energy. Understanding the origin of cosmic rays is a research field on its own. Some cosmic rays are galactic in origin, some are extragalactic. For most energies, supernova remnants – pulsars and the like – seem the best candidates. However, the source of the highest energy particles is still unknown; black holes might be involved in their formation.

Cosmic rays are probably the only type of radiation discovered without the help of shadows. But in the meantime, such shadows have been found. In a beautiful experiment performed in 1994, the shadow thrown by the Moon on high energy cosmic rays (about 10 TeV) was measured, as shown in Figure 105. When the position of the shadow is compared with the actual position of the Moon, a shift is found. And 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 Ref. 158 of antiprotons in cosmic rays between 0 % and 30 %. 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 Page 175 a mixture of many types of particles, as shown in Table 13. They arrive from outside the atmosphere as a mixture of which the largest fraction are protons, followed by α particles, iron and other nuclei. And, as mentioned above, most rays do not originate from the Sun. In other words, nuclei can thus travel alone over large distances. In fact,

FIGURE 106 An aurora borealis, produced by charged particles in the night sky (© Jan Curtis).

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 at the origin of cosmic rays are still a topic of research.

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*. In fact, the *neutrino* flux is many orders of magnitude higher than the flux of charged rays, but does not have any effect on human bodies.

Page 256

Vol. III, page 218

Cosmic rays have several effects on everyday life. Through the charges they produce in the atmosphere, they are probably responsible for the start and 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, also start the Kelvin generator.

Vol. III, page 19

If the magnetic field of the Earth would not exist, we might get sick from cosmic rays. The magnetic field diverts most rays towards the magnetic poles. Also both the upper and lower atmosphere help 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.

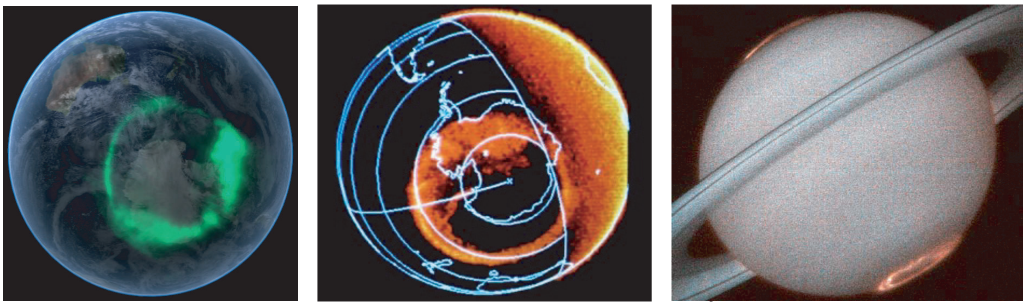


FIGURE 107 Two aurorae australes on Earth, seen from space (a composed image with superimposed UV intensity, and a view in the X-ray domain) and a double aurora on Saturn (all NASA).

Cosmic rays also produce beautifully coloured flashes inside the eyes of cosmonauts; they regularly enjoy these events in their trips. (And they all develop cataracts as a consequence.) But cosmic rays are not only dangerous and beautiful. They are also useful. If cosmic rays would not exist, 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. (See Figure 106 or 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. Aurorae are also seen in the rest of the solar system. Aurorae due to core magnetic fields have been observed on Jupiter, Saturn, Uranus, Neptune, Earth, Io and Ganymede. For an example, see Figure 107. Aurorae due to other mechanisms have been seen on Venus and Mars.

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 α 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. Vol. III, page 226

NUCLEI DECAY – MORE ON RADIOACTIVITY

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

$$N(t) = N(0)\,\mathrm{e}^{-t/\tau}\ . \qquad (54)$$

The parameter τ, the so-called *life time* or *decay time*, depends on the type of nucleus emitting the rays. Life times can vary from far 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. Obviously, formula (54) 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. Though in principle, the quantum Zeno effect could appear for small times t, for the case of radioactivity it has not yet been observed. Page 48

Instead of the life-time, often the half-life is used. The *half-life* is the time during which radioactivity decreases to *half* the starting value. Can you deduce how the two times are related? Challenge 128 s

Radioactivity is the decay of unstable nuclei. Most of all, radioactivity allows us 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 (54) you can then determine how many atoms have been in the mass 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 exponential decay (54) and the release of energy is typical of *metastable* systems. In 1903, Rutherford and Soddy discovered what the state of lower energy is for α and β emitters. In these cases, radioactivity changes the emitting atom; it is a spontaneous transmutation of the atom. An atom emitting α or β rays changes its chemical nature. Radioactivity thus implies, for the case of nuclei, the same result that statistical mechanics of gases implies for the case of atoms: they are quantum particles with a structure that can change over time. Vol. I, page 406

- In *α decay* – or *alpha decay* – the radiating nucleus emits a (doubly charged) helium nucleus, also called an *α particle*. The kinetic energy is typically a handful of MeV. After the emission, the nucleus has changed to a nucleus situated two places earlier in the periodic system of the elements. α decay occurs mainly for nuclei that are rich in protons. An example of α decay is the decay of the ^{238}U isotope of uranium.
- In *β decay* – or *beta decay* – a neutron transforms itself into a proton, emitting an electron – also called a *β particle* – and an antineutrino. Also β decay changes the chemical nature of the atom, but to the place following the original atom in the periodic table of the elements. Example of β emitters are radiocarbon, ^{14}C, ^{38}Cl, and ^{137}Cs, the isotope expelled by damaged nuclear reactors. We will explore β decay below. A variant is the β^+ decay, in which a proton changes into a neutron and emits a neutrino and Page 240

a positron. It occurs in proton-rich nuclei. An example is ^{22}Na. Another variant is electron capture; a nucleus sometimes captures an orbital electron, a proton is transformed into a neutron and a neutrino is emitted. This happens in ^{7}Be. Also bound β decay, as seen in ^{187}Re, is a variant of β decay.
- In *γ decay* – or *gamma decay* – the nucleus changes from an excited to a lower energy state by emitting a high energy photon, or *γ particle*. In this case, the chemical nature is not changed. Typical energies are in the MeV range. Due to the high energy, γ rays ionize the material they encounter; since they are not charged, they are not well absorbed by matter and penetrate deep into materials. γ radiation is thus by far the most dangerous type of (environmental) radioactivity. An example of γ decay is ^{99m}Tc. A variant of γ decay is *isomeric transition*. Still another variant is *internal conversion*, observed, for example, in ^{137m}Ba.
- In *neutron emission* the nucleus emits a neutron. The decay is rare on Earth, but occurs in the stellar explosions. Most neutron emitters have half-lives below a few seconds. Examples of neutron emitters are ^{5}He and ^{17}N.
- The process of *spontaneous fission* was discovered in 1940. The decay products vary, even for the same starting nucleus. But ^{239}Pu and ^{235}U can decay through spontaneous fission, though with a small probability.
- In *proton emission* the nucleus emits a proton. This decay is comparatively rare, and occurs only for about a hundred nuclides, for example for ^{53m}Co and ^{4}Li. The first example was discovered only in 1970. Around 2000, the simultaneous emission of two protons was also observed for the first time.
- In 1984, *cluster emission* or *heavy ion emission* was discovered. A small fraction of ^{223}Ra nuclei decay by emitting a ^{14}C nucleus. This decay occurs for half a dozen nuclides. Emission of ^{18}O has also been observed.

Ref. 159

Many combined and mixed decays also exist. These decays are studied by nuclear physicists. Radioactivity is a common process. As an example, 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). Why is this not dangerous?

Challenge 129 s

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

Challenge 130 e

Vol. IV, page 113

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. 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 30 years to discover that radioactive decays, like all decays, are quantum effects. All decays are triggered by the statistical fluctuations of the vacuum, more precisely, by the quantum fluctuations of the vacuum. 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 α decay with the tunnelling effect. He found that the

FIGURE 108 A modern accelerator mass spectrometer for radiocarbon dating, at the Hungarian Academy of Sciences (© HAS).

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 in different nuclei.

Radiometric dating

As a result of the chemical effects of radioactivity, the composition ratio of certain elements in minerals allows us 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. Where there surprises? No. The times found with radiometric dating are consistent with the relative time scale that geologists had defined independently for centuries before the technique appeared. Radiometric dating confirmed what had been deduced before.

Ref. 161 Radiometric dating is a science of its own. An overview of the isotopes used, together Page 183 with their specific applications in dating of specimen, is given in Table 14. The table shows how the technique of radiometric dating has deeply impacted astronomy, geology, evolutionary biology, archaeology and history. (And it has reduced the number of violent Ref. 160 believers.) Radioactive life times can usually be measured to within one or two per cent

TABLE 14 The main 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}Re	^{187}Os	42 Ga	rhenium–osmium method	rocks, lunar soil, meteorites
^{176}Lu	^{176}Hf	37 Ga	lutetium–hafnium method	rocks, lunar soil, meteorites
^{40}K	^{40}Ar	1.25 Ga	potassium–argon & argon–argon method	rocks, lunar soil, meteorites
^{40}K	^{40}Ca	1.25 Ga	potassium–calcium method	granite dating, not precise
^{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
^{235}U	^{207}Pb	0.7 Ga	uranium–lead method, lead–lead method	rocks, lunar soil, meteorites
^{234}U	^{230}Th	248 ka	uranium–thorium method	corals, stalagmites, bones, teeth
^{230}Th	^{226}Ra	75.4 ka	thorium-radon method	plant dating
^{26}Al	^{26}Mg	0.72 Ma	supernova debris dating, cosmogenic	checking that nucleosynthesis still takes place in the galaxy
^{10}Be	^{10}B	1.52 Ma	cosmogenic radiometric dating	ice cores
^{60}Fe	^{60}Ni	2.6 Ma (not 1.5 Ma)	supernova debris dating	deep sea crust
^{36}Cl	^{36}Ar	0.3 Ma	cosmogenic radiometric dating	ice cores
^{53}Mn	^{53}Cr	3.7 Ma	cosmogenic radiometric dating	meteorites, K/T boundary
^{182}Hf	^{182}W	9 Ma	cosmogenic radiometric dating	meteorites, sediments
^{14}C	^{14}N	5730 a	radiocarbon method, cosmogenic	wood, clothing, bones, organic material, wine
^{137}Cs	^{137}Ba	30 a	γ-ray counting	dating food and wine after nuclear accidents
^{210}Pb		22 a	γ-ray counting	dating wine
^{3}H	^{3}He	12.3 a	γ-ray counting	dating wine

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 to measure half-lives of thousands of millions of years to high precision? Challenge 131 s

Radiometric dating was even more successful in the field of ancient history. With the *radiocarbon dating method* historians determined the age of civilizations and the age of human artefacts.* Many false beliefs were shattered. In some belief communities the shock is still not over, even though over hundred years have passed since these results became known. Ref. 160

Radiocarbon dating uses the β decay of the radioactive carbon isotope ^{14}C, which 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 life time value of a few thousand years is particularly useful to date historic material. Therefore, 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. Such small amounts can be detached from most specimen without big damage. Accelerator mass spectroscopy showed that numerous religious relics are forgeries, such as a cloth in Turin, and that, in addition, several of their wardens are crooks.

Researchers have even developed an additional method to date stones that uses radioactivity. Whenever an α ray is emitted, the emitting atom gets a recoil. If the atom is part of a crystal, the crystal is damaged by the recoil. In many materials, 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 the liquid material from volcanic eruptions has become rock.

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 was beyond all previous estimates and expectations: the oldest rocks and the oldest meteorites, studied independently using different dating methods, are 4570(10) million years old. From this data, the age of the Earth is estimated to be 4540(50) million years. The Earth is indeed *old*. Ref. 162

But if the Earth is so old, why did it 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 imagery that many cultures ascribe to hell: fire and suffering. Because of the high temperature of lava, hell is inevitably depicted as a hot place located at the centre of the Earth. A striking

---

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

FIGURE 109 The lava sea in the volcano Erta Ale in Ethiopia (© Marco Fulle).

example is the volcano Erta Ale, shown in Figure 109. But why is lava still hot, after so many 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 magma nor eject any lava; hell would not be hot. Challenge 132 ny

The solution to the riddle is provided by radioactivity: the centre of the Earth contains an oven that is fuelled with an estimated 8 to 10 TW by radioactive uranium ^{235}U and ^{238}U, with 8 to 10 TW by radioactive thorium ^{232}Th and with around 4 TW by radioactive potassium ^{40}K. Ref. 163 The radioactivity of these elements, and a few others to a minor degree, keeps the centre of the Earth glowing. More precise investigations, taking into account the decay times and measured material concentrations, Page 183 show that this mechanism indeed explains the internal heat of the Earth. (In addition, the decay of radioactive potassium is the origin for the 1 % of argon found in the Earth's atmosphere.)

In short, radioactivity keeps lava hot. Radioactivity is the reason that we depict hell as hot. This brings up a challenge: why is the radioactivity of lava and of the Earth in general not dangerous to humans? Challenge 133 s

NUCLEI CAN FORM COMPOSITES

Nuclei are highly unstable when they contain more than about 280 nucleons. Nuclei with higher number of nucleons inevitably decay into smaller fragments. In short, heavy nuclei are unstable. But when the mass is above 10^{57} nucleons, they are stable again: such systems are 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 γ rays, in the range between 1 keV (or 0.2 fJ) and more than 20 MeV (or 3.3 pJ). The emission of γ rays by nuclei 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.

In short, the energies of the emitted and absorbed γ ray photons define the 'colour' of the nucleus. The γ ray spectrum can be used, like all colours, to distinguish nuclei from each other and to study their motion. In particular, the spectrum 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 nuclear motion inside solids or to detect the gravitational Doppler effect.

The study of γ-rays also allows us 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 nucleons per $\mathrm{fm^3}$, and a skin thickness of about 2.4 fm, where their density decreases. Nuclei are thus small clouds, as illustrated in Figure 110.

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. 164

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

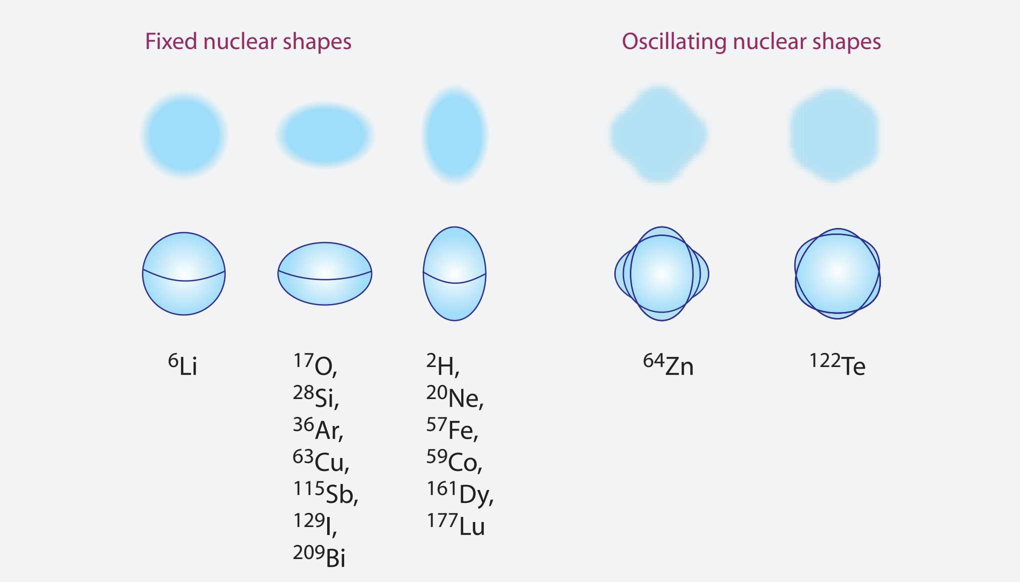


FIGURE 110 Various nuclear shapes – fixed: spherical, oblate, prolate (left) and oscillating (right), shown realistically as clouds (above) and simplified as geometric shapes (below).

octupole oscillations, are shown in Figure 110. In addition, non-spherical nuclei can also rotate. Several rapidly spinning nuclei, with a spin of up to $60\hbar$ and more, are known. They usually slow down step by step, emitting a photon and reducing their angular momentum at each step. Recently it was even discovered that nuclei can also have bulges that rotate around a fixed core, a bit like the tides that rotate around the Earth. Ref. 165

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 has shown that the whole range of observed phenomena can be reduced to these four fundamental processes. Each process is a type of motion. And in each process, the main interest is the comparison of the start and the end situations; the intermediate situations are less interesting. Nuclear interactions thus lack the complex types of motion which characterize everyday life. That is also the main 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. Eye observation, or seeing something, is a scattering experiment, as eye observation 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. In particular, bound motion determines shape and changes of shape of compounds: hadrons and nuclei.

Decay is obviously the basis of radioactivity. Nuclear decay can be due to the electromagnetic, the strong or the weak nuclear interaction. Decay allows studying 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, β 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

Vol. IV, page 118

The first man thought to have made transuranic elements, the physics genius Enrico Fermi, received the Nobel Prize in Physics 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 though, a new chemical element was named after her.)

When protons or neutrons are 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. In short, Fermi, Hahn, Meitner and Strassmann had observed reactions such as:

$$^{235}\mathrm{U} + \mathrm{n} \rightarrow {}^{143}\mathrm{Ba} + {}^{90}\mathrm{Kr} + 3n + 170\,\mathrm{MeV}\ . \qquad (55)$$

Meitner called the splitting process *nuclear fission*. The amount of energy liberated in fission is unusually large, millions of times larger than in a chemical interaction of an atom. In addition, several neutrons are emitted, which in turn can lead to the same process; fission 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.

Nuclear reactions are typically triggered by neutrons, protons, deuterons or γ particles. Apart from triggering fission, neutrons are used to transform lithium into tritium, which is used as (one type of) fuel in fusion reactors; and neutrons from (secondary) cosmic rays produce radiocarbon from the nitrogen in the atmosphere. Deuterons impinging on tritium produce helium in fusion reactors. Protons can trigger the transformation of lithium into beryllium. Photons can knock alpha particles or neutrons out of nuclei.

All nuclear reactions and decays are *transformations*. In each transformation, already the ancient 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

$$\mathrm{A} + \mathrm{B} \rightarrow \mathrm{C} + \mathrm{D} \tag{56}$$

implies that the reactions

$$\mathrm{A} + \overline{\mathrm{C}} \rightarrow \overline{\mathrm{B}} + \mathrm{D} \tag{57}$$

or

$$\overline{\mathrm{C}} + \overline{\mathrm{D}} \rightarrow \overline{\mathrm{A}} + \overline{\mathrm{B}} \tag{58}$$

or, if energy is sufficient,

$$\mathrm{A} \rightarrow \mathrm{C} + \mathrm{D} + \overline{\mathrm{B}}\ , \tag{59}$$

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 for protons and neutrons is 1 (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 number conservation is one reason for the danger of radioactivity, fission and fusion. The concept of baryon number was introduced by Ernst Stückelberg (b. 1905 Basel, d. 1984 Geneva), an important physicist who discovered several other concepts of particle physics, including Feynman diagrams before Feynman himself. Baryon number was renamed when Abraham Pais (b. 1918 Amsterdam, d. 2000 Copenhagen) introduced the terms 'lepton' and 'baryon'.

FIGURE 111 The destruction of four nuclear reactors in 2011 in Fukushima, in Japan, which rendered life impossible at a distance of 30 km around it (courtesy Digital Globe).

FIGURE 112 The explosion of a nuclear bomb: an involved method of killing many children in the country where it explodes and ruining the economic future of many children in the country that built it.

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* that can lead either to an explosion or to 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 almost incredibly absurd. The first such weapons were built during the Second World War, with the help of 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 (then) 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, rivers, lakes 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. The place is so radioactive that today it is even dangerous simply to fly over that island! Ref. 166

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. Fortunately, in the meantime, all countries have agreed to perform their nuclear tests underground.

Radioactivity is dangerous to humans, because it disrupts the processes inside living cells. Details on how radioactivity is measured and what effects on health it produces are provided below. Page 194

Not only nuclear bombs, also peaceful nuclear reactors are dangerous. The reason was

---

* 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 15 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}

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 α rays become radioactive in turn. They found that α rays transformed aluminium into radioactive phosphorus:

$$ {}^{27}_{13}\mathrm{Al} + {}^{4}_{2}\alpha \rightarrow {}^{30}_{15}\mathrm{P} + {}^{1}_{0}\mathrm{n} \ . \qquad (60)$$

In fact, almost all materials become radioactive when irradiated with alpha particles, neutrons or γ rays. As a result, radioactivity itself can only be contained with difficulty. After a time span that depends on the material and the radiation, any box that contains radioactive material has itself become radioactive. The ‘contagion’ stops only for very small amounts of radioactive material.

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 reactors now need to be dismantled. 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 companies cross that border.

In fact, one important nuclear reactor is (usually) not dangerous to humans: the Sun. We explore it shortly.

Page 200

CURIOSITIES AND CHALLENGES ON NUCLEI AND RADIOACTIVITY

Nowadays, nuclear magnetic resonance is also used to check the quality of food. For example, modern machines can detect whether orange juice is contaminated with juice from other fruit and can check whether the fruit were ripe when pressed. Other machines can check whether wine was made from the correct grapes and how it aged.

* *

FIGURE 113 A machine to test fruit quality with the help of nuclear magnetic resonance (© Bruker).

Ref. 171 Magnetic resonance machines pose no danger; but they do have some biological 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 such peripheral nerve stimulation yet.

* *

The amount of radioactive radiation is called the *dose*. The unit for the radioactive dose is one *gray*: it is the amount of radioactivity that deposits the energy 1 J on 1 kg of matter: $1\,\mathrm{Gy} = 1\,\mathrm{J/kg}$. A *sievert*, or 1 Sv, is the unit of radioactive dose *equivalent*; it is adjusted to humans by weighting each type of human tissue with a factor representing the impact of radiation deposition on it. Three 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 of 1 mSv; a CAT scan 8 mSv. For other measurement examples, see Table 15. Ref. 167

The *amount of radioactive material* is measured by the number of nuclear decays per second. One decay per second is called one *becquerel*, or 1 Bq. An adult human body typically contains 9 kBq, the European limit for food, in 2011, varies between 370

TABLE 16 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
Smoking 60 cigarettes a day	26 to 120 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 leads to death for 50 % of untreated patients

and 600 Bq/kg. The amount released by the Hiroshima bomb is estimated to have been between 4 PBq and 60 PBq, the amount released by the Chernobyl disaster was between 2 and 12 EBq, thus between 200 and 500 times larger. The numbers for the various Russian radioactive disasters in the 1960s and 1970 are similarly high. The release for the Fukushima reactor disaster in March 2011 is estimated to have been 370 to 630 PBq, which would put it at somewhere between 10 and 90 Hiroshima bombs.

The SI units for radioactivity are now common around the world; in the old days, 1 sievert was called 100 rem or 'Röntgen equivalent man'; the SI unit for dose, 1 gray, replaces what used to be called 100 rd or Rad. The SI unit for exposure, 1 C/kg, replaces the older unit 'röntgen', with the relation $1\,\mathrm{R} = 2.58 \cdot 10^{-4}\,\mathrm{C/kg}$. The SI unit becquerel replaces the curie (Ci), for which 1 Ci = 37 GBq.

* *

FIGURE 114 A dated image of Lake Karachay and the nuclear plant that was filling it with radioactivity (© Unknown).

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; they seem to be related to the formation process of lightning. Ref. 168

* *

Chain reactions are quite common in nature, and are not limited to the nuclear domain. *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 α emitter inside the human body are sufficient to cause lung cancer. Another example is polonium. Polonium 210 is present in tobacco leaves that were grown with artificial fertilizers. In addition, tobacco leaves filter other radioactive substances from the air. Polonium, lead, potassium and the other radioactive nuclei found in tobacco are the main reason that *smoking produces cancer*. Table 16 shows that the dose is considerable and that it is by far the largest dose absorbed in everyday life.

* *

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. In the 1970s, walking on the shore of the lake for an hour led 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; they were kept secret. Today, in contrast to Figure 114, the lake is partly filled with concrete – but not covered with it, as is often assumed. Ref. 169

* *

All lead is slightly radioactive, because it contains the ^{210}Pb isotope, a β 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 lead with a low radioactivity level. 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. The occurrence of *natural* nuclear reactors have been predicted in 1956 by Paul Kuroda. In 1972, the first such 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. All work without refuelling for between one and thirty years.

* *

Radioactivity also has forensic uses. On many surfaces, it is hard to make finger prints visible. One method is to put the object in question in an atmosphere of radioactive iodine or radioactive sulphur dioxide. The gases react with the substances in finger prints. The fingerprints have thus become radioactive. Looking at the scintillation signals of the prints – a method called *autoradiography* – then allows imaging the fingerprint simply by laying a photographic film or an equivalent detector over the object in question.

* *

In contrast to massive particles, massless particles cannot decay at all. There is a simple reason for it: massless particles do not experience time, as their paths are 'null'. A particle that does not experience time cannot have a half-life. (Can you find another argument?) Challenge 134 s

* *

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, they were able to see the bones in their own hand and arms. How can this be? Challenge 135 s

* *

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 at least 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.

* *

Nuclear bombs are terrible weapons. To experience their violence but also the criminal actions of many military people during the tests, have a look at the pictures of explosions. Ref. 170 In the 1950s 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 radioactive dating methods are used to date wine, and more are in development. A few are included in Table 14. Page 183

* *

A few rare 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

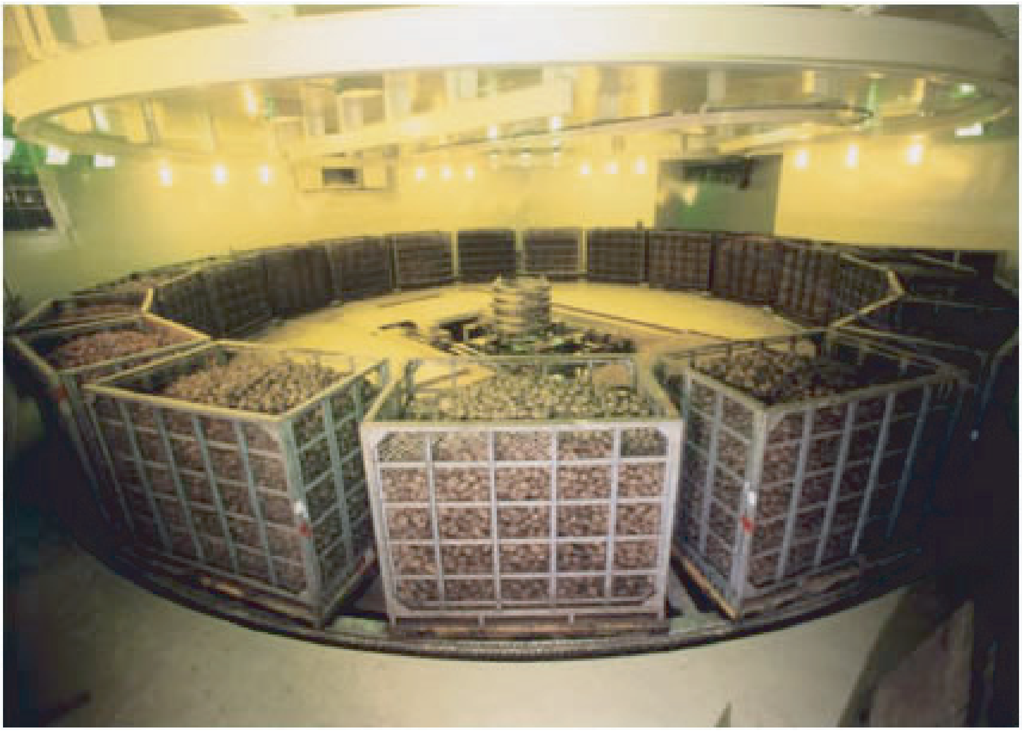


FIGURE 115 Potatoes being irradiated at the Shihorocho Agricultural Cooperative Isotope Irradiation Center in Japan. Good appetite!

of 27 GPa. On the other hand, these effects are predicted (and measured) to be negligible for nuclei of larger mass. A few additional nuclides show similar, but smaller effects. Ref. 159

The most interesting effect on nuclei is laser-induced fissioning of ^{238}U, which occurs for very high laser intensities. Page 259

* *

Both γ radiation and neutron radiation can be used to image objects without destroying them. γ rays have been used to image the interior of the Tutankhamun mask. Neutron radiation, which penetrates metals as easily as other materials, has been used to image, even at film speed, the processes inside car engines.

* *

γ rays are used in Asia to irradiate food. This is forbidden other countries, such as Germany. For example, γ rays are used to irradiate potatoes, in order to prevent sprouting. An example is given in Figure 115. It is better not to work there. In fact, over 30 countries allowed the food industry to irradiate food. For example, almost all spice in the world are treated with γ rays, to increase their shelf life. However, the consumer is rarely informed about such treatments.

* *

β rays with 10 MeV and γ rays are used in many large factories across the world to sterilize medical equipment, medical devices, toys, furniture and also to kill moulds in books and in animal food. (See www.bgs.eu for an example.)

* *

The non-radioactive isotopes ^{2}H – often written simply D – 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 $^{2}H_2{}^{18}O$ 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 $^{2}H_2O$. Measurements on the urine allow one 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.

* *

The percentage of the ^{18}O isotope in the water of the Earth's oceans can be used to deduce where the water came from. This was told in the first volume of our adventure. Vol. I, page 370

* *

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 275

Summary on nuclei

Atomic nuclei are composed of protons and neutrons. Their diameter is between one and a few femtometres, and they have angular momentum. Their angular momentum, if larger than zero, allows us 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 colours that are determined by their spectra. Nuclei can decay, can scatter, can break up and can react among each other. Nuclear reactions can be used to make bombs, power plants, generate biological mutations and to explore the human body. And as we will discover in the following, nuclear reactions are at the basis of the working of the Sun and of our own existence.

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 fascinating to offer: by exploring nuclei, we learn to understand the Sun, the stars, the early universe, the birth of matter and our own history. Ref. 172

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 in nature – the others being gravitation, electromagnetism and the weak nuclear interaction. Despite its strength, we do not experience the strong nuclear interaction in everyday life, because its range is limited to distances of a few femtometres, i.e., to a few proton diameters. Despite this limitation, the strong interaction tells a good story about the burning of the Sun and about the flesh and blood we are made of.

The Sun

At present, the Sun emits 385 YW of light. The amount was first measured by Claude Pouillet at the start of the nineteenth century. The power would be sufficient to melt away, every year, a volume of ice 500 times larger than the volume of the Earth. Challenge 136 e

Where does the huge energy emitted by the Sun come from? If it came from burning coal, the Sun would stop burning after a few thousands of years. When radioactivity was discovered, researchers tested the possibility that this process might be at the heart of the Sun's shining. However, even though radioactivity – or the process of fission that was discovered later – is able to release more energy than chemical burning, the composition of the Sun – mostly hydrogen and helium – makes this impossible.

The origin of the energy radiated by the Sun was clarified in 1929 by Fritz Houtermans, Ref. 173 Carl Friedrich von Weiszäcker, and Hans Bethe: the Sun burns by *hydrogen fusion*. Fusion is the composition of a large nucleus from smaller ones. In the Sun, the central fusion reaction

$$4\,{}^{1}\mathrm{H} \rightarrow {}^{4}\mathrm{He} + 2\,e^{+} + 2\,\nu + 4.4\,\mathrm{pJ} \qquad (61)$$

[**] 'Learning is anticipated joy about yourself.'

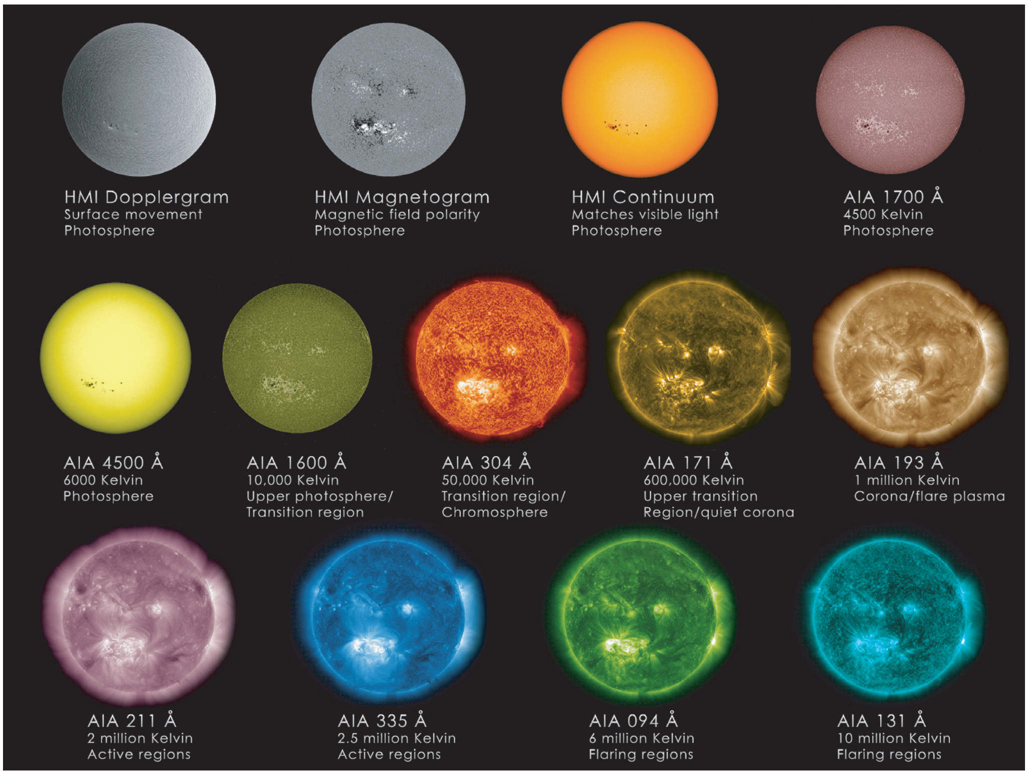


FIGURE 116 The Sun emits radiation at different wavelengths. Note that almost all images are shown in a single, false colour selected just for visual appeal. The collections does not show the radio wave images, which also show the solar spots, but with much lower resolution. (Courtesy NASA)

converts hydrogen nuclei into helium nuclei. The reaction is called the *hydrogen–hydrogen cycle* or *p–p cycle*. The hydrogen cycle is the result of a continuous cycle of three separate nuclear reactions:

$$\begin{aligned} \mathrm{p}+\mathrm{p} &\rightarrow \mathrm{d}+e^{+}+\nu \quad \text{(a weak nuclear reaction involving the deuteron)} \\ \mathrm{d}+\mathrm{p} &\rightarrow {}^{3}\mathrm{He}+\gamma \quad \text{(a strong nuclear reaction)} \\ {}^{3}\mathrm{He}+{}^{3}\mathrm{He} &\rightarrow \alpha+2\,\mathrm{p}+\gamma\,. \end{aligned} \tag{62}$$

We can also write the p-p cycle as

$$\begin{aligned} {}^{1}\mathrm{H}+{}^{1}\mathrm{H} &\rightarrow {}^{2}\mathrm{H}+e^{+}+\nu \quad \text{(a weak nuclear reaction)} \\ {}^{2}\mathrm{H}+{}^{1}\mathrm{H} &\rightarrow {}^{3}\mathrm{He}+\gamma \quad \text{(a strong nuclear reaction)} \\ {}^{3}\mathrm{He}+{}^{3}\mathrm{He} &\rightarrow {}^{4}\mathrm{He}+2\,{}^{1}\mathrm{H}+\gamma\,. \end{aligned} \tag{63}$$

In total, four protons are thus fused to one helium nucleus, or alpha particle; if we include the electrons, we can say that four hydrogen atoms are fused to one helium atom. The fusion process emits neutrinos and light with a total energy of 4.4 pJ (26.7 MeV). This is

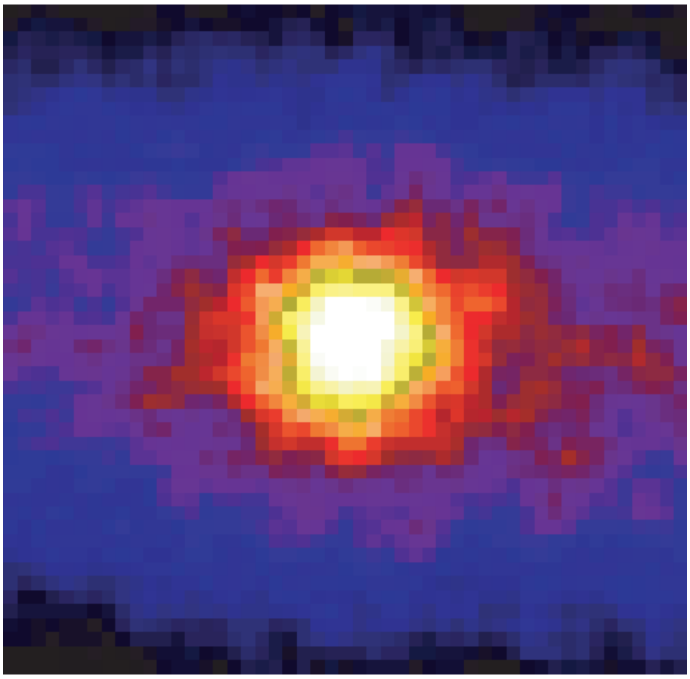


FIGURE 117 The Sun also emits neutrinos. Their intensities are shown here in a false colour image, taken through the whole Earth from an underground experiment, with a 503-day exposure, at energies from 7 to 25 MeV. However, due to scattering processes, the bright spot is several times the size of the Sun. (© Robert Svoboda)

the energy that makes the Sun shine. Most of the energy is emitted as light; around 10 % is carried away by neutrinos. The latter part is illustrated in Figure 117.

The first of the three reactions of equation (62) is due to the weak nuclear interaction. This transmutation and the normal β decay have the same first-order Feynman diagram. Challenge 137 e The weak interaction avoids that fusion happens too rapidly and ensures that the Sun will shine still for some time. Indeed, in the Sun, with a luminosity of 385 YW, there are Ref. 174 thus only 10^{38} fusions per second. This allows us to deduce that the Sun will last another handful of Ga (Gigayears) before it runs out of fuel.

The simplicity of the hydrogen-hydrogen cycle 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. The second reaction of the proton cycle takes a few seconds and the third about one million years.

The fusion reaction (62) takes place in the centre of the Sun, in the so-called *core*. Fortunately for us, the high energy γ photons generated in the Sun's centre are 'slowed' down by the outer layers of the Sun, namely the *radiation zone*, the *convection zone* with its involved internal motion, and the so-called *photosphere*, the thin layer we actually see. The last layer, the atmosphere, is not visible during a day, but only during an eclipse, as shown in Figure 120. More precisely, the solar atmosphere consists of the *temperature minimum*, the *chromosphere*, the *transition region*, the *corona* and the *heliosphere*.

During the elaborate slowing-down process inside the Sun, the γ photons are pro-

FIGURE 118 A photograph of the Sun at a visible wavelength of around 677 nm, by the SOHO space probe, showing a few sunspots (ESA and NASA).

gressively converted to visible photons, mainly through scattering. Scattering takes time. In the Sun, it takes along time: the sunlight of today was in fact generated at the time of Ref. 175 the Neandertals: a typical estimate is about 170 000 years ago. In other words, the average effective speed of light inside the Sun is estimated to be around 300 km/year! After these one hundred and seventy thousand years, the photons take another 8.3 minutes to reach the Earth and to sustain the life of all plants and animals.

Motion in and on the Sun

Vol. III, page 148 In its core, the Sun has a temperature of around 15 MK. At its surface, the temperature Challenge 138 e is around 5.8 kK. (Why is it cooler?) Since the Sun is cooler on its surface than in its centre, the Sun is not a homogeneous ball, but an inhomogeneous structure. If you want to experience the majestic beauty of the Sun, watch the stunning video www.youtube.

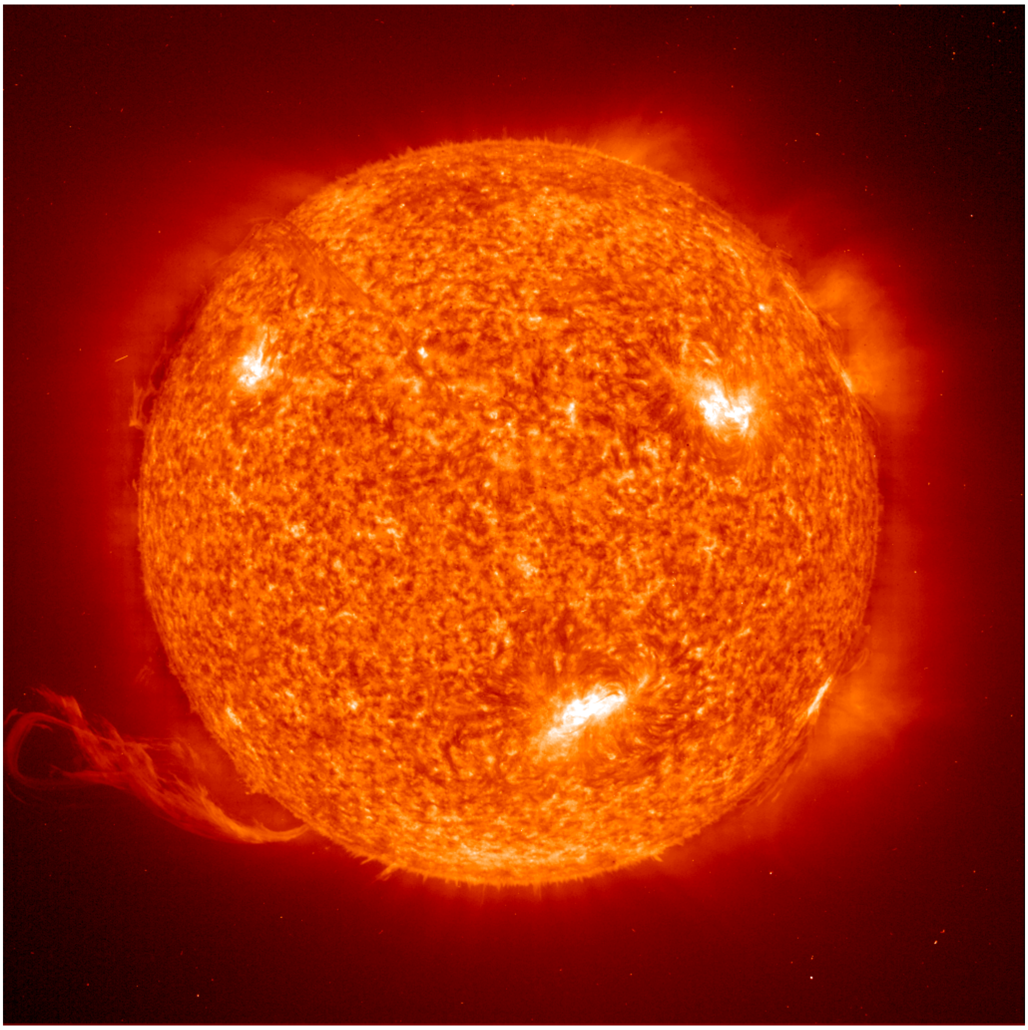


FIGURE 119 A photograph of the Sun at the extreme ultraviolet wavelength of 30.4 nm, thus in false colour, again by the SOHO space probe, showing solar prominences (ESA and NASA).

com/watch?v=ipvfwPqh3V4 that shows the Sun's surface over a two-week period. The inhomogeneity of the Sun's structure and surface is due to the convection processes induced by the temperature gradient. The convection, together with the rotation of the Sun around its axis, leads to fascinating structures that are shown in Figure 119 and the following ones: *solar eruptions*, including flares and coronal mass ejections, and *solar spots*.

In short, the Sun is not a static object. The matter in the Sun is in constant motion. An impressive way to experience the violent processes it contains is to watch the film shown in Figure 122, which shows the evolution of a so-called *solar flare*. Many solar eruptions, such as those shown in the lower left corner in Figure 119 or in Figure 123, eject matter far into space. When this matter reaches the Earth,* after being diluted by the journey, it affects our everyday environment. Such *solar storms* can deplete the higher atmosphere

---

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

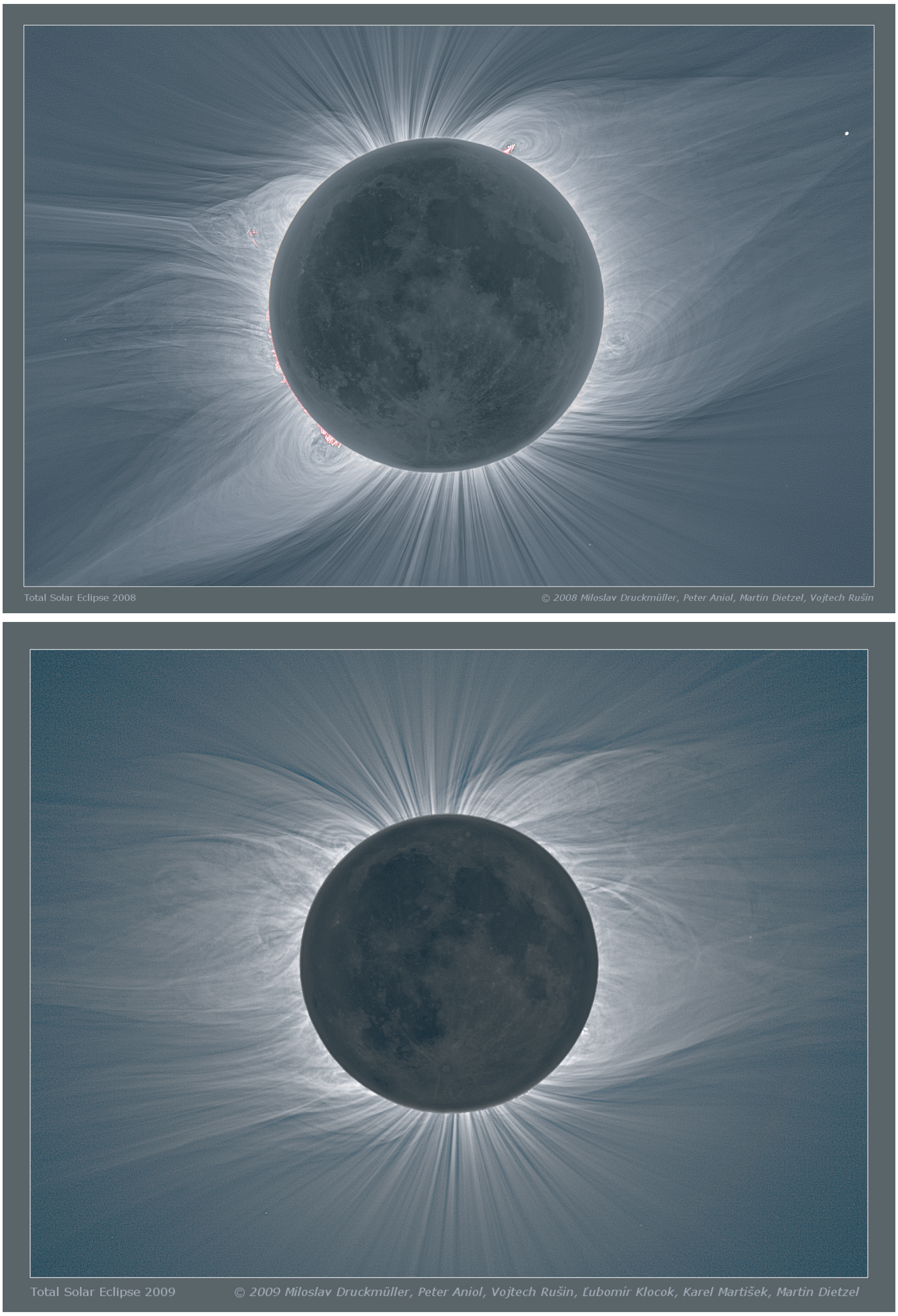


FIGURE 120 The complex details of the corona of the Sun during the 2008 eclipse in Bor Udzuur in Mongolia and the 2009 eclipse on the Marshall Islands. The images are digital compositions of several dozen photographs chosen to reproduce the experience of looking at the eclipse through a small telescope. The structures also allow to locate the solar poles. The top image includes protuberances. (Top image © Miloslav Druckmüller, Martin Dietzel, Peter Aniol, Vojtech Rušin; bottom image © Miloslav Druckmüller, Peter Aniol, Vojtech Rušin, Ľubomír Klocok, Karel Martišek and Martin Dietzel)

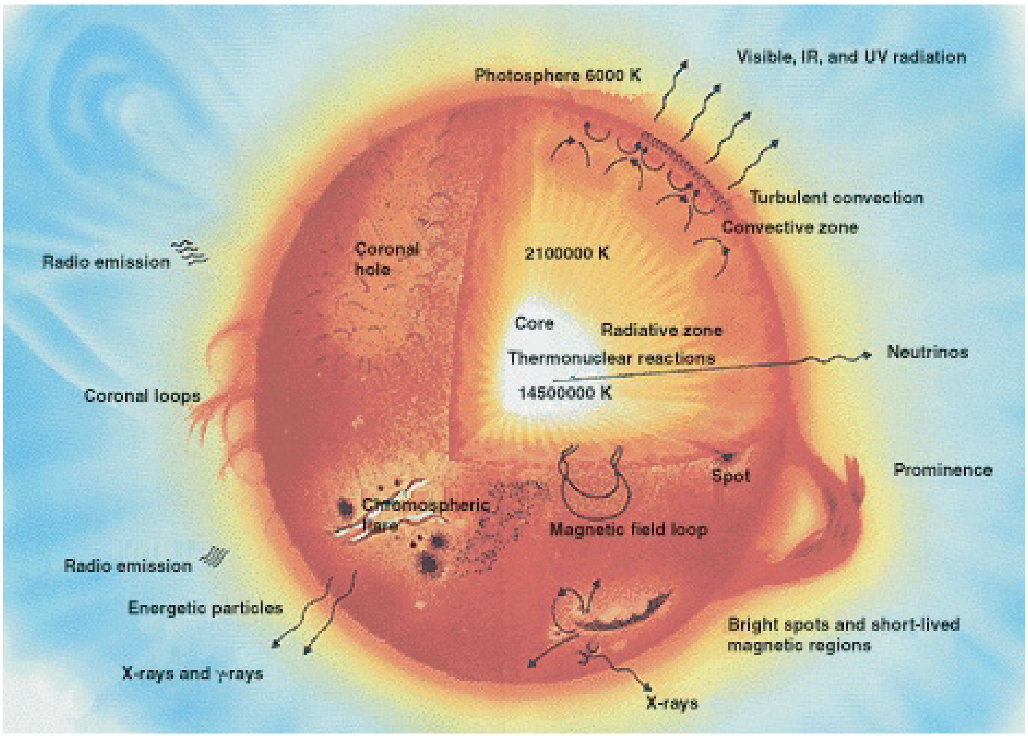


FIGURE 121 A drawing of the interior of the Sun (courtesy NASA).

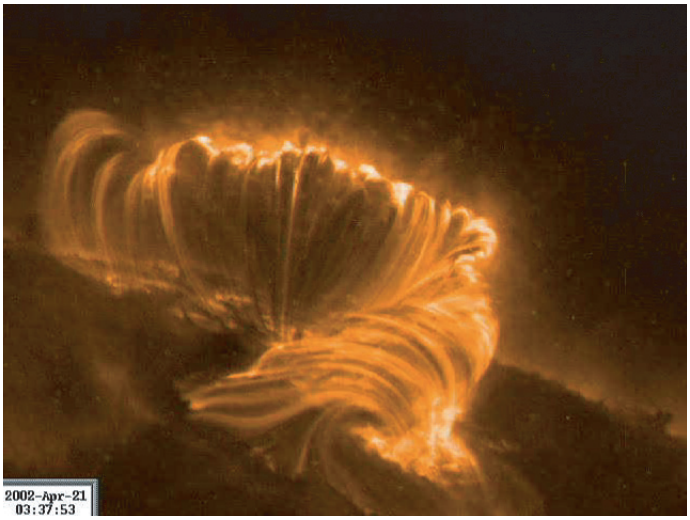


FIGURE 122 The evolution of a solar flare observed by the TRACE satellite (QuickTime film courtesy NASA).

and can thus possibly even trigger usual Earth storms. Other effects of solar storms are the formation of auroras and the loss of orientation of birds during their migration; this happens during exceptionally strong solar storms, because the magnetic field of the Earth

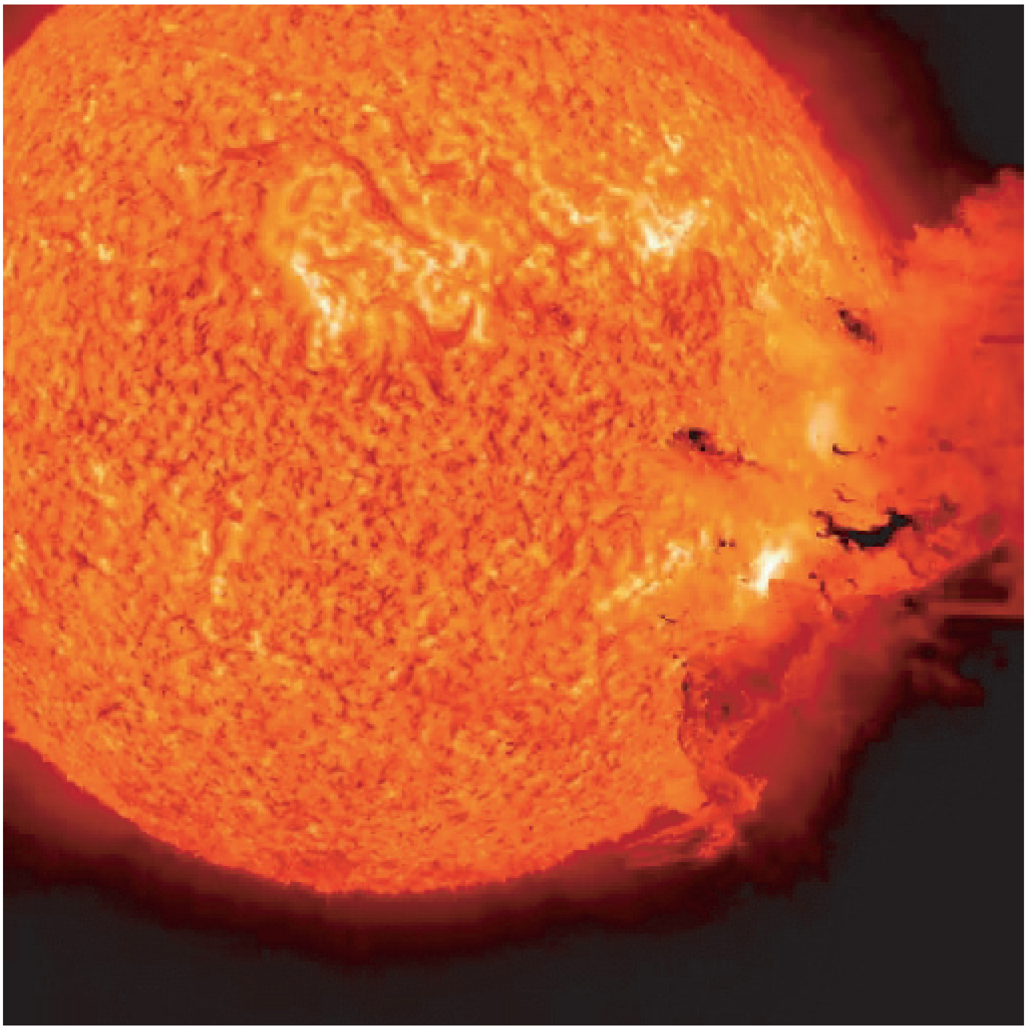


FIGURE 123 A spectacular coronal mass ejection observed on June 7, 2011 by the Solar Dynamic Observatory satellite (QuickTime film courtesy NASA).

is disturbed in these situations. A 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 power network has been redesigned to withstand such events.

How can the Sun's surface have a temperature of 6 kK, whereas the Sun's *corona* – the thin gas emanating from and surrounding the Sun that is visible during a total solar eclipse, as shown in Figure 120 – reaches one to three million Kelvin on average, with localized peaks inside a flare of up to 100 MK? 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 119 or above the dark spots in Figure 118. As a result, the particles of the corona are accelerated and heat the corona to temperatures that are a thousand times higher than those at the surface of the Sun.

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 nuclei of iron ^{56}Fe are produced. For nuclei beyond this nucleus, as shown in Figure 124, the binding energies per nucleon then decrease again; thus fusion is not energetically possible. It turns out that the heavier nuclei found on Earth and across the universe were formed through *neutron capture*. In short, nuclei below iron are made through fusion, nuclei above iron are made through neutron capture. And for the same reason, nuclei release energy through *fusion* when the result is lighter than iron, and release energy through *fission* when the starting point is above iron.

The different stars observed in the sky** can be distinguished by the type of fusion nuclear reaction that dominates them. Most stars, in particular young or light stars, run hydrogen fusion. But that is not all. There are several types of hydrogen fusion: the direct hydrogen–hydrogen (p–p) cycle, as found in the Sun and in many other stars, and the various CNO cycle(s) or Bethe-Weizsäcker cycle(s).

Page 200

The hydrogen cycle was described above and can be summarized as

$$4\,^1\mathrm{H} \rightarrow {}^4\mathrm{He} + 2\,e^+ + 2\,\nu + 4.4\,\mathrm{pJ} \qquad (64)$$

---

* Friedrich Houtermans (1903–1966) was one of the most colourful physicists of his time. He lived in Austria, England, the Soviet Union, Germany and the United States. He was analyzed by Sigmund Freud, imprisoned and tortured by the NKWD in Russia, then imprisoned by the Gestapo in Germany, then worked on nuclear fission. He worked with George Gamow and Robert Atkinson.

** To find out which stars are in the sky above you at present, see the www.surveyor.in-berlin.de/himmel website.

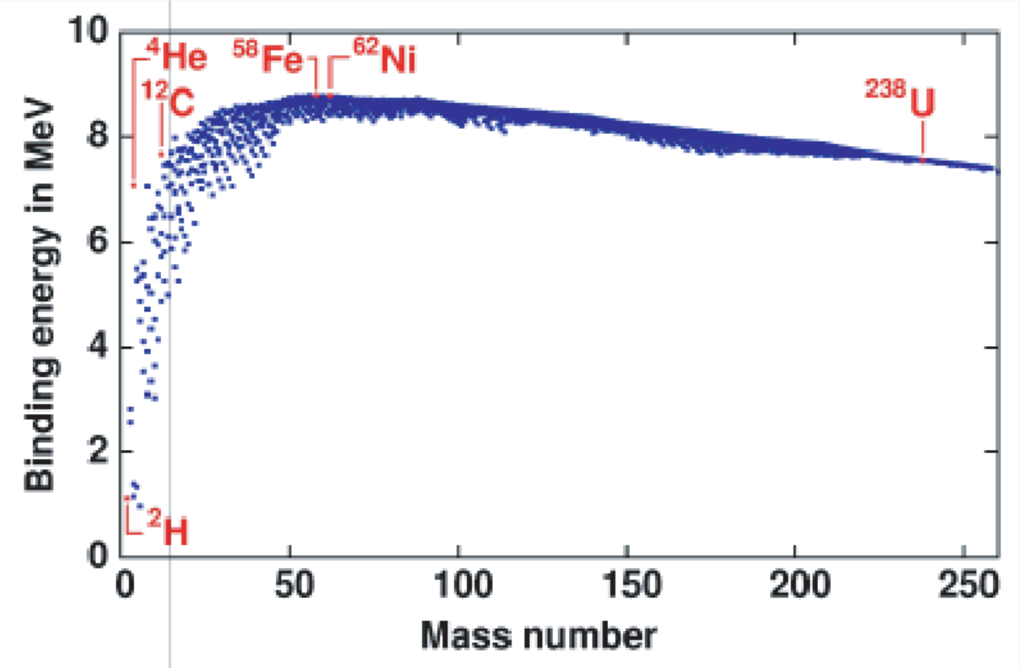


FIGURE 124 Measured values of the binding energy *per nucleon* in nuclei. The region on the left of the maximum, located at ^{58}Fe, is the region where *fusion* is energetically possible; the right region is where *fission* is possible (© Max Planck Institute for Gravitational Physics).

or, equivalently,

$$4\,\mathrm{p} \rightarrow \alpha + 2\,e^{+} + 2\,\nu + 26.7\,\mathrm{MeV}\ . \qquad (65)$$

But this is not the only way for a star to burn. If a star has heavier elements inside it, the hydrogen fusion uses these elements as catalysts. This happens through the so-called Bethe-Weizsäcker cycle or CNO cycle, which runs as

$$\begin{aligned} {}^{12}\mathrm{C} + {}^{1}\mathrm{H} &\rightarrow {}^{13}\mathrm{N} + \gamma \\ {}^{13}\mathrm{N} &\rightarrow {}^{13}\mathrm{C} + \mathrm{e}^{+} + \nu \\ {}^{13}\mathrm{C} + {}^{1}\mathrm{H} &\rightarrow {}^{14}\mathrm{N} + \gamma \\ {}^{14}\mathrm{N} + {}^{1}\mathrm{H} &\rightarrow {}^{15}\mathrm{O} + \gamma \\ {}^{15}\mathrm{O} &\rightarrow {}^{15}\mathrm{N} + \mathrm{e}^{+} + \nu \\ {}^{15}\mathrm{N} + {}^{1}\mathrm{H} &\rightarrow {}^{12}\mathrm{C} + {}^{4}\mathrm{He} \qquad (66) \end{aligned}$$

The end result of the cycle is the *same* as that of the hydrogen cycle, both in nuclei and in energy. The Bethe-Weizsäcker 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?) Inside the Sun, due to the comparatively low temperature of a few tens of million kelvin, the Bethe-Weizsäcker cycle (and its variations) is not as important as the hydrogen cycle.

Challenge 139 s

The proton cycle and the Bethe-Weizsäcker cycle are not the only options for the burning of stars. Heavier and older stars than the Sun can also shine through other fusion reactions. In particular, when no hydrogen is available any more, stars run *helium burning*:

$$3\,^{4}\mathrm{He} \rightarrow {}^{12}\mathrm{C}\,. \qquad (67)$$

This fusion reaction, also called the *triple-α process*, 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 apes, pigs and people. The prediction of this resonance by Fred Hoyle is one of the few predictions in physics made from 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.

Vol. III, page 337

The studies of star burning processes also explain why the Sun and the stars do not collapse. In fact, the Sun and most stars are balls of hot gas, and the *gas pressure* due to the high temperature of its constituents prevents their concentration into a small volume. For other types of stars – especially those of high mass such as red giants – the *radiation pressure* of the emitted photons prevents collapse; for still other stars, such as neutron stars, the role is taken by the *Pauli pressure*.

The nuclear reaction rates at the interior of a star are extremely sensitive to its 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-α reaction rate is proportional to T^{40}; these strong dependencies imply that stars usually shine with constancy over long times, often thousands and millions of years, because 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; some stars change in brightness every few years. And even the Sun shows such effects. In the 1960s, 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 studying what is happening inside stars, even separately in each of the layers they consist of.

By the way, it is still not clear how much the radiation of the Sun changes 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 intensity of the Suns, the so-called *solar constant* might have important consequences for climate research; however, the issue is still open.