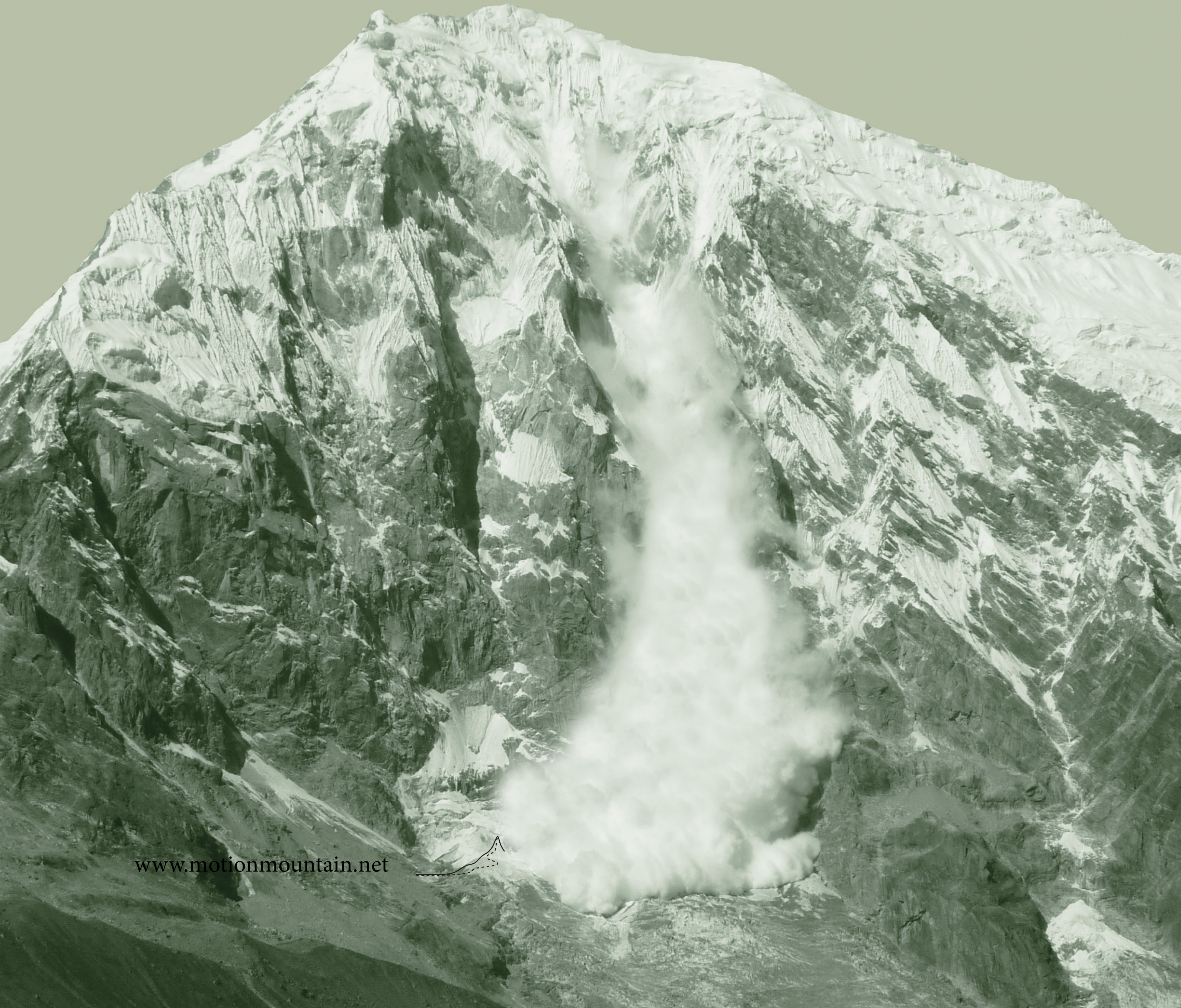
Christoph Schiller

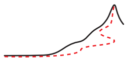
MOTION MOUNTAIN

THE ADVENTURE OF PHYSICS – VOL.III

LIGHT, CHARGES AND BRAINS

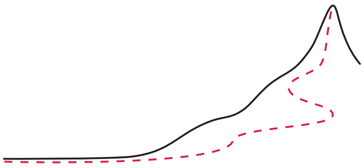


www.motionmountain.net



Christoph Schiller

Motion Mountain



The Adventure of Physics
Volume III

Light, Charges and Brains

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 the best ones when exploring everything electric. They lead from the weighing of electric current to the use of magnetic fields to heal bone fractures and up to the use of light to cut metals and the understanding of the human brain.

In the structure of physics, shown in Figure 1, motion due to electricity is the most fascinating aspect of the starting point at the bottom. Indeed, almost everything around us is due to electric processes. The present introduction to electricity, magnetism, light and the brain is the third of a six-volume overview of physics that arose from a threefold aim that I have pursued since 1990: to present motion in a way that is simple, up to date and captivating.

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

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

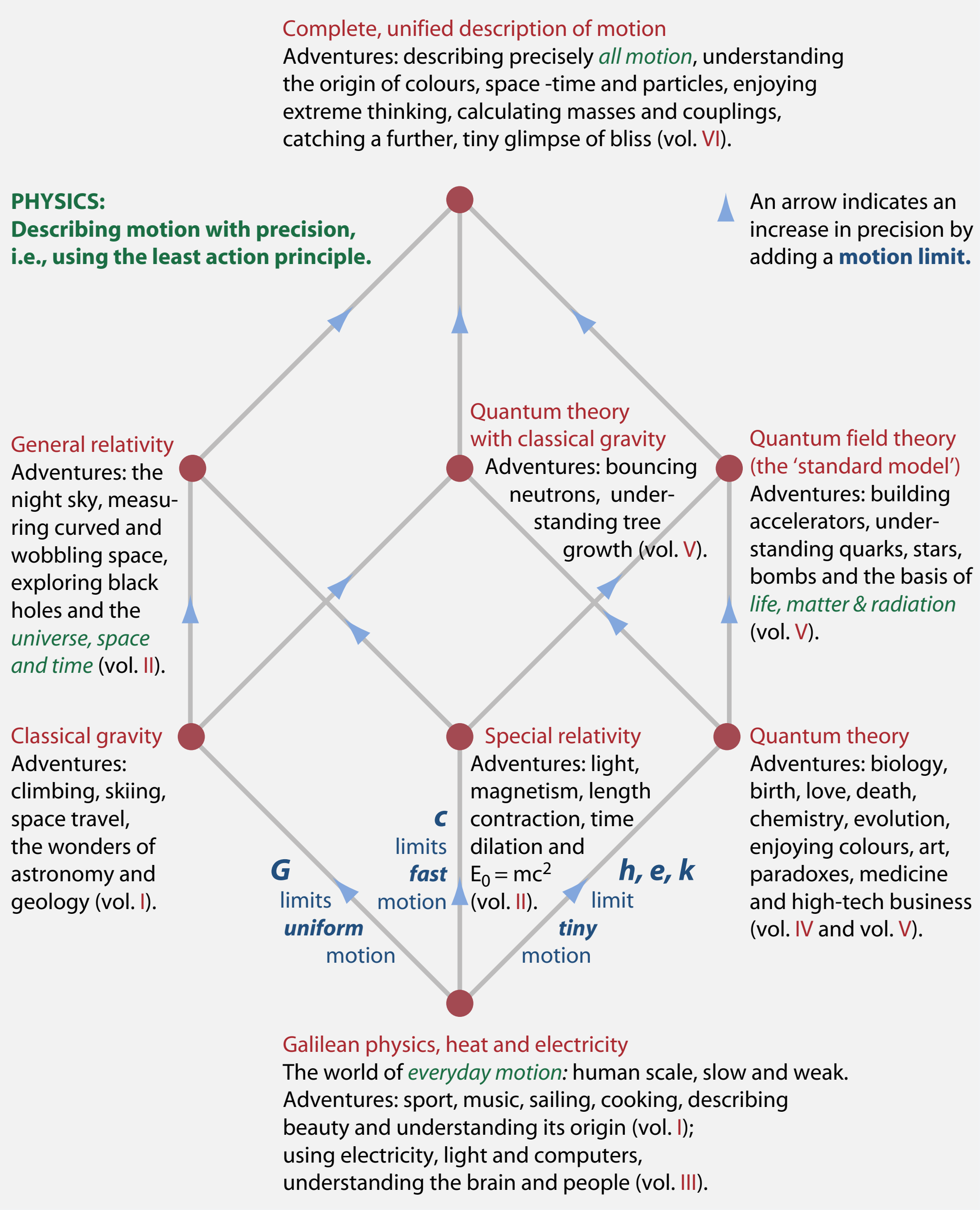
In order to be *captivating*, the text tries to startle the reader as much as possible. Reading a book on general physics should be like going to a magic show. We watch, we are astonished, we do not believe our eyes, we think, and finally we understand the trick. When we look at nature, we often have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



Complete, unified description of motion
Adventures: describing precisely *all motion*, understanding the origin of colours, space -time and particles, enjoying extreme thinking, calculating masses and couplings, catching a further, tiny glimpse of bliss (vol. VI).

PHYSICS:
Describing motion with precision, i.e., using the least action principle.

An arrow indicates an increase in precision by adding a **motion limit.**

General relativity
Adventures: the night sky, measuring curved and wobbling space, exploring black holes and the *universe, space and time* (vol. II).

Quantum theory with classical gravity
Adventures: bouncing neutrons, understanding tree growth (vol. V).

Quantum field theory (the 'standard model')
Adventures: building accelerators, understanding quarks, stars, bombs and the basis of *life, matter & radiation* (vol. V).

Classical gravity
Adventures: climbing, skiing, space travel, the wonders of astronomy and geology (vol. I).

Special relativity
Adventures: light, magnetism, length contraction, time dilation and $E_0 = mc^2$ (vol. II).

Quantum theory
Adventures: biology, birth, love, death, chemistry, evolution, enjoying colours, art, paradoxes, medicine and high-tech business (vol. IV and vol. V).

Galilean physics, heat and electricity
The world of *everyday motion:* human scale, slow and weak.
Adventures: sport, music, sailing, cooking, describing beauty and understanding its origin (vol. I);
using electricity, light and computers, understanding the brain and people (vol. III).

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.

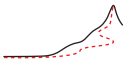
On behalf of all readers, thank you in advance for your input. For a particularly useful contribution you will be mentioned – if you want – in the acknowledgements, receive a reward, or both.

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



LIGHT, CHARGES AND BRAINS

In our quest to learn how things move,
the experience of hiking and other motion
leads us to discover that images are produced by charges,
that charges move, accumulate and interact,
and that there is a smallest charge in nature.
We understand what love has to do with magnets and amber,
why the brain is such an interesting device,
and what distinguishes a good from a bad lie.

Chapter 1

LIQUID ELECTRICITY, INVISIBLE FIELDS AND MAXIMUM SPEED

What is *light*? The study of relativity left us completely in the dark, even though we had embarked in it precisely to find an answer to that question. True, we have learned how the motion of light compares with that of objects. We also learned that light is a moving entity that cannot be stopped, that light provides the speed limit for any type of energy, and that light is our measurement standard for speed. However, we haven't learned anything about the nature of light itself, nor about *colours*, nor about how rain drops[**] and other matter produces them.

A second question is open: what is *contact*? We still do not know. In our exploration of relativity we learned that all interactions, including contact, are due to exchange of something. But of what? We only learned that truly mechanical interactions do not exist. What is the nature of contact?

Vol. II, page 83

A third question also arises: how do we *sense* contact or touch? What are *sensors* and how is their output, the data, processed in the brain or in machines? Not only the brain, also all other data processing systems use electricity. What is *data* and what is *electricity*?

The answer to the questions about the nature of light, contact and the brain is *not* related to gravitation. If we make a list of motors found in this world, we notice that gravitation hardly describes any of them. Neither the motion of sea waves, fire and earthquakes, nor that of a gentle breeze is caused by gravity. The same applies to the motion of light in a rainbow or to the motion of muscles. Have you ever listened to your own heart beat with a stethoscope? You can also use, as many medical doctors do now, a mobile phone to record your heart beat.) Without having done so, you cannot claim to have experienced the mystery of motion. Your heart has about 3000 million beats in your lifetime. Then it stops.

Vol. I, page 233

Challenge 2 e

It was one of the most astonishing discoveries of science that the origin of heart beats, fire, light and thought itself is connected to observations made thousands of years ago using two strange stones. These stones show

▷ All those examples of motion that are called *mechanical* in everyday life are, without exception, of *electrical* origin.

In particular, the solidity, the softness and the impenetrability of matter are due to internal electricity. But also the emission of light, the formation of colours and the work-

---

[**] The photograph of a circular rainbow on page 14 was taken in 2006 from the Telstra Tower in Canberra (© Oat Vaiyaboon).

FIGURE 2 Objects surrounded by fields: amber (c. 1 cm) attracts sawdust, lodestone (c. 1 cm) attracts iron filings and a mobile phone (c. 10 cm) attracts other mobile phones and people (© Wikimedia, Philips).

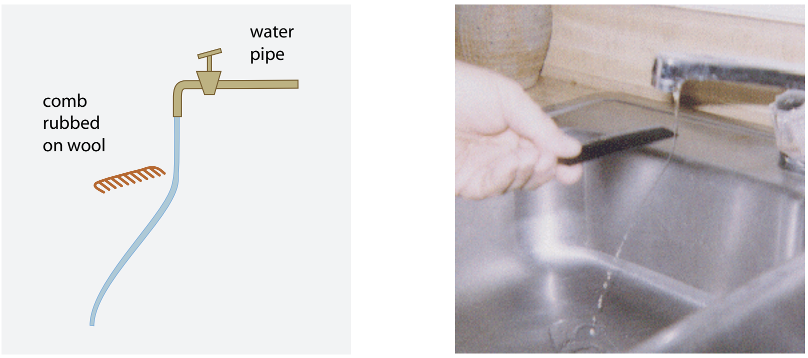


FIGURE 3 How to amaze kids, especially in dry weather (photo © Robert Fritzius).

Ref. 1 ing of our nerves and brains are due to electrical processes. As these aspects are part of everyday life, we can leave aside all complications due to gravity and curved space-time.

Exploring light, contact and the brain implies to explore how magicians levitate object. Indeed, the most productive way to study electrical motion is to start, as in the case of gravity, with those types of motion which are generated without any contact between the bodies involved. This can happen in three ways.

Fields: amber, lodestone and mobile phones

You can always surprise children with the effect shown in Figure 3: a comb rubbed on wool deviates running tap water. The same effect can be produced with an air-filled rubber balloon rubbed on wool. Everybody can deviate water streams without any contact.

The Greeks had already observed this effect a long time ago. In fact, the story of electricity starts with trees. Trees have a special relation to electricity. When a tree is cut, a viscous resin appears. With time it solidifies and, after millions of years, it forms *amber*. When amber is rubbed with a cat fur, it acquires the ability to attract small objects, such as saw dust or pieces of paper. This was already known to Thales of Miletus, one of the original seven sages, in the sixth century BCE. The same observation can be made with many other polymer combinations, for example with combs and hair, with soles of the shoe on carpets, and with dust and a lens or a cathode ray tube inside an old television. Another interesting effect can be observed when a rubbed comb is put near a burning candle. (Can you imagine what happens?) Challenge 3 s

Another part of the story of electricity involves *lodestone*, an iron mineral found in certain caves around the world, e.g. in a region (still) called Magnesia in the Greek province of Thessalia, and in some regions in central Asia. When two stones of this mineral are put near each other, they attract or repel each other, depending on their relative orientation. In addition, lodestone attracts objects made of cobalt, nickel or iron.

Today we also find various small objects in nature with more sophisticated properties, such as the one shown on the right of Figure 2. Some of these objects allow you to talk with far away friends, others unlock car doors, still others enable you to switch on a television.

In short, in nature there are situations where bodies exert influence on others *at a distance*. The space surrounding a body exerting such an influence is said to contain a field. A *(physical) field* is an entity that manifests itself by accelerating other bodies in a given region of space.

▷ A field is space that changes momenta.

If you prefer, *a field is space that exerts forces.* Or again, a field is space with some extra structure. Despite this extra structure, fields, like space, are invisible. The three objects just mentioned produce three types of fields.

1. The field around amber – called ἤλεκτρον in Greek, from a root meaning 'brilliant, shining' – is called an *electric field*. The name is due to a proposal by the famous physician and part-time physicist William Gilbert (b. 1544 Colchester, d. 1603 London). Objects surrounded by a permanent electric field are called *electrets*. Electrets are not so common; among others, they are used in certain loudspeaker systems. Electrets can be certain crystals or polymers.
2. The field surrounding the mineral found in Magnesia is called a *magnetic field* and Ref. 2 the objects producing a permanent field are called *magnets*. Most magnets, but not all, are made from metals.
3. The field around a mobile phone is called a *radio* field or, as we will see later, an *electromagnetic* field. In contrast to the previous fields, it oscillates over time. We will find out later that many other objects are surrounded by such fields, though these are often very weak. Objects that emit oscillating fields, such as mobile phones or lamps, are called radio transmitters or electromagnetic emitters. Certain radio transmitters, as we will see, are already familiar from everyday life: lamps and lasers.

Experiments show that fields have *no mass* and no material support. Fields influence bodies over a distance. Since fields are invisible, to make them imaginable, we need to colour them. Ways to colour *electric* fields are shown in Figure 4. The colourings are inspired by the experiments with seeds or dust. Visualizations for magnetic and radio fields follow below. These figures are the best way to *visualize* electric fields; also the researcher who first proposed the field concept, Michael Faraday, used such images.

Exploring visualizations of fields, we note that we can visualize electric fields either as

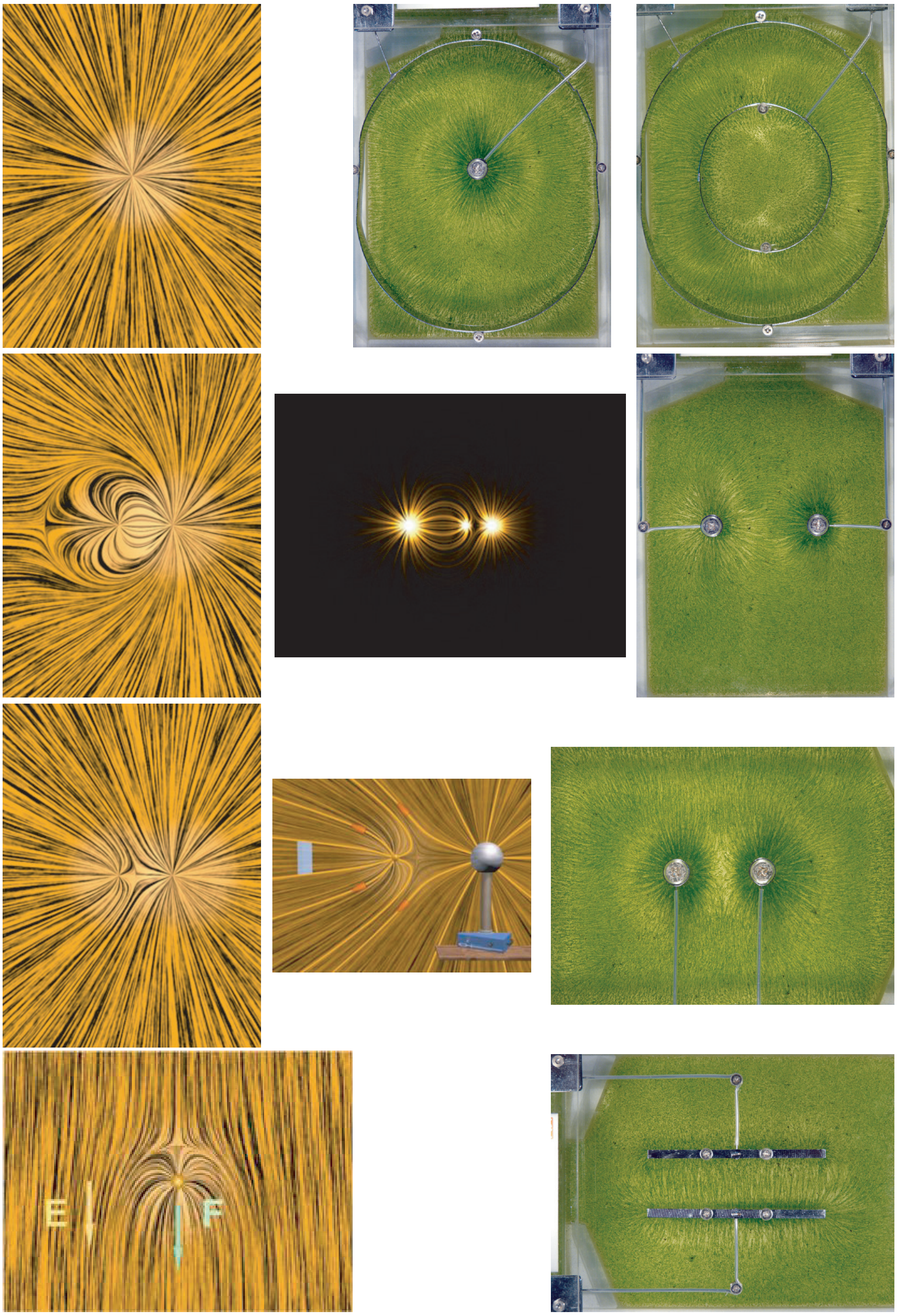


FIGURE 4 Visualizing what is invisible with computer graphics (left) and with seeds in oil (right): *an electric field is space with a structure*. Top: the field around a point or spherical charge; second row: two or three charges of different signs; third row: two charges of the same sign; bottom: a charge in an external field E, and the field between two plates. The charge will feel a force F directed along the so-called *electric field lines*; the density of the lines gives the intensity of the field and thus the strength of the force (© MIT, Eli Sidman, MIT).

FIGURE 5 Lightning: a picture taken with a moving camera, showing its multiple strokes (© Steven Horsburgh).

a tiny arrow or vector attached to every point of space, or as a bundle of lines in every region of space. Both visualizations are useful. We will encounter further visualizations below.

For a long time, electric, magnetic and radio fields were rarely noticed in everyday life. Indeed, in the past, most countries had laws that did not allow producing such fields! Still today, laws severely restrict the properties of machines that use and produce such fields. These laws require that for any device that moves, produces sound, or creates moving pictures, fields need to remain *inside* the device. Also for this reason a magician moving an object on a table via a hidden magnet still surprises and entertains his audience. To feel the fascination of fields more strongly, we take a deeper look into a few experimental results.

HOW CAN ONE MAKE LIGHTNING?

Everybody has seen a lightning flash or has observed the effect it can have on striking a tree. Obviously lightning is a moving phenomenon. Photographs such as that of Figure 5 show that the tip of a lightning flash advance with an average speed of around 600 km/s. But *what* is moving? To find out, we have to find a way of making lightning for ourselves. In 1995, the car company Opel accidentally rediscovered an old and simple method of achieving this.

Opel engineers had inadvertently built a spark generating mechanism into their cars; when filling the petrol tank, sparks were generated, which sometimes lead to the explosion of the fuel at the petrol station. Opel had to recall 2 million vehicles. Ref. 3

What had the engineers done wrong? They had unwittingly copied the conditions for a spark-generating device which anyone can build at home and which was originally invented by William Thomson:* the *Kelvin generator*. Repeating his experiment today, we would take two water taps, four empty bean or coffee cans, of which two have been

---

* William Thomson (b. 1824 Belfast, d. 1907 Largs), important physicist and professor at Glasgow University. He worked on the determination of the age of the Earth, showing that it was much older than 6000 years,

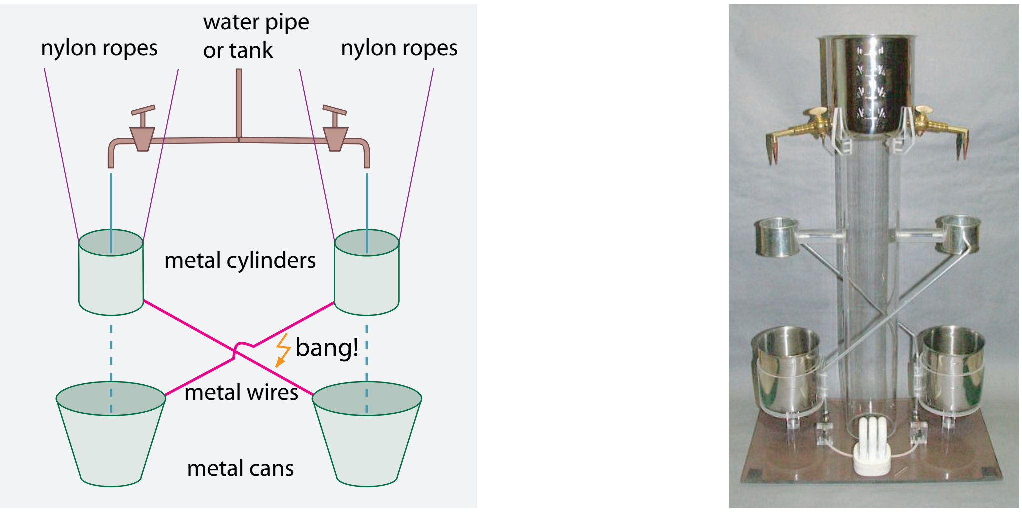


FIGURE 6 A simple Kelvin generator; the one on the right lights a fluorescent light bulb using dripping water (photograph © Harald Chmela).

Ref. 4 opened at both sides, some nylon rope and some metal wire. Putting this all together as shown in Figure 6, and letting the water flow, we find a strange effect: large sparks periodically jump between the two copper wires at the point where they are nearest to each other, giving out loud bangs. Can you guess what condition for the flow has to be realized for this to work? And what did Opel do to repair the cars they recalled? Challenge 4 s

If we stop the water flowing in a Kelvin generator just before the next spark is due, we find that both buckets are able to attract sawdust and pieces of paper. The generator thus does the same that rubbing amber does, just with more bang for the buck(et). Both buckets, and the attached metal pieces, are thus surrounded by electric fields. The fields increase with time, until the spark jumps. Just after the spark, the buckets are (almost) without surrounding electric field. Obviously, the flow of water somehow collects something on each bucket; today we call this *electric charge*. We also say that such bodies are *electrically charged*. This and other experiments also show that charge can *flow* in metals. When the electric fields are high enough, charge can also flow through air, leading to sparks or lightning.

We also find that the two buckets are always surrounded by *two different types* of electric fields: bodies that are attracted by one bucket are repelled by the other. The universal genius Charles Dufay (b. 1698 Paris, d. 1739 Paris) discovered:

---

as several sects believed, but also (falsely) maintained that the Earth was much younger than geologists and Darwin (correctly) had deduced. He strongly influenced the development of the theory of magnetism and electricity, the description of the aether, and thermodynamics. He propagated the use of the term ‘energy’ as it is used today, instead of the confusing older terms. He was one of the last scientists to propagate mechanical analogies for the explanation of phenomena, and thus strongly opposed Maxwell’s description of electromagnetism. It was mainly for this reason that he did not receive a Nobel Prize. He was also one of the minds behind the laying of the first transatlantic telegraphic cable. Victorian and religious to his bones, when he was knighted, he chose the name of a small brook near his home as his new name; thus he became Baron Kelvin of Largs. Therefore the unit of temperature obtained its name from a small Scottish river.

▷ There are *two different types* of electric charge.

In a long and careful series of experiments he confirmed that *all* materials he could get hold of can be charged electrically, and that all charges can be classified into two types. He was the first to show: Ref. 5

▷ Bodies of the *same* charge *repel* each other, and bodies of *different* charge *attract* each other.

Dufay showed in detail that all experiments on electricity can be explained with these statements. Dufay called the two types of charges 'vitreous' and 'resinous'. Unfortunately, Dufay died at a young age. Nevertheless, his results spread quickly. A few years later, Georg Bose used them to develop the first electrifying machine, which then made the exploration of sparks and the science of electricity fashionable across Europe.*

Twenty years after Dufay, in the 1750s, the politician and part-time physicist Benjamin Franklin (b. 1706 Boston, d. 1790 Philadelphia) proposed to call the electricity created on a glass rod rubbed with a dry cloth *positive* instead of vitreous, and that on a piece of amber *negative* instead of resinous. Thus, instead of two types of electric charge, he proposed that

▷ There is really only *one* type of charge.

Bodies can either have too much or too little of it. With the new terms, bodies with charges of the same sign repel each other, bodies with opposite charges attract each other; charges of opposite sign flowing together cancel each other out. Large absolute values of charge imply large charge effects. It then took over a hundred years for these concepts to be unanimously accepted.

In summary, *electric effects are due to the flow of charges.* Now, all flows take time. How fast is electricity? A simple way to measure the speed of electricity is to produce a small spark at one end of a long metal wire, and to observe how long it takes until the spark appears at the other end of the wire. In practice, the two sparks are almost simultaneous; the speed one measures is much higher than everything else we observe in our environment. How would you measure the speed? And why did different researchers get very different speed values in this experiment? Challenge 5 s The result of these experiments is that the speed of electricity is often a large percentage of the speed of light – though never faster than it.

Sparks, electric arcs and lightning are similar. Are they flows of charge? In 1752, experiments performed in France, following a suggestion by Benjamin Franklin, published in London in 1751, showed that one can indeed draw electricity from a thunderstorm via a long rod.** Thunderstorm clouds are surrounded by electric fields. These French exper-

---

* In fact, the fashion still goes on. Today, there are many additional ways to produces sparks or even arcs, i.e., sustained sparks. There is a sizeable subculture of people who build such high voltage generators as a hobby at home; see, for example, the website www.kronjaeger.com/hv. There is also a sizeable subculture of people who do this professionally, paid by tax money: the people who build particle accelerators.

** The details of how lightning is generated and how it propagates are still a topic of research. An introduction is given on page 218.

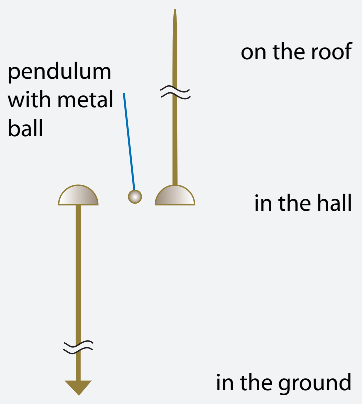


FIGURE 7 Franklin's personal lightning rod – a copy of Gordon's electric chime – is one of the many experiments that shows strikingly that charge can flow.

iments made Franklin famous worldwide; they were also the start of the use of lightning Ref. 6 rod all over the world. Later, Franklin had a lightning rod built through his own house, but of a somewhat unusual type, as shown in Figure 7. This device, invented by Andrew Gordon, is called an *electric chime*. Can you guess what it did in his hall during bad Challenge 6 s weather, all parts being made of metal, and why? (Do not repeat this experiment; any device attached to a lightning rod can kill.)

In summary, *electric fields start at bodies* – provided they are charged. Charging can be achieved by rubbing and other processes. There are two charge signs, negative and positive. *Charge can flow:* it is then called an electric *current*. The worst conductors of current are polymers; they are called *insulators* or *dielectrics*. A charge put on an insulator remains at the place where it was put. In contrast, metals are good conductors; a charge placed on a conductor spreads all over its surface. The best conductors are silver and copper. This is the reason that at present, after two hundred years of use of electricity, the highest concentration of copper in the world is below the surface of Manhattan. Also air is usually an insulator. However, charges can flow through air if the electric field is strong enough; this produces a spark or, when the spark is large, a lightning bolt.

ELECTRIC CHARGE

Because all experiments with electric charge can be explained by calling the two charges positive and negative, we deduce that some bodies have more, and some less charge than an uncharged, *neutral* body. Electric charges thus only flow when two differently charged bodies are brought into contact. Now, if charge can flow and accumulate, we must be able to somehow measure its amount. Obviously, the *amount* of electric charge on a body, usually abbreviated q, must be defined via the influence the body, say a piece of sawdust, feels when subjected to a field. Charge is thus defined by comparing it to a standard reference charge. For a charged body of mass m accelerated in a field, its charge q is determined by the relation

$$\frac{q}{q_{\text{ref}}} = \frac{\mathrm{d}\mathbf{p}/\mathrm{d}t}{\mathrm{d}\mathbf{p}_{\text{ref}}/\mathrm{d}t} , \qquad (1)$$

FIGURE 8 A simple set-up to confirm electric charge conservation: if rubbed fur is moved from the first pot to the second, the charge taken away from the first pot is transferred to the second, as shown by the two electrometers (© Wolfgang Rueckner).

TABLE 1 Properties of *classical* electric charge: a scalar density.

Electric charges	Physical property	Mathematical name	Definition
Can be distinguished	distinguishability	element of set	Page 285
Can be ordered	sequence	order	Vol. IV, page 224
Can be compared	measurability	metricity	Vol. IV, page 236
Can change gradually	continuity	completeness	Vol. V, page 364
Can be added	accumulability	additivity	Vol. I, page 81
Can be separated	separability	positive or negative	
Have no orientation	scalar	number	Page 291
Do not change	conservation	invariance	$q = \text{const}$

i.e., by comparing its momentum change with the momentum change of the reference charge. Charge thus determines the motion of bodies in electric fields in the same way that mass determines the motion of bodies in gravitational fields. Charge is therefore the second intrinsic property of bodies, after mass, that we discover in our walk.

In practice, electric charge is measured with electrometers. A few such devices are shown in Figure 9. The main experimental properties of electric charge that are discovered when experimenting with electrometers are listed in Table 1.

The unit of charge, the *coulomb*, is defined through a standard flow through metal wires, as explained in Appendix A. This is possible because all experiments show

▷ Charge is *conserved*, *flows* and can *accumulate*.

In other words, if the electric charge of a physical system changes, the reason always is that charge is flowing into or out of the system. This can be checked easily with two metal pots connected to two electrometers, as shown in Figure 8. *Charge thus behaves like a fluid substance.* Therefore we are forced to use for its description a scalar quantity

Ref. 7

FIGURE 9 Various electrometers: a self-made electrometer based on a jam pot, an ancient (opened) high precision Dolezalek electrometer, the *Ampullae of Lorenzini* of a shark, and a modern digital electrometer (© Harald Chmela, Klaus Jost at www.jostimages.com, Advantest).

q, which can take positive, vanishing, or negative values on a physical body.

Describing charge as a scalar quantity reproduces the behaviour of electrical charge in all everyday situations. However, as in the case of all previously encountered classical concepts, some of the experimental results for electrical charge in everyday situations from Table 1 will turn out to be only approximate. More precise experiments will require a revision of the idea of continuous change of charge value. Nevertheless, no counter-example to charge conservation has as yet been observed.

In summary, *electric charge is a scalar quantity that describes the origin of electric fields. Electric charge is conserved.* There is no way to destroy or create electric charge. We mentioned above that objects without electric charge are called *neutral*. Also neutral bodies are influenced by electric fields. This happens because a charged object that is brought near a neutral body polarizes it. *Electrical polarization* is the separation of the positive and negative charges onto different regions of a body. For this reason, neutral objects, such as hair or a water stream, are usually attracted to a charged body, such as a rubbed comb. Both insulators and conductors can be polarized; and polarization occurs for single molecules, everyday bodies and whole stars.

TABLE 2 Values of electrical charge observed in nature.

Observation	Charge
Smallest measured non-vanishing charge	$1.6 \cdot 10^{-19}$ C
Charge per bit in computer memory	down to 10^{-15} C
Charge in small capacitor	10^{-7} C
Charge flow in average lightning stroke	1 C to 100 C
Charge stored in a fully charged car battery	0.2 MC
Charge of planet Earth	−1 MC
Charge separated by modern power station in one year	$3 \cdot 10^{11}$ C
Total charge of positive (or negative) sign observed in universe	$10^{60\pm1}$ C
Total charge observed in universe	0 C

ELECTRIC FIELD STRENGTH

Charges produce attraction and repulsion on other charges. Equivalently, charges change momenta; charges exert forces on other charges. This happens over large distances. Experiments that explore energy and momentum conservation show that the best description of these interactions is as told so far: a charge produces a field, the field then acts on a second charge.

Experiments such as those of Figure 4 show:

▷ The *electric field* forms lines in space.

As a consequence, the electric field behaves like a small arrow fixed at each point $\boldsymbol{x}$ in space. Electric fields are described by a direction and a magnitude. The local direction of the field is given by the local direction of the field line – the tangent of the field line. The local magnitude of the field is given by the local density of the field lines. The direction and the magnitude do not depend on the observer. In short

▷ The *electric field* $\boldsymbol{E}(\boldsymbol{x})$ is a *vector* field.

Experiments show that it is best defined by the relation

$$q\boldsymbol{E}(\boldsymbol{x}) = \frac{\mathrm{d}\boldsymbol{p}(\boldsymbol{x})}{\mathrm{d}t} \qquad (2)$$

taken at every point in space $\boldsymbol{x}$. The definition of the electric field is thus based on how it *moves* charges. In general, the electric field is a vector

$$\boldsymbol{E}(\boldsymbol{x}) = (E_x, E_y, E_z) \qquad (3)$$

Challenge 7 e

and is measured in multiples of the unit N/C or V/m.

The definition of the electric field assumes that the test charge q is so small that it does not disturb the field E. We sweep this issue under the carpet for the time being. This is

TABLE 3 Some observed electric fields.

Observation	Electric field
Field 1 m away from an electron in vacuum	Challenge 9 s
Field values sensed by sharks	down to 0.5 µV/m
Cosmic noise	10 µV/m
Field of a 100 W FM radio transmitter at 100 km distance	0.5 mV/m
Field inside conductors, such as copper wire	0.1 V/m
Field just beneath a high power line	0.1 to 1 V/m
Field of a GSM antenna at 90 m	0.5 V/m
Field inside a typical home	1 to 10 V/m
Field of a 100 W bulb at 1 m distance	50 V/m
Ground field in Earth's atmosphere	100 to 300 V/m
Field inside thunder clouds	up to over 100 kV/m
Maximum electric field in air before sparks appear	1 to 3 MV/m
Electric fields in biological membranes	10 MV/m
Electric fields inside capacitors	up to 1 GV/m
Electric fields in petawatt laser pulses	10 TV/m
Electric fields in U^{91+} ions, at nucleus	1 EV/m
Maximum practical electric field in vacuum, limited by electron pair production	1.3 EV/m
Maximum possible electric field in nature (corrected Planck electric field $c^4/4Ge$)	$1.9 \cdot 10^{62}$ V/m

a drastic move: we ignore quantum theory and all quantum effects in this way; we come back to it below. Page 248

The definition of the electric field also assumes that space-time is flat, and it ignores all issues due to space-time curvature.

By the way, does the definition of electric field just given assume a charge speed that is far less than that of light? Challenge 8 s

To describe the motion due to electricity completely, we need a relation explaining how charges *produce* electric fields. This relation was established with precision (but not for the first time) during the French Revolution by Charles-Augustin de Coulomb, on his private estate.* He found that around any small-sized or any spherical charge Q *at rest* there is an electric field. At a position $\boldsymbol{r}$, this electric field $\boldsymbol{E}$ is given by

$$\boldsymbol{E}(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}\frac{\boldsymbol{r}}{r} \quad \text{where} \quad \frac{1}{4\pi\varepsilon_0} = 9.0\,\mathrm{GV\,m/C}\,. \qquad (4)$$

Later we will extend the relation for a charge in motion. The bizarre proportionality constant is universally valid. The constant is defined with the so-called *permittivity of free*

---

* Charles-Augustin de Coulomb (b. 1736 Angoulême, d. 1806 Paris), engineer and physicist, provided, with his careful experiments on electric charges, a firm basis for the study of electricity.

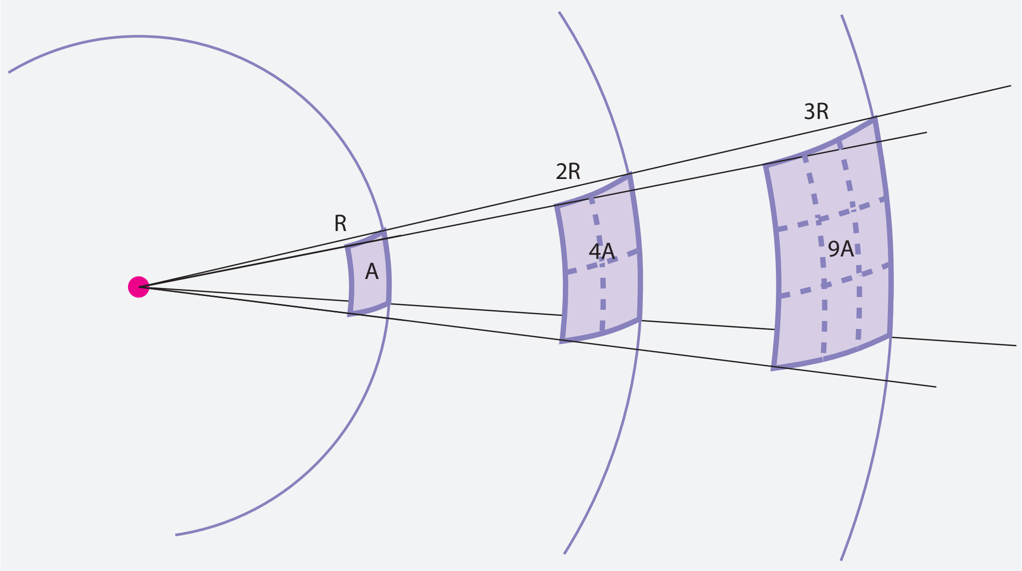


FIGURE 10 A visualization of Coulomb's formula and Gauss' law.

space ε_0 and is due to the historical way the unit of charge was defined first.* The essential point of the formula is the decrease of the field with the square of the distance; can you imagine the origin of this dependence? A simple way to picture Coulomb's formula is illustrated in Figure 10. Challenge 10 s

The two previous equations allow us to write the interaction between two charged bodies as

$$\frac{\mathrm{d}\boldsymbol{p}_1}{\mathrm{d}t} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}\frac{\boldsymbol{r}}{r} = -\frac{\mathrm{d}\boldsymbol{p}_2}{\mathrm{d}t}\,, \qquad (5)$$

where $\mathrm{d}\boldsymbol{p}$ is the momentum change and $\boldsymbol{r}$ is the vector connecting the two centres of mass. This is the famous expression for electrostatic attraction and repulsion. It also due to Coulomb. The relation is valid only for charged bodies that are either of *small size* or *spherical*, and most of all, only for bodies that are *at rest* with respect to each other and to the observer. The exploration of interactions among charges at rest is called *electrostatics*.

Electric fields accelerate charges. As a result, in everyday life, electric fields have two main properties: they contain energy and they can polarize bodies. The energy content is due to the electrostatic interaction between charges. The strength of this interaction is considerable. For example, it is the basis for the force of our muscles. Muscular force is a macroscopic effect of Coulomb's relation (5). Another example is the material strength of steel or diamond. As we will discover, all atoms are held together by electrostatic attraction. To convince yourself of the strength of electrostatic attraction, answer the fol-

---

* Other definitions of this and other proportionality constants to be encountered later are possible, leading to *unit systems* different from the SI system used here. The SI system is presented in detail in Appendix A. Among the older competitors, the Gaussian unit system often used in theoretical calculations, the Heaviside–Lorentz unit system, the electrostatic unit system and the electromagnetic unit system are the most important ones. Ref. 8

TABLE 4 Properties of the classical electric field: a (polar) vector at every point in space.

ELECTRIC FIELDS CAN	PHYSICAL PROPERTY	MATHEMATICAL NAME	DEFINITION
Attract bodies	accelerate charges	coupling	equation (4)
Repel bodies	accelerate charges	coupling	equation (4)
Be distinguished	distinguishability	element of set	Page 285
Change gradually	continuum	real vector space	Vol. I, page 80, Vol. V, page 364
Point somewhere	direction	vector space, dimensionality	Vol. I, page 80
Be compared	measurability	metricity	Vol. IV, page 236
Be added	additivity	vector space	Vol. I, page 80
Have defined angles	direction	Euclidean vector space	Vol. I, page 81
Exceed any limit	infinity	unboundedness	Page 286
Change direction under reflection	polarity	parity-odd vector	Page 90
Keep direction under time reversal	polarity	time-even vector	Page 90

lowing: What is the force between two boxes with a gram of protons each, located on the two poles of the Earth? Try to guess the result before you calculate the astonishing value. Challenge 11 s

The electric attraction is thus much stronger than the gravitational attraction. What is the ratio between the two? Challenge 12 e

Coulomb's relation for the field around a charge can be rephrased in a way that helps to generalize it to non-spherical bodies. Take a closed surface, i.e., a surface than encloses a certain volume. Then the integral of the electric field over this surface A, the electric flux, is the enclosed charge Q divided by ε_0:

$$\oint_{\text{closed surface}A} \boldsymbol{E}\, \mathrm{d}\boldsymbol{A} = \frac{Q}{\varepsilon_0} \,. \tag{6}$$

This mathematical relation, called *Gauss's 'law'*,* is equivalent the result of Coulomb.

---

* Carl-Friedrich Gauß (b. 1777 Braunschweig, d. 1855 Göttingen) was, together with the Leonhard Euler, the most important mathematician of all times. A famous child prodigy, when he was 19 years old, he constructed the regular heptadecagon with compass and ruler (see www.mathworld.wolfram.com/Heptadecagon.html). He was so proud of this result that he put a drawing of the figure on his tomb. Gauss produced many results in number theory, topology, statistics, algebra, complex numbers and differential geometry which are part of modern mathematics and bear his name. Among his many accomplishments, he produced a theory of curvature and developed non-Euclidean geometry. He also worked on electromagnetism and astronomy.

Gauss was a difficult character, worked always for himself, and did not found a school. He published little, as his motto was: pauca sed matura. As a consequence, when another mathematician published a new result, he regularly produced a notebook in which he had noted the very same result already years before. These notebooks are now available online, at www.sub.uni-goettingen.de.

FIGURE 11 Various types of charge pumps: a bicycle dynamo, an alternator in a power station, a Wimshurst machine, an electric eel, a voltaic cell, a leaf and a solar cell (© Wikimedia, Q-Cells).

Challenge 13 s

(Note that in the simplified form given here, it is valid only for static situations.) Since inside conductors the electrical field is zero, Gauss's relation implies, for example, that if a charge q is surrounded by an uncharged metal sphere, the *outer* surface of the metal sphere shows the same charge q.

Challenge 14 e

Vol. V, page 122

Challenge 15 s

Do uncharged, neutral bodies attract one other? In first approximation they do not. But when the question is investigated more precisely, we will find that they can attract one other. Can you find the conditions for this to happen? In fact, the conditions are quite important, as our own bodies, which are made of neutral molecules, are held together in this way.

Pumping charge

Owing to the high strength of electromagnetic interactions, separating charges is not an easy task. This is the reason that electrical effects have only been commonly used for about a hundred years. Humanity had to wait for practical and efficient devices to be invented for separating charges and putting them into motion: to use electric effects, we need *charge pumps*. Some devices are shown in Figure 11. Can you explain whether batteries or any other of these devices are sources of charges? Challenge 16 s

Of course, every charge pump requires energy. Batteries in mobile phones and the ion channels in living cells use chemical energy to do the trick. Thermoelectric elements, as used in some watches, use the temperature difference between the wrist and the air to separate charges; solar cells use light, piezoelectric elements use stress and dynamos or Kelvin generators use kinetic energy.

What is electricity?

The term *electricity* is also used as the name for a field of inquiry. Usually, the term is used to refer to electric current. In general, the term is used to refer to the effects of electric charges, of their motion and their fields.

In fact the vocabulary issue hides a deeper question: what is the nature of electric charge? In order to solve this extremely difficult issue, we start with the following question.

Can we detect the inertia of electricity?

If electric charge really is something *flowing* through metals, we should be able to observe the effects shown in Figure 12: electric charge should fall, should have inertia and should be separable from matter. And indeed, each of these effects has been observed. For example, when a long metal rod is kept vertically, we can measure an electrical potential difference, a *voltage*, between the top and the bottom. In other words, we can measure the *weight* of electricity in this way. Similarly, we can measure the potential difference between the ends of an accelerated rod. Ref. 9 Alternatively, we can measure the potential difference between the centre and the rim of a rotating metal disc. The last experiment was, in fact, the way in which the ratio q/m for currents in metals was first measured with precision. The value for the inertia of electricity is

$$q/m \approx -1.8(2) \cdot 10^{11}\ \mathrm{C/kg} \qquad (7)$$

for all metals, with small variations in the second digit. The minus sign is due to the definition of charge. In short, electrical charge in metals has mass, though a very small one.

If electric charge has mass, whenever we switch on an electrical current, we get a *recoil*. Ref. 10 This simple effect can easily be measured and confirms the mass to charge ratio just given. Also, the emission of current into air or into vacuum is observed; in fact, every cathode ray tube inside an old television used this principle to generate the beam producing the picture. Ref. 11 The emission works best for metal objects with sharp, pointed tips. The rays created this way – we could say that they are ‘free’ electricity – are called *cathode rays*. Within a few per cent, they show the same mass to charge ratio as expression (7). This correspondence thus shows that charges move almost as freely in metals as in air; this is the reason that metals are such good conductors of electric current.

If electric charge *falls* inside vertical metal rods, we can make the astonishing deduction that cathode rays should not be able to fall through a vertical metal tube. As we

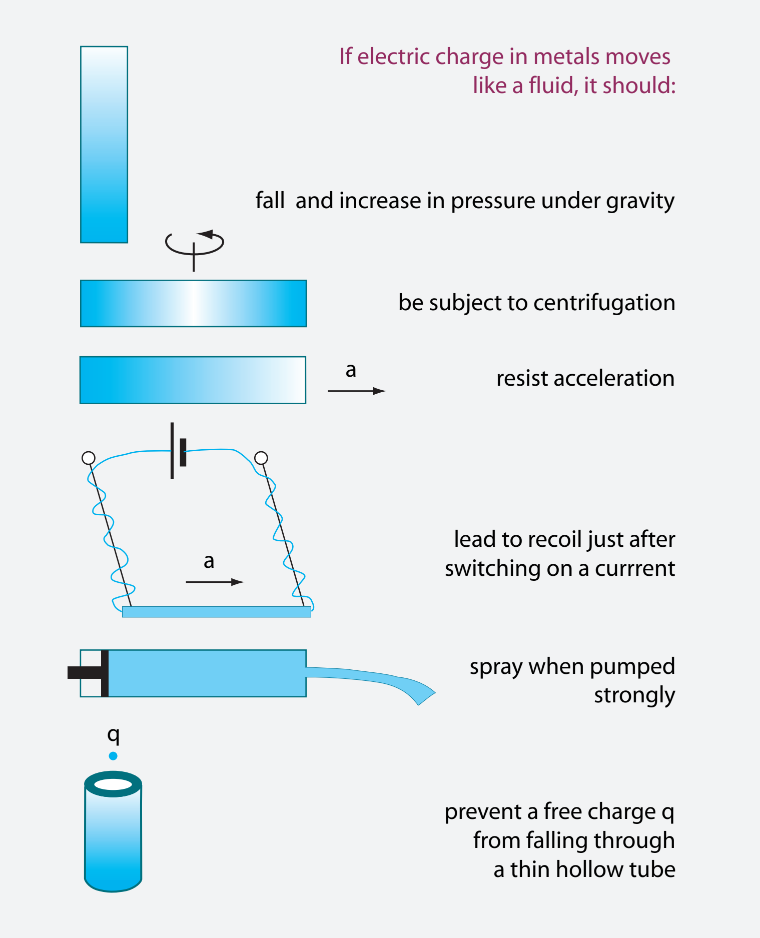


FIGURE 12 Consequences of the flow of electricity, as discussed in the text.

will see later, cathode rays consist of free electrons. The name 'electron' is due to George Stoney. Electrons are the smallest and lightest charges moving in metals; they are, usually – but not always – the 'atoms' of electricity. In particular, electrons conduct electric current in metals. The charge of an electron is small, 0.16 aC, so that flows of charge typical of everyday life consist of huge numbers of electrons; as a result, electrical charge effectively behaves like a continuous fluid. The particle itself was discovered and presented in 1897 by Johann Emil Wiechert (b. 1861 Tilsit, d. 1928 Göttingen) and, independently, three months later, by Joseph John Thomson (b. 1856 Cheetham Hill, d. 1940 Cambridge).

Cathode rays should not be able to fall through a vertical metal tube because the acceleration by the electrical field generated by the displaced electricity in the metal tube and the gravitational acceleration cancel. Thus electrons should not be able to fall through a long thin cylinder. This would not be the case if electricity in metals did not behave like a fluid. The experiment has indeed been performed, and a reduction of the acceleration of free fall for electrons of 90 % has been observed. Can you imagine why the ideal value of 100 % is not achieved?

Challenge 17 e

Ref. 12

Challenge 18 s

Precision experiments with charges ejected from metals show that they have a charge to mass ratio of

$$q/m = -1.758\,820\,150(44) \cdot 10^{11}\ \mathrm{C/kg} \tag{8}$$

The particles with this property are called *electrons*. Other types of charges, with different charge-to-mass ratio, also exist in nature. Examples are the *ions* found in batteries and leaves, the *muons* found in cosmic rays, and the *mesons* produced in particle accelerators. We will meet these particles later in our adventure.

Since electric current behaves like a liquid, we should be able to measure its speed. The first to do so, in 1834, was Charles Wheatstone. In a famous experiment, he used a wire of a quarter of a mile length to produce three sparks: one at the start, one at the middle, and one at the end. He then mounted a rapidly moving mirror on a mechanical watch. By noting how much the three spark images were shifted against each other on a screen, he determined the speed to be 0.45 Gm/s, though with a large measurement error. Latter, more precise measurements showed that the speed is always below 0.3 Gm/s, and that it depends on the metal and the type of insulation of the wire. The high value of the speed convinced many people to use electricity for transmitting messages. In fact, these experiments measure the *signal speed* of electromagnetic waves carried by metal wires. The actual speed of electric charges is much lower, as shown below. Page 249 A modern version of the signal speed experiment, for computer fans, uses the 'ping' command from the UNIX operating system. Ref. 13 The 'ping' command measures the time for a computer signal to reach another computer and return back. If the cable length between two computers is known, the signal speed can be deduced. Just try. Challenge 19 e

The speed of electricity is *too slow* for many people. Computer chips could be faster if it were higher. And computers that are connected to stock exchanges are located as near as possible to the stock exchange, because the time advantage the short communication distance (including the delay inside switching chips) provides is essential for getting a good financial performance in certain trading markets. Ref. 14

In summary, experiments show that *all charges have mass*. And like all massive bodies, *charges move slower than light*. Charge is a property of matter; images and light have no charge.

FEELING ELECTRIC FIELDS

Why is electricity dangerous to humans? The main reason is that the human body is controlled by 'electric wires' itself. As a result, electricity applied to human bodies from the outside interferes with the internal signals. This has been known since 1789. In that year the medical doctor Luigi Galvani (b. 1737 Bologna, d. 1798 Bologna) discovered that electrical current makes the muscles of a dead animal contract. The famous first experiment used frog legs: when electricity was applied to them, they twitched violently. Subsequent investigations confirmed that all nerves make use of electrical signals. Using electricity, one can make fresh corpses move, for example. Nerves are the 'control wires' of animals. We will explore nerves in more detail below. Page 51

Being electrically controlled, all mammals can sense strong electric fields. Humans can sense fields as low as 10 kV/m, when hair stands on end. In contrast, several animals can sense much weaker electric (and magnetic) fields. This ability is called *electroreception*. Sharks, for example, can detect fields down to 0.5 μV/m using special sensors, the

TABLE 5 Some observed electric current values.

Observation	Current
Smallest current ever measured (for one moving electron)	3 aA
Human nerve signals	20 µA
Lethal current for humans	as low as 20 mA, typically 100 mA
Current drawn by a train engine	600 A
Current in a lightning bolt	10 to 100 kA
Highest current produced by humans	20 MA
Current inside the Earth, at the origin of its magnetic field	*c.* 100 MA
Maximum possible current in nature (corrected Planck electric current $e\sqrt{c^5/4\hbar G}$)	1.5 YA

TABLE 6 Some sensors for electrical current.

Measurement	Sensor	Range
Conventional 20 euro multimeter	voltage drop over resistor	up to *c.* 3 A
Feeling threshold	human nerve	felt from 0.1 mA upwards
Reversible muscle contraction without danger	human nerve	up to 10 mA over long times, or up to 200 mA for at most 10 ms
Rhythm change	human heart	heart stops when about 20 mA flow through it
Strong muscle contraction with some damage	human nerve	up to 100 mA over long times, or up to 1 A for at most 200 ms
Smoke emission, strong burns	human flesh	from 1 A
Fire	trees	from 1 kA
Electric eel *Electrophorus electricus*	built-in	up to 1 A and 500 V

Page 24 *Ampullae of Lorenzini*, which are found around their mouth. Sharks use them to detect the field created by prey moving in water; this allows them to catch their prey even in the dark.

The muscles in living prey generate electric fields. Various water animals have developed electric field sensors to detect prey in water which is too muddy to see through. The salamander is an example, as is the platypus (*Ornithorhyncus anatinus*), the famous duck-billed mammal can also sense electric fields; but they achieve only sensitivities of the order of mV/m. In fact, only few mammals are known to be able to sense small electric fields: apart from the the platypus also the echydnas can sense electric fields with their beaks. In 2011, it was discovered that the Guiana dolphin, *Sotalia guianensis*, can sense fields as low as 0.5 mV/m with organs on their snout. It is conjectured that other dolphins also have the ability.

Numerous fish, the so-called strongly and *weakly-electric fish*, are able to *generate* electric fields in order to achieve even better prey detection.* This approach is used, for example, by the elephantnose fish (*Gnathonemus petersii*). The achieved sensitivity is below 2 mV/m. In fact, various electric fish use *time-varying* electric dipole fields to communicate! They tell each other their species, their sex, their identity, and communicate about courtship, aggression, appeasement and dangers. The frequencies they use are in the range between a few and 200 Hz, and the fields are dipole fields created between the anterior and posterior sections of their bodies.

Ref. 15

Ref. 16

The most fearsome – and the most ugly – electric animal is the electric eel, *Electrophorus electricus*. It can be 2 m long and weigh up to 20 kg. Because electric fields have stronger effects in air than in water, when a prey wades into its territory, the eel often jumps out of the water and against the prey, so that it can kill more easily using its built-in 500 V and 1 A high-voltage, high-current producing organ. It is able to kill horses in this way.

No land animal has special sensors for weak electric fields, because any electric field in air is strongly damped when it encounters a water-filled animal body.** Indeed, the usual atmosphere has a low, vertical electric field of around 100 V/m; inside the human body this field is damped to the μV/m range, which is far less than an animal's internal electric fields. In other words, humans do not have sensors for low electric fields because they are land animals. (Do humans have the ability to sense electric fields in water? Nobody seems to know.) However, there are a few exceptions. You might know that some older people can sense approaching thunderstorms in their joints. This is due the coincidence between the electromagnetic field frequency emitted by thunderclouds – around 100 kHz – and the resonant frequency of nerve cell membranes.

Challenge 20 r

Page 108

The water content of the human body also means that the electric fields in air that are found in nature are rarely dangerous to humans. But whenever humans consciously sense electric fields, such as when high voltage makes their hair stand on end, the situation is potentially dangerous.

The high impedance of air also means that, in the case of time-varying electromagnetic fields, humans are much more prone to be affected by the magnetic component than by the electric component.

Plants also sense and even produce electric fields. Inside many large plants, electrical signals are exchanged, for example, to inform about insect damage. In 2016, researchers finally discovered the molecular mechanism with which plant cells sense electric fields. It was known for a long time that flowers are often negatively charged. In 2013, it was shown that bees are able to sense these fields. Bees are usually positively charged, due to aerodynamic effects. The negative charge of the plants also makes the pollen stick better to the bee.

FIGURE 13 Various types of magnets and effective magnets: the needle in a compass, some horseshoe magnets, two galaxies, the magnetic organ of a dove, the Earth, a lifting magnet, and the Sun. (© Wikimedia, Shambhavi, Anthony Ayiomamitis, NASA).

TABLE 7 Searches for magnetic monopoles, i.e., for magnetic charges, in over 140 experiments.

Search	Magnetic charge
Smallest magnetic charge suggested by quantum theory	$g = \frac{h}{e} = \frac{eZ_0}{2\alpha} = 4.1\,\mathrm{pWb}$
Search in minerals, from mountains to the deep ocean	none, only dipoles Ref. 17
Search in meteorites and moon minerals	none, only dipoles Ref. 17
Search in cosmic rays	none (one false alarm in the 1970s), only dipoles Ref. 17
Search with particle accelerators	none, only dipoles Ref. 17

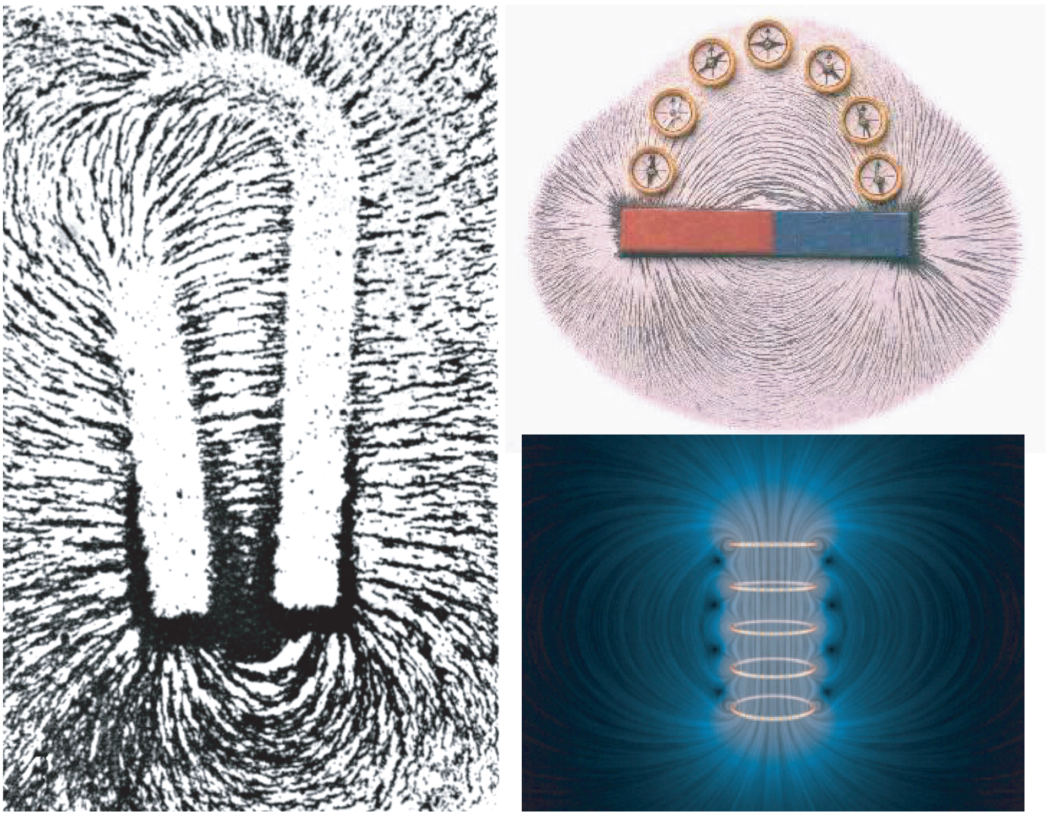


FIGURE 14 Visualizing magnetic fields around magnets and coils – with iron filings, with compass needles and iron filings, and with computer graphics (© Wikimedia, MIT).

MAGNETS AND OTHER MAGNETIC MATERIALS

The study of magnetism progressed across the world independently of the study of electricity. Towards the end of the twelfth century, the compass came into use in Europe. At that time, there still were heated debates on whether it pointed to the north or

---

[*] It took until the year 2000 for technology to make use of the same effect. Nowadays, airbag sensors in cars often use electric fields to sense whether the person sitting in the seat is a child or an adult, thus changing the way that the airbag behaves in an accident.

[**] Though a few land animals that swim a lot under water have electric field sensors.

TABLE 8 Some observed magnetic fields.

Observation	Magnetic field
Lowest measured magnetic field (e.g., fields of the Schumann resonances)	1 fT
Magnetic field produced by brain currents	0.1 pT to 3 pT
Magnetic field produced by single muscle action	1 pT
Intergalactic magnetic fields	1 pT to 10 pT
Magnetic field in the human chest, due to heart currents	100 pT
Magnetic field of our galaxy	0.5 nT
Magnetic field due to solar wind	0.2 to 80 nT
Magnetic field directly below high voltage power line	0.1 to 1 μT
Magnetic field of Earth	20 to 70 μT
Magnetic field inside home with electricity	0.1 to 100 μT
Magnetic field near mobile phone	100 μT
Magnetic field that influences visual image quality in the dark	100 μT
Magnetic field near iron magnet	100 mT
Solar spots	1 T
Magnetic fields near high technology permanent magnet	max 1.3 T
Magnetic fields that produces sense of coldness in humans	5 T or more
Magnetic fields in particle accelerator	10 T
Maximum static magnetic field produced with superconducting coils	22 T
Highest static magnetic fields produced in laboratory, using hybrid magnets	45 T
Highest *pulsed* magnetic fields produced without coil destruction	76 T
Pulsed magnetic fields produced, lasting about 1 μs, using imploding coils	1000 T
Field of white dwarf	10^4 T
Fields in petawatt laser pulses	30 kT
Field of neutron star	from 10^6 T to 10^{11} T
Quantum critical magnetic field	4.4 GT
Highest field ever measured, on magnetar and soft gamma repeater SGR-1806-20	0.8 to $1 \cdot 10^{11}$ T
Estimated magnetic field near atomic nucleus	1 TT
Maximum possible magnetic field in nature (corrected Planck magnetic field $c^3/4Ge$)	$6.3 \cdot 10^{53}$ T

the south. Then, in 1269, the military engineer Pierre de Maricourt (b. 1219 Maricourt, d. 1292 unknown) published his study of magnetic materials. He found that every magnet has *two* points of highest magnetization, and he called them *poles*. He found that even after a magnet is cut, the resulting pieces always retain two poles: when the stone is left free to rotate, one points to the north and the other to the south. Ref. 18

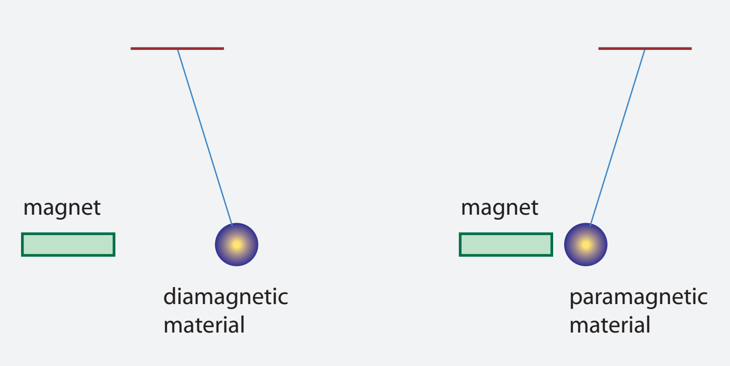


FIGURE 15 The two basic types of magnetic material behaviour (tested in an *inhomogeneous* field): diamagnetism and paramagnetism.

▷ All magnets are dipoles.

The two poles are called the *north pole* and the *south pole*. Maricourt also found that

▷ Like poles repel, and unlike poles attract.

As a consequence, the magnetic north pole of the Earth is the one near the south pole, and vice versa.

Magnets are surrounded by magnetic fields. Magnetic fields, like electric fields, can be visualized with field lines. Figure 14 shows some ways to do this. We directly note the main difference between magnetic and electric field lines: magnetic field lines have no beginning and no ends, whereas electric field lines do. (However, magnetic field lines are usually not closed; this only happens in very special cases.) The direction of the field lines gives the direction of the magnetic field, and the density of the lines gives the magnitude of the field.

Many systems in nature are magnets, as shown in Figure 13. The existence of two magnetic poles is valid for all magnets in nature: molecules, atoms and elementary particles are either dipoles or non-magnetic.

▷ There are no magnetic monopoles.

Magnetic field lines could start or end at a magnetic monopole – if one existed. Despite the promise of eternal fame, no magnetic monopole has ever been found. The searches are summarized in Table 7.

Magnets have a second important property, shown in Figure 15: magnets, through their magnetic field, transform non-magnetic materials into magnetic ones. There is thus a *magnetic polarization*, similar to the electric polarization. The amount of polarization depends on the material; some values are given in Table 9.

— Certain materials, the so-called *diamagnetic materials*, are *repelled* by magnets, though usually only by weak forces.
— Others, the so-called *paramagnetic* materials, are *attracted* to magnets.
— Some important materials, the *ferromagnetic materials*, such as steel, *retain* the in-

TABLE 9 The magnetic properties of materials – for static fields at room temperature.

Material	Relative magnetic permeability μ_r
Diamagnetic materials $\mu_r < 1$, repelled by magnets	
Type I superconductors	0
Highly oriented pyrolitic graphite	0.999 55
Bismuth	0.999 83
Graphite	0.999 84
Gold	0.999 966
Copper	0.999 9936
Water	0.999 9912
Usual animals and plants	like water
Paramagnetic materials $\mu_r > 1$, attracted by magnets	
Air, oxygen	1.000 0019
Biomagnetic particles in living organisms	1.000 006
Aluminium	1.000 022
Platinum	1.000 26
Ferromagnetic materials $\mu_r \gg 1$, able to form magnets	
SmCo	*c.* 1.04
NdFeB	*c.* 1.15
Cobalt	80 to 200
Nickel	100
Iron	300 to 10 000
Permalloy	*c.* 8 000
Ferrites	up to 15 000
μ-metal	up to 140 000
Amorphous metals	up to 500 000

duced magnetic polarization: they become permanently magnetized. This happens when the atoms in the material get aligned by an external magnet. Ferromagnetic materials are used to produce permanent *magnets* – thus artificial lodestone.

Magnetic materials are essential for the industrial production of electric current and are part of most devices that use electricity.

How do animals feel magnetic fields?

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

TABLE 10 The dielectric properties of materials – for static fields at room temperature.

Material	Relative electric permittivity ε_r
Dielectric materials	
Vacuum	1
Air	1.0006
Teflon	2.1
Graphite	10 to 15
Silicon dioxide	3.9
Silicon	11.7
Methanol	30
Water	80.1
Titanium dioxide	86-173
Paraelectric materials	
Strontium titanate (a perovskite)	310
Barium strontium titanate (a perovskite)	500
Ferroelectric materials $\varepsilon_r \gg 1$, able to form electrets	
Lithium niobate (below 1430 K)	...
Barium titanate	1 250 to 10 000
Ferroelectric polymers	up to 100 000
Calcium copper titanate	over 250 000

Note: the values of the electric permittivity depend on the frequency of the applied field and on the temperature. The values given here are only for static electric fields at room temperature. Values for higher frequencies or other temperatures show strong variations. Page 74

Ref. 19

It is known that honey bees, sharks, pigeons, the sandhill crane, various other birds, salmon, trout, sea turtles, dolphins and certain bacteria can feel magnetic fields. One speaks of the ability for *magnetoreception*. All these life forms use this ability for navigation. The most common detection method is the use of small magnetic particles inside a cell; the cell then senses how these small built-in magnets move in a magnetic field. The magnets are tiny, typically around 50 nm in size. These small magnets are used to navigate along the magnetic field of the Earth. For higher animals, the variations of the magnetic field of the Earth, 20 to 70 µT, produce a landscape that is similar to the visible landscape for humans. They can remember it and use it for navigation.

In fact, migrating birds like the sandhill crane (*Grus canadensis*) seem to have *two* ways to sense magnetic fields. First of all, they have small iron crystals located inside neurons that provide a magnetic map that is used for local navigation. For a long time, it was thought that these neurons were located in the skin above the beak. In recent

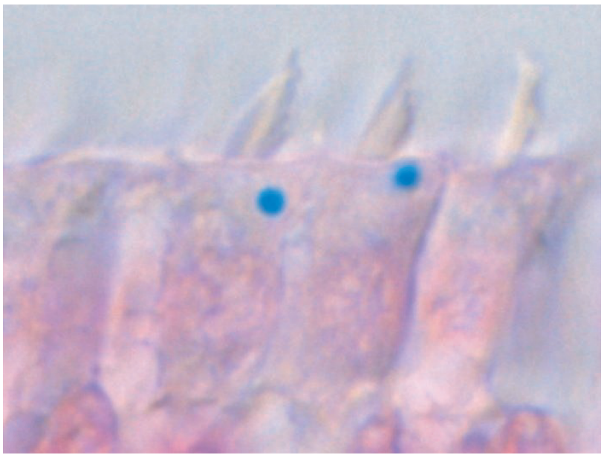


FIGURE 16 Stained cells from the inner ear of pigeons; the used chemical gives iron particles a blue colour. The magnetic particles, one in each cell, lie just beneath the hairs (© Institute of Molecular Pathology, Vienna).

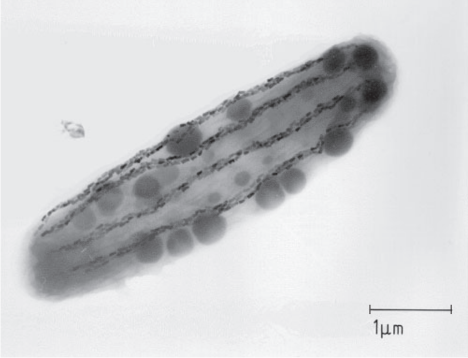


FIGURE 17 The magnetotactic bacterium *Magnetobacterium bavaricum* with its magnetosomes (© Marianne Hanzlik).

years, it finally appeared that this often-cited 'fact' was a collective mistake; the true magnetic sensor particles are probably located in the neurons inside the ears of the birds, Ref. 20 just below the hairs, as shown in Figure 16. The second magnetic sense of migrating birds is an inclination compass that tell them the angle between the magnetic field lines and the vertical. This system is based on magnetically sensitive protein molecules, so-called *cryptochromes*. The mechanism is located in the eye and is based on blue light. This second magnetic sense, which is still not properly understood, is used by birds to Ref. 21 decide the general direction in which to fly.

Can humans feel static magnetic fields? So far, there is no definite answer. Magnetic microcrystals are present in the human brain, but whether humans can feel magnetic Challenge 21 r fields is still an open issue. Maybe you can devise a way to test the this possibility?

In contrast, oscillating or pulsed magnetic fields can be felt by humans. There is anecdotal evidence that 0.2 T oscillating at 170 kHz leads to numbness in fingers for a few days. Beneficial effects of pulsed fields on well-being are also claimed, but are questionable; on the other hand, oscillating magnetic fields have positive effect on bone fracture healing.

Magnetism and electricity

Are magnetism and electricity related? In the early 19th century, François Arago* discovered that they were. He explored a ship that had survived a bad thunderstorm. At that time, ships where made of wood. The ship had been struck by lightning; as a result, the ship needed a new compass. Thus lightning has the ability to demagnetize compasses. Arago knew that lightning is an electrical phenomenon. He concluded that magnetism and electricity must be related.

In short, magnetism must be related to the *motion* of electric charges. If magnetism is related to motion, it must be possible to use magnetism and electricity to move matter.

How can one make a motor?

> Communism is the power of the local councils plus electricification of the whole country.
>
> Lenin.**

The reason for Lenin's famous statement were two discoveries. One was made in 1820 by Hans Christian Oersted*** and the other in 1831 by Michael Faraday.**** The consequences of these experiments changed the world completely in less than one century.

On the 21st of July of 1821, Hans Christian Oersted published a leaflet, in Latin, which took Europe by storm. Oersted had found – during a lecture demonstration to his students – that when a current is sent through a wire, a nearby magnet is put into motion. In other words, he found

▷ The flow of electricity can *move* bodies.

Due to Oersted's leaflet, everybody in Europe with a bit of dexterity started to experiment with electricity. Further experiments show that *two* wires in which charges flow attract or repel each other, depending on whether the currents are parallel or antiparallel. These and other experiments show that

▷ Wires that carry an electric current behave like magnets.

---

* François Arago (b. 1786 Estagel, d. 1853 Paris) was physicist and politician; he was a friend of Alexander von Humboldt.

** Lenin (b. 1870 Simbirsk, d. 1924 Gorki), founder of the Union of Soviet Socialist Republics, in 1920 stated this as the centre of his development plan for the country. In Russian, the local councils of that time were called soviets.

*** Hans Christian Oersted (b. 1777 Rudkøbing, d. 1851 Copenhagen) physicist and professor, founded the school that later became the Technical University Denmark.

**** Michael Faraday (b. 1791 Newington Butts, d. 1867 London) was born to a simple family, without schooling, and of deep and naive religious ideas. As a boy he became assistant to the most famous chemist of his time, Humphry Davy (b. 1778 Penzance, d. 1829 Geneva). Faraday had no mathematical training, but became an influential physicist and late in his life he even became member of the Royal Society. A modest man, he refused all other honours in his life. He worked on chemical topics, the atomic structure of matter and, most of all, he developed the idea of (magnetic) fields and field lines. He used fields to describe all his numerous experimental discoveries about electromagnetism, such as the Faraday effect. Fields were later described mathematically by Maxwell, who at that time was the only person in Britain to take over Faraday's field concept.

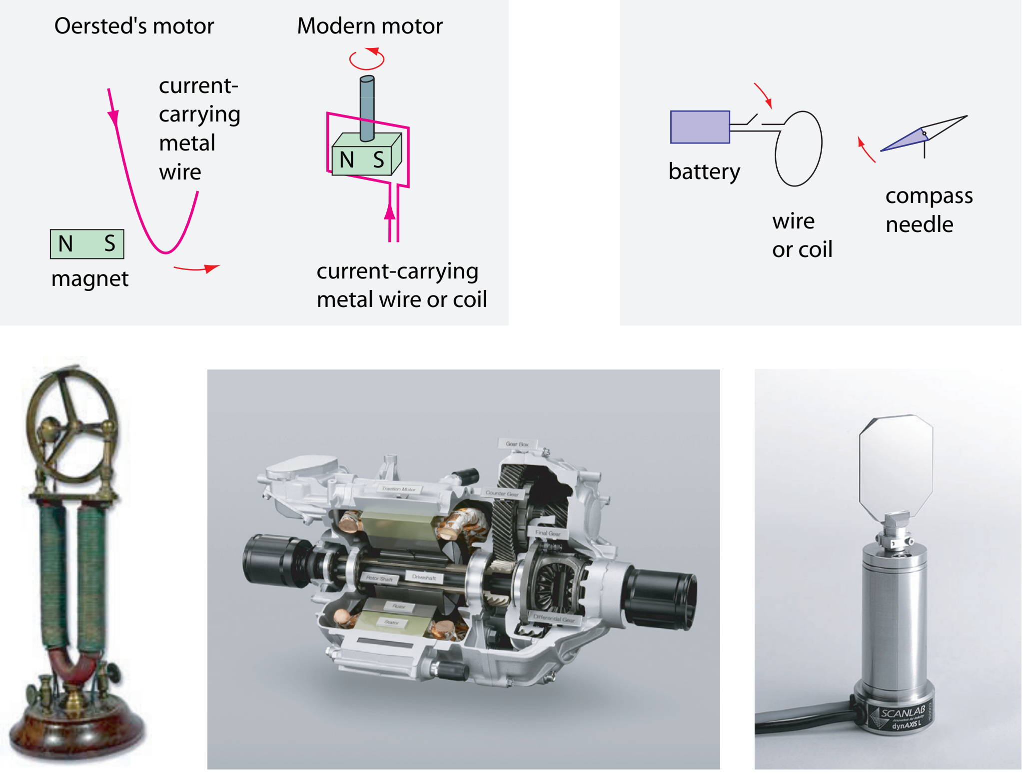


FIGURE 18 An old and a modern version of electric motor, and a mirror galvanometer with limited rotation range used for steering laser beams. Sizes are approximately 20 cm, 50 cm and 15 cm (© Wikimedia, Honda, Wikimedia).

In fact, the opposite is also true: if we imagine tiny currents moving in circles inside magnets, we get a unique description for all magnetic fields observed in nature. In other words, Oersted had found the definite proof that electricity can be turned into magnetism.

Shortly afterwards, Ampère* found that *coils* increase these effects dramatically compared to wires.

▷ Coils behave like small magnets.

---

* André-Marie Ampère (b. 1775 Lyon, d. 1836 Marseille), physicist and mathematician. Autodidact, he read the famous *Encyclopédie* as a child; in a life full of personal tragedies, he wandered from maths to chemistry and physics, worked as a school teacher, and published nothing of importance until 1820. Then the discovery of Oersted reached all over Europe: electrical current can deviate magnetic needles. Ampère worked for years on the problem, and in 1826 published the summary of his findings, which lead Maxwell to call him the 'Newton of electricity'. Ampère named and developed many areas of electrodynamics. In 1832, he and his technician also built the first *dynamo*, or rotative current generator. Of course, the unit of electrical current is named after him.

Ampère had two cats, which he liked dearly, a large one and a small one. When he was doing his experiments in his laboratory, they wanted to come in, and when they were in, they soon wanted to go out. One day he was fed up. He made two holes in his door, a large one and a small one.

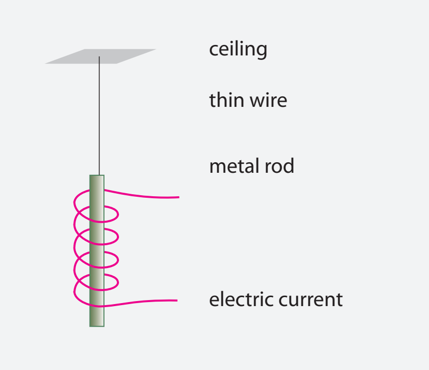


FIGURE 19 Current makes a metal rod rotate.

In particular, current-carrying coils, like magnets, always have two poles, usually called the north and the south pole. Opposite poles attract, like poles repel each other. Ampère was so proud of his discovery that he invented a special name for electrically conducting coils; he called them *solenoids*.

As is well known, the Earth is itself a large magnet, with its magnetic north pole near the geographic south pole, and vice versa. Every compass shows this. However, the magnetic field of the Earth is *not* due to a solid permanent magnet inside it. The Earth's solid core, at 6 ± 1 kK, is too hot to be a permanent magnet; instead, the magnetic field is due to circulating currents in the outer, liquid core. The Earth is thus more similar to a solenoid than to a magnet! By the way, the power to keep the geodynamo running is estimated to be between 200 and 500 GW and is due to the heat in the centre of the Earth. We explore the geodynamo below. Page 224

All the relations between electricity and magnetism can be used to make *electric motors*. First, electric current in a coil is used to generate a magnetic field; then the field is used to move a magnet attached to the motor axis. The details on how to do this effectively depend on the size of the motor one is building; they form a science on its own: electric engineering. Figure 18 shows some examples of electric motors.

WHICH CURRENTS FLOW INSIDE MAGNETS?

Magnetic monopoles do not exist. Therefore, all magnetic fields in nature are due to moving electric charges. But that is strange; if all magnetic fields are due to the motion of charges, this must be also the case inside lodestone, or inside a usual permanent magnet. Can this be shown?

In 1915, two men in the Netherlands found a simple way to prove that in any permanent magnet, charges are moving. They suspended a metal rod from the ceiling by a thin thread and then put a coil around the rod, as shown in Figure 19. They predicted that the tiny currents inside the rod would become aligned by the magnetic field of the coil. As a result, they expected that a current passing through the coil would make the rod turn around its axis. Indeed, when they sent a strong current through the coil, the rod rotated. (As a result of the current, the rod was magnetized.) Today, this effect is called the *Einstein–de Haas effect* after the two men who imagined, measured and explained it.* Ref. 22

---

* Wander Johannes de Haas (b. 1878 Lisse, d. 1960 Bilthoven) was a physicist who is most known for two

The effect thus shows that even in the case of a permanent magnet, the magnetic field is due to the internal motion of charges. The magnitude of the Einstein–de Haas effect also shows that the moving particles are electrons. Twelve years later, in 1927, it became clear that the angular momentum responsible for the effect is a mixture of orbital and spin angular momentum; in fact, the electron spin plays a central role in the effect. We will explore electron spin in the volumes on quantum theory. In short,

▷ Magnetic poles are due to the rotation axis of the charges.

In particular, a magnet has two poles because rotation axes have two ends.

Permanent magnets are made from ferromagnetic materials. Their permanent magnetization is due to the alignment of microscopic rotational motions. Due to this connection, an even more surprising effect can be predicted: Rotating a piece of non-magnetized ferromagnetic material should magnetize it, because the tiny rotating currents would then be aligned along the axis of rotation. This effect has indeed been observed; it is called the *Barnett effect* after its discoverer. Like the Einstein–de Haas effect, the magnitude of the Barnett effect can also be used to determine the gyromagnetic ratio of the electron. In short, also the Barnett effect proves that the spins of electrons (usually) play a larger role in magnetism than their orbital angular momentum.

Ref. 23

Vol. IV, page 107

DESCRIBING MAGNETIC FIELDS

All experiments show that the magnetic field has a given direction in space, and a magnitude common to all (resting) observers, whatever their orientation. We are thus tempted to describe the magnetic field by a vector. However, this would be wrong, since a magnetic field does not behave like an arrow when placed before a mirror. Imagine that a system produces a magnetic field directed to the right. You can take any system, a coil, a machine, etc. Now build or imagine a second system that is the exact mirror version of the first: a mirror coil, a mirror machine, etc. The magnetic system produced by the mirror system does not point to the left, as maybe you expected: it still points to the right. (Check by yourself.) In simple words, magnetic fields do *not* fully behave like arrows.

Challenge 22 e

In other words, it is *not* completely correct to describe a magnetic field by a vector $\boldsymbol{B} = (B_x, B_y, B_z)$, as vectors behave like arrows. The magnetic field is a *pseudovector* or *axial vector*; angular momentum and torque are also examples of such quantities. The precise way is to describe the magnetic field by the quantity*

$$\boldsymbol{B} = \begin{pmatrix} 0 & -B_z & B_y \\ B_z & 0 & -B_x \\ -B_y & B_x & 0 \end{pmatrix} , \qquad (9)$$

---

additional magneto-electric effects named after him, the *Shubnikov–de Haas effect* (the strong increase of the magnetic resistance of bismuth at low temperatures and high magnetic fields) and the *de Haas–van Alphen effect* (the diamagnetic susceptibility of bismuth at low temperatures is a periodic function of the magnetic field).

* The quantity $\boldsymbol{B}$ was not called the 'magnetic field' until recently. We follow here the modern, logical definition, which supersedes the traditional one, where $\boldsymbol{B}$ was called the 'magnetic flux density' or 'magnetic induction' and another quantity, $\boldsymbol{H}$, was called – incorrectly, but for over a century – the magnetic field. This quantity $\boldsymbol{H}$ will not appear in this walk, but it is important for the description of magnetism in materials.

TABLE 11 Properties of the classical magnetic field: an axial vector.

Magnetic fields can	Physical property	Mathematical name	Definition
Attract currents	deflect charges	coupling	equation (10)
Repel currents	deflect charges	coupling	equation (10)
Be distinguished	distinguishability	element of set	Page 285
Change gradually	continuum	real vector space	Vol. I, page 80, Vol. V, page 364
Point somewhere	direction	vector space, dimensionality	Vol. I, page 80
Be compared	measurability	metricity	Vol. IV, page 236
Be added	additivity	vector space	Vol. I, page 80
Have defined angles	direction	Euclidean vector space	Vol. I, page 81
Exceed any limit	infinity	unboundedness	Page 286
Keep direction under reflection	axiality	parity-even vector, pseudovector	Page 90
Change direction under time reversal	axiality	time-odd vector	Page 90

called an *antisymmetric tensor*.

The *magnetic field* is defined by the acceleration it imparts on moving charges. This acceleration is observed to follow

$$\boldsymbol{a} = \frac{q}{m} \boldsymbol{v} \times \boldsymbol{B} \tag{10}$$

for a charge q with mass m. The relation is often called *Lorentz acceleration*, after the important physicist Hendrik A. Lorentz* who first stated it clearly.** The Lorentz acceleration, also called the *Laplace acceleration*, defines the magnitude and the direction of the magnetic field $\boldsymbol{B}$. The unit of the magnetic field is called tesla and is abbreviated T. We have $1\,\mathrm{T} = 1\,\mathrm{N\,s/C\,m} = 1\,\mathrm{V\,s/m^2} = 1\,\mathrm{V\,s^2/A\,m}$.

The magnetic field is defined and measured by its influence moving charges. Let us

---

Vol. II, page 40 * For more details about Hendrik A. Lorentz (b. 1853 Arnhem, d. 1928 Haarlem), see the volume on relativity.

Vol. I, page 115 ** The expression $\boldsymbol{v} \times \boldsymbol{B}$ is the vector product of the two vectors. The most practical way to calculate the vector product $\boldsymbol{v} \times \boldsymbol{B}$ component by component is given by the determinant

$$\boldsymbol{v} \times \boldsymbol{B} = \begin{vmatrix} \boldsymbol{e}_x & v_x & B_x \\ \boldsymbol{e}_y & v_y & B_y \\ \boldsymbol{e}_z & v_z & B_z \end{vmatrix} \quad \text{or, more sloppily} \quad \boldsymbol{v} \times \boldsymbol{B} = \begin{vmatrix} + & - & + \\ v_x & v_y & v_z \\ B_x & B_y & B_z \end{vmatrix} . \tag{11}$$

This is easy to remember and easy to perform, both with letters and with numerical values. (Here, $\boldsymbol{e}_x$ is the unit basis vector in the x direction.) Written out, it is equivalent to the relation

$$\boldsymbol{v} \times \boldsymbol{B} = (v_y B_z - B_y v_z, B_x v_z - v_x B_z, v_x B_y - B_x v_y) \tag{12}$$

which is harder to remember.

TABLE 12 Some sensors for static and quasistatic magnetic fields.

MEASUREMENT	SENSOR	RANGE
Voltage	Hall probe	up to many T
Induced electromotive force (voltage)	doves	from a few nT
Bone growth stimulation	piezoelectricity and magnetostriction of bones	from 50 mT
Induced electromotive force (voltage)	human nerves	from a few T
Sensations in thorax and shoulders	human nerves	strong switched gradients
Sharks	induced voltage when waving left to right	a few nT
Plants	unclear	small effects on growth

explore the definition. Does the definition of magnetic field given here assume a charge speed much lower than that of light? Challenge 23 s

The definition of the magnetic field assumes, like that of the electric field, that the test charge q is so small that it does not disturb the field $\boldsymbol{B}$ to be measured. Again, we ignore this issue, which means that we ignore all quantum effects, until later in our adventure. Page 248

The definition of the magnetic field also assumes that space-time is flat, and it ignores all issues due to space-time curvature.

The Lorentz acceleration is the fundamental effect that a magnetic field has on a moving charge. The Lorentz acceleration is the effect at the root of any electric motor. An electric motor is a device that uses a magnetic field as efficiently as possible to accelerate charges flowing in a wire. Through the motion of the charges, the wire is then also moved. In an electric motor, electricity is thus transformed into magnetism and then into motion. The first efficient electric motors were built already in the 1830s.

Moving charges produce magnetic fields. Like for the electric field, we need to know how the *strength* of a magnetic field is determined by a moving charge. Experiments such as Oersted's show that the magnetic field of a point-like charge q moving with velocity $\boldsymbol{v}$ produces a field $\boldsymbol{B}$ given by

$$\boldsymbol{B}(\boldsymbol{r}) = \frac{\mu_0}{4\pi} q \, \frac{\boldsymbol{v} \times \boldsymbol{r}}{r^3} \quad \text{where} \quad \frac{\mu_0}{4\pi} = 10^{-7}\,\mathrm{N/A^2} \,. \tag{13}$$

This is called *Ampère's 'law'*. Again, the strange factor $\mu_0/4\pi$ is due to the historical way in which the electrical units were defined. The constant μ_0 is called the *permeability of the vacuum* and is defined by the fraction of newton per ampere squared given in the formula. It is easy to see that the magnetic field has an intensity given by $\boldsymbol{v}\boldsymbol{E}/c^2$, where $\boldsymbol{E}$ is the electric field measured by an observer moving *with* the charge. This is one of the many hints that magnetism is a relativistic effect. Challenge 24 e

We note that equation (13) is valid only for small velocities and accelerations. Can you find the general relation? Challenge 25 s

ELECTROMAGNETISM

In 1831, Michael Faraday discovered an additional piece of the jigsaw puzzle formed by electricity and magnetism, one that even the great Ampère had overlooked. He found that

▷ A *moving* magnet causes a current flow in an electrical circuit.

Magnetism can thus be turned into electricity. This important discovery allowed the production of electrical current flow by generators, so-called *dynamos*, using water power, wind power or steam power. In fact, the first dynamo was already built in 1832 by Ampère and his technician. Dynamos jump-started the use of electricity throughout the world. Behind every electrical wall plug there is a dynamo somewhere.

Oersted had found that electric current can produce magnetic fields. Faraday had found that magnetic fields could produce electric currents and electric fields. Electric and magnetic fields are thus two aspects of the same phenomenon: *electromagnetism*. It took another thirty years to unravel the full description.

Additional experiments show that magnetic fields also lead to electric fields when one changes to a moving viewpoint. You might check this on any of the examples of Figures 18 to 44.

▷ Magnetism is relativistic electricity.

Electric and magnetic fields are partly transformed into each other when switching from one inertial reference frame to the other. Magnetic and electrical fields thus behave like space and time, which are also mixed up when changing from one inertial frame to the other. In such a case, the theory of special relativity thus tells us that there must be a single concept, an *electromagnetic field*, describing them both. Investigating the details, one finds that the electromagnetic field $\boldsymbol{F}$ surrounding charged bodies has to be described by an antisymmetric 4-*tensor*

$$\boldsymbol{F}^{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix} \text{ or } \boldsymbol{F}_{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix} . \qquad (14)$$

Obviously, the electromagnetic field $\boldsymbol{F}$, and thus every component of these matrices, depends on space and time. Above all, the matrices show that electricity and magnetism are two faces of the same effect.* In addition, since electric fields appear only in the topmost row and leftmost column, the expressions show that in everyday life, for small speeds, electricity and magnetism *can* be separated. (Why?) Challenge 26 s

Using relativistic notation, the electromagnetic field is thus defined through the 4-

---

* Actually, the expression for the field contains everywhere the expression $1/\sqrt{\mu_0\varepsilon_0}$ instead of the speed of light c. We will explain the reason for this substitution shortly.

acceleration $\boldsymbol{b}$ that it produces on a charge q of mass m and 4-velocity $\boldsymbol{u}$:

$$\begin{aligned} m\boldsymbol{b} &= q\boldsymbol{F}\boldsymbol{u} \quad \text{or, equivalently, in 3-vector notation} \\ \mathrm{d}E/\mathrm{d}t &= q\boldsymbol{E}\boldsymbol{v} \quad \text{and} \quad \mathrm{d}\boldsymbol{p}/\mathrm{d}t = q(\boldsymbol{E}+\boldsymbol{v}\times\boldsymbol{B}) \; . \end{aligned} \tag{15}$$

The expressions show how the power $\mathrm{d}E/\mathrm{d}t$ (the letter E denotes energy, whereas $\boldsymbol{E}$ denotes the electric field) and the three-force $\mathrm{d}\boldsymbol{p}/\mathrm{d}t$ depend on the electric and magnetic fields.* The 4-vector expression and the 3-vector expression describe the same content; the simplicity of the first one is the reason for the involved matrices (14) describing the electromagnetic field $\boldsymbol{F}$.

We stress that the extended *Lorentz relation* (15) is the *definition* of the electromagnetic field $\boldsymbol{F}$, since the field is defined as that 'stuff' which accelerates charges. In particular, all devices that put charges into motion, such as batteries and dynamos, as well as all devices that are put into motion by flowing charges, such as electric motors and muscles, are described by this relation. That is why this relation is usually studied, in the 3-vector form, already in secondary school. The Lorentz relation describes all cases in which the motion of objects can be seen by the naked eye or felt by our senses, such as the movement of an electrical motor in a high speed train, in a lift and in a dental drill, the motion of the picture generating electron beam in a cathode ray tube inside an old television, or the travelling of an electrical signal in a cable and in the nerves of the body. Ref. 24, Ref. 25

In summary, we found that the interaction between charges can be described in two statements: First, charges *produce* electric and magnetic fields; second, charges *are affected* by electric and magnetic fields. Charges move and the fields depend on time. Their study is thus called *electrodynamics*.

THE INVARIANTS AND THE LAGRANGIAN OF ELECTROMAGNETIC FIELDS**

The electromagnetic field tensor $\boldsymbol{F}$ is an *antisymmetric* 4-tensor. (Can you write down the relation between $\boldsymbol{F}^{\mu\nu}$, $\boldsymbol{F}_{\mu\nu}$ and $\boldsymbol{F}^{\mu}{}_{\nu}$?) Like any antisymmetric tensor, the electromagnetic field has two *invariants*, i.e., two deduced properties that are the same for every observer. The first invariant is the expression Challenge 27 e

$$B^2 - E^2/c^2 = \frac{1}{2}\mathrm{tr}\,\boldsymbol{F}^2 \tag{17}$$

---

* In component notation, using the convention to sum over Greek indices that appear twice, the definition of the Lorentz force is

$$\begin{gathered} mb^{\mu} = m\frac{\mathrm{d}u^{\mu}}{\mathrm{d}\tau} = q\boldsymbol{F}^{\mu}{}_{\nu}u^{\nu} \quad \text{or} \\ m\frac{\mathrm{d}}{\mathrm{d}\tau}\begin{pmatrix} \gamma c \\ \gamma v_x \\ \gamma v_y \\ \gamma v_z \end{pmatrix} = q\begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix}\begin{pmatrix} \gamma c \\ \gamma v_x \\ \gamma v_y \\ \gamma v_z \end{pmatrix} . \end{gathered} \tag{16}$$

** This section can be skipped at first reading.

and the second invariant is the product

$$4\boldsymbol{EB} = -c \operatorname{tr} \boldsymbol{F}^{*}\boldsymbol{F} . \tag{18}$$

Challenge 28 s

Can you confirm the two invariants, using the definition of trace tr as the sum of the diagonal elements?

The first invariant expression, $B^2 - E^2/c^2 = \frac{1}{2}\operatorname{tr} \boldsymbol{F}^2$, turns out to be (proportional to) the Lagrangian density of the electromagnetic field. In particular, this first invariant is a scalar. This first invariant implies that if E is larger, smaller or equal to cB for one observer, it also is for all other observers. Like for all intensive quantities that evolve, the Lagrangian is proportional to the *square* of the intensive quantity. The minus sign in the expression is the same minus sign that appears also in $c^2t^2 - x^2$: it results from the mixing of electric and magnetic fields that is due to boosts.

The Lagrangian density can be used to define the classical action of the electromagnetic field:

$$S = \int \frac{\varepsilon_0}{2}E^2 - \frac{1}{2\mu_0}B^2 \mathrm{d}t\mathrm{d}V . \tag{19}$$

Vol. IV, page 47

Page 76

As usual, the action measures the *change* occurring in a system; it thus defines the amount of change that occurs when an electromagnetic field moves. (The expression for the change, or action, of a moving light beam reduces to the product of its intensity and total phase change.) The action of an electromagnetic field thus increases with its intensity and with its frequency. As usual, this expression for the action can be used to describe the motion of the electromagnetic field by using the *principle of least action*. Indeed, the principle implies the evolution equations of the electromagnetic field, which are called *Maxwell's field equations of electrodynamics*. This approach is the simplest way to deduce them. We will discuss the field equations in detail shortly.

The second invariant of the electromagnetic field tensor, $4\boldsymbol{E}\cdot\boldsymbol{B} = -c \operatorname{tr} \boldsymbol{F}^{*}\boldsymbol{F}$, is a pseudoscalar; it describes whether the angle between the electric and the magnetic field is acute or obtuse for all observers.*

---

* There is in fact a third Lorentz invariant, far less known. It is specific to the electromagnetic field and is a combination of the field and its vector potential:

$$\begin{aligned}\kappa_3 &= \frac{1}{2}A_\mu A^\mu \boldsymbol{F}_{\rho\nu}\boldsymbol{F}^{\nu\rho} - 2A_\rho \boldsymbol{F}^{\rho\nu}\boldsymbol{F}_{\nu\mu}A^\mu \\ &= (\boldsymbol{A}\cdot\boldsymbol{E})^2 + (\boldsymbol{A}\cdot\boldsymbol{B})^2 - |\boldsymbol{A}\times\boldsymbol{E}|^2 - |\boldsymbol{A}\times\boldsymbol{B}|^2 + 4\frac{\varphi}{c}(\boldsymbol{A}\cdot\boldsymbol{E}\times\boldsymbol{B}) - \left(\frac{\varphi}{c}\right)^2(E^2+B^2) .\end{aligned} \tag{20}$$

Ref. 26

Page 86

Challenge 29 s

This expression is Lorentz (but not gauge) invariant; knowing it can help clarify unclear issues, such as the lack of existence of waves in which the electric and magnetic fields are parallel. Indeed, for plane monochromatic waves all three invariants *vanish* in the Lorentz gauge. Also the quantities $\partial_\mu j^\mu$, $j_\mu A^\mu$ – j being the electric current – and $\partial_\mu A^\mu$ are Lorentz invariants. (Why?) The last one, the frame independence of the divergence of the four-potential, reflects the invariance of gauge choice. The gauge in which the expression is set to zero is called the *Lorentz gauge*.

THE USES OF ELECTROMAGNETIC EFFECTS

The application of electromagnetic effects to daily life has changed the world. For example, the installation of electric lighting in city streets has almost eliminated the previously so common night assaults. These and all other electrical devices exploit the fact that charges can flow in metals and, in particular, that electromagnetic energy can be transformed

- into mechanical energy – as done in loudspeakers, motors and muscles;
- into light – as in lamps, lasers, glass fibres, glow worms, giant squids and various deep ocean animals;
- into heat – as in electric ovens, blankets, tea pots and by electric eels to stun and kill prey;
- into chemical effects – as in hydrolysis, battery charging, electroplating and the brain;
- into coldness – as in refrigerators and Peltier elements, but in no known living system;
- into radio wave signals – as in radio and television, but in no known living system;
- into stored information – as in magnetic records, computers, animal and human memory.

Due to all these options, electrical light, lasers, batteries, electric motors, refrigerators, radio, telephones, X-rays, television and computers have changed human life completely in less than one century.

Electromagnetic effects are thus useful to perform something at a specific place and time, thus to realize actuators. In addition, electromagnetic effects are useful so capture information from the environment, thus to realize sensors. Page 231

Many of these uses of electromagnetism also occur in biological systems. However, no biological system makes use of X-rays, though. (Why?) No living being seems to use electric cooling. (Why?) And could there be biological systems that communicate via radio waves? Challenge 30 s Challenge 31 s

HOW DO NERVES WORK?

Nerves are wonders. Without nerves, we would not experience pleasure, we would not experience pain, we would not see and we would not hear. Without nerves, we would not live. But how do nerves transport signals?

Page 32 In 1789, as mentioned above, Luigi Galvani discovered that nerves transport electric signals, by doing experiments with frog legs. Are nerves wires? One and a half centuries after Galvani it became clear that nerves, more precisely, nerve axons, do not conduct electricity using electrons, as metal wires do, but by using *ions*. Nerve signals propagate using the motion of sodium Na^+ and potassium K^+ ions through the cell membrane of the nerve. The resulting signal speed is between 0.5 m/s and 120 m/s, depending on the type of nerve. (Nerve axons coated with myelin, a protein that acts as an electric insulator, are faster than uncoated axons.) The signal speed is sufficient for the survival of most species – it helps the body to run away in case of danger.

Nerves differ from wires in another aspect: they cannot transmit constant voltage signals, but only signal *pulses*. The first, approximate model for this behaviour was presented in 1952 by Hodgkin and Huxley. Using observations about the behaviour of potassium Ref. 27

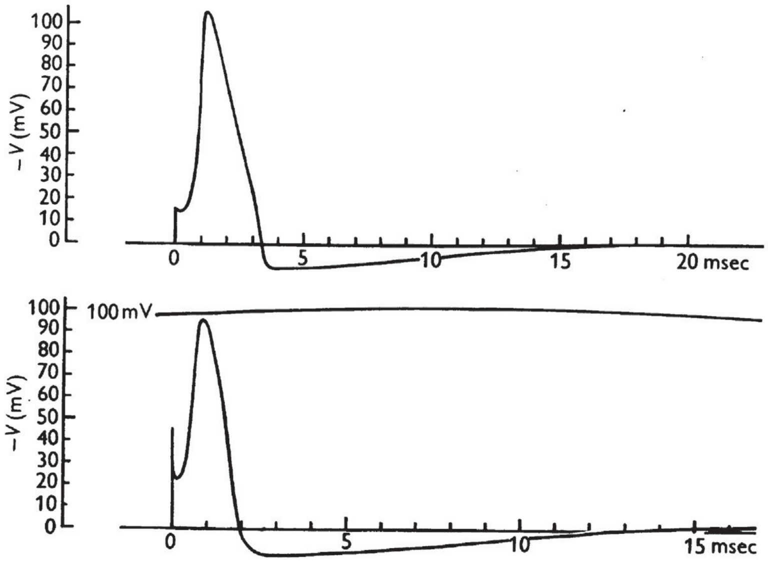


Fig. 13. Upper curve: solution of eqn. (26) for initial depolarization of 15 mV, calculated for 6° C. Lower curve: tracing of membrane action potential recorded at 9·1° C (axon 14). The vertical scales are the same in both curves (apart from curvature in the lower record). The horizontal scales differ by a factor appropriate to the temperature difference.

FIGURE 20 The electrical signals calculated (above) and measured (below) in a nerve, following Hodgkin and Huxley.

and sodium ions, they deduced an elaborate evolution equation that describes the voltage V in nerves, and thus the way the signals propagate. The equation reproduces the characteristic voltage spikes measured in nerves, shown in Figure 20.

The precise mechanism with which ions cross the membranes, using so-called *channel proteins*, was elucidated only twenty years later. Despite this huge body of work, and even though Hodgkin and Huxley received the Nobel Prize for Medicine for their work, the model cannot be correct. The model does not explain the reversibility of the propagation process, the observed thickness change of the nerve during propagation nor the excitation of nerves by simple deformation or temperature changes; most of all, the model does not explain the working of anaesthetics. The detailed working of nerves remained unknown.

Ref. 28 Only around the year 2000 did Thomas Heimburg and his team discover the way signals propagate in nerves. He showed that a nerve pulse is an electromechanical solitonic wave of the cylindrical membrane. In the cylindrical membrane, the protein structure changes from liquid to solid and back to liquid. This short, slightly thicker ring of solid proteins propagates along the cylinder: that is the nerve pulse. In short, the nerve pulse does not make proteins move, but makes the region of solidity move. The model is shown in Figure 21. (The term 'solid' has a precise technical meaning in two-dimensional systems and describes a specific ordered state of the molecules.) This propagation model explains all the properties of nerve pulses that were unexplained before. In particular, it explains that anaesthetics work because they dissolve in the membrane and thus block the formation and the propagation of the rings. All quantitative predictions of the model

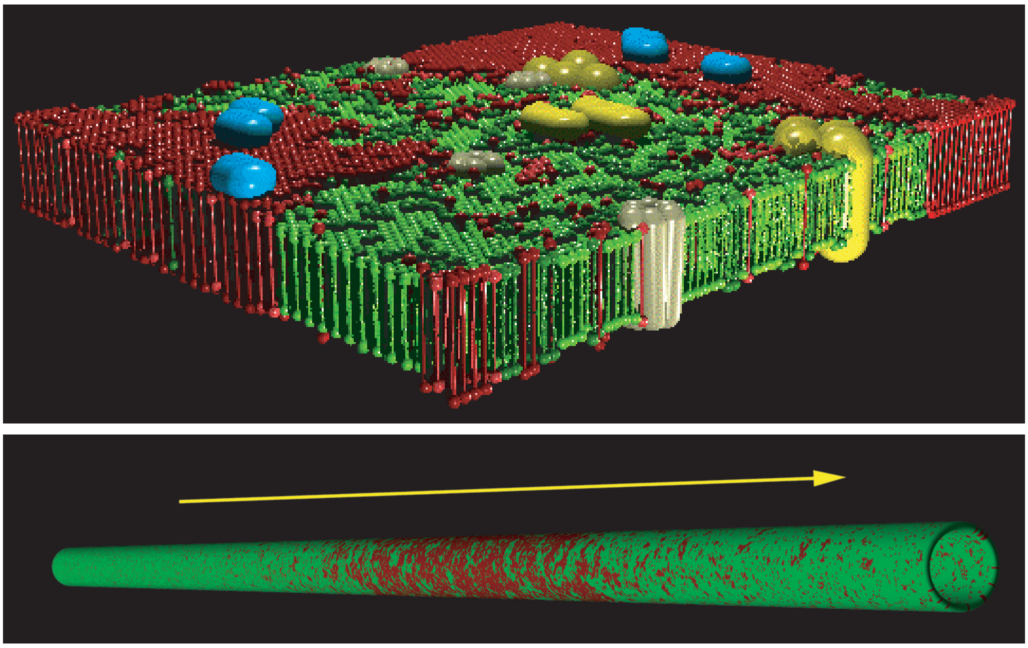


FIGURE 21 Top: A biomembrane, with solid-type lipids (red), liquid lipids (green) and various dissolved proteins (yellow, blue, white). Bottom: a nerve pulse propagating as a two-dimensional phase transformation liquid/solid/liquid along a cylindrical nerve membrane (© Thomas Heimburg/Wiley-VCH).

match observations.

In summary, nerve signals are electromechanical pulses; they are a mixture of current and sound waves. The electromechanical model of nerves explains how signals propagate, how pain is felt and why no pain is felt during anaesthesia.

Interestingly, the electromechanical model of nerve pulse propagation does not (yet) explain why we loose consciousness during anaesthesia. This is an additional process that takes place in the brain. It is known that loss of consciousness is related to the change of brain waves, but the details are still a topic of research. Brains still have wonderful properties to be explored.

How motors prove relativity to be right

> “The only mathematical operation I performed in my life was to turn the handle of a calculator.”
>
> Michael Faraday

All electric motors are based on the result that electric currents interact with magnetic fields. The simplest example is the attraction of two wires carrying parallel currents. This observation alone, made in 1820 by Ampère, is sufficient to make motion larger than a certain maximal speed impossible. The argument is beautifully simple. Ref. 29

We change the original experiment and imagine two long, electrically charged rods of mass m, moving in the same direction with velocity v and separation d. An observer

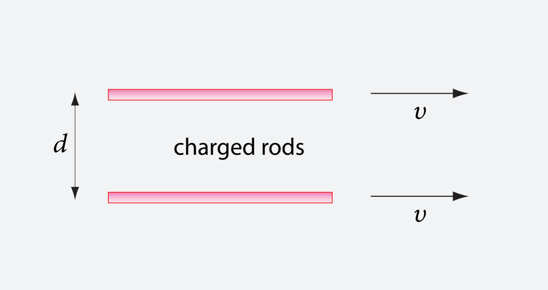


FIGURE 22 Charged rods moving in parallel illustrate the relativistic aspect of magnetism, as explained in the text.

Challenge 32 e

moving with the rods would see an electrostatic repulsion between the rods given by

$$ma_e = -\frac{1}{4\pi\varepsilon_0}\frac{2\lambda^2}{d} \qquad (21)$$

where λ is the charge per length of the rods. A second, *resting* observer sees two effects: the electrostatic repulsion and the attraction discovered by Ampère. The second observer therefore observes

Challenge 33 e

$$ma_{em} = -\frac{1}{4\pi\varepsilon_0}\frac{2\lambda^2}{d} + \frac{\mu_0}{2\pi}\frac{\lambda^2 v^2}{d} \,. \qquad (22)$$

This expression must be consistent with the observation of the first observer. This is the case only if both observers find repulsions. It is easy to check that the second observer sees a repulsion, as does the first one, only if

$$v^2 < \frac{1}{\varepsilon_0\mu_0} = c^2 \,. \qquad (23)$$

This maximum speed c, with a value of 0.3 GM/s, is thus valid for any object carrying charges. But *all* everyday objects contain charges: there is thus a maximum speed for matter.

Are you able to extend the argument for a maximum speed to neutral particles as well? We will find out more on this limit velocity, which we know already, in a minute.

Challenge 34 d

Another argument for magnetism as a relativistic effect is the following. In a wire with electrical current, the charge is zero for an observer at rest with respect to the wire: the wire is *neutral* for that observer. The reason is that the charges enter and exit the wire at the same time for that observer. Now imagine an observer who flies along the wire. The entrance and exit events do not occur simultaneously any more; the wire is *charged* for a moving observer. (The charge depends on the direction of the observer's motion.) Now imagine that the moving observer is electrically charged. He will be attracted or repelled by the wire, because for him, the wire is charged. The moving observer will say that the attraction is due to the *electric* field of the wire. The observer at rest will also note the attraction or repulsion of the moving observer, but since for him, the wire is neutral, he will deduce that moving charges experience a force – possibly with a slightly different

TABLE 13 Voltage values observed in nature.

Observation	Voltage
Smallest measured voltage	*c.* 10 fV
Human nerves	70 mV
Volta cell	1 V
Voltaic cell ('battery')	1.5 V
Mains in households	230 V or 110 V
Electric eel	100 to 600 V
Tramway supply	500 V
Sparks when rubbing a polymer pullover	1 kV
Electric fence	0.7 to 10 kV
Train supply	15 kV
Ignition plug in cars	15 kV
Colour television cathode ray tube	30 kV
X-ray tube	30 to 200 kV
Electron microscopes	0.5 kV to 3 MV
Stun gun	65 to 600 kV
Lightning stroke	10 to 100 MV
Record accelerator voltage	1 TV
Maximum possible voltage in nature, the corrected Planck voltage $\sqrt{c^4/16\pi\varepsilon_0 G}$	$5.2 \cdot 10^{26}$ V

value, but this is a technicality – due to the electric current in the wire; the observer at rest will thus say that a wire with a current is surrounded by a *magnetic* field which only produces an effect on charges that move.

In summary, electric effects are due to more or less static electric charges and to their electric fields; magnetism, magnetic effects and magnetic fields are due to *moving* electric charges.* The existence of magnetic fields is a relativistic consequence of the existence of electric fields. In particular, magnetism is *not* due to particles with magnetic charges. Such particles, called magnetic monopoles, do not exist. (Magnetic charges can be introduced as a mathematical tool, though, for the description of materials.) The strength of magnetism, used in any running electric motor, including your electric toothbrush, proves relativity right: there is a maximum speed in nature for all masses and charges. Both electric and magnetic fields carry energy and momentum. They are two faces of the same coin.

Page 96

Page 249

* 'Electrons move in metal with a speed of about 1 μm/s; thus if I walk with the same speed along a cable carrying a constant current, I should not be able to sense any magnetic field.' What is wrong with this argument?

Challenge 35 d

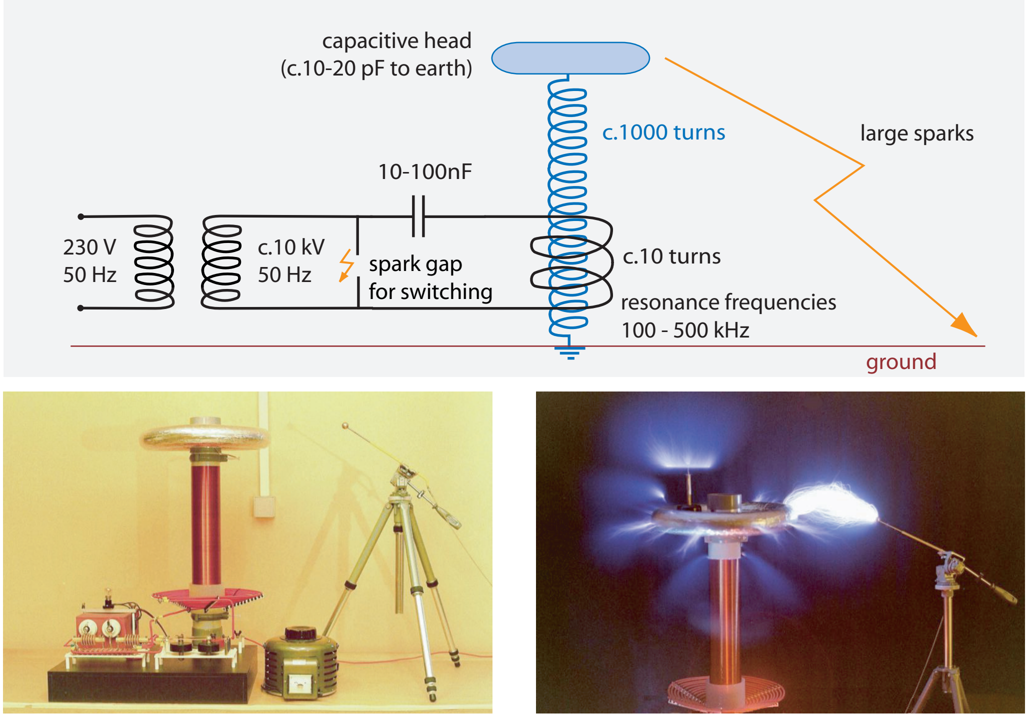


FIGURE 23 The schematics, the realization and the operation of a Tesla coil, including spark and corona discharges (photographs © Robert Billon).

Curiosities and fun challenges about things electric and magnetic

> “Alii vero et facta mirati et intellecta assecuti.[*]”
> Augustine of Hippo

Before we study the motion of an electromagnetic field in detail, let’s have some fun with electricity.

* *

Nowadays, having fun with sparks is straightforward. Tesla coils, named after Nikola Tesla[**] are the simplest devices that allow long sparks to be produced at home. Attention: this is dangerous; that is the reason that such devices cannot be bought (almost) anywhere. The basic diagram and an example is shown in Figure 23. Tesla coils look like

---

[*] ‘Others however marvelled about the facts and understood their meaning.’ Augustine, Sermon 98, 3. Augustine of Hippo (b. 354 Tagaste, d. 430 Hippo Regius) is an influential moral theologian. Despite this, he did not take care of his extramarital son, nor of his son’s mother, because his own mother had forbidden him to do so.

[**] Никола Тесла (b. 1856 Smiljan, d. 1943 New York City), engineer and inventor. He invented and promoted the polyphase alternating current system, the alternating current electric motor, wireless communication, fluorescent lighting and many other applications of electricity. He is also one of the inventors of radio. The SI unit of the magnetic field is named after him. A flamboyant character, his ideas were sometimes unrealistic; for example he imagined that Tesla coils could be used for wireless power transmission.

large metal mushrooms (to avoid unwanted discharges) and plans for their construction can be found on numerous websites or from numerous enthusiast's clubs, such as www.stefan-kluge.de.

∗ ∗

In 1722, George Graham discovered, by watching a compass needle, that the magnetic field of the Earth shows daily variations. Can you imagine why these variations occur? Challenge 36 s

∗ ∗

If even knocking on a wooden door is an electric effect, we should be able to detect fields when doing so. Can you devise an experiment to check this? Challenge 37 d

∗ ∗

Birds come to no harm when they sit on unprotected electricity lines. Nevertheless, one almost never observes any birds on tall, high voltage lines of 100 kV or more, which transport power across longer distances. Why? Challenge 38 s

∗ ∗

How can you distinguish a magnet from an non-magnetized metal bar of the same size and material, using no external means? Challenge 39 s

∗ ∗

In the basement of a house there are three switches that control three light bulbs in the first floor. You are in the basement and are allowed to go to the first floor only once. How do you find out which switch controls which bulb? Challenge 40 s

∗ ∗

How do you wire up a light bulb to the mains and three switches so that the light can be switched on at any of the switches and off at any other switch? And for four switches? Nobody will take a physicist seriously who is able to write Maxwell's equations but cannot solve this little problem. Challenge 41 s

∗ ∗

The first appliances built to generate electric currents were large rubbing machines. Then, in 1799 Alessandro Volta (b. 1745 Como, d. 1827 Como) invented a new device to generate electricity and called it a *pile*; today its basic element is called a *(voltaic) cell*, a *primary cell*[*] or, less correctly, a *battery*. (Correctly speaking, a battery is a collection of cells, as the one found in a car.) Voltaic cells are based on chemical processes; they provide much more current and are smaller and easier to handle than electrostatic machines. The invention of the battery changed the investigation of electricity so profoundly that Volta became world famous. At last, a simple and reliable source of electricity was available for use in experiments; unlike rubbing machines, cells and piles are compact, work in all weather conditions and make no noise.

An apple or a potato or a lemon with a piece of copper and one of zinc inserted is one

---

[*] A *secondary cell* is a rechargeable cell.

FIGURE 24 A common playground effect (© Evan Keller).

of the simplest possible voltaic cells. It provides a voltage of about 1 V and can be used to run digital clocks or to produce clicks in headphones. Volta was also the discoverer of the charge 'law' $q = CU$ for capacitors (C being the capacity, and U the voltage) and the inventor of the high sensitivity capacitor electroscope. A modest man, nevertheless, the unit of electrical potential, or 'tension', as Volta used to call it, was deduced from his name. A 'battery' is a large number of voltaic cells; the term was taken from an earlier, almost purely military use.* A battery in a mobile phone is just an elaborated replacement for a number of apples or potatoes.

* *

Voltaic cells exist in all biological cells. For halobacteria, the internal voltaic cells are even essential to survival. Living in saltwater, internal voltaic cells help them to avoid death due to osmosis.

* *

An famous challenge: Do full and empty alkaline batteries, e.g., of the AA type, behave differently or the same when falling on a stone (hard) floor? Challenge 43 s

* *

What happened in Figure 24? Why are most of such pictures taken in good weather and with blond children? Challenge 44 d

* *

A PC or a telephone can communicate without wires, by using radio waves. Why are these and other electrical appliances not able to obtain their *power* via radio waves, thus eliminating power cables? Challenge 45 s

---

* A pile made of sets of a zinc plate, a sheet of blotting paper soaked with salt water and a copper coin is

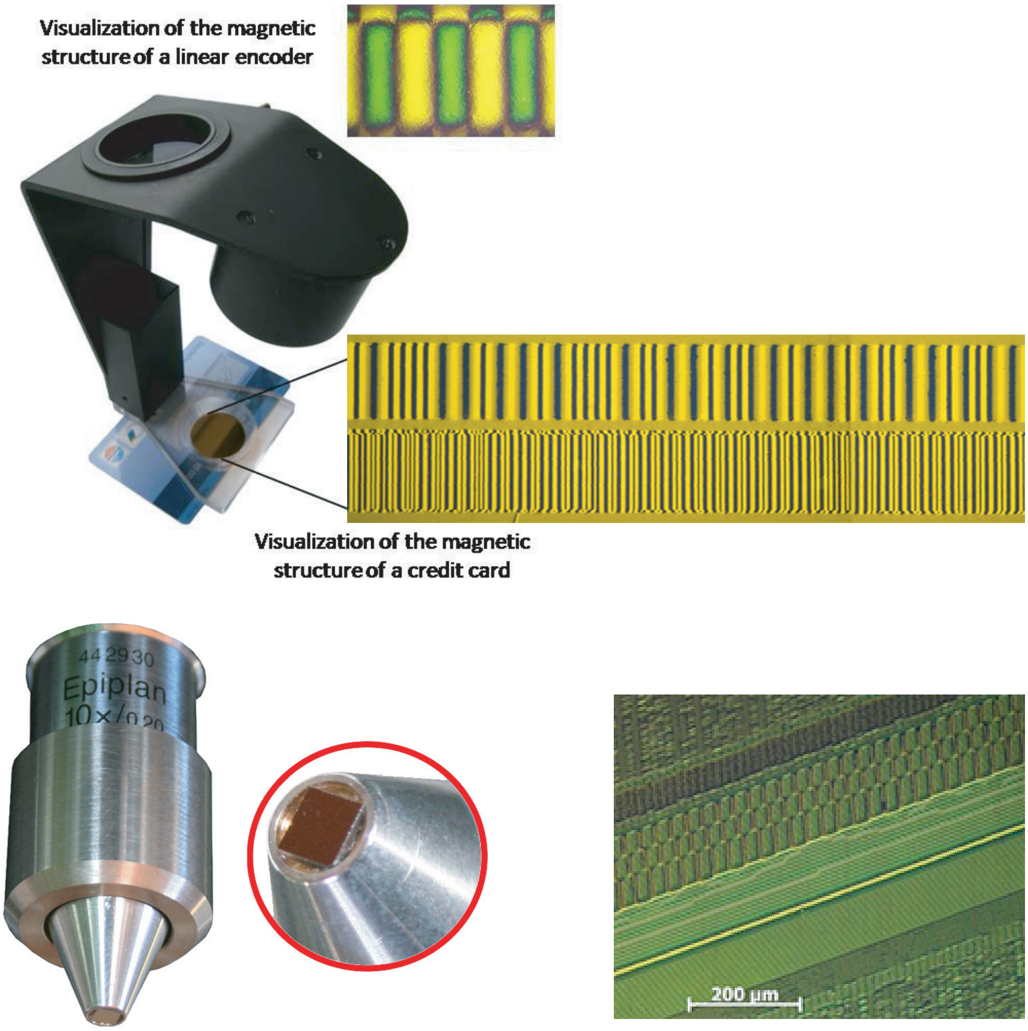


FIGURE 25 Top: how to see the information stored in the magnetic stripe on a credit card without any electronics, just using a lens, a polarizer and a magneto-optic layer; bottom: how to see the information on a hard disk in the same way, by adding a simple coated glass plate to a polarizing microscope (© Matesy).

* *

Magnetic storage looks far less mysterious if it is visualized. Figure 25 shows how simply with can be done. The method also allows taking films. What happens inside a metal when it is magnetized? The beautiful films at www.youtube.com/watch?v=HzxTqQ40wSU and www.youtube.com/watch?v=LFC6tbbMUaA, taken by Hendryk Richert of Matesy, show how the magnetization regions change when a magnet is approached to a piece of metal. Also these films have been made with a simple microscope, using as only help a polarizer and a layer of yttrium iron garnet deposited on glass.

Challenge 42 e easily constructed at home and tested with a calculator or a digital watch.

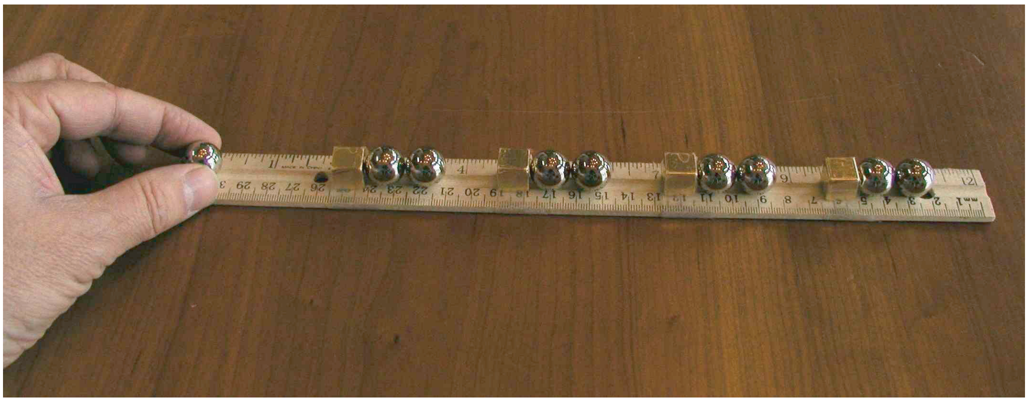


FIGURE 26 A Gauss rifle, made with a few steel balls and four magnets attached to a ruler with scotch tape (© Simon Quellen Field).

* *

Also plants react to magnetic fields. In particular, different magnetic fields yield different growth patterns. The mechanisms, related to the cryptochrome system, are still a subject of research.

* *

Magnets can be used to accelerate steel balls. The most famous example is the *Gauss rifle* shown in Figure 26. If the leftmost ball is gently rolled towards the first magnet, the third ball is strongly kicked away. Then the process repeats: the speed increases even more for the fifth, the seventh and the ninth ball. The experiment never fails to surprise whoever sees it for the first time. Where does the momentum of the final ball come from? Challenge 46 e

* *

Objects that are not right–left symmetric are called *chiral*, from the Greek word for 'hand'. Can you make a mirror that does *not* switch chirality (i.e., does not 'switch left and right')? In two different ways? Challenge 47 s

* *

An adhesive tape roll is a dangerous device. Pulling the roll quickly leads to light emission (through triboluminescence) and to small sparks. It is suspected that several explosions in mines were triggered when such a spark ignited a combustible gas mixture.

* *

Take an envelope, wet it and seal it. After letting it dry for a day or more, open it in the dark. At the place where the two sides of paper are being separated from each other, the envelope glows with a blue colour. Why? Is it possible to speed up the test using a hair dryer? Challenge 48 s

* *

A charge in an electric field feels a force. In other words, electric field produce a *poten-*

FIGURE 27 A dangerous hobby, here demonstrated by Robert Krampf (© Wikimedia).

tial energy for charges. Since energy is conserved, electric potential energy can be transformed into kinetic energy or in thermal energy. What do these possibilities allow doing? What do they prevent from doing? Challenge 49 e

* *

Electromagnetism is full of surprises and offers many effects that can be reproduced at home. The internet is full of descriptions of how to construct Tesla coils to produce sparks, coil guns or rail guns to shoot objects, electrostatic machines to make your hair stand on end and much more. If you like experiments, just search for these terms. Some people earn their living by showing high voltage effects on stage, such as long discharges from their fingers or hair. A well-known example is Robert Krampf, also called ‘Mr. Electricity’, at thehappyscientist.com. Do not emulate these performers; it is rarely told that several of them have suffered dangerous accidents while doing so.

* *

The moving discharges seen in so many displays, called *plasma globes*, are produced in a glass bowl filled with helium, neon or another inert gas at low pressure, typically 0.1 to 10 kPa, an applied voltage of 5 to 10 kV and usually a frequency of 30 to 40 kHz. At these conditions, the ion temperature of the discharges is room temperature, so that there is no danger; the electron temperature, which cannot be felt, is around 20 000 K. Approaching the hand to the sphere changes the electric potential and this also the shape of the discharges. If you approach a fluorescent tube to such a set-up, it will start glowing; and by moving your finger on the tube, you can ‘magically’ change the glow region. The internet is full of information on plasma globes. Ref. 30

* *

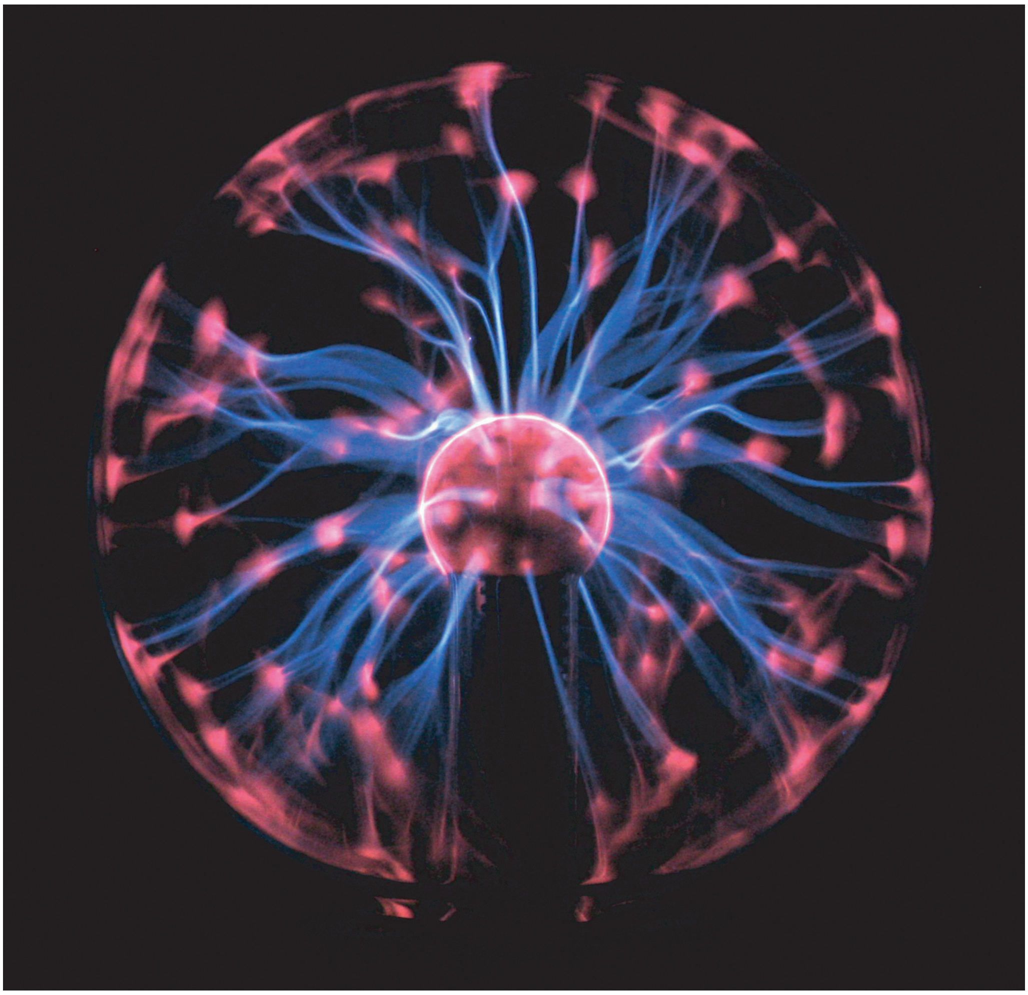


FIGURE 28 A low pressure glass sphere, or plasma globe, with a diameter of 30 cm and a built-in high voltage generator, showing its characteristic electric discharges. In a usual plasma globe, the discharges move around – slowly and irregularly. (© Philip Evans).

A high voltage can lead to current flow through air, because air becomes conductive in high electric fields. In such discharges, air molecules are put in motion. As a result, one can make objects that are attached to a pulsed high tension source lift up in the air, if one optimizes this air motion so that it points downwards everywhere. The high tension is thus effectively used to accelerate ionized air in one direction and, as a result, an object will move in the opposite direction, using the same principle as a rocket. An example is shown in Figure 29, using the power supply of a PC monitor. (Watch out: danger!) Numerous websites explain how to build these so-called lifters at home; in Figure 29, the bottle and the candle are used as high voltage insulator to keep one of the two thin high voltage wires (not visible in the photograph) high enough in the air, in order to avoid discharges to the environment or to interfere with the lifter's motion. Unfortunately, the majority of websites – not all – give incorrect or confused explanations

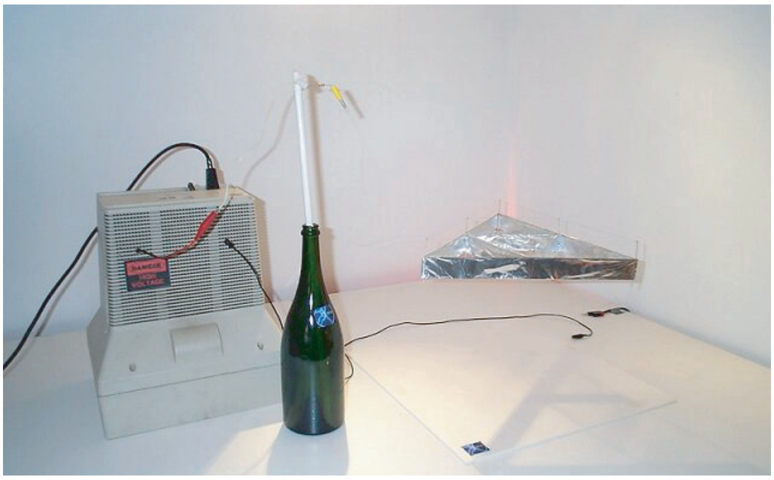


FIGURE 29 Lifting a light object – covered with aluminium foil – using a high tension discharge (© Jean-Louis Naudin at www.jlnlabs.org).

of the phenomenon. These websites thus provide a good challenge for one to learn to distinguish fact from speculation. Challenge 50 e

* *

The electric effects produced by friction and by liquid flow are usually small. However, in the 1990s, a number of oil tankers disappeared suddenly. The sailors had washed out the oil tanks by hosing sea water onto the tank walls. The spraying led to charging of the tank; a discharge then led to the oil fumes in the tank igniting. This led to an explosion and subsequently the tankers sank. Similar accidents also happen regularly when chemicals are moved from one tank to another.

* *

Rubbing a plastic spoon with a piece of wool charges it. Such a charged spoon can be used to extract pepper from a salt–pepper mixture by holding the spoon over the mixture. Challenge 51 s Why?

* *

When charges move, they produce a magnetic field. In particular, when ions inside the Earth move due to heat convection, they produce the Earth's magnetic field. When the ions high up in the atmosphere are moved by solar wind, a geomagnetic storm appears; its field strength can be as high as that of the Earth itself. In 2003, an additional mechanism was discovered. When the tides move the water of the oceans, the ions in the salt water produce a tiny magnetic field; it can be measured by highly sensitive magnetometers in satellites orbiting the Earth. After two years of measurements from a small satellite it was possible to make a beautiful film of the oceanic flows. Figure 30 gives an Ref. 31 impression.

* *

The magnetic field of the Earth is clearly influenced by the Sun. Figure 31 shows the

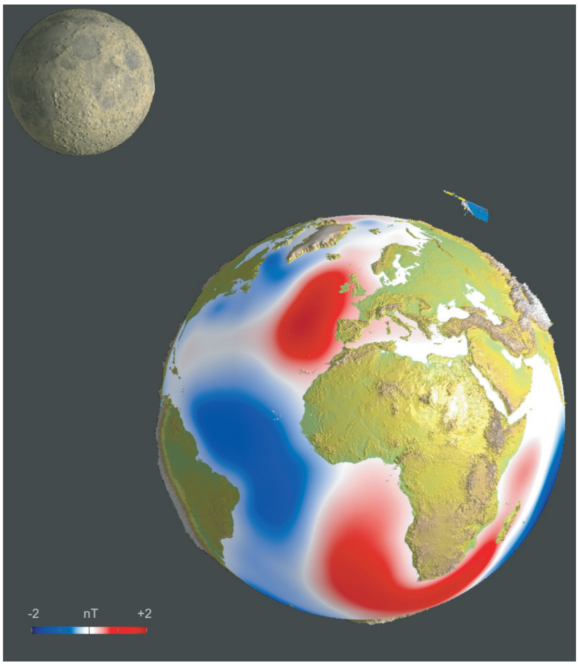


FIGURE 30 The magnetic field due to the tides (© Stefan Maus).

details of how the stream of charged particles from the Sun, the *solar wind*, influences the field lines and a several processes occurring in the higher atmosphere. Figure 32 shows the effects. The details of these fascinating processes are still a subject of research.

* *

The names electrode, electrolyte, ion, anode and cathode were suggested by William Whewell (b. 1794 Lancaster, d. 1866 Cambridge) on demand of Michael Faraday; Faraday had no formal education and asked his friend Whewell to form two Greek words for him. For anode and cathode, Whewell took words that literally mean 'upward street' and 'downward street'. Faraday then popularized these terms, like the other words mentioned above.

* *

Challenge 52 s The shortest light pulse produced so far had a duration of 100 as. To how many wavelengths of green light would that correspond?

* *

How long can batteries last? At Oxford University, in Clarendon Hall, visitors can watch a battery-operated electric bell that is ringing since 1840. The two batteries, two Zamboni piles, produce a high voltage and low current, sufficient to keep the bell ringing. Several other similar devices, using Zamboni piles, have worked in Italy with the same batteries for over 100 years.

* *

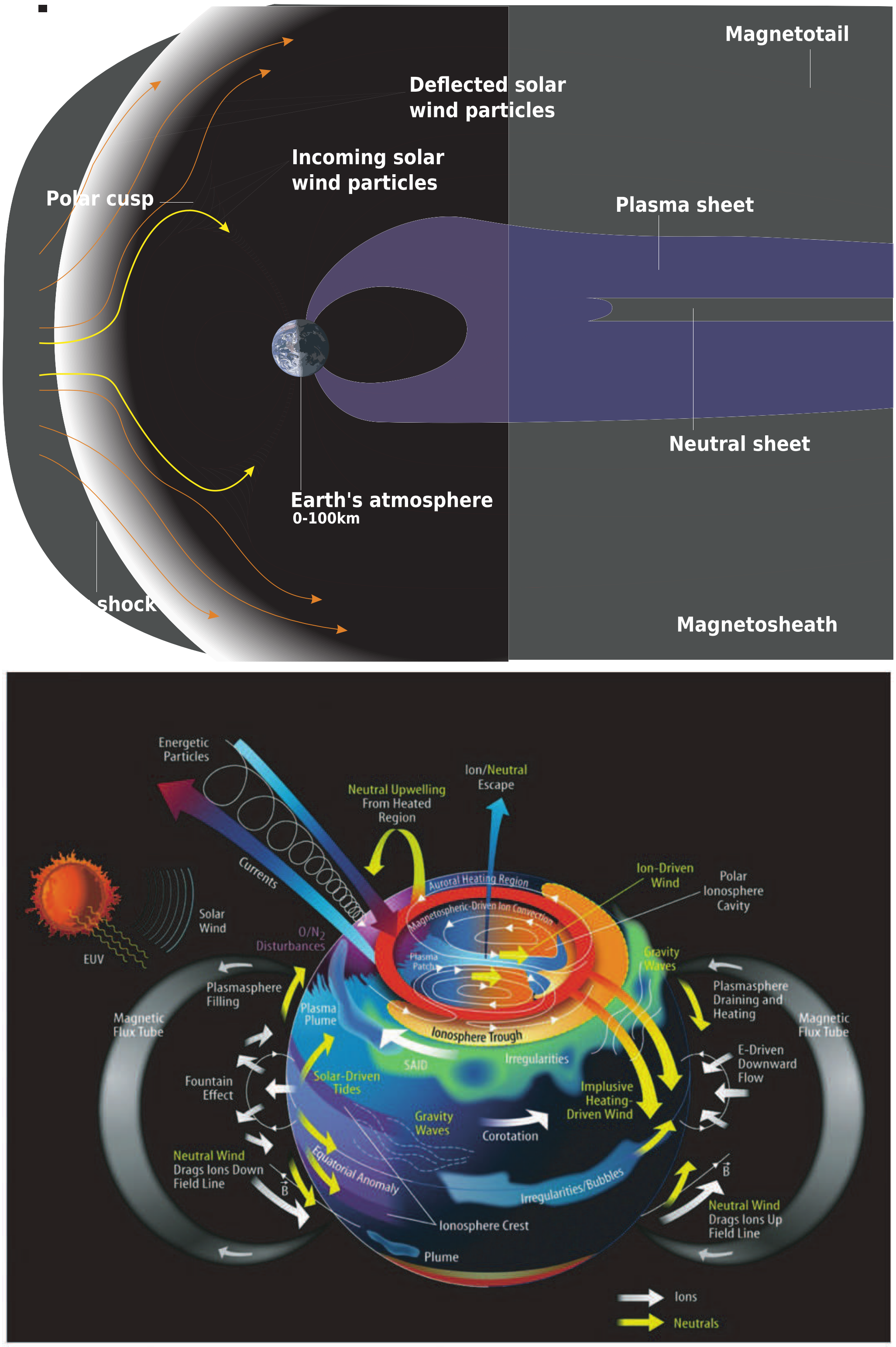


FIGURE 31 Top: the interaction of the solar wind and the Earth’s magnetic field. Bottom: the magnetic environment of the Earth (courtesy NASA).

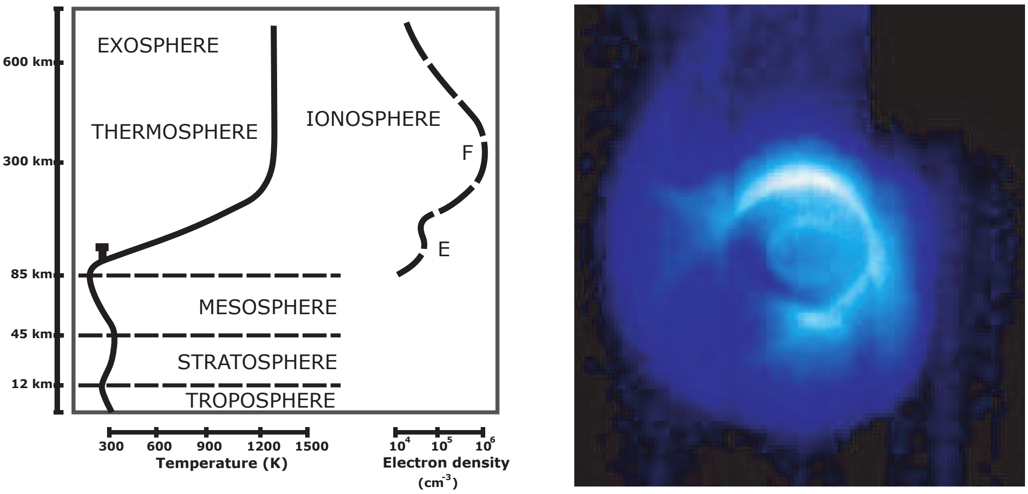


FIGURE 32 The names of the layers around the Earth and a photograph of the cold plasma, or magnetosphere, surrounding the Earth, taken in the extreme ultraviolet, and showing both the ring at the basis of each aurora and a tail pointing towards the Sun (courtesy NASA).

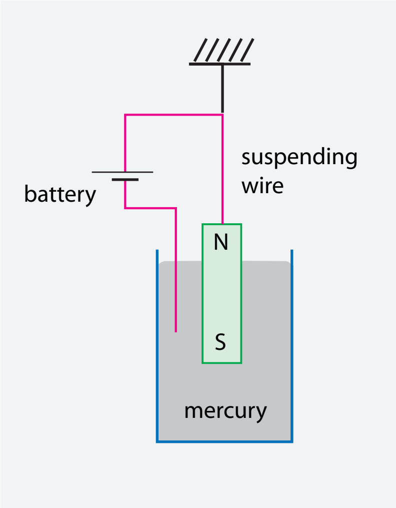


FIGURE 33 A unipolar motor.

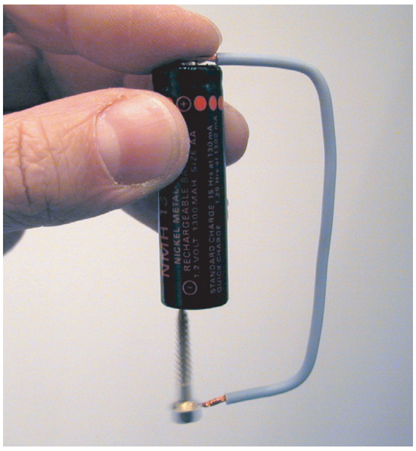


FIGURE 34 The simplest motor (© Stefan Kluge).

Challenge 53 s Why do we often see shadows of houses and shadows of trees, but never shadows of the electrical cables hanging over streets?

* *

Challenge 54 s How would you measure the speed of the tip of a lightning bolt? What range of values do you expect?

* *

Ref. 32 One of the simplest possible electric motors was discovered by Faraday in 1831. A magnet suspended in mercury will start to turn around its axis if a current flows through it. (See Figure 33.) In addition, when the magnet is forced to turn, the device (often also called Barlow's wheel) also works as a current generator; people have even tried to generate domestic current with such a system! Can you explain how it works? Challenge 55 s

The modern version of this motor makes use of a battery, a wire, a conductive samarium–cobalt magnet and a screw. The result is shown in Figure 34.

* *

Ref. 33 The magnetic field of the Earth has a dipole strength of $7.8 \cdot 10^{22}\,\mathrm{A\,m^2}$. It shields us, together with the atmosphere, from lethal solar winds and cosmic radiation particles, by deflecting them to the poles. Today, a lack of magnetic field would lead to high radiation on sunny days; but in the past, its lack would have prevented the evolution of the human species. We owe our existence to the magnetic field of the Earth. At present, the magnetic field decreases by about 5 % per century. It seems that it might disappear temporarily in 1500 years; it is unclear whether this will lead to an increase of the cosmic radiation hitting the Earth's surface, or if the solar wind itself will take over the shielding effect.

* *

Comparing electricity with water is a good way of understanding electronics. Figure 35 shows a few examples that even a teenager can use. Can you fill in the correspondence for the coil, and thus for a transformer? Challenge 56 s

The picture also includes the *transistor*. This device, as the hydraulic component shows, can be used to control a large current by using a small current. Therefore, transistors can be used as *switches* and as *amplifiers*. This is the reason that all electronic circuits, from radios to mobile phones and computers – make heavy use of transistors. A modern mobile phone or computer typically contains several million transistors, mostly assembled inside so-called *integrated circuits*. The design of these devices is a science on its own.

* *

There is even a way to push the previous analogy in another direction: it is possible to produce a mathematically consistent analogy between electric circuits and continuous fields. The required circuits are infinite grids or meshes in all directions of space, and are called *mimetic discretizations*. If you like to think in electric terms, you might enjoy pursuing this. Just search for the term on the internet.

* *

The ionosphere around the Earth has a resonant frequency of 7 Hz; for this reason any apparatus measuring low frequencies always gets a strong signal at this value. Can you give an explanation of the frequency? Challenge 57 s

* *

The Kirlian effect, which allows one to make such intriguingly beautiful photographs, is not a property of objects, but a result of the applied time-varying electric field.

* *

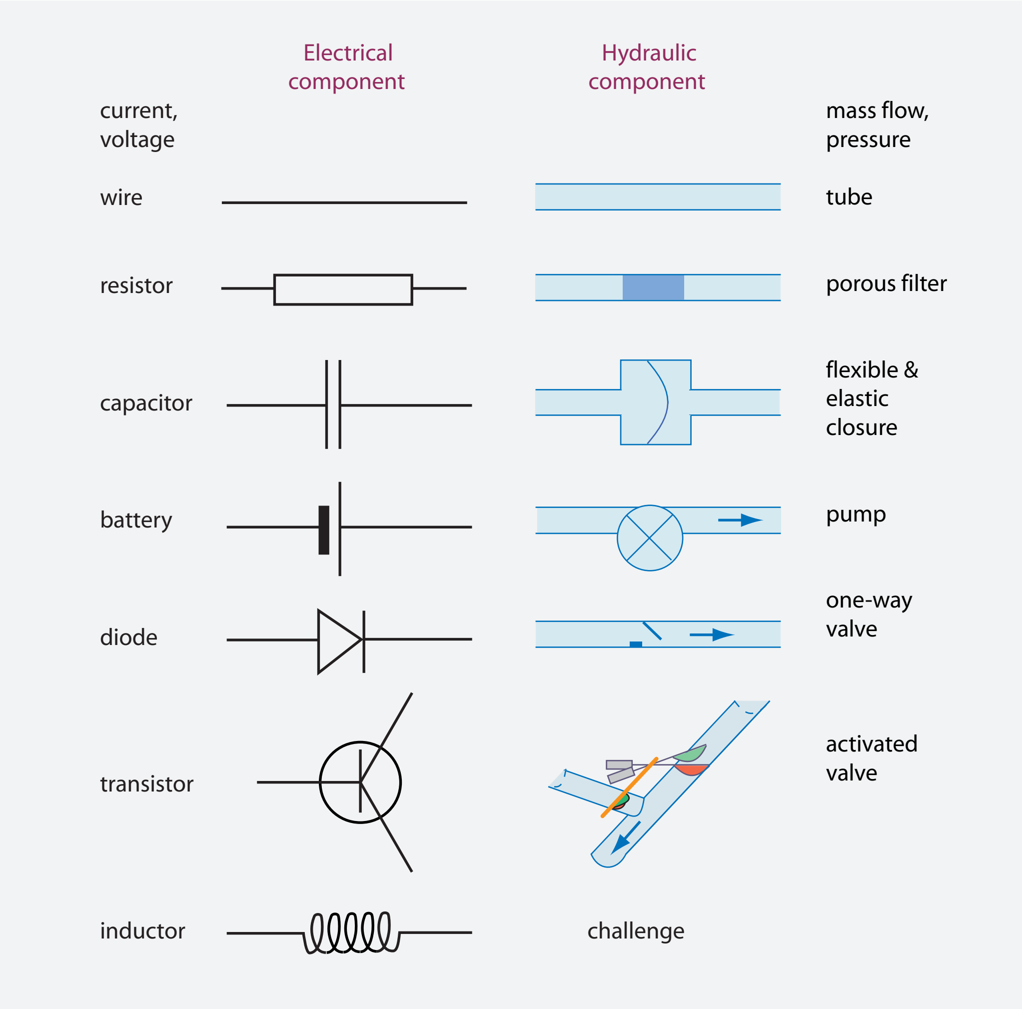


FIGURE 35 The correspondence of electronics and water flow.

At home, electricity is mostly used as alternating current. In other words, no electrons actually flow through cables; as the drift speed of electrons in copper wires is of the order of 1 μm/s, electrons just move back and forward by 20 nm. Nothing flows in or out of the cables! Why do the electricity companies require a real flow of money in return, instead of being satisfied with a back and forth motion of money?

Page 249

Challenge 58 e

* *

Do electrons and protons have the same charge? Experiments show that the values are equal to within at least twenty digits. How would you check this?

Challenge 59 ny

* *

Charge values are velocity-independent, even near the speed of light. How would you confirm this?

Challenge 60 ny

FIGURE 36 The floating bed problem: while the left model, with a length of around 40 cm and a floating height of a few centimetres, exists and has been admired by many, the scaled-up, real-size version on the right is impossible (© Janjaap Ruissenaars at www.UniverseArchitecture.com). The two images on the right are *not* photographs: they show a dream, not reality. Why?

* *

Magnets can be used, even by school children, to climb steel walls. Have a look at the www.physicslessons.com/TPNN.htm website.

* *

Can magnets be used to make a floating bed? In 2006, a Dutch architect presented to the public a small model of a beautiful floating bed, shown on the left of Figure 36, kept floating in the air by permanent magnets. To prevent that the model bed falls over, it is fastened to the ground by four ropes. On his website, the architect also offers a real-size version of the same bed, for a price of over one million US dollars. However, the images of the scaled up bed – the only two images that exist – are not photographs, but computer graphics, as this dream bed is impossible. Why? Challenge 61 s

* *

Extremely high magnetic fields have strange effects. At fields of 10^{10} T, vacuum becomes effectively birefringent, photons can split and coalesce, and atoms get squeezed. Hydrogen atoms, for example, are estimated to get two hundred times narrower in one direction. Fortunately, these conditions exist only in specific neutron stars, called *magnetars*. Page 111

* *

Ohm's 'law', the observation that for almost all materials the current I is proportional to the voltage U, is

$$U \sim I \quad \text{or} \quad \frac{U}{I} = R = \text{const.} \tag{24}$$

and is due to a school teacher. Georg Simon Ohm (b. 1789 Erlangen, d. 1854 Munich), was a school teacher and physicist. He explored the validity of the proportionality in great depth and for many materials; in those days, such measurements were difficult to perform. Ohm discovered that the proportionality applies to most materials and to many current levels, as long as the temperature, the material density and the charge densities remain constant. The proportionality is thus not valid for situations with sparks or in semiconductors. But it is valid for most solid conductors, in particular for metals.

Ohm's efforts were recognized only late in his life, and he eventually was promoted to professor at the Technical University in Munich. Later the unit of *electrical resistance* R – this is the official name for the proportionality factor between voltage, which is measured in *volt*, and current, which measured in *ampere* – was named after him. One *ohm* is defined and written as $1\,\mathrm{V/A}=1\,\Omega$.

Today, Ohm's relation is easy to measure. Recently, even the electrical resistance of single atoms has been measured: in the case of xenon it turned out to be about $10^5\,\Omega$. It was also found that lead atoms are ten times more conductive than gold atoms. Can you imagine why? Ref. 34 Challenge 62 ny

* *

Since many decades, Ohm's 'law' is taught in secondary school until every pupil in a class has lost his interest in the matter. For example, the electric power P transformed into heat in a resistor is given

$$P = UI = I^2 R = \frac{U^2}{R} \; . \qquad (25)$$

Vol. I, page 354 We mentioned this relation already earlier on; have a look. Now you know everything that needs to be known on the topic. Above all, the expression for electric power in a resistor describes electric heating, for example the heating in a modern kitchen stove or in a coffee machine.

* *

Ohm's 'law', so simple it seems, has many fascinating mathematical aspects. For example, in 1958, the Dutch physicist J.L. van der Pauw proved an astonishing formula and method that allows measuring the specific resistance ρ of material layers of *any* shape. One only needs to attach four gold wires to the layer anywhere on its border. The specific resistance is then given by the expression shown in Figure 37. Can you imagine how the formula is deduced? (This is not an easy problem.) Challenge 63 d The formula reduced the workload in laboratories across the world by a significant amount; before the formula had been discovered, in every experiment, researchers also had to produce separate, dedicated samples that allowed measuring the specific resistance of the material they were investigating.

* *

A good way to make money is to produce electricity and sell it. In 1964, a completely new method was invented by Fletcher Osterle. The method was presented to a larger public in a beautiful experiment in 2003. Ref. 35 Larry Kostiuk and his group took a plate of glass, added

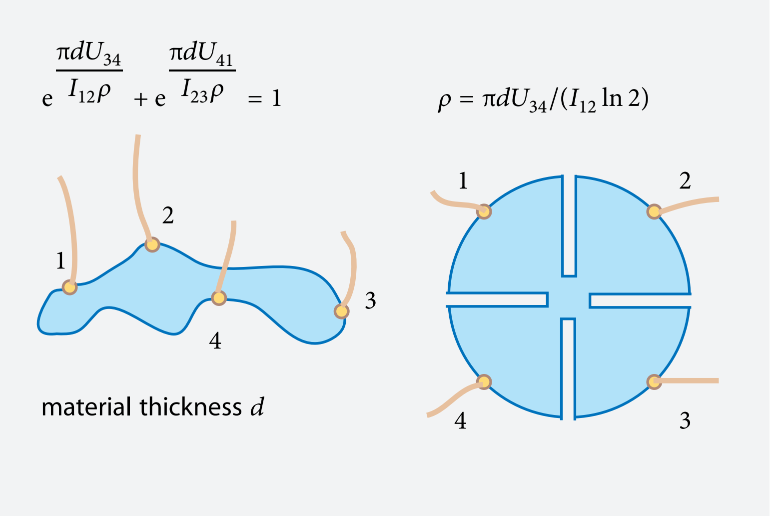


FIGURE 37 Can you deduce Van der Pauw's formula for the specific resistance ρ of homogeneous layers of any shape (left) or its special case for a symmetrical shape (right)?

a conducting layer on each side, and then etched a few hundred thousand tiny channels through the plate: 450 000 microchannels, each around 15 μm in diameter, in the 2 cm diameter plate. When water is made to flow through the channels, a current is generated. The contacts at the two conducting plates can be used like battery contacts and generated 1.5 μA of electric current.

This simple device uses the effect that glass, like most insulators, is covered with a charged layer when it is immersed in a liquid. Can you imagine why a current is generated? Challenge 64 s Unfortunately, the efficiency of electricity generation is only about 1 %, making the method much less interesting than a simple blade wheel powering a dynamo.

* *

For beautiful animations about magnetic and electric fields, see the website web.mit.edu/8.02t/www/802TEAL3D/visualizations.

* *

Electrostatics is sometimes counter-intuitive. Take an isolated, conducting sphere of radius R, and a point charge located outside the sphere, both with the same charge. Even though charges of equal sign repel each other, at small distances from the sphere, the point charge is attracted to the sphere. Why? At which distance d do they repel? Challenge 65 s

* *

Gallium arsenide semiconductors can be patterned with so-called quantum dots and *point contacts*. These structures allow one to count single electrons. This is now routinely done in several laboratories around the world.

* *

Ref. 36 The charges on two capacitors in series are not generally equal, as naive theory states. For perfect, leak-free capacitors the voltage ratio is given by the inverse capacity ratio

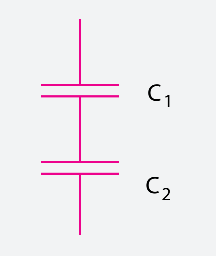


FIGURE 38 Capacitors in series.

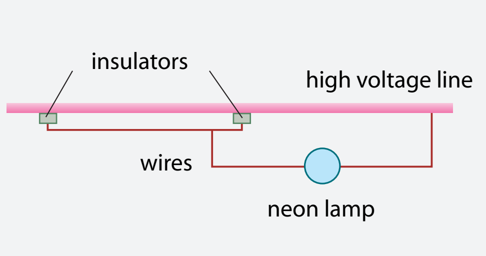


FIGURE 39 A neon lamp hanging from a high voltage line.

$V_1/V_2 = C_2/C_1$, due to the equality of the electric charges stored. This is easily deduced from Figure 38. However, in practice this is only correct for times between a few and a few dozen minutes. Why? Challenge 66 s

* *

On certain high voltage cables leading across the landscape, small neon lamps, called *balisors*, shine when the current flows, as shown in Figure 39. You can see them from the train when riding from Paris to Roissy airport. How do they work? Challenge 67 s

* *

During rain or fog, high-voltage lines often make noises; sometimes they even *sing*. What is going on? Challenge 68 s

* *

Electric polarizability is the property of matter responsible for the deviation of water flowing from a tap caused by a charged comb. Page 16 It is defined as the strength of electric dipole induced by an applied electric field. The definition simply translates the observation that many objects acquire a charge when an electric field is applied. Incidentally, how precisely combs get charged when rubbed, a phenomenon called *electrification*, is still one of the mysteries of modern science.

* *

A pure magnetic field cannot be transformed into a pure electric field by change of observation frame. The best that can be achieved is a state similar to an *equal mixture* of magnetic and electric fields. Can you provide an argument elucidating this relation? Challenge 69 s

* *

Calculating resistance of infinite grids is one of the most captivating problems in electricity, as shown in Figure 40. Can you find the solution? Challenge 70 ny

* *

To every limit value in nature there is a corresponding indeterminacy relation. This is also valid also for electricity and the lower charge limit. Indeed, there is an indeterminacy

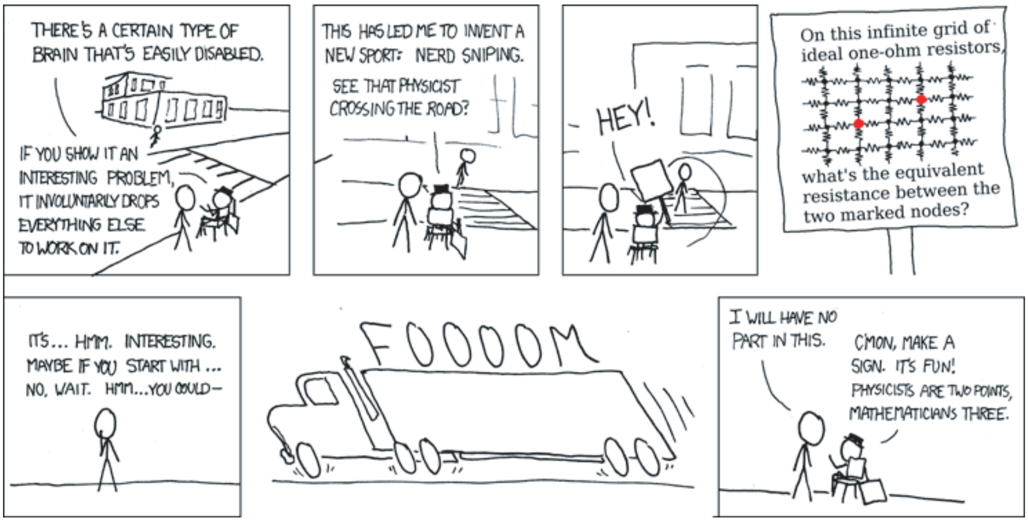


FIGURE 40 An electrical problem that is not easy (© Randall Munroe).

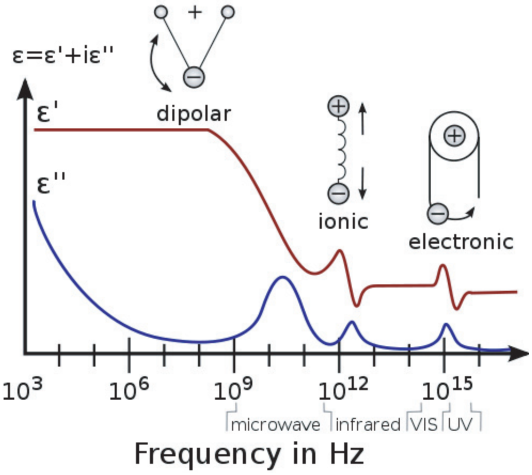


FIGURE 41 The change of the relative permittivity (real and imaginary) with frequency for an abstract material (mix), and the general processes responsible for the different domains (© Kenneth Mauritz).

relation for capacitors, of the form

$$\Delta C \, \Delta U \geqslant e \qquad (26)$$

where e is the positron charge, C capacity and U potential difference. There is also an indeterminacy relation between electric current I and time t

$$\Delta I \, \Delta t \geqslant e \, . \qquad (27)$$

Ref. 37 Both these relations may be found in the literature.

* *

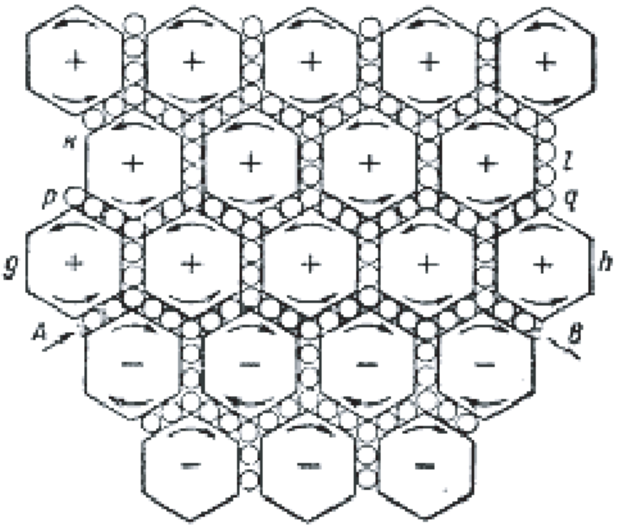


FIGURE 42 Maxwell's unsuccessful model of the vacuum.

Electric properties of materials, in contrast to their magnetic properties, vary strongly with the frequency of the applied electric field. Figure 41 illustrates how the permittivity changes with frequency, and which microscopic processes are at the basis of the property at a specific frequency. The graph is only schematic: it shows features from different materials combined together. In nature, the real and imaginary parts of the permittivity are related by the so-called *Kramers-Kronig relations*, which are important for many material topics related to wave phenomena. The two curves in the graph do not follow them completely.

* *

If an axis rotates, one can attach a magnet to its end. With such a rotating magnet an extremely cheap tachymeter can be realized. How?

Challenge 71 e

* *

In Maxwell's 1861 paper on electromagnetism, he includes Figure 42 as a model of magnetic and electric fields of the vacuum. What is the biggest problem of this model of the vacuum?

Challenge 72 s

* *

Ref. 38

For how long can silicon-based integrated circuits be made smaller and smaller? The opinions on this matter differ. Optimistic predictions, often called Moore's 'law', alternate with predictions that from 2011 onwards, the size reduction will be moderate due to the high cost of the required equipment. For example, the next generation of wafer steppers, the most expensive machines in the production of silicon chips, must work in the extreme ultraviolet – usually 13 nm – in order to achieve small transistor sizes. At this wavelength air is an absorber, and lenses have to be replaced by mirrors. It is unclear whether this will be technically and economically feasible. Future will tell.

* *

In the 1990s, microscope images showed, surprisingly, that the tusks of narwhals are full of nerve endings. Thus the tusk may be a sensory organ. However, the details and the exact use of the organ is not understood. How would you find out? Challenge 73 s

A summary: three basic facts about electricity

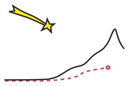
The experiments we have described so far show three basic results:

▷ Electric charges affect, thus exert force on other charges.
▷ Electric charges are conserved.
▷ Charges, like all matter, move slower than light.

From these three statements – the definition of charge, the conservation of charge, and the invariance of the speed of light – we can deduce every aspect of classical electrodynamics. An alternative summary would be: *charges are conserved; their effects obey relativity.*

Ref. 39 In particular, the Lagrangian of electrodynamics and Maxwell's field equations can be deduced from these three statements; they describe the way that charges *produce* any electric, magnetic or electromagnetic field. Also the Lorentz force can be deduced; it describes how the motion of massive charges and the motion of the electromagnetic field is related.

Ref. 39 The proof of the connection between charge conservation and the field equations can be given mathematically; we do not present it here, because the algebra is somewhat Ref. 40 involved. The essential connection to remember is: all of electrodynamics follows from the properties of charges that we have discovered so far.



Chapter 2

THE DESCRIPTION OF ELECTROMAGNETIC FIELD EVOLUTION

Electric and magnetic fields change: simply said, they move. How exactly does this happen? In the 1860s, James Clerk Maxwell[**] collected all experimental knowledge he could find, and deduced the precise description of electromagnetic field motion. Twenty years later, Heaviside and Hertz extracted the main points of Maxwell ideas from his difficult papers written in unusual quaternion notation and called their summary *Maxwell's theory of the electromagnetic field.*

Vol. IV, page 232

The motion of the electromagnetic field is described by a set of evolution equations. In the relativistic description, the set consists of *two* equations, in the non-relativistic case of *four* equations. All observations of classical electrodynamics follow from these equations. In fact, if quantum effects are properly taken into account, *all* electromagnetic effects of nature are described.

The first field equation of electrodynamics

The first relativistic field equation of electrodynamics is the precise statement that electromagnetic fields *originate at charges*, and nowhere else. It can be written[***]

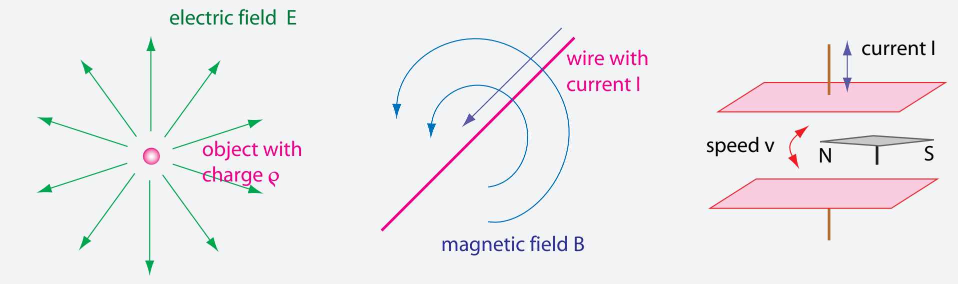
$$d\boldsymbol{F} = j\mu_0$$

or, equivalently

$$\nabla\cdot\boldsymbol{E} = \frac{\rho}{\varepsilon_0} \quad \text{and} \quad \nabla\times\boldsymbol{B} - \frac{1}{c^2}\frac{\partial\boldsymbol{E}}{\partial t} = \mu_0\boldsymbol{j}\ . \tag{28}$$

[**] James Clerk Maxwell (b. 1831 Edinburgh, d. 1879 Cambridge) is one of the most important and influential physicists. He founded electromagnetism by theoretically unifying electricity and magnetism, as described in this chapter. His work on thermodynamics forms the second pillar of his activity. In addition, he studied the theory of colours and developed the colour triangle; he was one of the first people to make a colour photograph. He is regarded by many as the greatest physicist ever. Both 'Clerk' and 'Maxwell' were his family names.

[***] There is a certain freedom in writing the equations, because different authors absorb different combinations of the constants c and μ_0 into the definitions of the quantities F, A and j. The one given here is the most common version. The equations can be generalized to cases where the charges are not surrounded by vacuum, but located inside matter. We will not explore these situations in our walk because, as we will discover later on, the seemingly special case of vacuum in fact describes all of nature.



Charges are sinks or sources of electric field lines.

Currents have magnetic vortex field lines wrapped around them.

Changing electric fields produce magnetic fields.

FIGURE 43 The first of Maxwell's field equations of electrodynamics illustrated in three drawings.

Each of these two equivalent ways* to write the first Maxwell equation makes a simple statement:

▷ *Electrical charges carry the electromagnetic field.* They carry it along with them.

For example, this first equation describes the attraction of dust by electrically charged objects and the working of electromagnets.

This first field equation is equivalent to the three basic observations illustrated in Figure 43: *Coulomb's 'law'* on the attraction and repulsion of charges, *Ampère's 'law'* on the attraction and repulsion of current-carrying wires, and *Maxwell's addition,* the observation that changing electric fields induce magnetic effects. More precisely, if we know where charges are and how they move, we can determine the electromagnetic field $\boldsymbol{F}$ that they generate. *Static charges*, described by a density ρ, produce *electrostatic* fields, and moving charges, described by a 3-current density $\boldsymbol{j}$, produce a mix of electric and magnetic fields. *Stationary currents* produce *magnetostatic* fields. In general, moving charges produce moving fields.

The first field equation also contains the *right hand rule* for magnetic fields around wires, through the vector product. And as already mentioned, the equation also states, most clearly in its last form, that changing electric fields induce magnetic fields. The effect is essential in the primary side of transformers. The small factor $1/c^2$ implies that the effect is small; therefore coils with *many* windings or *strong* electric currents are needed to produce or detect the effect.

Challenge 74 e

---

* In component form, the first equation can be written

$$d_\mu \boldsymbol{F}^{\mu\nu} = j^\nu \mu_0 = (\rho c, \boldsymbol{j})\mu_0 = (\rho_0 \gamma c, \rho_0 \gamma \boldsymbol{v})\mu_0 \quad \text{or}$$

$$(\partial_t/c, \partial_x, \partial_y, \partial_z)\begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix} = \mu_0(\rho c, \boldsymbol{j}) \ . \qquad (29)$$

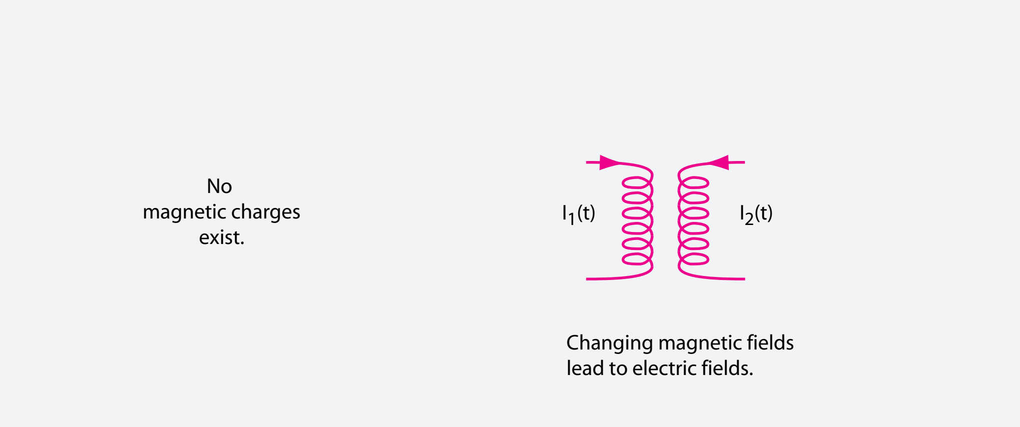


FIGURE 44 The second field equation of electrodynamics.

THE SECOND FIELD EQUATION OF ELECTRODYNAMICS

The second of Maxwell's field equations, illustrated in Figure 44, expresses the observation that in nature there are *no* magnetic charges, i.e., that magnetic fields have no sources. As a result, the equation also gives a precise description of how changing magnetic fields create electric fields, and vice versa – often called *Faraday's 'law'*. The second of Maxwell's equations for electrodynamics can be written

$$d\,{}^*\boldsymbol{F} = 0 \quad \text{with} \quad {}^*\boldsymbol{F}^{\rho\sigma} = \tfrac{1}{2}\varepsilon^{\rho\sigma\mu\nu}\boldsymbol{F}_{\mu\nu}$$

or, equivalently

$$\nabla\cdot\boldsymbol{B} = 0 \quad \text{and} \quad \nabla\times\boldsymbol{E} = -\frac{\partial\boldsymbol{B}}{\partial t}\ . \qquad (30)$$

The second field equation* thus expresses the *lack of sources for the dual field tensor* ${}^*\boldsymbol{F}$. In other words,

▷ *In nature there are no magnetic charges, i.e., no magnetic monopoles.*

---

* In component form, the second Maxwell equation can be written

$$d_\mu{}^*\boldsymbol{F}^{\mu\nu} = 0 \quad \text{or}$$

$$(\partial_t/c, \partial_x, \partial_y, \partial_z)\begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z/c & -E_y/c \\ B_y & -E_z/c & 0 & E_x/c \\ B_z & E_y/c & -E_x/c & 0 \end{pmatrix} = (0,0,0,0) \quad \text{or}$$

$$\varepsilon^{\sigma\mu\nu\rho}\partial_\mu\boldsymbol{F}_{\nu\rho} = 0 \quad \text{or}$$

$$\partial_\mu\boldsymbol{F}_{\nu\rho} + \partial_\nu\boldsymbol{F}_{\rho\mu} + \partial_\rho\boldsymbol{F}_{\mu\nu} = 0\ . \qquad (31)$$

Page 91

We note that the dual tensor ${}^*\boldsymbol{F}$ follows from the field tensor $\boldsymbol{F}$ by substituting E/c by B and B by $-E/c$. This is the so-called *duality transformation*. More on this duality below.

There are no sources for magnetic fields. The second field equation thus states that cutting a magnet with a north and a south pole in any way always produces pieces with *two* poles, never a piece with a single pole.

Since there are no magnetic charges, magnetic field lines have *no beginning and no end*; not only the magnetic field lines induced by charges, no, *all* magnetic field lines have no beginning and no end. For example, field lines continue inside magnets. The lack of beginnings and ends is expressed mathematically by stating that the *magnetic flux* through a *closed* surface S – such as a sphere or a cube – *always vanishes*: $\int_S \boldsymbol{B}\,\mathrm{d}\boldsymbol{A} = 0$. In other words, all field lines that enter a closed volume also leave it.* No magnetic flux leaves a volume. This is often called the *magnetic Gauss 'law'*.

Furthermore, the second field equation expresses

▷ *Changes in magnetic fields produce electric fields.*

This effect is used in the secondary side of transformers and in dynamos. The cross product in the expression implies that an electric field generated in this way – also called an *electromotive field* – has no start and end points. The electromotive field lines thus can run in circles: in most practical cases they run along electric circuits. In short, an electric field can have vortices (like the magnetic field), but only when there is a changing magnetic field. The minus sign is essential to ensure energy conservation (why?) and has a special name: it is called *Lenz's rule*. Challenge 75 ny

In practice, the second Maxwell equation is always needed together with the first. Can you see why? Challenge 76 ny

THE VALIDITY AND THE ESSENCE OF MAXWELL'S FIELD EQUATIONS

We saw above that Lorentz' evolution equation

$$\begin{aligned} & m\boldsymbol{b} = q\boldsymbol{F}\boldsymbol{u} \\ & \text{or, equivalently} \\ & \mathrm{d}E/\mathrm{d}t = q\boldsymbol{E}\boldsymbol{v} \quad \text{and} \quad \mathrm{d}\boldsymbol{p}/\mathrm{d}t = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \end{aligned} \tag{32}$$

describes how charges move given the motion of the fields. Together with Lorentz' evolution equation, the two Maxwell's evolution equations (28) and (30) describe *all* electromagnetic phenomena occurring on everyday scales, from mobile phones, car batteries, to personal computers, lasers, lightning, holograms and rainbows. In other words, this description of electromagnetic fields is complete for everyday life. Only quantum effects and the effects of curved space-time are not included.

Maxwell's equations seem very complex. But we should not forget that they contain only *four* basic ideas.

---

* In contrast to what is often said and written in physics books, magnetic field lines are, in general, *not* closed lines; they are *not*, in general, loops or vortex lines. Closed magnetic field lines occur only for straight wires; they are not even loops for simple helical coils. In fact, in all usual, non-academic situations, magnetic field lines start and end at spatial infinity. Ref. 41

Magnetic field *lines* are a mathematical tool, they do not provide a completely useful description of the magnetic field. The magnetic field is best described by its *vector field*.

1. Electric charges follow Coulomb's 'law'.
2. Electric charges moves slower than light.
3. Electric charges are conserved.
4. Magnetic charges do not exist.

If we want to be really simplistic, Maxwell's equations are just the relativistic formulation of Coulomb's 'law'. Indeed, as we have seen before, Maxwell's equations follow from charge conservation alone. Ref. 39

Maxwell's equations remain fascinating to this day. Their applications are numerous, from industry to life-saving medicine, from toys and music to materials science, fusion research and astronomy. Transport, telecommunication, computers, electronics, most jobs, human life and practically all of its pleasures depend on electricity and magnetism. Already in 1899, after Heinrich Hertz put Maxwell's equations into modern form, he said and wrote:

> Man kann diese wunderbare Theorie nicht studieren, ohne bisweilen die Empfindung haben, als wohne den mathematischen Formeln selbständiges Leben und eigener Verstand inne, als seien dieselben klüger als wir, klüger sogar als ihre Erfinder, als gäben sie mehr heraus, als seinerzeit in sie hineingelegt wurde. *

When Ludwig Boltzmann wrote his book about electromagnetism in 1893, he added the following lyrical motto at the beginning of the chapter on Maxwell's equations:

> War es ein Gott der diese Zeichen schrieb,
> Die mit geheimnisvoll verborgnen Trieb
> Die Kräfte der Natur um mich enthüllen
> Und mir das Herz mit stiller Freud erfüllen?**

Indeed, the Maxwell's formulae have retained their fascination. New applications are still being found and developed every year, all over the world.

In this adventure, will not explore many applications of the field equations. We leave most of them aside and continue directly towards our aim to understand the connection between electromagnetic fields, everyday motion and the motion of light. In fact, the electromagnetic field has an important property that we mentioned right at the start: the field itself can move. In particular, the field can carry energy, linear momentum and angular momentum.

---

* 'One cannot study this wonderful theory without sometimes having the feeling that these mathematical formulae contain independent life and their own intelligence, that they are smarter than us, smarter even than their discoverers, and that they give us more than was originally put into them.

** 'Was it a god who wrote these signs / which with secret hidden drive / uncover nature's forces around me / and fill my heart with silent joy?' These four lines by Boltzmann are a paraphrase of four lines from Goethe's *Faust*.

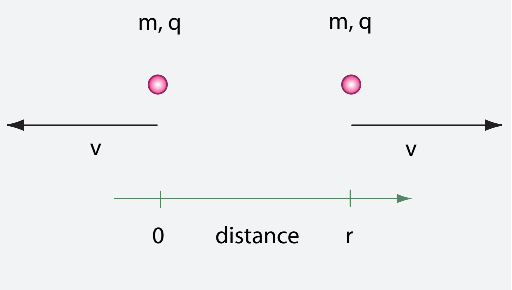


FIGURE 45 Charged particles after a collision.

COLLIDING CHARGED PARTICLES

Electromagnetic fields move. A simple experiment clarifies the meaning of motion for fields: When two charged particles collide, their total momentum is *not* conserved. Let us check this.

Imagine two particles of identical mass and identical charge just after a collision, when they are moving away from one another. The situation is illustrated in Figure 45. Imagine also that the two masses are large, so that the acceleration due to their electrical repulsion is small. For an observer at the centre of gravity of the two, each particle feels an acceleration from the electric field of the other. This electric field E is given by the so-called *Heaviside formula* Challenge 77 ny

$$E = \frac{q\,(1-v^2/c^2)}{4\pi\varepsilon_0 r^2} \ . \tag{33}$$

In other words, the total system has a vanishing total momentum for this observer.

Take a second observer, moving with respect to the first with velocity v, so that the first charge will be at rest. Expression (33) leads to two *different* values for the electric fields, one at the position of each particle. In other words, the system of the two particles is not in inertial motion, as we would expect; the total momentum is not conserved for this observer. The missing momentum is small, but where did it go? Ref. 42 Challenge 78 s

This at first surprising effect has even been put in the form of a theorem by Van Dam and Wigner. They showed that, for a system of particles interacting at a distance, the total particle energy–momentum cannot remain constant in all inertial frames. Ref. 43

The total momentum of the system is conserved only because

▷ The electromagnetic field itself also carries some momentum.

In short, momentum is conserved in the experiment, but some of it is carried by the field. The precise amount depends on the observer.

Two colliding charged particles thus show us that electromagnetic fields have momentum. If electromagnetic fields have momentum, they are able to *strike* objects and to be struck by them. As we will show below, light is also an electromagnetic field. Thus we should be able to move objects by shining light on to them. We should even be able to suspend particles in mid air by shining light on to them from below. Both predictions are correct, and some experiments will be presented shortly. Page 120

We conclude that any sort of field leading to particle interactions must carry both energy and momentum, as the argument applies to all such cases. In particular, it applies to nuclear interactions. Indeed, in the quantum part of our adventure we will even find an additional result: all fields are themselves composed of particles. The energy and momentum of fields then become an obvious state of affairs. In short, it makes sense to say that electromagnetic fields move, because they carry energy and momentum.

What is contact?

The exploration of collisions, together with the result that matter consists of charged particles, allows us to deduce

▷ Everyday contact is the exchange of electromagnetic fields.

In particular, we learn that actual contact does not exist in everyday life.

▷ In everyday contact, *nothing actually touches* anything else.

We have to bury a dream that has guided thinkers for centuries: the world is not mechanical. All processes around us are either electric or gravitational.

The gauge field – the electromagnetic vector potential[*]

The study of moving fields is called *field theory* and electrodynamics is the prime example. (The other classical example is fluid dynamics; moving electromagnetic fields and moving fluids are very similar mathematically.) Field theory is a beautiful topic; field lines, equipotential lines and vortex lines are some of the concepts introduced in this domain. They fascinate many.[**] However, in this mountain ascent we keep the discussion focused on motion.

We have seen that fields force us to extend our concept of motion. Motion is not only the change in state of objects and of space-time, but also the *change in state of fields*. We therefore need, also for fields, a complete and precise description of their state.

The observations using amber and magnets have shown us that *electromagnetic fields possess energy and momentum*. Fields can impart energy and momentum to particles. The experiments with motors have shown us that objects can add energy and momentum to fields. We therefore need to define a *state function* which allows us to define energy and momentum for electric and magnetic fields. And since electric and magnetic fields transport energy, their motion must follow the speed limit in nature.

Hertz and Heaviside defined the state function of fields in two standard steps. The first step is the definition of the *(magnetic) vector potential*, which describes the momentum

---

[*] This section can be skipped at first reading.

Challenge 79 s

[**] What is the relation, for static fields, between field lines and (equi-) potential surfaces? Can a field line cross a potential surface twice? For more details on topics such as these, see the *free* textbook by Bo Thidé, *Electromagnetic Field Theory*, on his www.plasma.uu.se/CED/Book website. And of course, in English, have a look at the texts by Schwinger and by Jackson.

Ref. 1, Ref. 24

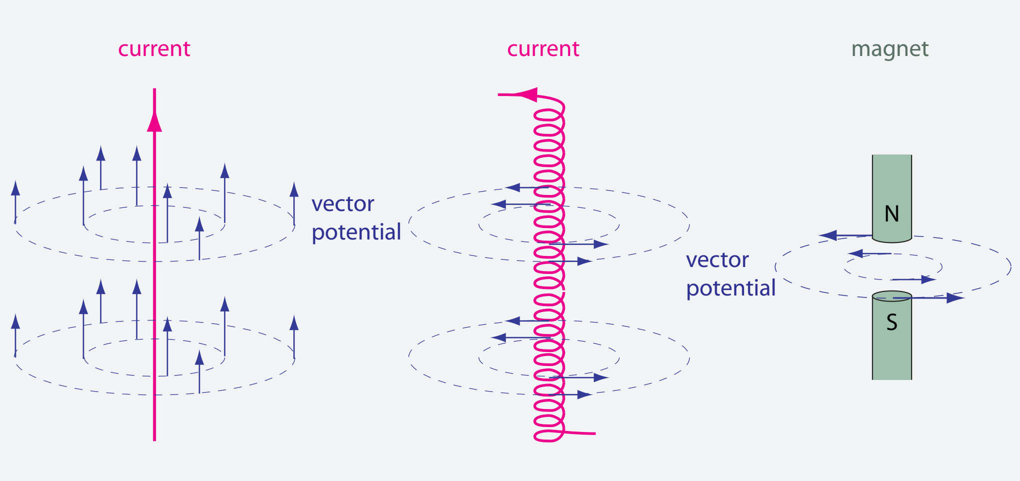


FIGURE 46 Vector potentials for selected situations.

Ref. 44 per charge that the field provides:

$$\boldsymbol{A} = \frac{\boldsymbol{p}}{q} . \tag{34}$$

When a charged particle moves through a magnetic potential $\boldsymbol{A}(\boldsymbol{x})$, its momentum changes by $q\Delta\boldsymbol{A}$; it changes by the difference between the potential values at the start and end points, multiplied by its charge. Owing to this definition, the vector potential has the property that

$$\boldsymbol{B} = \nabla \times \boldsymbol{A} = \text{curl } \boldsymbol{A} \tag{35}$$

i.e., that the magnetic field is the *curl* of the magnetic potential. In most other languages the curl is called the *rotation* and abbreviated rot. To visualize what the curl or rotation is, imagine that the field vectors are the velocity vectors of flowing air. Now put a tiny paddle-wheel at a point, as shown in Figure 47. If it turns, the curl is non-zero. The rotation speed of the paddle-wheel is maximal for some direction of the axis; this maximal speed defines both the magnitude and the direction of the curl at the point. (The right-hand rule is implied.) For example, the curl for the velocities of a rotating solid body is everywhere 2ω, or twice the angular velocity. Challenge 80 ny

Ref. 45 The vector potential for a long straight current-carrying wire is parallel to the wire; it has the magnitude Challenge 81 ny

$$A(r) = -\frac{\mu_0 I}{4\pi} \ln \frac{r}{r_0} , \tag{36}$$

which depends on the radial distance r from the wire and an integration constant r_0. This expression for the vector potential, pictured in Figure 46, shows how the moving current produces a linear momentum in the (electro-) magnetic field around it. In the case of a solenoid, the vector potential ‘circulates’ around the solenoid. The magnitude

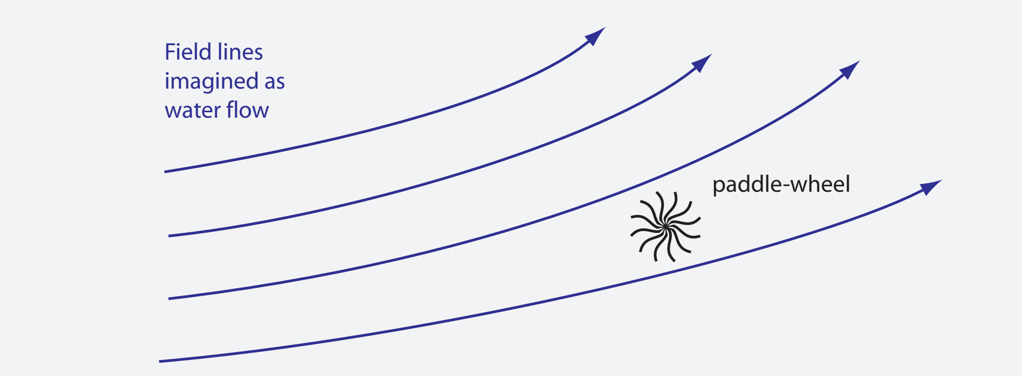


FIGURE 47 Visualizing the curl of a vector field. Imagine the field to be flowing air and check whether the small paddle-wheel rotates; if it does, the local curl is non-zero. The direction of the curl is the direction of the paddle-wheel axis that yields the highest rotation velocity.

obeys

$$A(\mathrm{r}) = -\frac{\Phi}{4\pi}\frac{1}{r} \;, \tag{37}$$

where Φ is the magnetic flux inside the solenoid. We see that, in general, the vector potential is *dragged along* by moving charges. The dragging effect decreases for larger distances. This fits well with the image of the vector potential as the momentum of the electromagnetic field.

This behaviour of the vector potential around charges is reminiscent of the way honey is dragged along by a spoon moving in it. In both cases, the dragging effect decreases with distance. However, the vector potential, unlike the honey, does *not* produce any friction that slows down charge motion. The vector potential thus behaves like a frictionless liquid.

Inside the solenoid, the magnetic field is constant and uniform. For such a field $\boldsymbol{B}$ we find the vector potential Challenge 82 e

$$\boldsymbol{A}(\mathrm{r}) = \frac{1}{2}\boldsymbol{B}\times\boldsymbol{r} \;. \tag{38}$$

In this case, the magnetic potential thus increases with increasing distance from the origin.[*] In the centre of the solenoid, the potential vanishes. The analogy of the dragged honey gives exactly the same behaviour.

However, there is a catch. The magnetic potential is *not* defined uniquely. If $\boldsymbol{A}(\boldsymbol{x})$ is a vector potential, then the different vector potential

$$\boldsymbol{A}'(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x}) + \nabla\Lambda \;, \tag{39}$$

where $\Lambda(t,\boldsymbol{x})$ is some scalar function, is *also* a vector potential for the same situation. (The magnetic field $\boldsymbol{B}$ stays the same, though.) Worse, can you confirm that the corres-

---

[*] This is only possible as long as the field is constant; since all fields drop again at large distances – because the energy of a field is always finite – also the vector potential drops at large distances.

Challenge 83 ny ponding (absolute) momentum values also change? This unavoidable ambiguity, called *gauge invariance* or *gauge symmetry*, is a central property of the electromagnetic field. We will explore it in more detail below.

Not only the momentum, but also the energy of the electromagnetic field is defined ambiguously. Indeed, the second step in the specification of a state for the electromagnetic field is the definition of the *electric potential* as the energy U per charge: Ref. 44

$$\varphi = \frac{U}{q} \tag{40}$$

In other words, the potential $\varphi(\boldsymbol{x})$ at a point $\boldsymbol{x}$ is the energy needed to move a unit charge to the point $\boldsymbol{x}$ starting from a point where the potential vanishes. The potential energy is thus given by $q\varphi$. From this definition, the electric field $\boldsymbol{E}$ is simply the *change* of the potential with position corrected by the time dependence of momentum, i.e.,

$$\boldsymbol{E} = -\nabla\varphi - \frac{\partial}{\partial t}\boldsymbol{A} , \tag{41}$$

Obviously, there is a freedom in the choice of the definition of the potential. If $\varphi(\boldsymbol{x})$ is a possible potential, then

$$\varphi'(\boldsymbol{x}) = \varphi(\boldsymbol{x}) - \frac{\partial}{\partial t}\Lambda \tag{42}$$

is also a potential function for the same situation. This freedom is the generalization of the freedom to define energy up to a constant. Nevertheless, the electric field $\boldsymbol{E}$ remains the same for all potentials.

Ref. 44 To be convinced that the potentials really are the energy and momentum of the electromagnetic field, we note that for a moving charge we have Challenge 84 ny

$$\begin{aligned} \frac{d}{\mathrm{d}t}\left(\frac{1}{2}mv^2 + q\varphi\right) &= \frac{\partial}{\partial t}q\left(\varphi - \boldsymbol{v}\boldsymbol{A}\right) \\ \frac{d}{\mathrm{d}t}\left(m\boldsymbol{v} + q\boldsymbol{A}\right) &= -\nabla q\left(\varphi - \boldsymbol{v}\boldsymbol{A}\right) , \end{aligned} \tag{43}$$

which show that the changes of generalized energy and momentum of a particle (on the left-hand side) are due to the change of the energy and momentum of the electromagnetic field (on the right-hand side).*

In relativistic 4-vector notation, the energy and the momentum of the field appear together in one quantity. The state function of the electromagnetic field becomes

$$A^\mu = (\varphi/c, \boldsymbol{A}) \tag{44}$$

and is called the *4-potential*. It is easy to see that the description of the field is complete,

---

* This connection also shows why the expression $P^\mu - qA^\mu$ appears so regularly in formulae; indeed, it plays a central role in the quantum theory of a particle in the electromagnetic field.

since we have

$$\boldsymbol{F} = d\,A \quad \text{or} \quad \boldsymbol{F}^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \quad (\text{and} \quad \boldsymbol{F}_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu)\,, \qquad (45)$$

which means that the electromagnetic field F is completely specified by the 4-potential A.* But as just said, the 4-potential itself is *not* uniquely defined. Indeed, any other equivalent 4-potential A' is related to A by the *gauge transformation*

$$A'^{\,\mu} = A^\mu + \partial^\mu \Lambda \qquad (46)$$

where $\Lambda = \Lambda(t, x)$ is any arbitrarily chosen scalar field. The new field A' leads to the *same* electromagnetic field, and to the same accelerations and evolutions. The 4-potential A is thus an *overdescription* of the physical situation as several *different* gauge choices correspond to the *same* physical situation.** Therefore we have to check that all measurement results are independent of gauge transformations, i.e., that all observables are gauge invariant quantities. Such gauge invariant quantities are, as we just saw, the fields $\boldsymbol{F}$ and ${}^*\boldsymbol{F}$, and in general all classical quantities. We add that many theoretical physicists use the term 'electromagnetic field' loosely for both the quantities $F^{\mu\nu}$ and A^μ.

There is a simple image, due to Maxwell, to help overcoming the conceptual difficulties of the vector potential. It turns out that the closed line integral over A_μ is gauge invariant, because

Challenge 86 e

$$\oint A^\mu \mathrm{d}x_\mu = \oint (A^\mu + \partial^\mu \Lambda)\mathrm{d}x_\mu = \oint A'^\mu \mathrm{d}x_\mu\,. \qquad (47)$$

In other words, if we picture the vector potential as a quantity allowing us to associate a number to a tiny ring at each point in space, we get a good, gauge invariant picture of the vector potential.***

Now that we have defined a state function that describes the energy and momentum of the electromagnetic field, let us look at what happens in more detail when electromagnetic fields move.

The Lagrangian of electromagnetism****

Instead of using the field and Lorentz equations, the motion of a charged particle and the related motion of the electromagnetic field can also be described using a Lagrangian. It is not hard to see that the action S_{CED} for a particle in classical electrodynamics can be

---

* The connection between A_μ and A^μ, the same as for every other 4-vector, was mentioned earlier on; can you restate it?

Challenge 85 e

** Choosing a function Λ is often called *choosing a gauge*; the 4-potential A is also called the *gauge field*. These strange terms have historic reasons and are now common to all of physics.

Ref. 46

*** In the part of the text on quantum theory we will see that the exponent of this expression, namely $\exp(iq \oint A_\mu \mathrm{d}x^\mu)/\hbar$, usually called the *phase factor*, can indeed be directly observed in experiments.

**** This section can be skipped at first reading.

Challenge 87 ny symbolically defined by*

$$S_{\mathrm{CED}} = -c^2 m \int \mathrm{d}\tau - \tfrac{1}{4\mu_0} \int \boldsymbol{F} \wedge {*\boldsymbol{F}} - \int j \wedge A \,, \tag{48}$$

which in index notation becomes

$$S_{\mathrm{CED}} = -mc \int_{-\infty}^{\infty} \sqrt{\eta_{\mu\nu} \frac{\mathrm{d}x_n^{\mu}(s)}{\mathrm{d}s} \frac{\mathrm{d}x_n^{\nu}(s)}{\mathrm{d}s}}\ \mathrm{d}s - \int_{\mathbf{M}} \left(\tfrac{1}{4\mu_0} \boldsymbol{F}_{\mu\nu} \boldsymbol{F}^{\mu\nu} + j_\mu A^\mu \right) \mathrm{d}^4 x \,, \tag{49}$$

or, in 3-vector notation

$$S_{\mathrm{CED}} = -c^2 m \int \mathrm{d}\tau + \int (q\boldsymbol{v}\boldsymbol{A} - q\varphi)\, \mathrm{d}t\mathrm{d}V + \int \left(\frac{\varepsilon_0}{2} \boldsymbol{E}^2 - \frac{1}{2\mu_0} \boldsymbol{B}^2 \right) \mathrm{d}t\mathrm{d}V \,. \tag{50}$$

The new part is the measure of the change – or action – due to the electromagnetic field. The pure field change is given by the term $\boldsymbol{F} \wedge {*\boldsymbol{F}}$, and the change due to interaction with matter is given by the term $j \wedge A$.

The least action principle, as usual, states that the change in a system is always as small as possible. The action S_{CED} leads to the evolution equations by requiring that the action be stationary under variations δ and δ' of the positions and of the fields which vanish at infinity. In other terms, the principle of least action requires that

$$\begin{aligned} \delta S = 0 \quad \text{when} \quad & x_\mu = x_\mu + \delta_\mu \quad \text{and} \quad A_\mu = A_\mu + \delta'_\mu \quad , \\ & \text{provided} \quad \delta x_\mu(\theta) \to 0 \quad \text{for} \quad |\theta| \to \infty \\ & \text{and} \quad \delta A_\mu(x_\nu) \to 0 \quad \text{for} \quad |x_\nu| \to \infty \,. \end{aligned} \tag{51}$$

Vol. I, page 254 Challenge 88 ny In the same way as in the case of mechanics, using the variational method for the two variables A and x, we recover the evolution equations for particle position and fields

$$b^\mu = \frac{q}{m} \boldsymbol{F}^\mu_\nu u^\nu \quad , \quad \partial_\mu \boldsymbol{F}^{\mu\nu} = j^\nu \mu_0 \quad , \quad \text{and} \quad \varepsilon^{\mu\nu\rho\sigma} \partial_\nu \boldsymbol{F}_{\rho\sigma} = 0 \,, \tag{52}$$

which we know already: they are the Lorentz relation and the two field equations. Obviously, they are equivalent to the variational principle based on S_{CED}. Both descriptions have to be completed by specifying *initial conditions* for the particles and the fields, as well as *boundary conditions* for the latter. We need the first and zeroth derivatives of the position of the particles, and the zeroth derivative for the electromagnetic field.

With the Lagrangian (48) all of classical electrodynamics can be described and understood. For the rest of our exploration of electrodynamics, we look at some specific topics from this vast field.

Ref. 48 * The product described by the symbol ∧, 'wedge' or 'hat', and the duality operator ∗ have a precise mathematical meaning. The background, the concept of *(mathematical) form*, carries us too far from our walk.

THE ENERGY–MOMENTUM TENSOR AND ITS SYMMETRIES OF MOTION

We know from classical mechanics that we get the definition of energy and momentum by using Noether's theorem. In particular, both the definition and the conservation of energy and momentum arise from the Lorentz symmetry of the Lagrangian. For example, we found that relativistic particles have an energy–momentum *vector*. At the point at which the particle is located, it describes its energy and momentum.

Since the electromagnetic field is not a localized entity, like a point particle, but an extended entity, a full description is more involved. In order to describe the energy–momentum of the electromagnetic field completely, we need to know the *flow* of energy and momentum at every point in space, separately *for each direction*. This makes a description with a *tensor* necessary, the so-called *energy–momentum tensor* $\boldsymbol{T}$ of the electromagnetic field. Vol. II, page 196

The electric field times a charge is the force on that charge, or equivalently, its momentum increase per time. The generalization for the full electromagnetic field $\boldsymbol{F}$, and for the full power–force (or 4-force) vector $\boldsymbol{K}$ is

$$F^{\mu\nu} j_\mu = K^\nu = \partial_\mu \boldsymbol{T}^{\mu\nu} \quad . \tag{53}$$

This short equation, which can also be derived from the Lagrangian, contains a lot of information. In particular, it expresses that every change in energy of the field is the sum of the energy radiated away (via the energy flow described by the Poynting vector $\boldsymbol{S}$) and of change in the kinetic energy of the charges. The equation also makes a similar statement on the momentum of the electromagnetic field.

The detailed parts of the energy–momentum tensor $\boldsymbol{T}$ are found to be

$$\begin{aligned} \boldsymbol{T}^{\mu\nu} &= \left(\begin{array}{c|c} \text{energy} & \text{energy flow or} \\ \text{density} & \text{momentum density} \\ \hline \text{energy flow or} & \text{momentum} \\ \text{momentum density} & \text{flow density} \end{array}\right) \\ &= \left(\begin{array}{c|c} u & \boldsymbol{S}/c = c\boldsymbol{p} \\ \hline c\boldsymbol{p} & T \end{array}\right) = \left(\begin{array}{c|c} (\varepsilon_0 E^2 + B^2/\mu_0)/2 & \varepsilon_0 c \boldsymbol{E} \times \boldsymbol{B} \\ \hline \varepsilon_0 c \cdot & -\varepsilon_0 E_i E_j - B_i B_j/\mu_0 \\ \boldsymbol{E} \times \boldsymbol{B} & 1/2\delta_{ij}(\varepsilon_0 E^2 + B^2/\mu_0) \end{array}\right) \end{aligned} \tag{54}$$

where $\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{B}/\mu_0$ is the *Poynting vector* that describes the energy flow density of the electromagnetic field. The energy–momentum tensor $\boldsymbol{T}$ obeys a continuity relation: it describes a conserved quantity.

We can sum up by stating that in nature, energy and momentum are conserved, if we take into account the momentum and energy of the electromagnetic field. And the energy–momentum tensor shows again that electrodynamics is a gauge invariant description: the energy and momentum values do not depend on gauge choices.

The energy–momentum tensor, like the Lagrangian, shows that electrodynamics is Challenge 89 e invariant under *motion inversion*. If all charges change direction of motion – a situation often confusingly called 'time inversion' – they move backwards along the same paths they took when moving forward. Every example of motion due to electric or magnetic causes can also take place backwards.

On the other hand, everyday life shows many electric and magnetic effects that are not time invariant, such as the breaking of bodies or the burning of electric light bulbs. Can you explain how this fits together? Challenge 90 s

We also note that charges and mass destroy a further symmetry of the vacuum that we mentioned in special relativity: only the vacuum is invariant under conformal transformations. In particular, only the vacuum is invariant under the spatial inversion $r \to 1/r$. Any other physical system does not obey conformal symmetry. Vol. II, page 89

To sum up, electrodynamic motion, like all other examples of motion that we have encountered so far, is deterministic, slower than c, reversible and conserved. This is no big surprise. Nevertheless, two other symmetries of electromagnetism deserve special mention.

ENERGY AND MOMENTA OF THE ELECTROMAGNETIC FIELD

All moving entities have energy, momentum and angular momentum. This also applies to the electromagnetic field. Indeed, the description so far allows us to write the *total* energy E_{nergy} of the electromagnetic field as

$$E_{\text{nergy}} = \frac{1}{4\pi} \int \frac{\varepsilon_0}{2} (\boldsymbol{E}^2 + c^2 \boldsymbol{B}^2) \, \mathrm{d}V \, . \tag{55}$$

Energy is thus quadratic in the fields.

For the total linear momentum $\boldsymbol{p}$ we obtain

$$\boldsymbol{p} = \frac{1}{4\pi} \int \varepsilon_0 \boldsymbol{E} \times \boldsymbol{B} \, \mathrm{d}V \, . \tag{56}$$

The expression inside the integral is the *momentum density*. The related vector $\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{B}/\mu_0$, is called the *Poynting vector** and describes the *energy flux*; it is a vector field and has the units $\mathrm{W/m^2}$. The Poynting vector is the momentum density divided by c^2; indeed, special relativity implies that the momentum and the energy flow for electromagnetic fields are related by a factor c^2. The Poynting vector thus describes the energy flowing per area per time, in other words, the power per area. As shown below, the Poynting vector is a part of the energy–momentum tensor. Page 88

Can you produce a graph of the Poynting vector field for a cable carrying direct current? For a transformer? Challenge 91 s

Ref. 47 For the total angular momentum we have

$$\boldsymbol{L} = \frac{\varepsilon_0}{4\pi} \int \boldsymbol{E} \times \boldsymbol{A} \, \mathrm{d}V = \frac{\varepsilon_0}{4\pi} \int \boldsymbol{r} \times (\boldsymbol{E} \times \boldsymbol{B}) \, \mathrm{d}V \, , \tag{57}$$

where $\boldsymbol{A}$ is the magnetic vector potential.

In summary, the electromagnetic field has linear and angular momentum and energy, with well-defined values. Nevertheless, for most everyday situations, the actual values

---

* John Henry Poynting (b. 1852 Monton, d. 1914 Birmingham) introduced the concept in 1884.

FIGURE 48 Which one is the original landscape? (NOAA).

Challenge 92 e are negligibly small, as you may want to check.

What is a mirror? Is nature parity-invariant?

We will study the strange properties of mirrors several times during our walk. We start with the simplest one first. Everybody can observe, by painting each of their hands in a different colour, that a mirror does *not* exchange right and left, as little as it exchanges up and down; however, a mirror does exchange right and left *handedness*. In fact, it does so by exchanging front and back.

Electrodynamics give a second answer: a mirror is a device that switches magnetic north and south poles but does not switch the sign of charges. Can you confirm this Challenge 93 s with a diagram?

But is it always possible to distinguish left from right? This seems easy: this text is quite different from a bɘɿoɿɿim version, as are many other objects in our surroundings. But take a simple landscape. Are you able to say which of the two pictures of Figure 48 is the original?

Astonishingly, it is actually impossible to distinguish an original picture of nature from its mirror image if it does not contain any human traces. In other words, everyday nature is somehow left–right symmetric. This observation is so common that all candidate exceptions have been extensively studied. Examples are the jaw movement of Vol. V, page 259 ruminating cows, the helical growth of plants, such as hops, the spiral direction of snail shells or the left turn taken by all bats when exiting their cave. The most famous example is the position of the heart. The mechanisms leading to this disposition are still being investigated. Recent research discovered that the oriented motion of the cilia on embryos, Vol. V, page 27 in the region called the *node*, determines the right–left asymmetry. We will explore the issue later on.

Most human bodies have more muscles on the right side for right-handers, such as Albert Einstein and Pablo Picasso, and correspondingly on the left side for left-handers, such as Charlie Chaplin and Peter Ustinov. This asymmetry reflects an asymmetry of the human brain, called lateralization, which is essential to human nature.

Another asymmetry of the human body is the hair whirl on the back of the head; the majority of humans have only one, and in 80 % of the cases it is left turning. But many Challenge 94 s people have more than one. Can you name additional body asymmetries?

The left–right symmetry of nature appears because everyday nature is described by gravitation and, as we will see, by electromagnetism. Both interactions share an important property: substituting all coordinates in their equations by the negative of their values leaves the equations unchanged. This means that for any solution of these equations, i.e., for any naturally occurring system, a mirror image is a possibility that can also occur naturally. Everyday nature thus cannot distinguish between right and left. Indeed, there are right *and* left handers, people with their heart on the left *and* others with their heart on the right side, etc.

To explore further this strange aspect of nature, try the following experiment: imagine you are exchanging radio messages with a Martian; are you able to explain to him what right and left are, so that when you meet, you are sure you are talking about the same thing? Challenge 95 s

Actually, the *mirror symmetry* of everyday nature – also called its *parity invariance* – is so pervasive that most animals cannot distinguish left from right in a deeper sense. Ref. 49 Most animals react to mirror stimuli with mirror responses. It is hard to teach them different ways to react, and it is possible almost only for mammals. The many experiments performed in this area gave the result that animals have symmetrical nervous systems, and possibly only humans show *lateralization*, i.e., a preferred hand and different uses for the left and the right parts of the brain.

To sum up this digression, classical electrodynamics is left–right symmetric, or parity invariant. Can you show this using its Lagrangian? Challenge 96 s

Why do metals provide good mirrors? Metals are strong absorbers of light. Any strong absorber has a metallic shine. This is true for metals, if they are thick enough, but also for dye or ink crystals. Any material that strongly absorbs a light wavelength also reflects it efficiently. The cause of the strong absorption of a metal are the electrons inside it; they can move almost freely and thus absorb most visible light frequencies; this leads to evanescent waves in the material and strong reflection. Strong reflection appears as soon as the absorption length is as low as about one wavelength. This is the reason that, for example, strong coffee, strong tea and dense alkali vapour work as mirrors. (However, strong reflection is also possible without strong absorption, as the ubiquitous dielectric multilayers show.)

Page 90 Here is a puzzle: a concave mirror shows an inverted image; so does a plane mirror if it is partly folded along the horizontal. What happens if this mirror is rotated around the line of sight? Challenge 97 s

WHAT IS THE DIFFERENCE BETWEEN ELECTRIC AND MAGNETIC FIELDS?

Obviously, the standard answer is that electric fields have sources, and magnetic fields do not; as a result, magnetic fields are small relativistic effects of importance only when charge velocities are high or when electrical fields cancel out.

For situations involving matter, fields can indeed be distinguished with their sources. Up to the present day, no particle with a magnetic charge, called a *magnetic monopole*, has ever been found, even though its existence is possible in several speculative models of particle physics. Vol. V, page 269 If found, the action (48) would have to be modified by the addition of a fourth term, namely the magnetic current density. However, no such particle has yet been detected, despite intensive search efforts.

In empty space, when matter is not around, it is possible to take a completely different view. In empty space the electric and the magnetic fields can be seen as two faces of the same quantity, since a transformation such as

$$\begin{aligned} \boldsymbol{E} &\rightarrow c\,\boldsymbol{B} \\ \boldsymbol{B} &\rightarrow -\boldsymbol{E}/c \end{aligned} \tag{58}$$

called (electromagnetic) *duality* transformation, transforms each vacuum Maxwell equation into the other. The minus sign is necessary for this. (In fact, there are even more such transformations; can you spot them?) Alternatively, the duality transformation transforms $\boldsymbol{F}$ into ${}^{*}\boldsymbol{F}$. In other words, in empty space we *cannot* distinguish electric from magnetic fields. In particular, it is impossible to say, given a field line in vacuum, whether it is a magnetic or an electric field line.

Challenge 98 s

Matter would be symmetric under duality only if magnetic charges, also called magnetic monopoles, could exist. In that case the transformation (58) could be extended to

$$c\rho_{\mathrm{e}} \rightarrow \rho_{\mathrm{m}} \quad , \quad \rho_{\mathrm{m}} \rightarrow -c\rho_{\mathrm{e}} \; . \tag{59}$$

For a long time, it was thought that duality can be used in the search for the final, unified theory of physics. However, this hope has evaporated. The reason for this failure can be traced back to a small but ugly fact: the electromagnetic duality transformation changes the sign of the Lagrangian, and thus of the action. Therefore, electromagnetic duality is not a real symmetry of nature, and thus does not help to reach a deeper understanding of electromagnetism.

Ref. 50

Challenge 99 e

Duality, by the way, is a symmetry that works *only* in Minkowski space-time, i.e., in space-times of 3 + 1 dimensions. Mathematically, duality is closely related to the existence of quaternions, to the possibility of interpreting Lorentz boosts as rotations in 3 + 1 dimensions, and last, but not least, to the possibility of defining other smooth mathematical structures than the standard one on the space R^4. These mathematical connections are mysterious for the time being; they somehow point to the special role that *four* space-time dimensions play in nature. More details will become apparent in the last volume of our adventure.

COULD ELECTRODYNAMICS BE DIFFERENT?

Ref. 39

We saw that electrodynamics is based on three ideas: the conservation of charge, the speed limit for charges and Coulomb's inverse square relation. Could any of these be wrong or need modification?

Experiments imply that the only candidate for modification is Coulomb's relation. Indeed, any interaction, such as Coulomb's relation (4), which acts, for one given observer, between two particles independently of 3-velocity, must depend on 3-velocity for other inertial observers.* Such an interaction must also depend on the 4-velocity, to ensure the requirement from special relativity that the 4-acceleration must be 4-orthogonal

---

* This can be deduced from special relativity, from the reasoning of page 53 or from the formula in the footnote of page 83 in volume II.

to the 4-velocity. The simplest case of such an interaction is an interaction in which the acceleration is proportional to the 4-velocity. Together with the request that the interaction leaves the rest mass constant, we then recover electrodynamics. Other interactions do not agree with experiment. Ref. 51

In fact, the requirements of gauge symmetry and of relativistic invariance make it impossible to modify electrodynamics. In short, it does not seem possible to have a behaviour different from $1/r^2$ for a classical interaction.

Maybe a tiny deviation from Coulomb's relation is possible? An inverse square dependence implies a vanishing mass of light and light particles, the photons. Is the mass really zero? The issue has been extensively studied. A massive photon would lead to a wavelength dependence of the speed of light in vacuum, to deviations from the inverse square 'law', to deviations from Ampère's 'law', to the existence of longitudinal electromagnetic waves and more. No evidence for these effects has ever been found. A summary of these studies shows that the photon mass is below 10^{-53} kg, maybe even below 10^{-63} kg. Some arguments are not universally accepted, thus the limit varies somewhat from researcher to researcher. Ref. 52

A small non-vanishing mass for the photon would change electrodynamics somewhat. The inclusion of a tiny mass poses no special problems, and the corresponding Lagrangian, the so-called *Proca Lagrangian*, has already been studied, just in case. Ref. 52

Strictly speaking, the photon mass cannot be said to vanish. In particular, a photon with a Compton wavelength of the radius of the visible universe cannot be distinguished from one with zero mass through any experiment. This gives a limit mass of 10^{-69} kg for the photon. Photons with such a small mass value would not invalidate electrodynamics as we know it. We note that at present, the experimental limits are still much larger. A surprise is still possible, in principle.

Interestingly, a non-zero mass of the photon would imply the lack of magnetic monopoles, as the symmetry between electric and magnetic fields would be broken. It is therefore important on the one hand to try to improve the experimental mass limit for photons, and on the other hand to explore whether the limit due to the universe's size has any implications for this issue. The question is still open.

In summary, it seems extremely difficult, if not impossible, to find modifications of electrodynamics that agree with experiment. Electrodynamics is fixed once for all.

THE BRAIN: THE TOUGHEST CHALLENGE FOR ELECTRODYNAMICS

Researchers working on classical electrodynamics still face a fascinating experimental and theoretical issue: understanding the process of thought. Researchers face two challenges in this domain. First, they must find ways to *model* the thought process. Second, the technology to *measure* the currents in the brain must be extended. In both domains, recent progress has been spectacular.

Important research has been carried out on many levels of thought modelling. For example, research using computer tomography, PET scans and MRI imaging has shown that the distinction between the *conscious* and the *unconscious* can be measured: it has a biological basis. Conscious and unconscious thoughts happen in different brain regions. Psychological processes, such as *repression* of unpleasant thoughts, can actually be observed in brain scans. Modellers of brain mechanisms are learning that various concepts

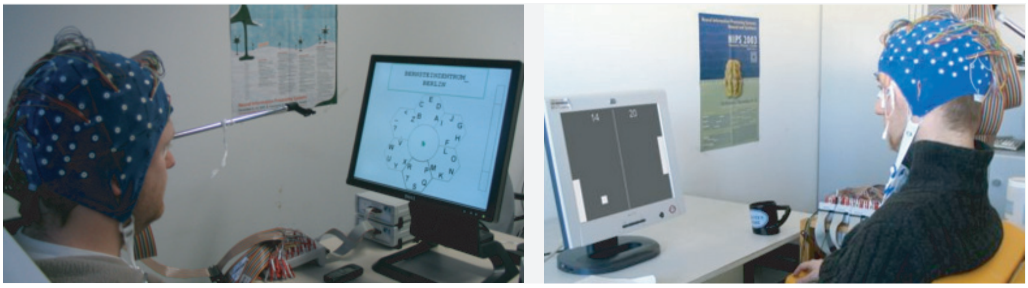


FIGURE 49 Typing a letter and playing video tennis using thought alone (© Fraunhofer FIRST).

of psychology are descriptions for actual physical processes. This research approach is still in its infancy, but very promising.

About the specific aspects of the working of the brain, such as learning, storage, recognition of shapes, location of sound sources or map formation, modern neurobiology and animal experimentation have allowed deducing models that make quantitative predictions. More on this will be told below.

Page 271

On the experimental side, research into magnetoencephalography devices is making rapid progress. The magnetic fields produced by brain currents are as low as 10 fT, which require sensors at liquid helium temperature and a good shielding of background noise. Improving the sensitivity and the spatial resolution of these systems is a central task. Also computer models and algorithms are making rapid progress.

The whole programme would be complete as soon as, in a distant future, a sensitive measuring apparatus could detect what is going on inside the brain and then could deduce or 'read' the thoughts of a person from these measurements. Thought reading might be the most complex of all challenges that science and technology are facing. Clearly, such a feat will require involved and expensive machinery, so that there is no danger for a misuse of the technique. There are good reasons to believe that full thought reading will never be possible in this way, due to the lack of localization of cognitive thought inside the brain and due to the variations in cognitive processing from one person to another. But the understanding and modelling of the brain will be a useful technology in a number of aspects of daily life, especially for the disabled.

On the path towards thought reading, the small progress that has been achieved so far is already fascinating. Wearing a cap full of electric contacts – a so-called *brain–computer interface* – and looking at a computer screen, it is now possible to type letters using the power of thought alone. Such a system is shown in Figure 49. The user controls the computer simply by *imagining* that he turns the arrow on the screen with his right hand. The brain currents created by the imagination process are read out and translated into computer commands by an electronic device. The system, based on neural network algorithms, works after only 20 minutes of training with a particular person. In this way, the system allows people who are fully paralysed to communicate with others again. The system is so fast that it allows playing 'mental video tennis' on a computer screen.

Ref. 53

Typing with thought alone is possible because the brain region responsible for the hand is near the skull, so that signals for hand rotation can be read out with sufficient spatial resolution by the electrodes on the cap. Researchers know that resolution limitations do not allow reading out the commands for single fingers in this way. For such high resolution tasks, electrodes still need to be *implanted* inside the relevant brain region. However, at present the functional lifetime for such electrodes is only a few months, so that the dream of controlling machines or even artificial limbs in this way is still distant.

Recent research with brain–computer interfaces suggests that in a not-too distant future a computer might be able to read out a secret number, such as a credit card PIN, Ref. 54 that a person is thinking about. The coming decades will surely yield more such research results.

Challenges and fun curiosities about electrodynamics

Not only animals, also plants can feel electric and magnetic fields. At least for magnetic fields, the sensors seem to use very similar mechanisms to those used by animals and bacteria.

* *

For everyday size – and larger – systems, electromagnetic motors are most effective. For microscopic sizes, electrostatic motors are more effective. They are used in sensors and small actuators. In contrast, large power systems use alternating current instead of direct current.

* *

If you calculate the Poynting vector for a charged magnet – or simpler, a point charge near a magnet – you get a surprising result: the electromagnetic energy flows in circles around the magnet. How is this possible? Where does this angular momentum come from? Challenge 100 s

Worse, any atom is an example of such a system – actually of two such systems. Why Ref. 55 is this effect not taken into account in calculations in quantum theory?

* *

Perfectly spherical electromagnetic waves are impossible in nature. Can you show this using Maxwell's equation of electromagnetism, or even without them? Challenge 101 s

* *

Mirrors exist in many forms. An important mirror for radio waves is the ionosphere; especially during the night, when certain absorbing layers disappear, the ionosphere allows receiving radio stations from far away. When the weather is favourable, it is possible to receive radio stations sending from the antipodes. Another radio mirror is the Moon; with modern receivers it is possible to receive radio signals and, since a few years, even television signals reflected by the Moon.

* *

In the past, textbooks often said that the Poynting vector, the electromagnetic energy flow, was not uniquely defined. Even Richard Feynman talks about this issue in his *Lectures on Physics*, in section 27-4. Can you show that there is no such ambiguity in the Poynting vector, and that those textbooks are all wrong? Challenge 102 s

* *

Ref. 56 No magnetic charges exist. More precisely, no particles with a single, non-zero magnetic charge exist. But we can introduce the mathematical quantity 'magnetic charge' nevertheless – it is usually called '*magnetic pole strength* – as long as we require that every object always has equal amounts of opposite magnetic charge values. With this condition, the magnetic charge is the divergence of the magnetization and obeys the magnetostatic Poisson equation, in a striking parallel to the electric case.

* *

Ref. 57 A recent object of research are solutions to the vacuum field equations that have *knotted* field lines. Such solutions do exist in theory, as shown by various authors. However, nobody has been able to realize such a solution in an experiment.

* *

Challenge 103 s Any wall plug is a dipole driven by an alternating electric field. Why does a wall plug, delivering 230 V or 100 V at 50 Hz or 60 Hz, not radiate electromagnetic fields?

* *

Challenge 104 e Why does a voltage transformer contain a ferromagnetic core?

* *

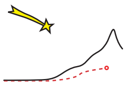
Challenge 105 s Are there electromagnetic motors in biological systems?

Summary on electromagnetic field motion

In summary, the electromagnetic field carries energy, linear momentum and angular momentum. It is thus appropriate to say that the electromagnetic field *moves*. The motion of the electromagnetic field is described by a least action principle, or equivalently, by Maxwell's equations.

The motion of the electromagnetic field can be visualized as the motion of its electric and its magnetic field lines. The motion of the fields conserves energy and momentum. The motion of electromagnetic fields is continuous, relative, reversible and mirror-invariant.

These results directly lead to ask: What is the nature of light?



Chapter 3

WHAT IS LIGHT?

Ref. 58

The nature of light has fascinated explorers of nature since at least the time of the ancient Greeks. The answer appeared in 1848, when Gustav Kirchhoff noted that the experimental values on both sides of the following equation agreed within measurement errors:

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \,. \tag{60}$$

This equality suggested the answer to the question asked two thousand years earlier:

▷ Light is an electromagnetic wave.

Ten years later, in 1858, Bernhard Riemann** proved mathematically that any electromagnetic wave in vacuum must propagate with a speed c given by the above equation. We note that the quantities on the right-hand side are electric and magnetic, while the quantity on the left-hand side is optical. The expression of Kirchhoff and Riemann thus unifies electromagnetism and optics. The modern value for the speed of electromagnetic waves, usually called c from Latin *celeritas*, is

$$c = 299\,792\,458\ \mathrm{m/s} \,. \tag{61}$$

Page 352

The value for c is an integer number, because the meter is nowadays *defined* in such a way as to exactly achieve this number.

In 1865, Maxwell summarized all data on electricity and magnetism collected in the previous 2500 years in his equations. Almost nobody read his papers, because he wrote them using quaternions. The equations were then simplified independently by Heinrich Hertz and Oliver Heaviside. They deduced the original result of Riemann: in the case of empty space, the equations of the electromagnetic potentials can be written as

$$\Box \boldsymbol{A} = 0 \quad \text{or, equivalently} \quad \varepsilon_0 \mu_0 \frac{\partial^2 \varphi}{\partial t^2} + \frac{\partial^2 A_x}{\partial x^2} + \frac{\partial^2 A_y}{\partial y^2} + \frac{\partial^2 A_z}{\partial z^2} = 0 \,. \tag{62}$$

** Bernhard Riemann (b. 1826 Breselenz, d. 1866 Selasca), important mathematician. A path-breaking mathematician, he also studied curved space, providing several of the mathematical and conceptual foundations of general relativity, but then died at an early age.

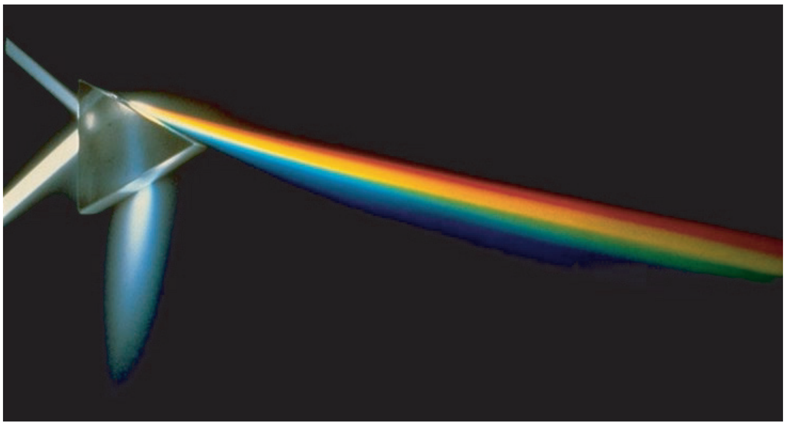


FIGURE 50 White light travelling through a glass prism (photograph by Susan Schwartzenberg, © Exploratorium www.exploratorium.edu).

Challenge 106 e

This evolution equation is a *wave equation*, because it admits solutions of the type

$$\boldsymbol{A}(t,\boldsymbol{x}) = \boldsymbol{A}_0 \sin(\omega t - \boldsymbol{kx} + \delta) = (A_{0x}, A_{0y}, A_{0z}) \sin(2\pi f t - 2\pi \boldsymbol{x}/\lambda + \delta) \ , \qquad (63)$$

which are commonly called harmonic *plane electromagnetic waves*. We recall that a *wave* in physics is any propagating imbalance, and that a *harmonic* wave is a wave described by a sine curve.

Vol. I, page 293

Such a harmonic plane electromagnetic wave in vacuum satisfies equation (62) for any value of *amplitude* A_0, of *phase* δ, and of *angular frequency* ω, provided the angular frequency and the *wave vector* $\boldsymbol{k}$ satisfy the relation

$$\omega(\boldsymbol{k}) = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} k \quad \text{or} \quad \omega(\boldsymbol{k}) = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \sqrt{\boldsymbol{k}^2} \ . \qquad (64)$$

The relation $\omega(\boldsymbol{k})$ between the angular frequency and the wave vector, the so-called *dispersion relation*, is the main property of any type of wave, be it a sound wave, a water wave, an electromagnetic wave, or any other kind.

The specific dispersion relation (64) is *linear* and implies a *phase velocity* c, the velocity with which wave crests and troughs move, given by $c = \omega/k = 1/\sqrt{\varepsilon_0 \mu_0}$, thus reproducing the result by Kirchhoff and Riemann.

Experiments in empty space confirm that the phase velocity c is independent of the frequency, amplitude or phase of the wave. This constant phase velocity c thus characterizes electromagnetic waves, and distinguishes them from all other types of waves in everyday life.

What are electromagnetic waves?

To get a clearer idea of electromagnetic waves, we explore their properties. The wave equation (62) for the electromagnetic field is *linear* in the field; this means that the sum of two allowed situations is itself an allowed situation. Mathematically speaking, any *superposition* of two solutions is also a solution. We therefore know that electromagnetic

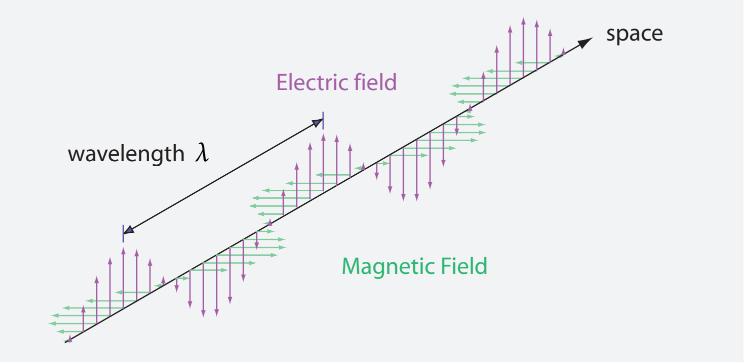


FIGURE 51 The general structure of a plane, monochromatic and linearly polarized electromagnetic wave at a specific instant of time.

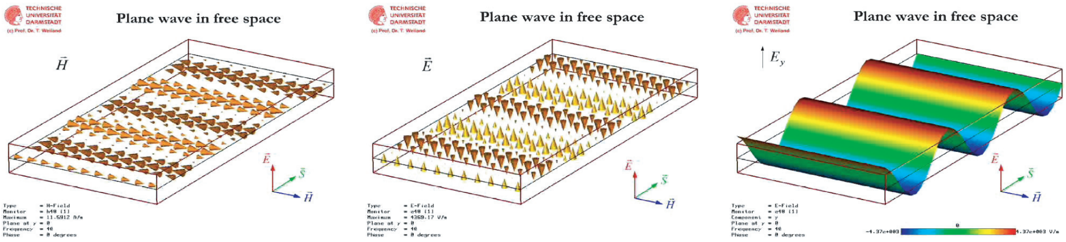


FIGURE 52 A plane, monochromatic and linearly polarized electromagnetic wave, showing the evolution of the electric field, the magnetic field, and again the electric field, in a further visualization (Mpg films © Thomas Weiland).

waves must show *interference*, as all linear waves do.

Linearity also implies that two waves can *cross* each other without disturbing each other, and that electromagnetic waves can travel undisturbed across static electromagnetic fields.

Linearity also means that every electromagnetic wave can be described as a superposition of harmonic, or pure sine waves, each of which is described by expression (63), with its own frequency, amplitude and phase. It thus makes sense to talk about the *spectrum* of electromagnetic waves, i.e., about the range of frequencies and their properties.

The simplest possible electromagnetic wave, the harmonic plane wave with *linear* Page 111 *polarization*, is illustrated in Figure 51. Note that for this simplest type of waves, the electric and the magnetic field are *in phase*. (Can you prove this experimentally and by calculation?) The surfaces formed by all points of maximal field intensity are parallel planes, spaced by (half the) wavelength; these planes move along the direction of the propagation with the phase velocity.

EXPERIMENTS WITH ELECTROMAGNETIC WAVES

After Riemann and Maxwell predicted the existence of electromagnetic waves, in the years between 1885 and 1889, Heinrich Hertz* discovered and studied them. He fabric-

---

* Heinrich Rudolf Hertz (b. 1857 Hamburg, d. 1894 Bonn), important theoretical and experimental physicist. The unit of frequency is named after him. Despite his early death, Hertz was a central figure in the

FIGURE 53 Heinrich Hertz (1857–1894).

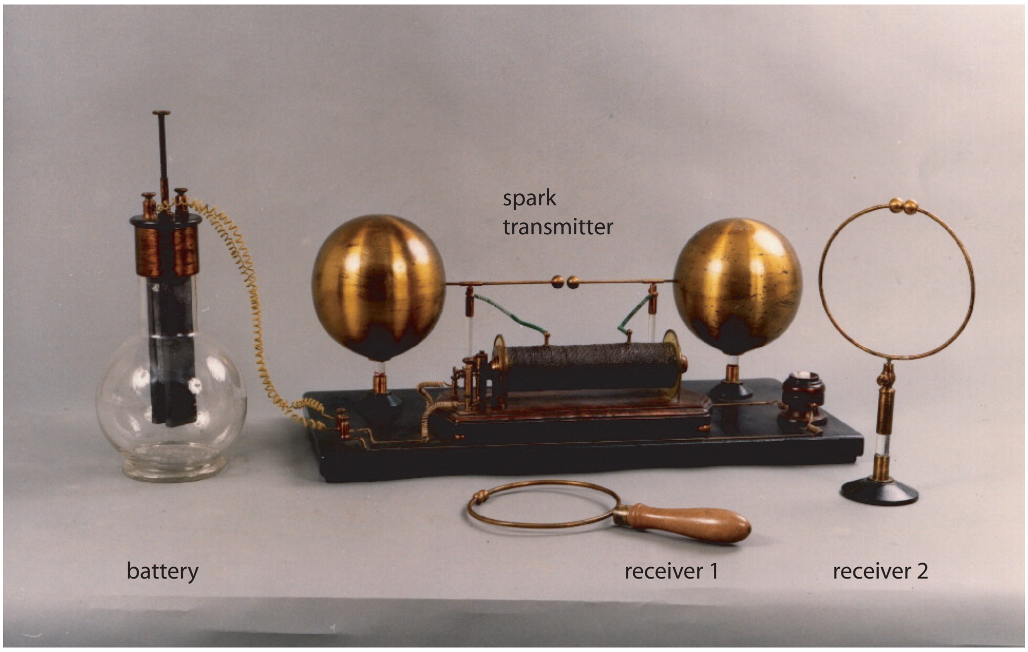


FIGURE 54 A reconstruction of one of the first transmitters and receivers of electromagnetic waves by Heinrich Hertz (© Fondazione Guglielmo Marconi).

ated a very simple transmitter and receiver for 2 GHz waves, shown in Figure 54. Such waves are still used today: cordless telephones and the last generation of mobile phones work at this frequency – though the transmitters and the receivers look somewhat differently nowadays. Such waves are now also called *radio waves*, since physicists tend to call all moving force fields *radiation*, recycling somewhat incorrectly a Greek term that originally meant 'light emission.'

Today Hertz's experiment can be repeated in a much simpler way. As shown in Figure 55, a budget of a few euro is sufficient to remotely switch on a light emitting diode with a gas lighter. (After each activation, the coherer has to be gently tapped, in order to get ready for the next activation.) Attaching longer wires as antennas and ground allows this set-up to achieve transmission distances up to 30 m.

---

development of electromagnetism, in the explanation of Maxwell's theory and in the unfolding of radio communication technology. More about him on page 236 in volume I.

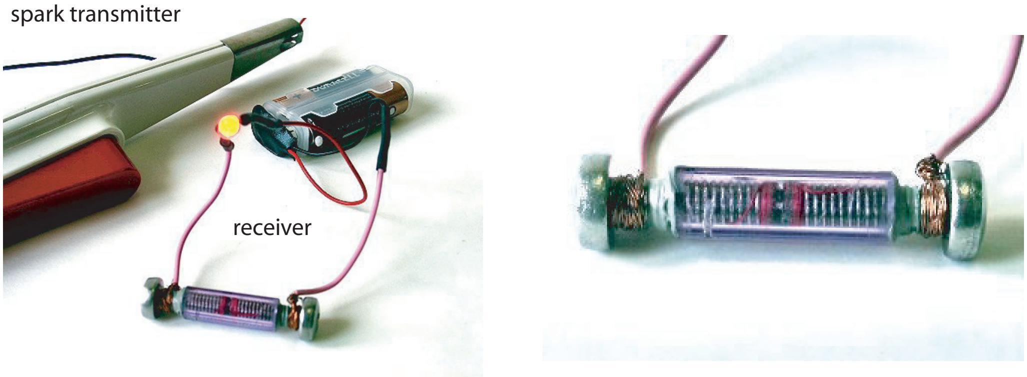


FIGURE 55 The simplest radio transmitter possible, a gas lighter and a wire, together with the simplest radio receiver possible, built from a battery pack, a light emitting diode, and a simple *coherer* made from a ball pen housing, two screws and some metal powder (© Guido Pegna).

Hertz also measured the *speed* of the waves he produced. In fact, you can also measure the speed at home, with a chocolate bar and a (older) kitchen microwave oven. A microwave oven emits radio waves at 2.5 GHz – not far from Hertz's value. Inside the oven, these waves form standing waves. Just put the chocolate bar (or a piece of cheese) in the oven and switch the power off as soon as melting begins. You will notice that the bar melts at regularly spaced spots. These spots are half a wavelength apart. From the measured wavelength value and the frequency, the speed of light and of radio waves simply follows as the product of the two.

If you are not convinced, you can measure the speed directly, by telephoning a friend on another continent, if you can make sure of using a satellite line (choose a low cost provider). There is about half a second additional delay between the end of a sentence and the answer of the friend, compared with normal conversation. In this half second, the signal goes up to the geostationary satellite, down again and returns the same way. This half second gives a speed of $c \approx 4 \cdot 36\,000\,\mathrm{km}/0.5\,\mathrm{s} \approx 3 \cdot 10^5\,\mathrm{km/s}$, which is close to the precise value. Radio amateurs who reflect their signals from the Moon can perform a similar experiment and achieve higher precision.

In summary: electromagnetic waves exist and move with the speed of light.

LIGHT AS A WAVE

The electromagnetic wave equation is not limited to radio waves; it has even more interesting stories to tell. Above all, the wave equation confirmed earlier predictions that *light* itself is an electromagnetic wave, albeit with a much higher frequency and much shorter wavelength than radio waves. We check this in two steps: we first show that light is a wave and then show that it is electromagnetic.

The first to suggest that light is a (kind of) *wave* was, around the year 1678, the important physicist Christiaan Huygens.[*] You can confirm that light is a wave with your

---

[*] Christiaan Huygens (b. 1629 's Gravenhage, d. 1695 Hofwyck) was one of the main physicists and mathematicians of his time. Huygens clarified the concepts of mechanics; he also was one of the first to show that light is a wave. He wrote influential books on probability theory, clock mechanisms, optics and astronomy.

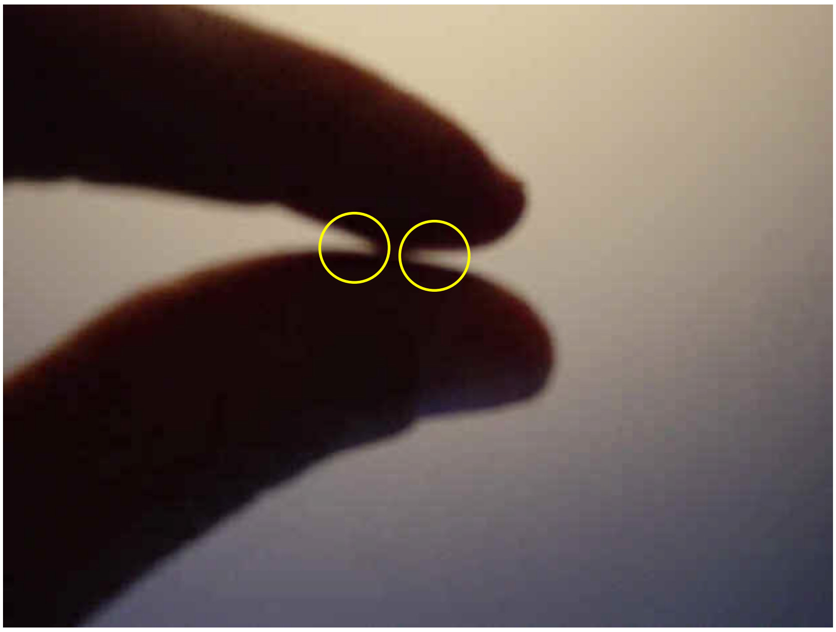


FIGURE 56 Diffraction lines can be seen between the fingers, if one looks carefully enough. (© Chuck Bueter)

FIGURE 57 The primary and secondary rainbow, and the supernumerary bows below the primary bow (© Antonio Martos and Wolfgang Hinz).

own fingers. Simply place your hand one or two centimetres in front of your eye, look towards the sky through the gap between the middle and the index finger and let the two fingers almost touch. You will see a number of dark lines crossing the gap. These lines are the interference pattern formed by the light behind the slit created by the fingers. Figure 56 shows an example. *Interference* is the name given to the effect and the amplitude patterns that appear when several waves superpose.* The interference patterns depend on the spacing between the fingers. This experiment therefore allows you to estimate the wavelength of light, and thus, if you know its speed, its frequency. Can you do this? Challenge 108 s

Historically, another effect was central in convincing researchers that light was a wave:

Among other achievements, Huygens showed that the Orion Nebula consists of stars, discovered Titan, the moon of Saturn, and showed that the rings of Saturn consist of rock. (This is in contrast to Saturn itself, whose density is lower than that of water.)

Challenge 107 s * Where does the energy go in an interference pattern?

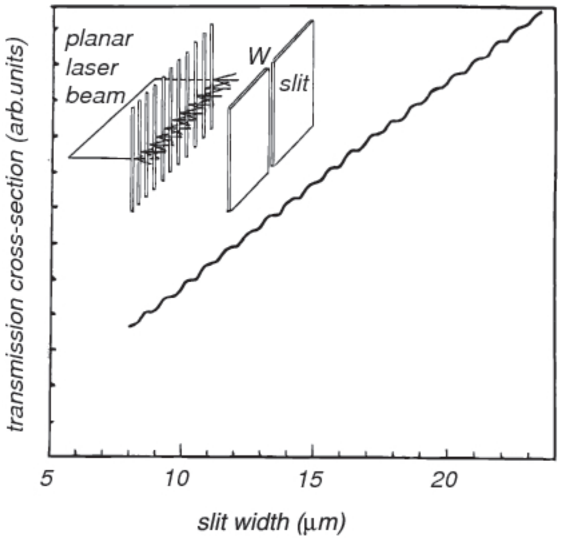


FIGURE 58 The light power transmitted through a slit as function of its width (© Nature).

supernumerary rainbows, the additional bows below the main or primary rainbow. If we look carefully at a rainbow, below the main red–yellow–green–blue–violet bow, we observe weaker, additional green, blue and violet bows. Depending on the intensity of the rainbow, several of these supernumerary rainbows can be observed. They are due to interference of light triggered by the water droplets, as Thomas Young showed around 1803.* Indeed, the repetition distance of the supernumerary bows depends on the radius and shape distribution of the average water droplets that form them. (Details about the normal rainbows are given below.) Both supernumerary rainbows and Thomas Young were essential to convince people that light is a wave. It seems that in those times scientists either did not trust their own eyes or fingers, or did not have any. Ref. 59 Page 130 Ref. 60 Page 125

There are many other ways in which the wave character of light can be made apparent. Maybe the most beautiful is an experiment carried out by a team of Dutch physicists in 1990. They simply measured the light transmitted through a *slit* in a metal plate. It turns out that the transmitted intensity depends on the width of the slit. Their surprising result is shown in Figure 58. Can you explain the origin of the unexpected intensity steps in the curve? Ref. 61 Challenge 109 ny

Interference of light is a common effect. It is commonly seen when lasers are used. A few examples are shown in Figure 59. Both white light interference and laser interference

---

* Thomas Young (b. 1773 Milverton, d. 1829 London), read the bible at two, spoke Latin at four; a doctor of medicine, he became a professor of physics. He introduced the concept of *interference* into optics, explaining Newtonian rings and supernumerary rainbows; he was the first person to determine light's *wavelength*, a concept that he also introduced, and its dependence on colour. He was the first to deduce the three-colour vision explanation of the eye and, after reading of the discovery of polarization, explained light as a transverse wave. In short, Young discovered most of what people learn at secondary school about light. He was a universal talent: he also worked on the deciphering of hieroglyphs, studied languages and introduced the term 'Indo-European', explored ship building and many engineering problems. Young collaborated with Fraunhofer and Fresnel. In Britain his ideas on light were not accepted, since Newton's followers crushed all opposing views. Towards the end of his life, his results were finally made known to the physics community by Fresnel and Helmholtz.

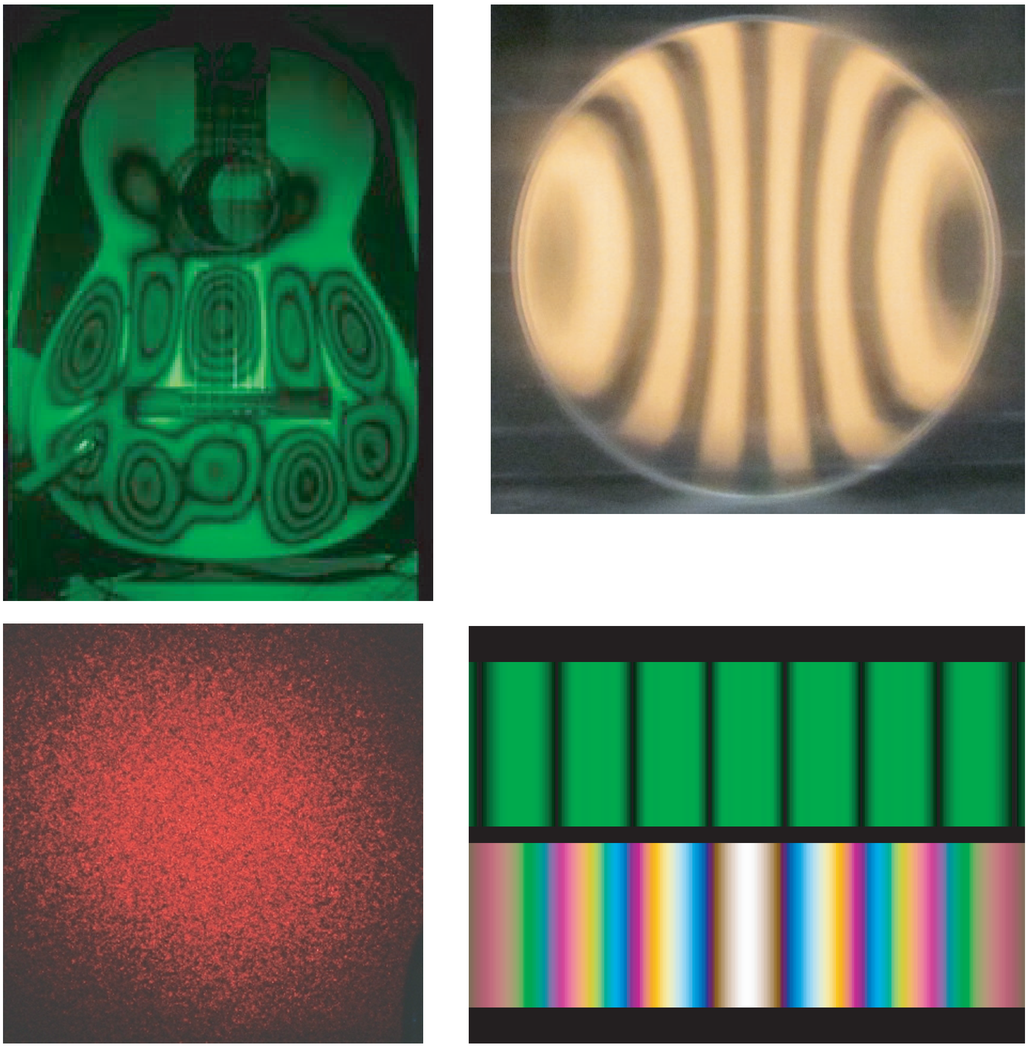


FIGURE 59 Some interference patterns: the interference that a playing guitar produces in laser holography that show how the body vibrates, the interference produced by a good parabolic telescope mirror of 27 cm diameter, a speckle laser pattern on a rough surface and the diffraction pattern produced by two parallel narrow slits illuminated with green light and with white light respectively (© Bernard Richardson, Cardiff University, Mel Bartels, Epzcaw and Dietrich Zawischa).

are used for measurements; nowadays, a whole industry makes use of interference effects.

Given an interference pattern like the green one in Figure 59, you may wish to calculate the distance between the lines, given the slit distance s, the colour and the distance d to the screen. (This experiment was used to determine the wavelength of the light for the first time.) Challenge 110 s

Another proof that light is a wave is the discovery of light polarization. We will explore it shortly. Numerous other experiments on the creation, detection and measurement of light waves were performed between the seventeenth and the twentieth century. For example, in 1800, William Herschel discovered *infrared light* using a prism and a

FIGURE 60 The same rainbow in the visible and in the infrared, showing how infrared comes before red (© Stefan Zeiger).

Challenge 111 s thermometer. (Can you guess how?) In 1801, Johann Wilhelm Ritter (b. 1776 Samitz, d. 1810 Munich) a more than colourful figure of natural Romanticism, discovered *ultraviolet light* using silver chloride, AgCl, and again a prism. Modern cameras can image infrared light, as shown beautifully in Figure 60.

At the end of the twentieth century a beautiful confirmation of the oscillations in light waves became possible. Using quite sophisticated experiments, researchers measured the Ref. 62 oscillation frequency of visible light *directly*. They actually managed to count how often light wave oscillate in a second! The frequency value, between 375 and 750 THz, is exactly as predicted. The frequency value is so high that its detection was impossible for a long time. But with these modern experiments the dispersion relation of light, $\omega = ck$, has Ref. 63 been confirmed in all its details, and to extremely high precision.

The result of all these experiments is: light waves, like all other waves, can be distinguished by their wavelength or frequency values. The most important categories are Page 108 listed in Table 14. For visible light, the wavelength lies between 0.4 μm, corresponding to violet, and 0.8 μm, corresponding to red. The wavelength of a visible harmonic light wave determines its *colour*.

Light is a wave. This statement also ends a discussion that led to intense debate in the Middle Ages: How narrow can a light beam be? A light beam *cannot* be arbitrarily narrow. The wave properties of light imply that any attempt to produce an extremely narrow beam of light, say by shining light on a tiny hole in a wall, produces a strongly divergent beam. A light beam cannot even have a sharp border. Also every attempt to concentrate light of a single wavelength on a tiny spot has its limits, as Figure 61 shows: with a factor of order 1, the product of the two transverse quantities wd equals that of the two longitudinal quantities λf. In short,

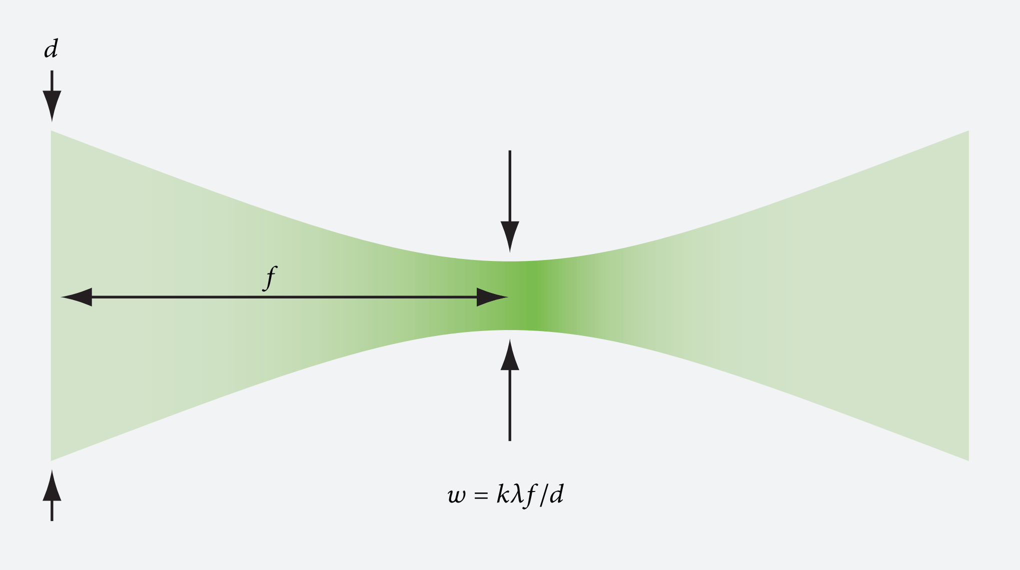


FIGURE 61 The focus of a converging light beam has a minimum size, the waist radius w, given by the wavelength and the geometry. The waist radius also depends on a number k, of order 1, that describes how the light intensity changes transversally to the beam. Note that the transition between the green beam and the background is never sharp, in contrast to the drawing.

▷ Light beams cannot be arbitrary narrow lines.

The diameter of a light beam is both determined and limited by the wavelength and by the geometric arrangement that produce it.

Light and other electromagnetic waves

The experiments mentioned so far showed that electromagnetic waves exist, that they move with the same speed as light, and that light is a wave. To confirm that light waves are indeed *electromagnetic* is more difficult. The most convincing proof would be to repeat Hertz's experiments for light. In Hertz's experiment, shown in Figure 54, the receiver is a simple open metal circle; when the wave – more precisely, its magnetic field – arrives, a spark is generated and the wave is thus detected. Page 100

In an almost incredible feat of miniaturization, in 2009, the research group of Kobus Kuipers managed to make metal rings much smaller than a micrometre, and repeat Hertz's experiment for light. An impression of their experiment is given in Figure 62. They could clearly discern the maxima and minima of waves, as well as their polarization. They thus showed that light is an electromagnetic wave in exactly the same way as Hertz did for radio waves. Ref. 64

Of course, people in the 19th century had less technology at their disposal and were not easily convinced. They had to look for other ways to show that light is electromagnetic in nature. Now, since the evolution equations of the electrodynamic field are linear, additional electric or magnetic fields alone do not influence the motion of light. On the other hand, we know that electromagnetic waves are emitted only by accelerated charges, Challenge 112 e

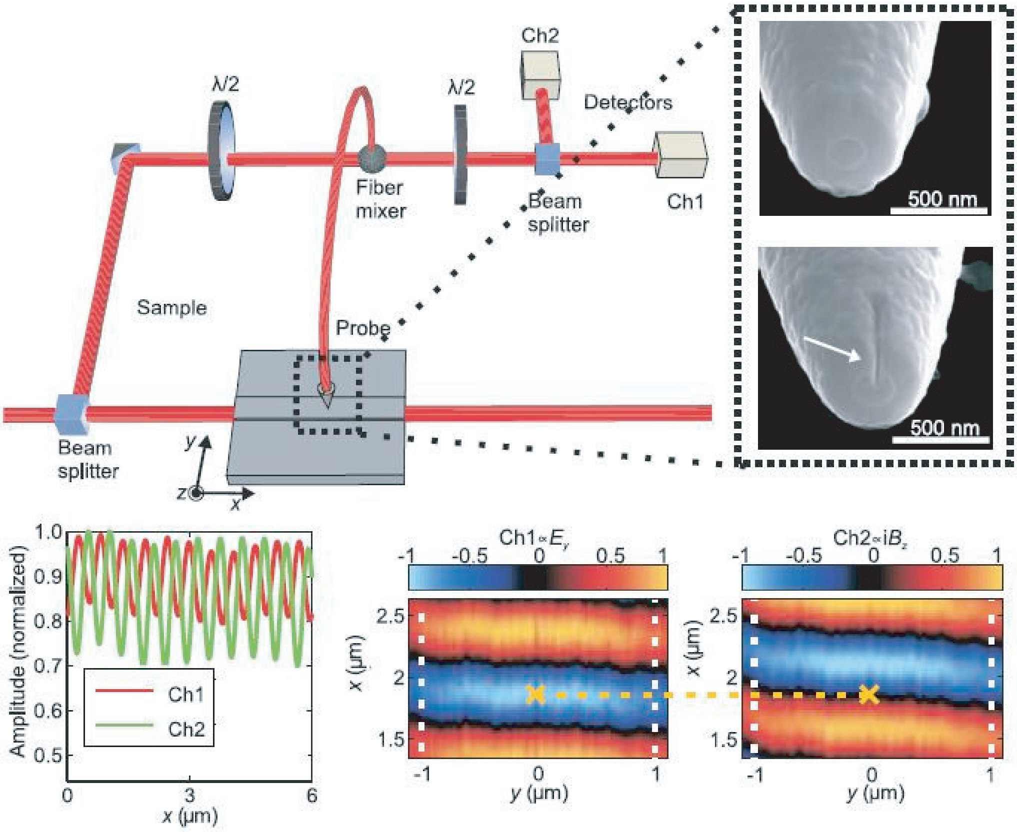


FIGURE 62 An experiment measuring the electric and magnetic field of light. Top left: the general set-up; top right: the antenna, indicated by an arrow; bottom: the measurement data (© Kobus Kuipers)

and that all light is emitted from matter. It thus follows that matter is full of electromagnetic fields and accelerated electric charges. This in turn implies that the influence of matter on light can be understood from its internal electromagnetic fields and, in particular, that subjecting matter to an *external* electromagnetic field should change the light it emits, the way matter interacts with light, or generally, the material properties as a whole.

Searching for effects of electricity and magnetism on matter has been a main effort of physicists for over a hundred years. For example, electric fields influence the light transmission of oil, an effect discovered by John Kerr in 1875.* Also the discovery that certain gases change colour when subject to a field yielded several Nobel Prizes for physics. With time, many more influences on light-related properties by matter subjected to fields were found. An extensive list is given below, in the table on page 231. It turns out that apart from a few exceptions the effects can *all* be described by the electromagnetic Lagrangian (48), or equivalently, by Maxwell's equations (52). In summary, classical electrodynamics indeed unifies the description of electricity, magnetism and optics; all phenomena

Page 87

---

* John Kerr (b. 1824 Ardrossan, d. 1907 Glasgow), was mathematician and physicist, as well as friend and collaborator of William Thomson.

in these fields, from the rainbow to radio and from lightning to electric motors, are found to be different aspects of the evolution of the electromagnetic field.

After two centuries of research, it became clear that light and radio waves form only a small section of the full *spectrum of electromagnetic waves*, which contains the waves from the smallest possible to the largest possible wavelengths. The full spectrum is given in the following table.

TABLE 14 The electromagnetic spectrum.

Frequency	Wavelength	Name	Main properties	Appearance	Use
$3\cdot10^{-18}$ Hz	10^{26} m	**Lower frequency limit**		see the section on cosmology	
< 10 Hz	> 30 Mm	**Quasistatic fields**		intergalactic, galactic, stellar and planetary fields, brain, electrical fish	power transmission, accelerating and deflecting cosmic radiation
		Radio waves		electronic devices	
10 Hz–50 kHz	30 Mm–6 km	ELW	go round the globe, penetrate into water, penetrate metal	nerve cells, electromechanical devices	power transmission, communication through metal walls, communication with submarines www.vlf.it
50–500 kHz	6 km–0.6 km	LW	follow Earth’s curvature, felt by nerves (‘bad weather nerves’)	emitted by thunderstorms	radio communications, telegraphy, inductive heating
500–1500 kHz	600 m–200 m	MW	reflected by night sky		radio
1.5–30 MHz	200 m–10 m	SW	circle world if reflected by the ionosphere, destroy hot air balloons	emitted by stars	radio transmissions, radio amateurs, spying
15–150 MHz	20 m–2 m	VHF	allow battery operated transmitters	emitted by Jupiter	remote controls, closed networks, tv, radio amateurs, radio navigation, military, police, taxi
150–1500 MHz	2 m–0.2 m	UHF	*idem*, line of sight propagation		radio, walkie-talkies, tv, mobile phones, internet via cable, satellite communication, bicycle speedometers
		Microwaves			

Frequency	Wavelength	Name	Main properties	Appearance	Use
1.5 – 15 GHz	20 cm–2 cm	SHF	*idem*, absorbed by water	night sky, emitted by hydrogen atoms	radio astronomy, used for cooking (2.45 GHz), telecommunications, radar
15 – 150 GHz	20 mm–2 mm	EHF	*idem*, absorbed by water		
		Infrared	allows night vision	emitted by every warm object	satellite photography of Earth, astronomy
0.3 – 100 THz	1000 –3 μm	IRC or far infrared		sunlight, living beings	seeing through clothes, envelopes and teeth
100 – 210 THz	3 μm–1.4 μm	IRB or medium infrared		sunlight	used for optical fibre communications for telephone and cable television
210 – 384 THz	1400–780 nm	IRA or near infrared	penetrates for several cm into human skin	sunlight, radiation from hot bodies	healing of wounds, rheumatism, sport physiotherapy, hidden illumination
375 – 750 THz	800–400 nm	**Light**	not (much) absorbed by air, detected by the eye (up to over 900 nm at sufficient power)	heat (‘hot light’), lasers & chemical reactions e.g. phosphor oxidation, fireflies (‘cold light’)	definition of straightness, enhancing photosynthesis in agriculture, photodynamic therapy, hyperbilirubinaemia treatment
384 – 484 THz	780–620 nm	Red	penetrate flesh	blood	alarm signal, used for breast imaging Ref. 65
	700 nm	Laboratory primary red		filtered tungsten lamp	colour reference for printing, painting, illumination and displays
484 – 511 THz	620–587 nm	Orange		various fruit	attracts birds and insects
511 – 525 THz	587–571 nm	Yellow		majority of flowers	*idem*; best background for reading black text

FREQUENCY	WAVELENGTH	NAME	MAIN PROPERTIES	APPEARANCE	USE
525 – 614 THz	571–488 nm	Green	maximum eye sensitivity	algae and plants	highest luminous efficiency response (‘felt brightness’) per light energy for the human eye
	546.1 nm	Laboratory primary green		mercury lamp	colour reference
614 – 692 THz	488– 433 nm	Blue		sky, gems, water	
	435.8 nm	Laboratory primary blue		mercury lamp	colour reference
692 – 789 THz	433– 380 nm	Indigo, violet		flowers, gems	
		Ultraviolet			
789 – 952 THz	380– 315 nm	UVA	penetrate 1 mm into skin, darken it, produce vitamin D, suppress immune system, cause skin cancer, destroy eye lens	emitted by Sun, stars, lasers and flames	seen by certain birds, integrated circuit fabrication
0.95 – 1.07 PHz	315–280 nm	UVB	*idem*, destroy DNA, cause skin cancer	*idem*	*idem*
1.07 – 3.0 PHz	280– 100 nm	UVC, VUV	form oxygen radicals from air, kill bacteria, penetrate 10 μm into skin	emitted by Sun, stars, lasers and welding arcs	disinfection, water purification, waste disposal, integrated circuit fabrication
3 –24 PHz	100–13 nm	EUV			sky maps, silicon lithography
		X-rays	penetrate materials	emitted by stars, plasmas and black holes	imaging human tissue
24 – 240 PHz	13–1.3 nm	Soft X-rays	*idem*	synchrotron radiation	*idem*
> 240 PHz or > 1 keV	< 1.2 nm	Hard X-rays	*idem*	emitted when fast electrons hit matter	crystallography, structure determination
> 12 EHz or > 50 keV	< 24 pm	**γ-rays**	*idem*	radioactivity, cosmic rays	chemical analysis, disinfection, astronomy
$2 \cdot 10^{43}$ Hz	$\approx 10^{-35}$ m	**Planck limit**		see the last volume of this series	

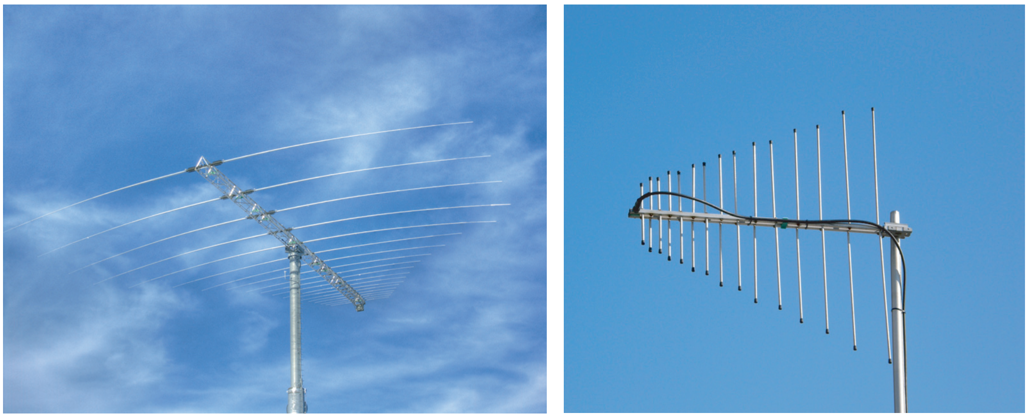


FIGURE 63 Antennas for horizontally and vertically polarized electromagnetic waves (© Martin Abegglen, K. Krallis).

POLARIZATION OF ELECTROMAGNETIC WAVES

Page 99 The electric field in light or in an electromagnetic wave looks like the amplitude of a water wave, generalized to three dimensions, as shown in Figure 51 and Figure 52. The same is valid for magnetic fields, and the two fields are perpendicular to each other.

One question about light and all other electromagnetic waves arises: In which spatial direction does the oscillation occur? The answer is hidden in the parameter $\boldsymbol{A}_0$ in expression (63), but shown in Figure 51 and Figure 52. Generally speaking, the fields in electromagnetic waves oscillate in directions *perpendicular* to their motion. Therefore, we follow:

▷ Even for identical frequency and phase, waves can still differ: they can have different *polarization* directions.

For example, the polarization of radio transmitters determines whether radio antennas of receivers have to be kept horizontal or vertical, as shown in Figure 63. For all electromagnetic waves, the polarization is defined, by convention, by the orientation of the *electric* field vector, because practically all effects of electromagnetic waves are due to the electric field.

Polarization is easily achieved also for light, e.g., by shining it through a stretched plastic film, called a polarizer, or by using glass, water or some special stones. After the physician and physicist Thomas Young understood that light is a transverse wave in 1803, Louis Malus discovered polarization by reflection in 1808 by Louis Malus (b. 1775 Paris, d. 1812 Paris). Malus discovered and described polarization when he explored the strange double images produced by calcite, a transparent crystal found in many minerals. Figure 64 shows two examples. Calcite ($CaCO_3$) splits light beams into two – it is *birefringent* – and polarizes them differently. That is the reason that calcite – or feldspar, ($KAlSi_3O_8$), which shows the same effect – is part of every crystal collection. If you ever get hold of a piece of transparent calcite or feldspar, do look through it at something written on paper, and rotate the crystal around the vertical. Its properties are intriguing. Challenge 113 e

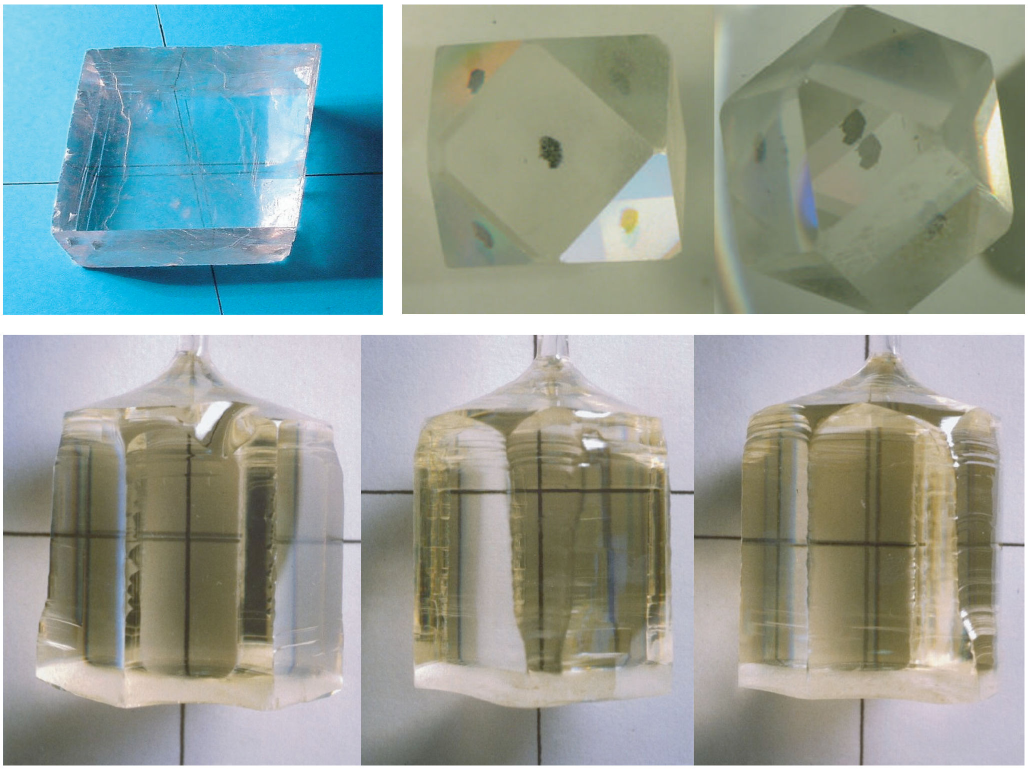


FIGURE 64 Birefringence in crystals: *calcite* lying on crossed lines (top left, crystal size around 4 cm), *rutile* lying on an ink spot, photographed along the optical axis (middle) and at an angle to it (top right, crystal size around 1 cm), and an octagonal sodium vanadate crystal doped with manganese, showing *three* different behaviours (bottom, crystal diameter 1.9 cm) (© Roger Weller/Cochise College, Brad Amos, Martin Pietralla).

(Can you show that *trirefringence*, if defined as the appearance of three images, cannot exist?) Challenge 114 d

When Malus discovered the polarization of light, he did not know yet that light was electromagnetic. But his discovery definitively settled the wave nature of light.

The light from the sky – not that from the Sun – is partially polarized. The polarization occurs when the light is scattered by the molecules in the air. The polarization is perpendicular to the direction towards the Sun, as illustrated in Figure 65. Ref. 66 The shape is easy to remember with the following connection: *A rainbow is polarized everywhere in tangential direction.* Photographers know that when the Sun is rising or setting, the sky is mainly polarized in north-south direction. This fact can make a lake or a digital watch look black when observed in the evening in northern or southern direction – at a certain observation angles.

Also the sunlight below water is partially polarized. David Brewster (1781 Jedburgh-1868 Allerly) discovered this effect in 1812. Brewster, who was clergyman and physicist, found that when a light beam is partially transmitted and partially reflected at an interface, the polarization changes. Figure 67 shows an extreme example. The effect is used in many optical devices.

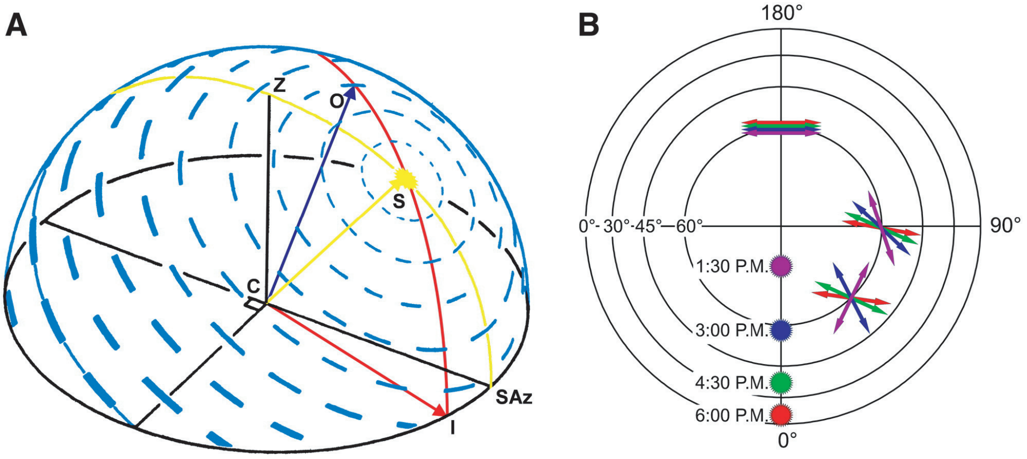


FIGURE 65 Left: the polarization of daylight in the clear sky as a solar elevation of 53°. The orientation and the thickness of the blue bars illustrate the orientation and degree of polarization of the electric field as seen by an observer in the center C of the sphere. The orientation is always perpendicular to a great circle (red) that is defined by connecting a given observation point in the sky O with the position of the Sun S. SAz indicates the solar azimuth of the Sun. Right: the zenithal projection of solar elevation and electric field orientation for different light colours at four times of August 1, at 23.4° N, 5.2° E. Circles represent elevation and the straight lines represent azimuth. The circular polarization pattern of the sky is used by photographers to modify sky photographs and by insects and birds to navigate. (© Keram Pfeiffer/Elsevier).

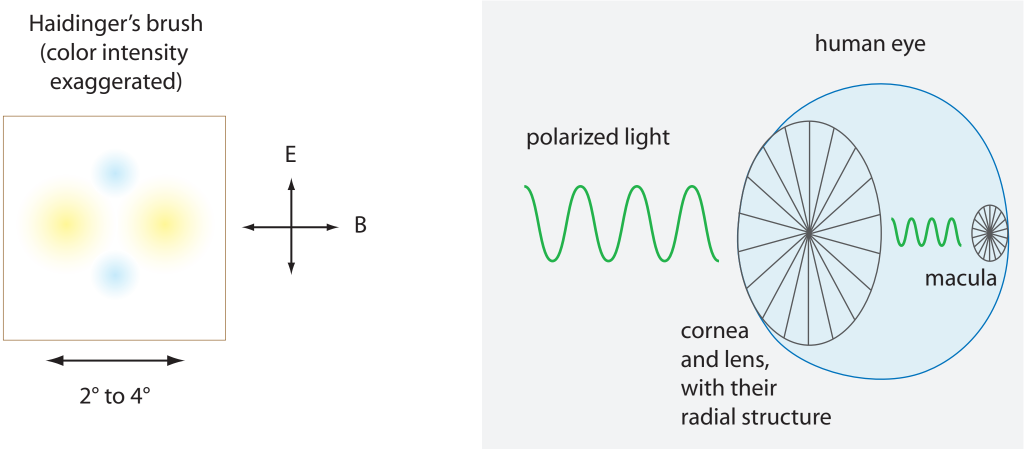


FIGURE 66 Haidinger's brush and its origin in the human eye.

Many insects, spiders, certain birds and certain shrimps can detect polarization with their eyes. Honey bees and many other insects use polarization to deduce the position of the Sun, even when it is hidden behind clouds, and use the effect for navigation. Some beetles of the genus *Scarabeus* even use the polarization of moonlight for navigation, and many insects use polarization of sunlight to distinguish water surfaces from mirages. (Can you find out how?) Ref. 67 Challenge 115 s

In 1844, the mineralogist Wilhelm Haidinger (b. 1795 Vienna, d. 1871 Dornbach)

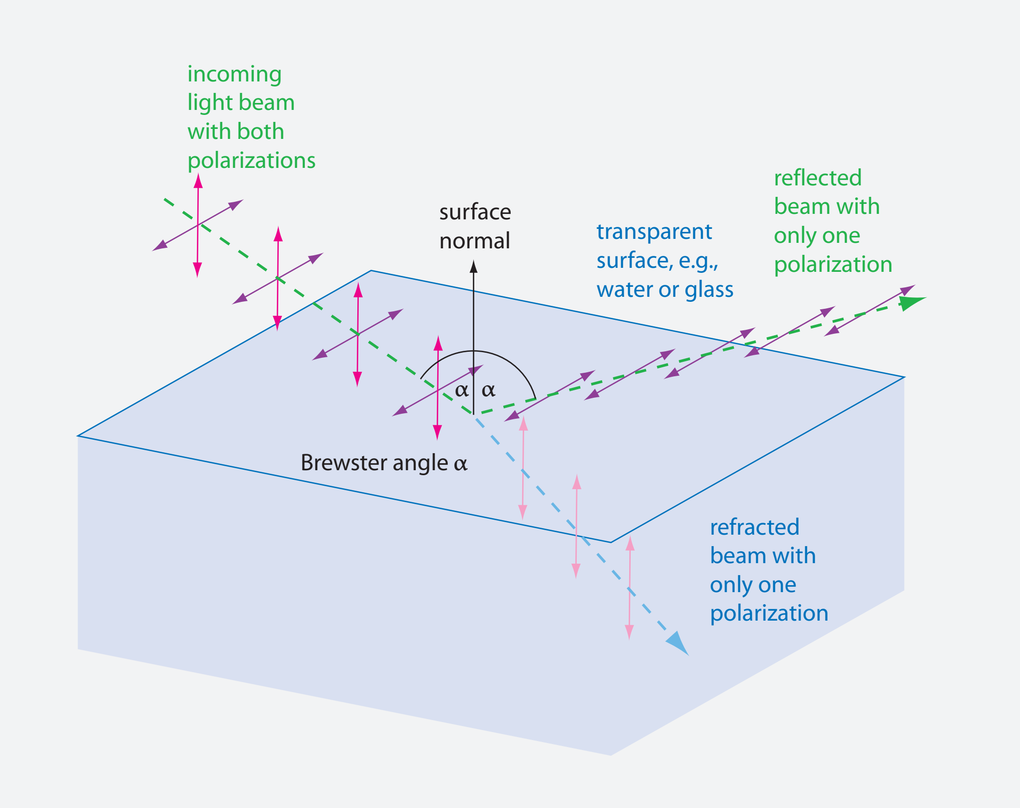


FIGURE 67 For every transparent material, at the so-called *Brewster angle*, only the horizontally polarized light is reflected; the vertically polarized light is then fully refracted. The Brewster angle is a material-dependent quantity. The value for water is, for most wavelengths, 53° and for glass 56(1)°, measured from the line that is normal to the surface.

Ref. 68 discovered that there is a way to observe the polarization of light with the unaided human eye. The best way to observe the effect is by looking at a distance of about an arms's length on a white LCD screen and slowly tilt your head. You will note an *extremely faint* yellow or yellow-blue pattern, about two fingers wide, that is superimposed on the white background. This pattern is called *polarization brush* or *Haidinger's brush*. A rough illustration is given in Figure 66. The weak effect disappears after a few seconds if the head stops rotating along the line of sight. Haidinger's brush is due to the birefringence of Ref. 69 the cornea and the lens of the human eye, together with the morphology of the macula lutea inside the eye. The cornea acts as a radially oriented, colour-dependent polarizer, whereas the yellow spot acts as a radially oriented analyser. In short, the human eye is indeed able to see the directions in which the electric and magnetic field of light are oscillating.

Haidinger's brush, being yellow, is also visible in the blue sky, provided that the air is clear. (Indeed, it is easily drowned out by multiple scattering, and therefore provides a test of atmospheric transparency.) In the sky, Haidinger's brush is barely the size of a thumbnail at arm's length. (The angular size is the angular size of the macula.) The yellow arm of the cross points to the Sun, if you look about 90° away from it, high in the sky. To see it really clearly, hold a polarizer (or polarizing sunglasses) upwards and look through it, and then rotate it about the line of sight.

When polarized light is directed to a transparent medium, the ratio between the reflected and the transmitted light intensity depends on the polarization. The transmitted intensity can be zero or near zero for certain critical combinations of angles and polarizations. When the engineers at the Mercedes Benz car company forgot this, it cost the company millions of Euros. Behind the windshield, one of their car models had a sensor that detects whether it is day or night. The photodiode sensor worked well, except when the weather was extremely good, with a blue sky and no clouds; in that case, the sensor gave "night" as output. The mystery was solved when people recognized that the geometry was near the Brewster angle, that in such weather, the light from the sky is polarized and had a low amount of infrared light, at which the – wrongly chosen – photodiode was most sensitive. As a result, tens of thousands of cars had to be repaired.

Note that all possible polarizations of light form a continuous set. However, a general plane wave can be seen as the superposition of two orthogonal, linearly polarized waves with different amplitudes and different phases. Mathematically, all linearly polarized electromagnetic waves with the same frequency and direction for a two-dimensional vector space.

Light can also be *unpolarized*. Unpolarized light is a mixture of light of various polarizations. Light from the Sun and from other hot sources is typically unpolarized, due to the Brownian motion of the emitting sources. *Partially* polarized light is a mixture of polarized and unpolarized light.

In summary, for a wave in three-dimensional space, there are two basic types of polarization. One often classifies them into horizontal and vertical polarization, or, with other terms, into parallel and perpendicular polarization. A generally polarized wave is a superposition of these two basis states. These are the so-called *linear* polarization states.

Interestingly, a generally polarized plane wave can also be seen as the superposition of right and left *circularly polarized waves*. An illustration of a circularly polarized wave is given in Figure 68. In nature, circular polarization is extremely rare. Firefly larvae emit circularly polarized light. The light reflected by many species of scarab beetles is
Ref. 70 circularly polarized, as is the case for various stomatopod crustaceans, such as the mantis shrimp. The latter – and probably the former – are also able to detect circularly polarized light.

THE RANGE OF ELECTROMAGNETIC RADIATION

Electromagnetic waves of lower frequency, or radio waves, are commonly used to transmit mobile phone signals as well as television, radio and satellite programs. Like light, radio waves are due to moving electrons. In everyday life, light is (usually) generated by electrons accelerated inside atoms or molecules. Radio waves, which have lower frequency and thus larger wavelength, are more easily generated by electrons that are accelerated in metals roughly of the size of the wavelength; such pieces of metal are called *antennas*.

Radio waves emitted by a hand-held device can carry signals round the Earth. In other words, radio waves have a large range. How is this possible? After all, a static

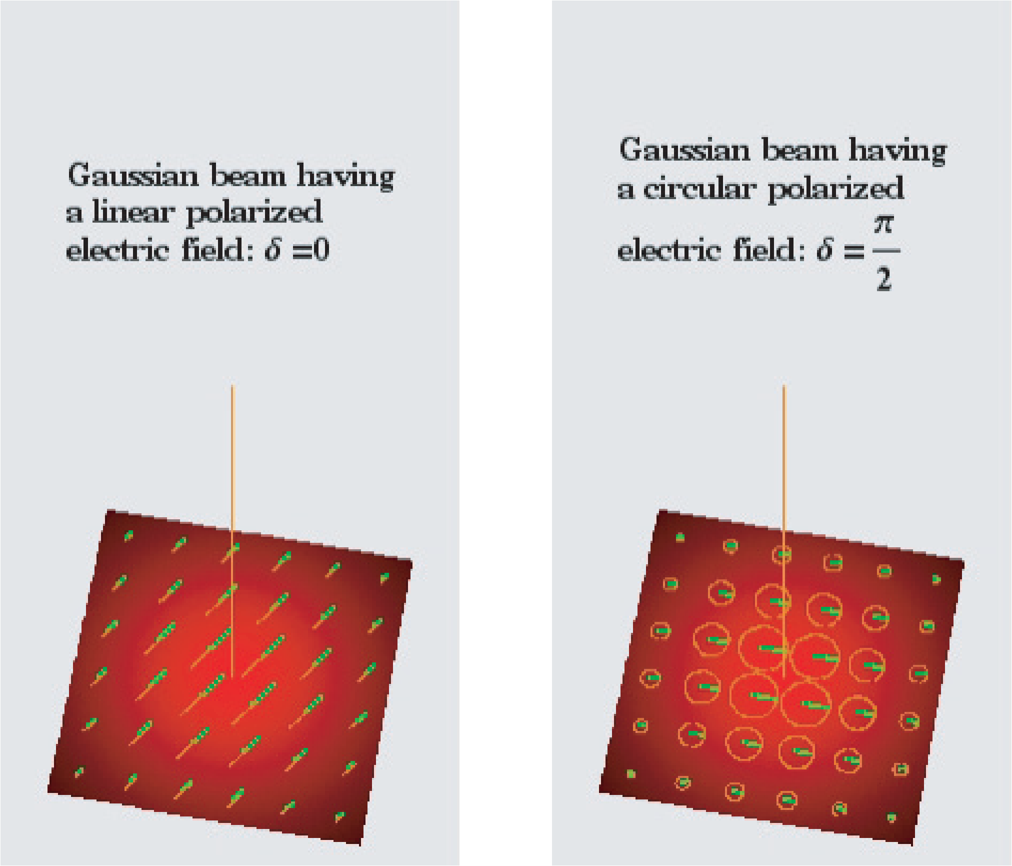


FIGURE 68 Left: the electric field of a Gaussian, linearly polarized electromagnetic wave (a beam); right: a Gaussian, circularly polarized beam (QuickTime film © José Antonio Díaz Navas).

electric field is usually unmeasureable after a distance of a dozen meters. It turns out that the field strength of radio waves decreases as $1/r$, where r is the distance from the source. The field strength thus decreases much more slowly than for static fields, which Ref. 71 decrease as $1/r^2$. Why is this the case?

The slow $1/r$ dependence of radio waves can be understood qualitatively from the drawing shown in Figure 69. It shows the electric field around a charged particle that undergoes the simplest possible accelerated motion: a bounce on a wall. In fact, the last, lower diagram is sufficient to show that the transverse field, given by the *kink* in the Challenge 116 d electric field lines, decreases as $1/r$. Can you deduce the dependence?

If we perform the construction of the field lines for a charge that undergoes *repeated* bounces, we get field lines with regularly spaced kinks that move away from the source. For a charge undergoing *harmonic* motion, we get the field lines shown in Figure 70. The figure thus shows the mechanism of the simplest antenna (or light source) one can imagine.

The magnitude of the transverse electric field can also be used to deduce the relation between the acceleration a of a charge q and the radiated electromagnetic power P. First,

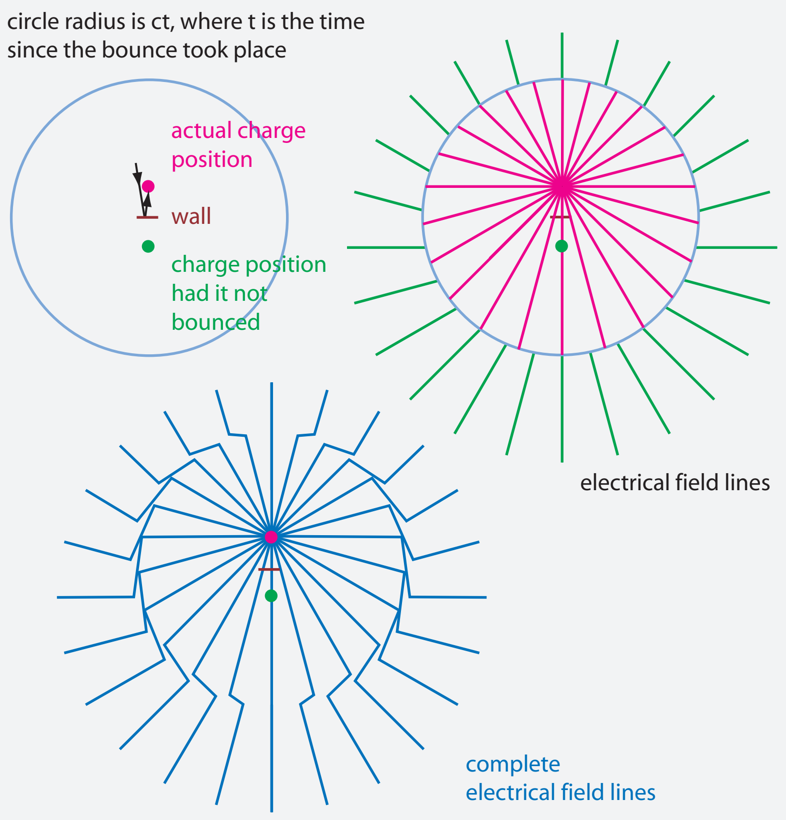


FIGURE 69 Constructing, in three steps, the electrical field around a charged particle bouncing from a wall.

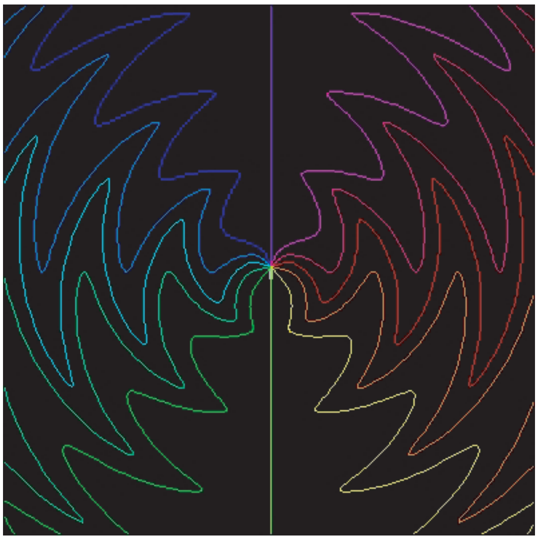


FIGURE 70 The electrical field around a particle oscillating in vertical direction (QuickTime film © Daniel Schroeder).

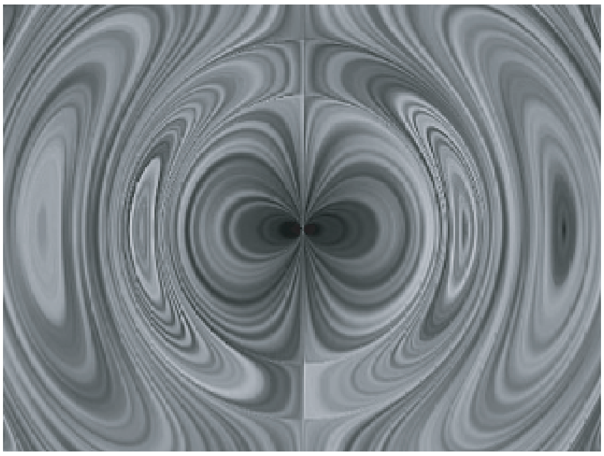


FIGURE 71 The electrical field around an oscillating *dipole* (QuickTime film © Daniel Weiskopf).

the transverse electric field (calculated in the last challenge) has to be squared, to give the local electric energy density. Then it has to be doubled, to include magnetic energy. Finally, we have to integrate over all angles; this gives a factor of 2/3. In total we get

$$P = \frac{q^2 a^2}{6\pi\varepsilon_0 c^3} \,. \tag{65}$$

The total radiated power P thus depends on the square of the acceleration and on the square of the charge that is being accelerated. This is the so-called *Larmor formula*. It shows why radio transmitters need power supplies and allows deducing how large they need to be. Note that Figure 69 and Figure 70 and also show that transmitter antennas have a *preferred* direction of power emission.

Usually, the source of electromagnetic radiation is described more accurately as an oscillating dipole. A visualization of the electric field in this case is given in Figure 71. At large distances, a wave section can be approximated as a plane wave.

In all cases, we find that the intensity of radio waves decrease slowly with distance and that radio communication is possible.

The slowness of progress in physics – and relativity

Gustav Kirchhoff's and Bernhard Riemann's expression from the 1850s for the speed of light and all other electromagnetic waves

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \tag{66}$$

is so strange that we should be intrigued whenever we see it. Something essential is missing. The expression states that the speed c is *independent* of the proper motion of the observer measuring the electromagnetic field and *independent* of the speed of the emitting source. In other words, the speed of light is predicted to be independent of the lamp speed and independent of the observer speed. This is indeed confirmed by all experiments, as explained in the volume on relativity. Vol. II, page 22

In addition, expression (66) implies that no observer can outrun light. In other words, light does *not* behave like a stream of bullets: the speed of bullet depends on the speed of the gun and of the target. A target can always outrun a bullet, if it moves rapidly enough. The speed of light is a *limit speed*.

Experiments confirm that also the speed of radio waves, of X-rays and of γ-rays is independent of the transmitter and the receiver. Experiments confirm that these speeds have the same value as the speed of light. All this is contained in expression (66). In short,

▷ The expression $c = 1/\sqrt{\varepsilon_0 \mu_0}$ shows that speed c is *invariant* and is the *limit energy speed* in nature.

Incredibly, *nobody* explored the consequences of this invariance until Lorentz and others started doing so in the 1890s, triggering Einstein until he settled the issues in 1905. The theory of relativity remained undiscovered for two generations! As in so many other cases, the progress of physics was much slower than necessary.

The invariance of the speed of light c is the essential point that distinguishes special relativity from Galilean physics. Since every electromagnetic device – such as every electric motor – makes use of expression (66), every electromagnetic device is a working proof of special relativity.

How does the world look when riding on a light beam?

At the end of the nineteenth century, the teenager Albert Einstein read a book series by Aaron Bernstein discussing the speed of light. Ref. 72 The book asked what would happen if an observer moved at the same speed as light. Einstein thought much about the issue, and in particular, asked himself what kind of electromagnetic field he would observe in that case. Einstein later explained that this Gedanken experiment convinced him already at that young age that *nothing* could travel at the speed of light, since the field observed would have a property not found in nature. Can you find out which one he meant? Challenge 117 s

Riding on a light beam situation would have strange consequences:

- You would have no mirror image, like a vampire.
- Light would not be oscillating, but would be a static field.
- Nothing would move, like in the tale of sleeping beauty.

But also at speeds *near* the velocity of light observations would be interesting. You would:

- see a lot of light coming towards you and almost no light from the sides or from behind; the sky would be blue/white in the front and red/black behind;
- observe that everything around happens very very slowly;
- experience the smallest dust particle as a deadly bullet.

Challenge 118 s Can you think of more strange consequences? It is rather reassuring that our planet moves rather slowly through its environment, when compared to the speed of light.

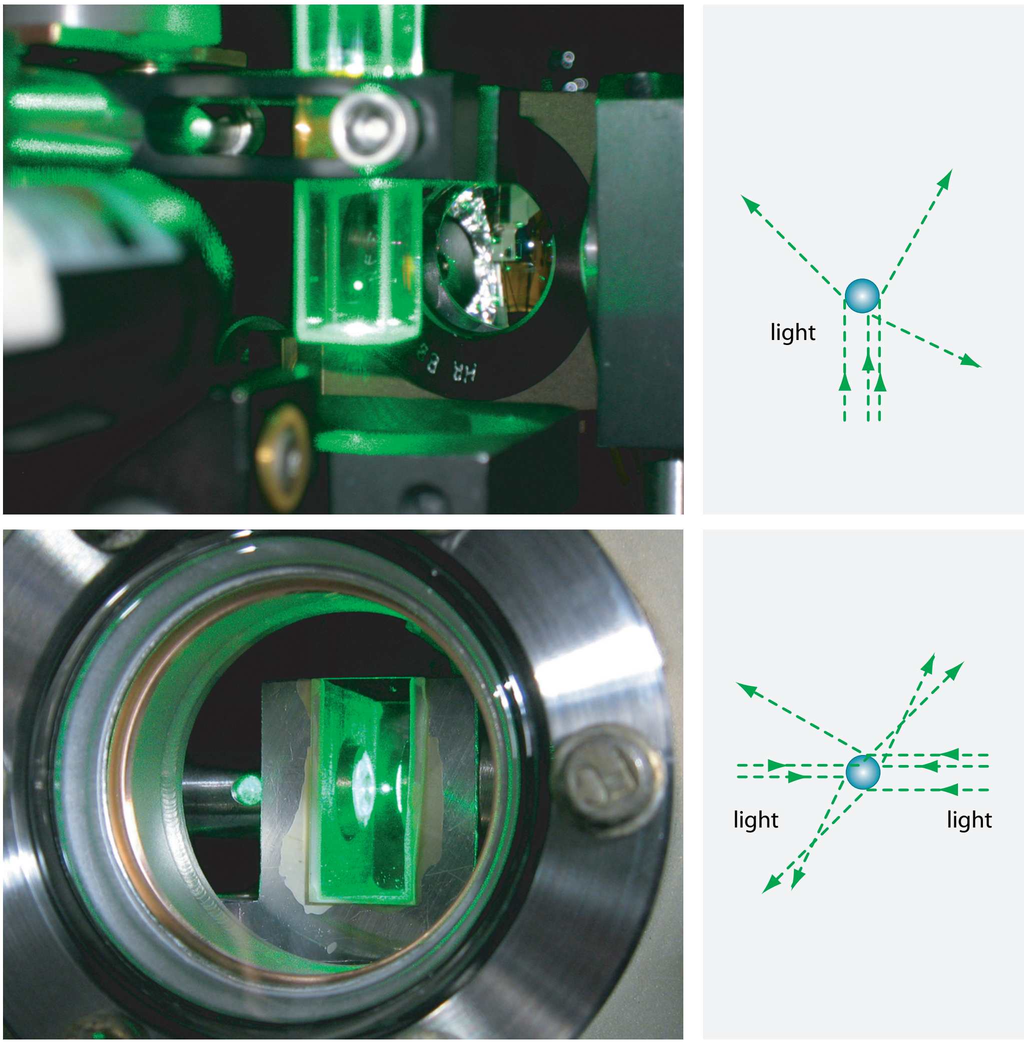


FIGURE 72 Levitating a small glass bead with a laser from below and with two opposed horizontal laser beams (© Mark Raizen, Tongcang Li).

CAN WE TOUCH LIGHT?

Ref. 73 If a little glass bead is put on top of a powerful laser, the bead remains suspended in mid-air, as shown in Figure 72.* This example of optical levitation proves that light has Vol. I, page 98 momentum. Therefore, contrary to what we said in the beginning of our mountain ascent, images *can* be touched! In fact, the ease with which objects can be pushed even has a special name. For planets and planetoids, it is called the *albedo*, and for general objects it is called the *reflectivity*, abbreviated as r.

---

* The heaviest object that has been levitated with a laser had a mass of 20 g; the laser used was the size of a building, and the method also made use of a few additional effects, such as internal shock waves, to keep the object in the air.

FIGURE 73 The tail of comet McNaught, photographed in Australia in 2007 (© Flagstaffotos).

Like each type of electromagnetic field, and like every kind of wave, light carries energy; the energy flow $\boldsymbol{T}$ per surface and time is Challenge 119 e

$$\boldsymbol{T} = \frac{1}{\mu_0} \boldsymbol{E} \times \boldsymbol{B} \quad \text{giving an average} \quad \langle T \rangle = \frac{1}{2\mu_0} E_{\text{max}} B_{\text{max}} . \tag{67}$$

Obviously, light also has a momentum P. It is related to the energy E by

$$P = \frac{E}{c} . \tag{68}$$

Challenge 120 e As a result, the pressure p exerted by light on a body is given by

$$p = \frac{T}{c}(1 + r) \tag{69}$$

where for black bodies we have a reflectivity $r = 0$ and for mirrors $r = 1$; other bodies have values in between. What is your guess for the amount of pressure due to sunlight on a black surface of one square metre? Is this the reason that we feel more pressure during the day than during the night? Challenge 121 s

If lasers are not available, rather delicate equipment is needed to detect the momentum or the radiation pressure of light. Already in 1619, Johannes Kepler had suggested in *De cometis* that the tails of comets exist only because the light of the Sun hits

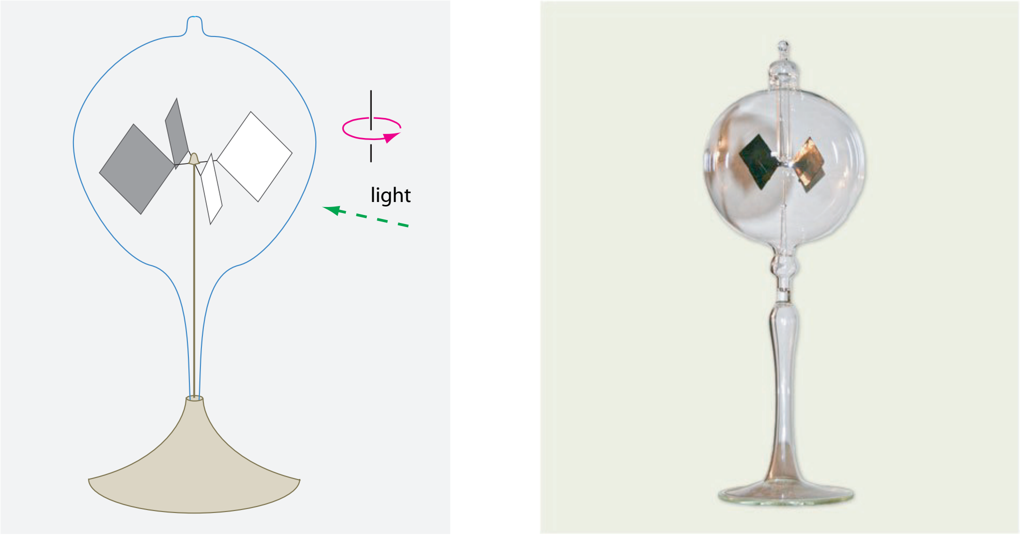


FIGURE 74 A commercial light mill turns *against* the light (Wikimedia).

the small dust particles that detach from it. For this reason, the tail always points *away* from the Sun, as you might want to check at the next opportunity. Today, we know that Kepler was right; but proving the hypothesis is not easy. Challenge 122 e

In order to detect the radiation pressure of light, in 1873, William Crookes* invented the *light mill radiometer*. The light mill consists of four thin plates, black on one side and shiny on the other, that are mounted on a vertical axis, as shown in Figure 74. However, when Crookes finished building it – it was similar to those sold in shops today – he found, like everybody else, that it turned in the wrong direction, namely with the shiny side towards the light! (Why is it wrong?) You can check it by yourself by shining a laser pointer on to it. The behaviour has been a puzzle for quite some time. Explaining it involves the tiny amount of gas left over in the glass bulb and takes us too far from the path of our adventure. It was only in 1901, with the advent of much better pumps, that the Russian physicist Pyotr Lebedew managed to create a sufficiently good vacuum to allow him to measure the light pressure with such an improved, true radiometer. Lebedew also confirmed the predicted value of the light pressure and proved the correctness of Kepler's hypothesis about comet tails. Today it is even possible to build tiny propellers that start to turn when light shines on to them, in exactly the same way that the wind turns windmills. Challenge 123 s Ref. 74 Ref. 75 Ref. 76

But light cannot only touch and be touched, it can also *grab*. In the 1980s, Arthur Ashkin and his research group developed actual *optical tweezers* that allow one to grab, suspend and move small transparent spheres of 1 to 20 μm diameter using laser beams. It is possible to do this through a microscope, so that one can also observe at the same time what is happening. This technique is now routinely used in biological research around Ref. 77

---

[*] William Crookes (b. 1832 London, d. 1919 London), chemist and physicist, discoverer of thallium, mistaken discoverer of other 'elements', convinced believer in spiritualism and president of the Society for Psychical Research. For this bizarre mix of achievements he was elected to the Royal Society and received numerous prizes and other honours.

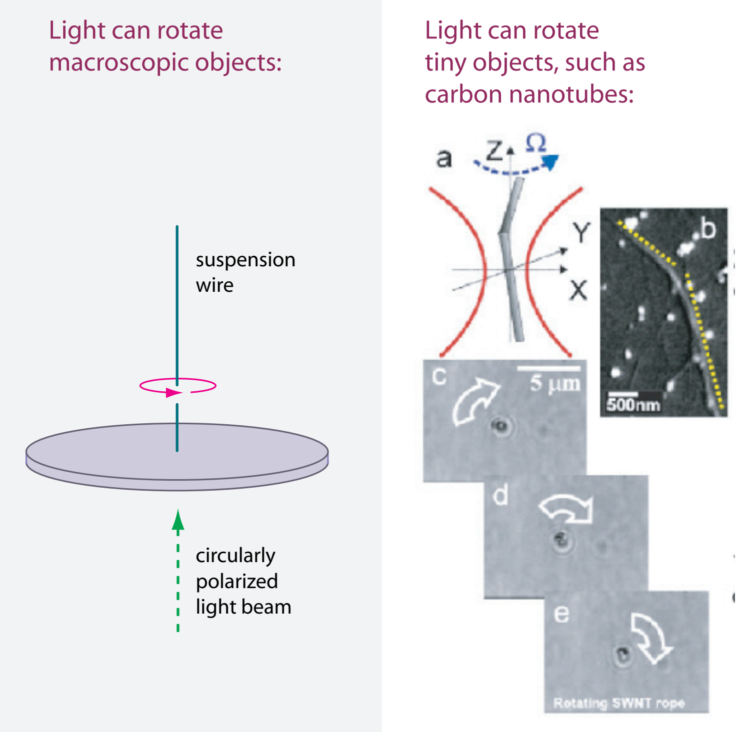


FIGURE 75 Light can rotate objects (© A.C. Ferrari)

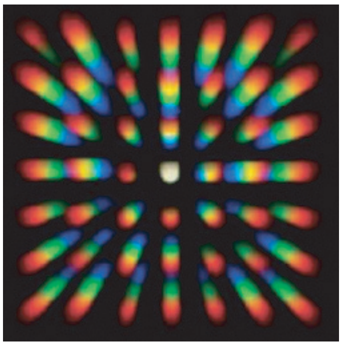


FIGURE 76 Umbrellas decompose white light: look at a small lamp through a black umbrella at night (© Wikimedia).

the world, and has been used, for example, to measure the force of single muscle fibres, by chemically attaching their ends to glass or Teflon spheres and then pulling them apart with such optical tweezers.

But that is not all. In the last decade of the twentieth century, several groups even Ref. 77 managed to *rotate* objects, thus realizing actual *optical spanners*. They are able to rotate particles at will in one direction or the other, by changing the optical properties of the laser beam used to trap the particle.

In fact, it does not take much to deduce that if light has linear momentum, circularly polarized light also has *angular* momentum. In fact, for such a wave the angular momentum $\boldsymbol{L}$ is given by

$$L = \frac{E}{\omega} , \tag{70}$$

Challenge 124 e where E is the energy. Equivalently, the angular momentum of a wave is $\lambda/2\pi$ times its linear momentum $\boldsymbol{p}$. Ref. 78 For light, this result was already confirmed in the early twentieth century: a light beam can put certain materials (which ones?) Challenge 125 ny into rotation; in liquids, this is now standard practice in laboratories. Ref. 79 Two examples are shown in Figure 75. Of course, the effect is even stronger with a laser beam. But already in the 1960s, a beautiful demonstration was performed with microwaves. A circularly polarized microwave beam from a maser – the microwave equivalent of a laser – can put a metal piece absorbing it into rotation. Indeed, for a beam with cylindrical symmetry, depending on the sense of rotation, the angular momentum is either parallel or antiparallel to the direction of propagation. All these experiments confirm that light also carries angular momentum, an effect which will play an important role in the quantum part of our mountain ascent.

We note that not for all waves in nature is angular momentum given by energy per angular frequency. This is only the case for waves made of what in quantum theory will be called spin 1 particles. For example, for gravity waves the angular momentum is *twice* this value, and they are therefore expected to be made of spin 2 particles.

What does this mean for the comet tails mentioned above? The issue was settled definitely in 1986. Ref. 80 A satellite was shot up to an altitude of 110 000 km and made to release a cloud of barium. The cloud was visible from the Earth, and it soon developed a tail that was visible from Earth: that was the first artificial comet. It turns out that comet tails shapes are partly due to hitting photons, but also partly to the solar wind and even to magnetic fields.

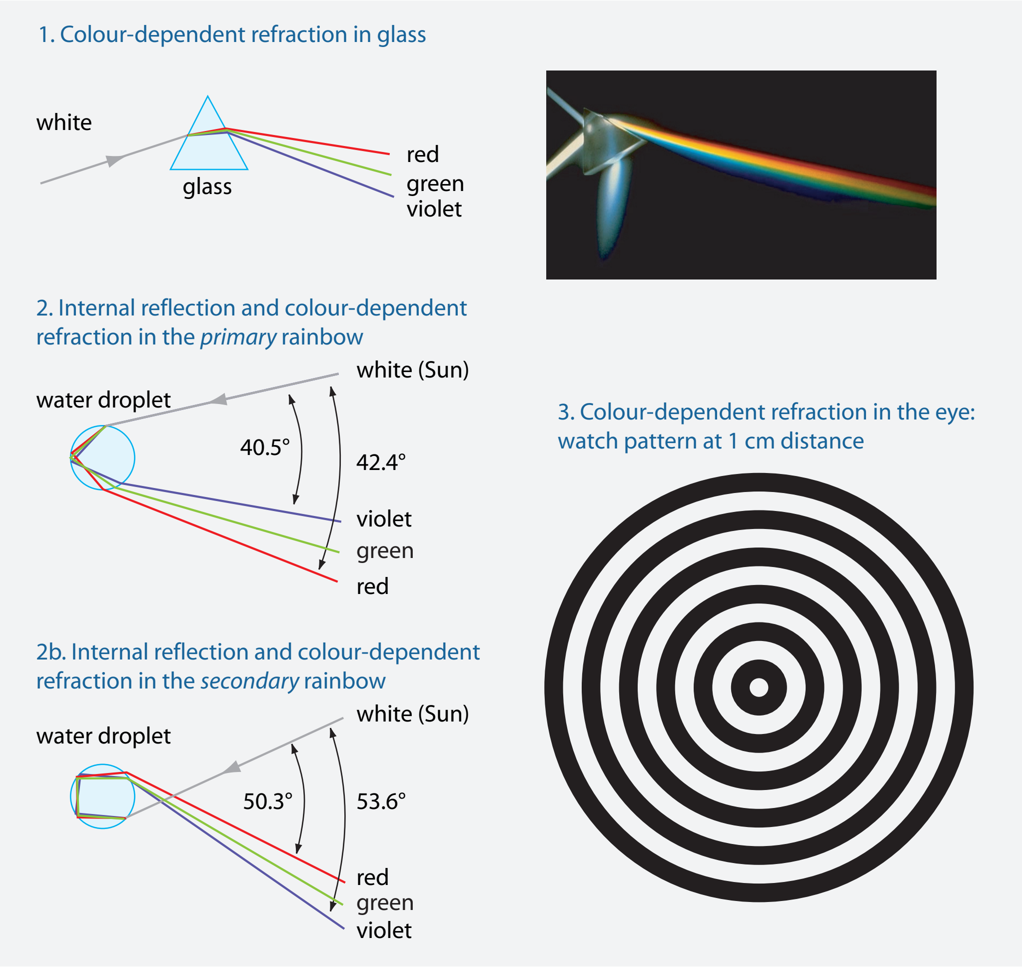
In summary, light can touch, light can rotate, and light can be touched. Obviously, if light can rotate bodies, it can also *be* itself rotated. Could you imagine how this can be achieved? Challenge 126 s

War, light and lies

From the tiny effects of equation (69) for light pressure we deduce that light is not an efficient tool for hitting objects. On the other hand, light is able to *heat up* objects, as we can feel in the sun or when the skin is touched by a laser beam of about 100 mW or more. For the same reason even cheap laser pointers are dangerous to the eye.

In the 1980s, and again in 2001, a group of people who had read too many science fiction novels managed to persuade the military – who also indulge in this habit – that lasers could be used to shoot down missiles, and that a lot of tax money should be spent on developing such lasers. Using the definition of the Poynting vector and a hitting time of about 0.1 s, are you able to estimate the weight and size of the battery necessary for such a device to work? What would happen in cloudy or rainy weather? Challenge 127 ny

Other people tried to persuade NASA to study the possibility of propelling a rocket using emitted light instead of ejected gas. Are you able to estimate that this is not feasible? Challenge 128 e



1. Colour-dependent refraction in glass

2. Internal reflection and colour-dependent refraction in the *primary* rainbow

2b. Internal reflection and colour-dependent refraction in the *secondary* rainbow

3. Colour-dependent refraction in the eye: watch pattern at 1 cm distance

FIGURE 77 Three proofs that *white* light is a mixture of colours (with exaggerated angle differences): prism decomposition, rainbow formation (simplified, as explained in the text) and the coloured borders seen on a circular black and white pattern (photograph by Susan Schwartzenberg, © Exploratorium www.exploratorium.edu).

WHAT IS COLOUR?

Challenge 129 s

We saw that radio waves of certain frequencies are visible. Within that range, different frequencies correspond to different colours. (Are you able to convince a friend about this?) But the story does not finish here. Numerous colours can be produced either by a single wavelength, i.e., by *monochromatic* light, or by a *mixture* of several different colours. For example, standard yellow can be, if it is pure, an electromagnetic beam of 575 nm wavelength or it can be a mixture of standard green of 546.1 nm and standard red of 700 nm. The eye cannot distinguish between the two cases; only spectrometers can. In everyday life, all colours turn out to be mixed, with the exceptions of those of yellow street lamps and of laser beams and of laboratory spectra. You can check this for yourself, using an umbrella or a compact disc: they decompose light mixtures, but they do not decompose pure colours, such as those from a laser pointer or an LED display.

Challenge 130 e Even the colours of the rainbows are impure, because they are mixed with the white light of the background sky and because the diameter of the Sun smears the spectrum.

In particular, *white light* is a mixture of a continuous range of colours with a specific intensity per wavelength. If you want to check that white light is a mixture of colours without any light source, simply hold the lower right-hand side of Figure 77 so close to your eye that you cannot focus the stripes any more. The unsharp borders of the white stripes have either a pink or a green shade. These colours are due to the imperfections of the human eye, its so-called *chromatic aberrations*. Chromatic aberrations have the consequence that not all light frequencies follow the same path through the lens of the eye, and therefore they hit the retina at different spots. This is the same effect that occurs in prisms or in water drops showing a rainbow.

The left-hand side of Figure 77 explains how rainbows form. Above all, the internal reflection inside the water droplets in the sky is responsible for throwing back the light coming from the Sun, whereas the wavelength-dependent refraction at the air–water surface is responsible for the different paths of each colour. The first two persons to verify this explanation were Theodoricus Teutonicus de Vriberg (*c.* 1240 to *c.* 1318), in Ref. 81 the years from 1304 to 1310 and, at the same time, the Persian mathematician Kamal al-Din al-Farisi. To check the explanation, they did something smart and simple that Challenge 131 e anybody can repeat at home. They built an enlarged water droplet by filling a thin spherical (or cylindrical) glass container with water; then they shone a beam of white light through it. Theodoricus and al-Farisi found exactly what is shown in Figure 77. With this experiment, each of them was able to reproduce the opening angle of the main or *primary* rainbow, its colour sequence, as well as the existence of a *secondary* rainbow, its observed angle and its inverted colour sequence.* All these rainbows are found in Page 102 Figure 57. Theodoricus's beautiful experiment is sometimes called the most important contribution of natural science in the Middle Ages.

By the way, the shape of the rainbow tells something about the shape of the water Challenge 133 s droplets. Can you deduce the connection?

Incidentally, the explanation of the rainbow given in Figure 77 is not complete. It assumes that the light ray hits the water droplet at a specific spot on its surface. If the light ray hits the droplet at other spots – technically, at other impact parameters, the rainbows appear at other angles; however, all those other rainbows wash out. Only the visible rainbow remains, because its deflection angles are extremal. The primary rainbow is, in fact, the coloured edge of a white disc. And indeed, the region above the primary bow is always darker than the region below it.

Water droplets are not the only prisms found in nature. At sunset, the atmosphere itself also acts as a prism, or more precisely, as a cylindrical lens affected by spherochromatism. Therefore, especially at sunset, the Sun is split into different images, one for each colour, which are slightly shifted with respect to each other; the total shift is about Ref. 84 1 % of the diameter. As a result, the rim of the evening Sun is coloured. If the weather is

---

Challenge 132 s * Can you guess where the ternary and quaternary rainbows are to be seen? There are rare reported sightings of them; only two or three photographs exist world-wide. The hunt to observe the fifth-order rainbow is Ref. 82 still open. (In the laboratory, bows around droplets up to the thirteenth order have been observed.) For more details, see the beautiful website at www.atoptics.co.uk. There are several formulae for the angles of the various orders of rainbows; they follow from straightforward geometric considerations, but are too involved to be given here.

FIGURE 78 A green flash above the setting Sun and one above the Moon, showing also the colour change of the Moon rim (© Andrew Young and Laurent Laveder/PixHeaven.net).

FIGURE 79 Milk and water simulate the evening sky (© Antonio Martos).

favourable, if the air is clear up to and beyond the horizon, and if the correct temperature profile is present in the atmosphere, a colour-dependent mirage will appear: for about a second it will be possible to see, after or near the red, orange and yellow images of the setting Sun, the green–blue image, sometimes even detached. This is the famous *green* Ref. 83 *flash* described by Jules Verne in his novel *Le Rayon-vert*. The green flash is often seen Ref. 84, Ref. 85 on tropical beaches, for example in Hawaii, and from the decks of ships in warm waters.

Even pure air splits white light. However, this effect is not due to dispersion, but to scattering. Wavelength-dependent scattering, mainly *Rayleigh scattering*, is the reason that the sky and distant mountains look blue and that the Sun looks red at sunset and sunrise. (The sky looks black even during the day from the Moon.) You can repeat this effect by looking through water at a black surface or at a lamp. Adding a few drops of milk to the water makes the lamp yellow and then red, and makes the black surface blue (like the sky seen from the Earth as compared to the sky seen from the Moon) as shown in Figure 79. More milk increases the effect. For the same reason, sunsets are especially

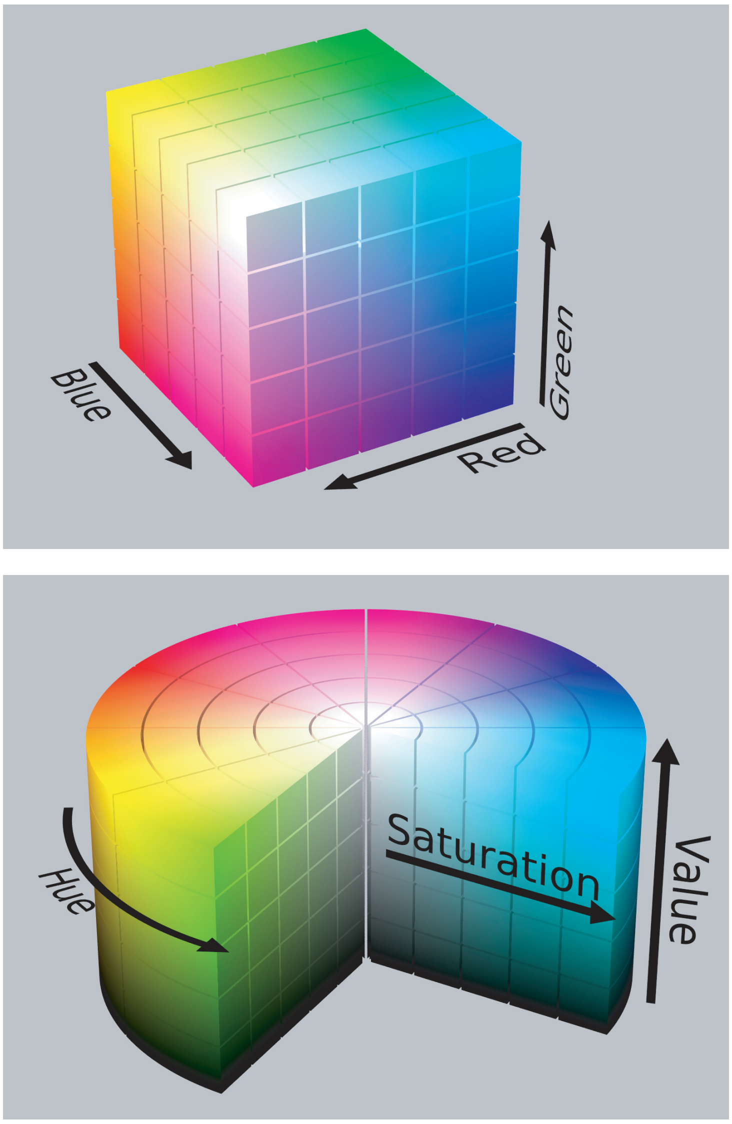


FIGURE 80 Two of the many ways to illustrate the set of all possible human colours: (top) as mixtures of red, green and blue values that increase along the three coordinate axes, and (bottom) using hue, saturation and brightness value coordinates (© SharkD).

red after volcanic eruptions.

Ref. 86

In the evening, however, the sky is blue for another, far less known reason: at the time around sunset, the sky is blue mainly because of the ozone layer. Ozone is a blue gas. Without ozone, the sky would be yellowish during sunsets.

In summary, light is, in general, a mixture of wavelengths. As a result, light wavelength or frequency are *not* sufficient to describe colour. Colour experts call *hue* that aspect of colour that matches most closely the change with wavelength. But every colour has two additional characteristics. For example, any given colour can be bright or dark; *brightness* is a second, independent property of colour. A third independent property of colour is its *saturation*; it expresses how strongly a colour differs from white. A strongly saturated colour is the opposite of a pale, or weakly saturated colour.

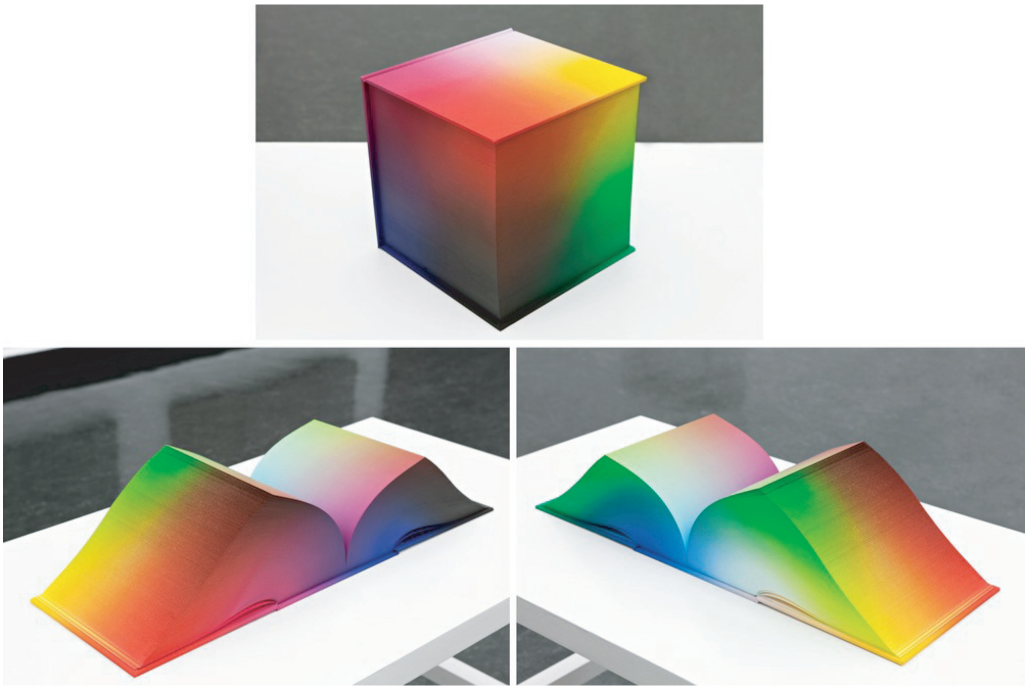


FIGURE 81 A unique colour book that illustrates, on each page and on all its outside surfaces, the three-dimensional colour space of humans (© Tauba Auerbach).

Human colour space is three-dimensional. Humans are thrichromatic. Figure 80 illustrates the point. Every colour we see is described by *three* independent parameters, because the human eye has three types of cones, thus three types of colour-sensitive cells. This is the reason that any colour selection scheme, for example on a computer, has – at least – three parameters that can be varied. A modern artist, Tauba Auerbach, even produced a beautiful book version of the colour space, shown in Figure 81. The number three is also the reason that every display has at least three different types of pixels. These three parameters do not need to be hue, saturation and brightness value. They can also be taken to be the intensities of red, green and blue. Many other colour properties can be used to describe colour, such as lightness, chroma, purity, luma and others. Also descriptions with four and more parameters – which then are not independent from each other – are used, especially in the printing industry.

Ref. 87

Many birds, reptiles, fish and various insects have four-dimensional colour spaces that include the ultraviolet; butterflies and pigeons have five-dimensional colour spaces, and other bird species have even higher-dimensional colour spaces. Mantis shrimps possibly have the most complex eyes in the animal kingdom, with up to twelve-dimensional colour spaces. (One species of mantis shrimps, *Gonodyctylus smithii*, can also detect circular and linear light polarization in complete detail.) In contrast to humans and apes, most mammals have only two-dimensional colour spaces. Also colour-blind persons can have lower-dimensional colour spaces. In other terms, the number of dimensions of the perceived colour space is not a property of light, nor a property of nature, but a specific property of our human eyes. *Colours in nature and colours perceived by humans differ.*

FIGURE 82 Exceptionally many supernumerary rainbows (© Denis Betsch).

There is no colour space in nature.

Colours in nature and colours in human perception differ in an additional way, discovered by linguists. In human language, colours have a natural *order*. All people of the world, whether they come from the sea, the desert or the mountains, order colours in the following sequence: 1. black and white, 2. red, 3. green and yellow, 4. blue, 5. brown, 6. mauve, pink, orange, grey and sometimes a twelfth term that differs from language to language. (Colours that refer to objects, such as aubergine or sepia, or colours that are not generally applicable, such as blond, are excluded in this discussion.) The precise discovery is the following: if a particular language has a word for any of these colours, then it also has a word for all the preceding ones. The result also implies that people use these basic colour classes even if their language does *not* have a word for each of them. These strong statements have been confirmed for over 100 languages. Ref. 88

Fun with rainbows

The width of the usual, primary rainbow is 2.25°, for the secondary rainbow it is about twice that value (which is one reason why it is less bright). The width is larger than the dispersion angle difference given in Figure 77 because the angular size of the sun, about 0.5°, has (roughly) to be added on top of the angle difference.

The finite size of droplets leads, via interference, to the supernumerary rainbows, as mentioned above. If the droplets are small and all of the same size, the number of super- Page 102

FIGURE 83 Five rare types of rainbows: a fogbow (top left), an irregular, split rainbow in a windy situation due to non-spherical rain drops (top right, shown with increased colour saturation), a six-fold rainbow (middle left), a red rainbow at sunset (middle right), and a moonbow, created by the Moon, not by the Sun, and brightened digitally (© Michel Tournay, Eva Seidenfaden, Terje Nordvik, Zhu XiaoJin and Laurent Laveder).

numerary rainbows increases, as Figure 82 shows strikingly.

If the droplets are extremely fine, the rainbow becomes white; it is then called a *fogbow*. Such bows are also often seen from aeroplanes. If the droplets are not round, for Ref. 60 example due to strong wind, one can get a so-called *irregular* or *twinned rainbow*. An example is shown in Figure 83.

Light from the rainbow is tangentially polarized. You can check that easily with polarizing sunglasses. During the internal reflection in the water droplets, as the reflection Challenge 134 e

FIGURE 84 A composite photograph showing the parhelia, the light pillars, the halo and the upper tangent arc formed by ice crystals in the air, if they are all are oriented in the same direction (© Phil Appleton).

FIGURE 85 A rare circumzenithal arc formed by hexagonal ice crystals in upper regions of the atmosphere (© Paul Gitto).

angle is very near to the angle at which total reflection sets in, light gets polarized. (Why does this lead to polarization?) Challenge 135 ny

If the air is full of ice crystals instead of droplets, the situation changes again. One can then get additional images of the sun in the direction of the sun. They are called *parhelia*, sometimes also Sun dogs. This happens most clearly with no wind, if the crystals are all oriented in the same direction. In that case one can take photographs such as the one shown in Figure 84.

Rare bows and other astonishing atmospheric effects are best explored on the website providing the 'optical picture of the day' at www.atoptics.co.uk/opod.htm. There one can find third- and fourth-order rainbows, fogbows that include supernumerary bows, lunar fogbows, rainbows whose secondary bow has supernumeraries, irregular rainbows, moonbows, circumzenithal arcs, Sun's halos, Sun's pillars, green flashes, and much more. The website presents the beauty of light in nature – and all effects are also explained in detail.

What is the speed of light? What is signal speed?

Physics talks about motion. Talking is the exchange of sound; and sound is an example of a signal.

▷ A *(physical) signal* is the transport of information using the transport of energy.

Vol. I, page 306 There are no signals without a motion of energy. Indeed, there is no way to store information without storing energy. To any signal we can thus ascribe a propagation speed. We call it the *signal speed*. The highest possible signal speed is also the maximal velocity of the general influences, or, to use sloppy language, the maximal velocity with which effects spread causes.

If the signal is carried by matter, such as by the written text in a letter, the signal velocity is the velocity of the material carrier. Experiments show that this speed is limited by the speed of light.

For a wave carrier, such as water waves, sound, light or radio waves, the situation is less evident. What is the speed of a wave? The first answer that comes to mind is the speed with which wave crests of a sine wave move. This already introduced *phase velocity* is given by the ratio between the frequency and the wavelength of a monochromatic wave, i.e., by

$$v_{\mathrm{ph}} = \frac{\omega}{k} \; . \tag{71}$$

For example, the phase velocity determines interference phenomena. Light in a vacuum has the same phase velocity $v_{\mathrm{ph}} = c$ for all frequencies. Are you able to imagine an experiment to test this to high precision? Challenge 136 s

On the other hand, there are cases where the phase velocity is *greater* than c, most notably when light travels through an absorbing substance, and when at the same time the frequency is near to an absorption maximum. In these cases however, experiments show that the phase velocity is *not* the signal velocity. For such situations, a better approximation to the signal speed is the *group velocity*, i.e., the velocity at which a group maximum will travel. This velocity is given by Ref. 89

$$v_{\mathrm{gr}} = \left.\frac{\mathrm{d}\omega}{\mathrm{d}k}\right|_{k_0} \; , \tag{72}$$

where k_0 is the central wavelength of the wave packet at which the derivative is taken.

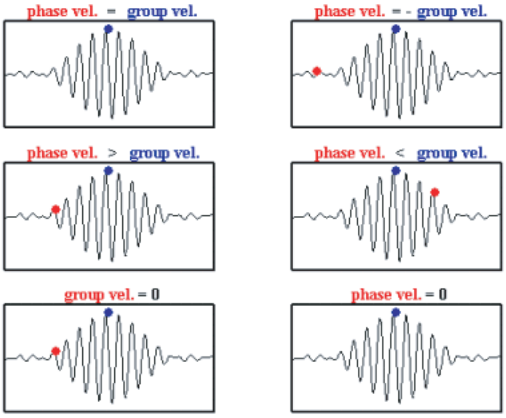


FIGURE 86 A visualisation of group velocity (dark blue) and phase velocity (bright red) for different types of waves (QuickTime film © ISVR, University of Southampton).

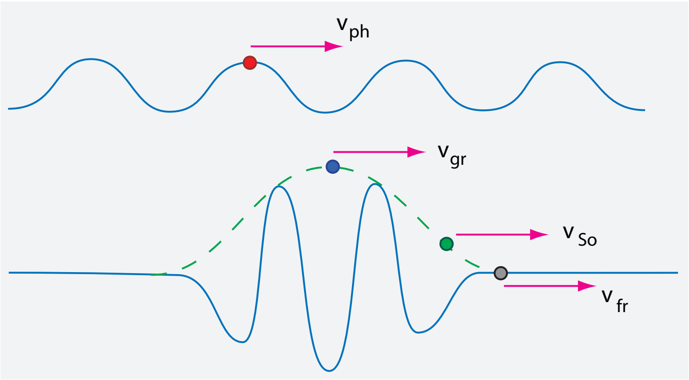


FIGURE 87 The definition of the important velocities in wave phenomena: the phase velocity, the group velocity, Sommerfeld's front velocity and the forerunner velocity.

We observe that $\omega = c(k)k = 2\pi v_{\mathrm{ph}}/\lambda$ implies the relation

$$v_{\mathrm{gr}} = \left.\frac{\mathrm{d}\omega}{\mathrm{d}k}\right|_{k_0} = v_{\mathrm{ph}} - \lambda \frac{\mathrm{d}v_{\mathrm{ph}}}{\mathrm{d}\lambda} \ . \qquad (73)$$

This means that the sign of the last term determines whether the group velocity is larger or smaller than the phase velocity. For a travelling group, as shown by the dashed line in Figure 87, this means that new maxima appear either at the end or at the front of the group. Experiments show that this is only the case for light passing *through matter*; for light *in vacuum*, the group velocity has the same value $v_{\mathrm{gr}} = c$ for all values of the wave vector magnitude k.

Challenge 137 ny

You should be warned that many publications are still propagating the incorrect statement that the group velocity *in a material* is never greater than c, the speed of light in vacuum. Actually, the group velocity in a material can be zero, infinite or even negative; this happens when the light pulse is very narrow, i.e., when it includes a wide range of frequencies, or again when the frequency is near an absorption transition. In many (but not all) cases the group is found to widen substantially or even to split, making it difficult to define precisely the group maximum and thus its velocity. Many experiments have confirmed these predictions. For example, the group velocity in certain materials has been measured to be *ten times* that of light. The refractive index then is smaller than 1. However, in all these cases the group velocity is *not* the same as the signal speed.* Ref. 90

What then is the best velocity describing signal propagation? Arnold Sommerfeld** almost solved the main problem in the beginning of the twentieth century. He defined the signal velocity as the velocity v_{So} of the *front slope* of the pulse. The definition is illustrated in Figure 87. The definition cannot be summarized in a formula, but it does have the property that it describes signal propagation for almost all experiments, in particular those in which the group and phase velocity are larger than the speed of light. When studying its properties, it was found that for no material is Sommerfeld's signal velocity greater than the speed of light in vacuum. Ref. 89

Sometimes it is conceptually easier to describe signal propagation with the help of the energy velocity. As previously mentioned, every signal transports energy. The *energy velocity* v_{en} is defined as the ratio between the energy flow density $\boldsymbol{S}$, i.e., the Poynting vector, and the energy density W, both taken in the direction of propagation. For electromagnetic fields – the only ones fast enough to be interesting for eventual superluminal signals – this ratio is

$$\boldsymbol{v}_{\text{en}} = \frac{\langle \boldsymbol{P} \rangle}{\langle W \rangle} \; . \tag{74}$$

However, as in the case of the front velocity, in the case of the energy velocity we have to specify the underlying averaging procedure, denoted by ⟨⟩, i.e., whether we mean the energy transported by the main pulse or by the front of it. In vacuum, neither speed is ever greater than the speed of light.*** (In general, the velocity of energy in matter has a value slightly different from Sommerfeld's signal velocity.) Ref. 89

In recent years, the progress in light detector technology, allowing one to detect even the tiniest energies, has forced scientists to take the fastest of all these energy velocities to describe signal velocity. Using detectors with the highest possible sensitivity we can use as signal the first point of the wave train whose amplitude is different from zero, i.e., the first tiny amount of energy arriving. This point's velocity, conceptually similar to Sommerfeld's signal velocity, is commonly called the *front velocity* or, to distinguish it

---

[*] In quantum mechanics, Erwin Schrödinger proved that the velocity of an electron is given by the group velocity of its wave function. Therefore the same discussion reappeared in quantum theory, as we will find out in the next volume of our mountain ascent. Vol. IV, page 94

[**] Arnold Sommerfeld (b. 1868 Königsberg, d. 1951 Munich) was a central figure in the spread of special and general relativity, of quantum theory, and of their applications. A professor in Munich, an excellent teacher and text book writer, he worked on atomic theory, on the theory of metals and on electrodynamics, and was the first to understand the importance and the mystery around 'Sommerfeld's famous fine structure constant.'

[***] Signals not only carry energy, they also carry negative entropy ('information'). The entropy of a transmitter increases during transmission. The receiver decreases in entropy (but less than the increase at the transmitter, of course). Ref. 92

Note that the negative group velocity implies energy transport against the propagation velocity of light. This is possible only in *energy loaded* materials. Ref. 93

Challenge 138 s even more clearly from Sommerfeld's case, the *forerunner velocity*. It is simply given by

$$v_{\mathrm{fr}} = \lim_{\omega \to \infty} \frac{\omega}{k} \, . \tag{75}$$

The forerunner velocity is *never* greater than the speed of light in a vacuum, even in materials. In fact it is precisely c because, for extremely high frequencies, the ratio ω/k is independent of the material, and vacuum properties take over.

▷ The forerunner velocity is the true signal velocity or the *true velocity of light*.

Using the forerunner speed, all discussions on light speed become clear and unambiguous.

To end this section, here are two challenges for you. Which of all the velocities of light is measured in experiments determining the velocity of light, e.g. when light is sent to the Moon and reflected back? And now a more difficult one: why is the signal speed of light inside matter less than the speed in vacuum, as all experiments show?

Challenge 139 s

Challenge 140 s

SIGNALS AND PREDICTIONS

When one person reads a text over the phone to a neighbour who listens to it and maybe repeats it, we speak of communication. For any third person, the speed of communication is always less than the speed of light. But if the neighbour already knows the text, he can recite it without having heard the readers' voice. To the third observer such a situation appears to imply motion that is faster than light. Prediction can thus *mimic* communication and, in particular, it can mimic faster-than-light (superluminal) communication. Such a situation was demonstrated most spectacularly in 1994 by Günter Nimtz, who seemingly transported music – all music is predictable for short time scales – through a 'faster-than-light' system. To distinguish between the two situations, we note that in the case of prediction, no transport of energy takes place, in contrast to the case of communication. In other words, the definition of a signal as a transporter of information is not as useful and clear-cut as the definition of a signal as a *transporter of energy*. In the above-mentioned experiment, no energy was transported faster than light. The same distinction between prediction on the one hand and signal or energy propagation on the other will be used later to clarify some famous experiments in quantum mechanics.

Ref. 91

> “If the rate at which physics papers are being published continues to increase, physics journals will soon be filling library shelves faster than the speed of light. This does not violate relativity since no useful information is being transmitted.”
>
> David Mermin

AETHER GOOD-BYE

Gamma rays, X-rays, light and radio waves are moving electromagnetic waves. All exist in empty space. What is oscillating when light travels? Maxwell himself called the

TABLE 15 Experimental properties of *flat*, classical vacuum, thus neglecting all quantum effects and all effects of general relativity.

Physical property	Experimental value
Permeability	μ_0 =1.3 μH/m
Permittivity	ε_0 =8.9 pF/m
Wave impedance/resistance	Z_0 = 376.7 Ω
Conformal invariance	applies
Spatial dimensionality	3
Topology	R^3
Friction on moving bodies	none
Components	none
Mass and energy content	none
Motion	none

oscillating 'medium' the *aether*. The properties of the oscillating medium that are measured in experiments are listed in Table 15. The strange numerical values are due to the definition of the units henry and farad. Page 353

Ref. 94 The last item of Table 15 is the most important: despite intensive efforts, nobody has been able to detect any *motion* of the so-called aether. In particular, there is no motion of the aether relative to the vacuum. In other words, even though the aether supposedly oscillates, it does not move. Together with the other data, all these results can be summed up in one sentence: there is no way to distinguish the aether from the vacuum. Challenge 141 e

Sometimes one hears that certain experiments or even the theory of relativity show that the aether does not exist. There is a lot of truth in this statement; in fact, experiments show something even more important:

▷ The aether is *indistinguishable* from the vacuum.

This statement is true in all cases. For example, we found out in the section on general relativity that a curved vacuum *can* move; but the aether still remains indistinguishable from it.* Also quantum field theory confirms the identity of aether and vacuum.

What then is oscillating in the case of electromagnetic waves? We now have a simple answer to this old question: the vacuum. The vacuum is the carrier, or carrier medium, of electromagnetic waves. The flat, Lorentz-invariant vacuum carries waves, even though it cannot move and it does not provide a favourite coordinate system. Flat vacuum is thus something special, and it is also acceptable to avoid the terms 'carrier' or 'medium' altogether. In some bizarre clubs it is even compulsory to do so. However, this avoidance

---

Ref. 95 * Historically, the term 'aether' has been used as an expression for several different ideas, depending on the author. First of all it was used for the idea that a vacuum is not empty, but *full*; secondly, that this fullness can be described by *mechanical models*, such as gears, little spheres, vortices, etc.; thirdly, it was imagined that the aether is a *substance*, similar to matter. All these ideas are put to rest by relativity. Nevertheless, these issues will reappear in the last part of our mountain ascent, when the description of the vacuum itself is explored.

is impossible in general relativity, as we have seen, and is equally impossible in quantum field theory, as we will find out.*

In short, experiments in the domain of special relativity have *abolished* the aether: it is a superfluous concept; the physical vacuum has many of the properties that were once ascribed to the aether. From now on, we will drop the concept of aether from our vocabulary. On the other hand, we have not yet finished the study of the vacuum; vacuum will keep us busy for the rest of our walk, starting with the part of our adventure that follows, the part on quantum physics. In fact, quantum physics shows that all experimental values in Table 15 require amendments. Ref. 95 Challenge 142 d

CHALLENGES AND FUN CURIOSITIES ABOUT LIGHT, POLARIZATION AND THE GEOMETRIC PHASE

Since light is a wave, something must happen if it is directed to a hole less than its wavelength in diameter. What exactly happens? Challenge 143 s

* *

On a sunny day at moderate latitudes on the Earth, sunlight has a power density of 1 kW/m^2. What is the corresponding energy density and what are the average electric and magnetic fields? Challenge 144 s

* *

Spectrally pure light is often called 'monochromatic'. Why is this a misnomer? Challenge 145 s

* *

Electrodynamics shows that light beams always push; they never pull. Can you confirm that 'tractor beams' are impossible in nature? Challenge 146 e

* *

It is well known that the glowing material in light bulbs is tungsten wire in an inert gas. This was the result of a series of experiments that began with the grandmother of all lamps, namely the cucumber. The older generation knows that a pickled cucumber, when attached to the 230 V of the mains, glows with a bright green light. (Be careful; the experiment is dirty and dangerous.)

* *

Light beams have an effective temperature and entropy. Though not often discussed nowadays, the thermodynamics of light has been explored in great detail by Max von Laue (b. 1879 Koblenz, d. 1960 Berlin) in the years between 1900 and 1906. Von Laue showed that usual light propagation in empty space is a reversible process and that the entropy of a beam indeed remains constant in this case. When light is diffracted, scattered or reflected diffusively, the effective temperature decreases and the entropy increases. The most interesting case is interference, where entropy usually increases, but Ref. 96

---

* In 2013, the German Physical Society published an official expert opinion stating that "electromagnetic waves do not need vacuum as carrier." The society also wants all physics teachers to tell this false statement to their pupils. Physicists all over the world are still laughing.

sometimes decreases.

* *

We saw that light has energy, linear momentum, angular momentum, entropy, temperature, pressure, chemical potential and, as we will see in the next volume, consists of quantons. It makes thus sense to state:

▷ Light is a substance.

Challenge 147 s Enjoy exploring this conclusion.

* *

The wave impedance of the vacuum of 376.7 Ω has practical consequences. If an electromagnetic wave impinges on a large, thin, resistive film along the normal direction, the numerical value of the film resistance determines what happens. If the film resistance is much *larger* than 376.7 Ω per square, the film is essentially transparent, and the wave will be *transmitted*. If the film resistance is much lower than 376.7 Ω per square, the film is essentially a short circuit for the wave, and the wave will be *reflected*. Finally, if the film resistance is *comparable* to 376.7 Ω per square, the film is impedance-matched and the wave will be *absorbed*.

* *

If the light emitted by the headlights of cars were polarized from the bottom left to the upper right (as seen from the car's driver) one could vastly improve the quality of driving at night: one could add a polarizer to the wind shield oriented in the same direction. As a result, a driver would see the reflection of his own light, but the light from cars coming towards him would be considerably dampened. Why is this not done in modern cars? Challenge 148 s

* *

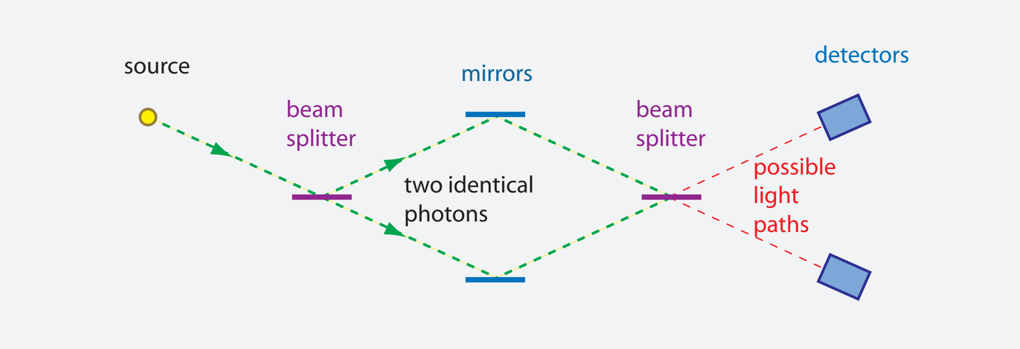
Could light have a tiny mass, and move with a speed just below the maximal speed possible in nature? Ref. 97 The question has been studied extensively. If light had mass, Maxwell's equations would have to be modified, the speed of light would depend on the frequency and on the source and detector speed, and longitudinal electromagnetic radiation would exist. Despite a promise for eternal fame, no such effect has been observed.

* *

A beam of light can be polarized. The direction of polarization can be changed by sending the light through materials that are birefringent, such as liquid crystals, calcite or stressed polymers. But polarization can also be changed with the help of mirrors. To achieve such a polarization change, the path of light has to be genuinely three-dimensional; the path must not lie in a plane.

To understand the rotation of polarization with mirrors, the best tool is the so-called *geometric phase*. The geometric phase is an angle that occurs in three-dimensional paths of any polarized wave. The geometric phase is a general phenomenon that appears both for light wave, for wave functions, and even for transverse mechanical oscillations. To visualize geometric phase, we look at the Figure 89.

The left image of Figure 89 can be seen as paper strip or a leather belt folded in space,



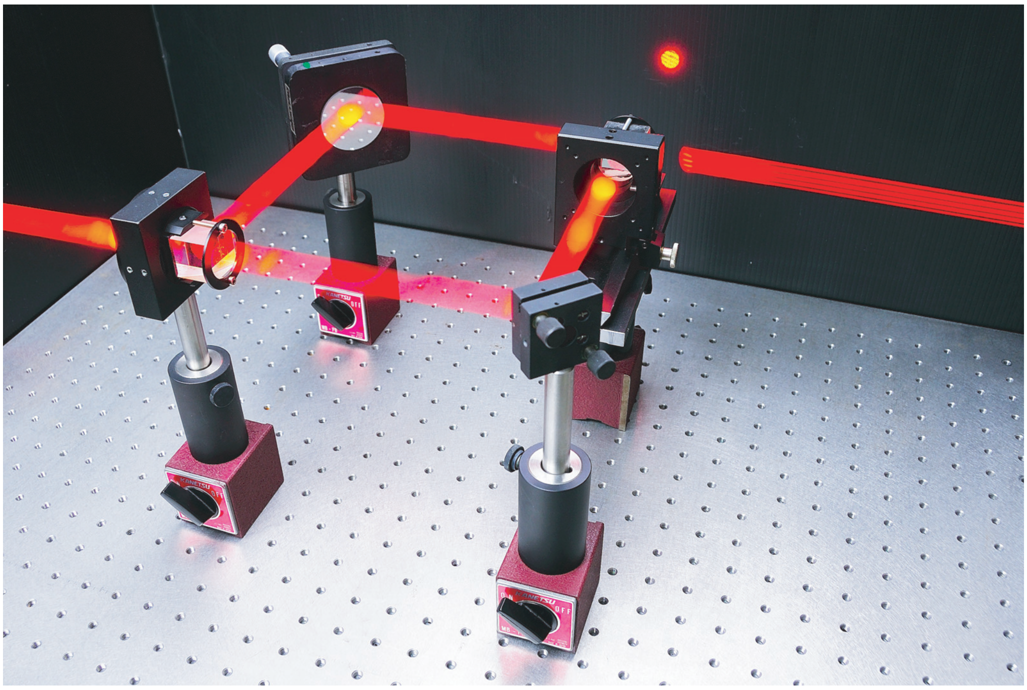


FIGURE 88 A conventional two-dimensional (Mach–Zehnder) interferometer, with sides of equal lengths, and its outputs A and B. Light exits in direction A, the direction of constructive interference (photo © Félix Dieu and Gaël Osowiecki).

with a bright and a dark coloured side. It is not a surprise that the orientation of the strip at the end differs from the start. Imagine to follow the strip with the palm of your hand flat on it, along its three-dimensional path. At the end of the path, your arm is twisted. This twist angle is the *geometric phase* induced by the path. Ref. 98

Instead of a hand following the paper strip, we now imagine that a polarized light beam follows the path defined by the centre of the strip. At the bends, mirrors change the motion of the light, but at each tiny advance, the polarization remains parallel to the polarization just before. One speaks of *parallel transport*. The result for light is the same as for the belt: At the end of the path, the polarization of the light beam has been rotated. In short, *parallel transport in three dimensions results in a geometric phase.* In particular, it is thus possible to rotate the polarization of a beam of light with the help of mirrors

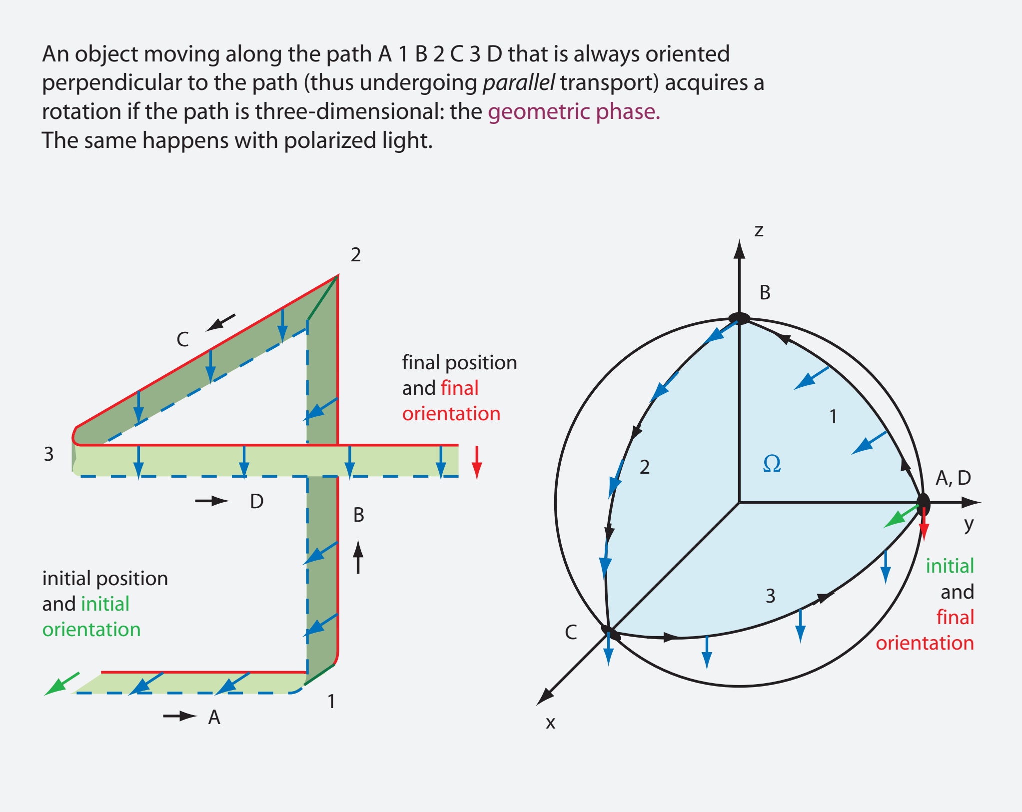


FIGURE 89 Left: a three-dimensional path traced by a pointed object that behaves like the polarization of light. The bends 1, 2 and 3 could be induced by mirrors. Right: the rotation angle of the polarization is given by the solid angle Ω, the geometric phase, enclosed by the path.

only.

Also transverse mechanical oscillations work in this way. When a Foucault pendulum oscillates, its path – a segment of a circle due to the rotation of the Earth – is three-dimensional. The direction of oscillation – akin to the polarization of the light or the orientation of the paper strip – changes along the path.

Vol. I, page 140

Since wave functions in quantum mechanics are also described by a transverse phase, they show similar effects when they follow three-dimensional paths. The Aharonov-Bohm effect is an example for a situation where a three-dimensional path leads to phase change.

Vol. IV, page 97

The other, right-hand drawing in Figure 89, illustrating the so-called *sphere of directions*, shows how to calculate the angle of rotation due to a specific path. *The geometric phase turns out to be the solid angle enclosed by the path.* In short, the geometric phase angle is given by the enclosed solid angle. With this result, the geometric phase has no mysteries any more. (For paths that are not closed on the sphere of directions, the calculation can still be carried out by suitably closing the path on the sphere.) A pretty case is the experiment in which polarized light is fed into a helically coiled optical fibre. In this case, the geometric phase is fixed by the length of the fibre and the pitch length of

Challenge 149 ny

the helix. Effects of the geometric phase have also been observed in molecules, in nuclei, neutron beams, in interferometers of all kind, in particle accelerators, in gyroscopes, in general relativity and in many other settings. Ref. 99

Historically, the geometric phase has been discovered independently by many people in different fields of physics. The researcher who understood its general importance in quantum physics was Michael Berry in 1983, but the phase was known in quantum physics, optics and mechanics long before, among others through the work in nuclear physics by Christopher Longuet-Higgins in the 1950s, through the work on light by the young genius Shivaramakrishnan Pancharatnam also in the 1950s, through the work on molecules by Alden Mead in the 1970s, and, of course, through the mentioned Foucault pendulum from 1851. Vol. I, page 244 But also the errors in the south-pointing carriage, which we mentioned before, are due to the geometric phase. Vol. I, page 206 Following Michael Berry, the phenomenon is now called the *geometric phase*. Older expressions, such as adiabatic phase, topological phase, quantal phase, Berry's phase and various other terms are not used any more.

After this excursion, here is a challenge of the real world. What is the smallest number of mirrors needed in a device to change the polarization of a light beam that exits the device in the same direction as it came in? Challenge 150 s

* *

In many optical systems – including laser systems and cameras – the polarization of light is controlled with the help of *waveplates*. They are made from a birefringent materials. A *half-wave* waveplate allows to rotate the polarization of a linearly polarized beam. If the waveplate is rotated by an angle α, the polarization of the beam is rotated by an angle 2α. A *quarter-wave* waveplate transforms linear polarization into circular polarization – and vice versa.

* *

Vikings had no magnetic compass and no clocks. Still, they were able to navigate precisely across the Atlantic Ocean over long times and distances. It seems that they used 'sunstones' as navigation devices, which most probably were birefringent crystals, such as calcite, cordierite or tourmaline. The Vikings probably had an orientable crystal mounted on their ship. With the crystal, a navigator could determine the position of the Sun and steer his ship accordingly. The exact method used is still matter of dispute; it might have been similar to the method used by bees or certain spiders, thus allowing to determine the position of the Sun also in cloudy weather or during twilight. Page 113 This allowed to navigate along constant latitude with sufficient precision, even for three weeks of travel. Ref. 100 The resulting uncertainties have been simulated numerically; but the method has yet to be tested on a real ship.

* *

An interferometer is a device that uses the interference of light to study the properties of a light beam. A common interferometer, the Mach–Zehnder interferometer, is shown in Figure 88. If all sides have equal length, light interferes constructively in the output direction A and destructively in the other output direction B. Thus light exits in direction A.

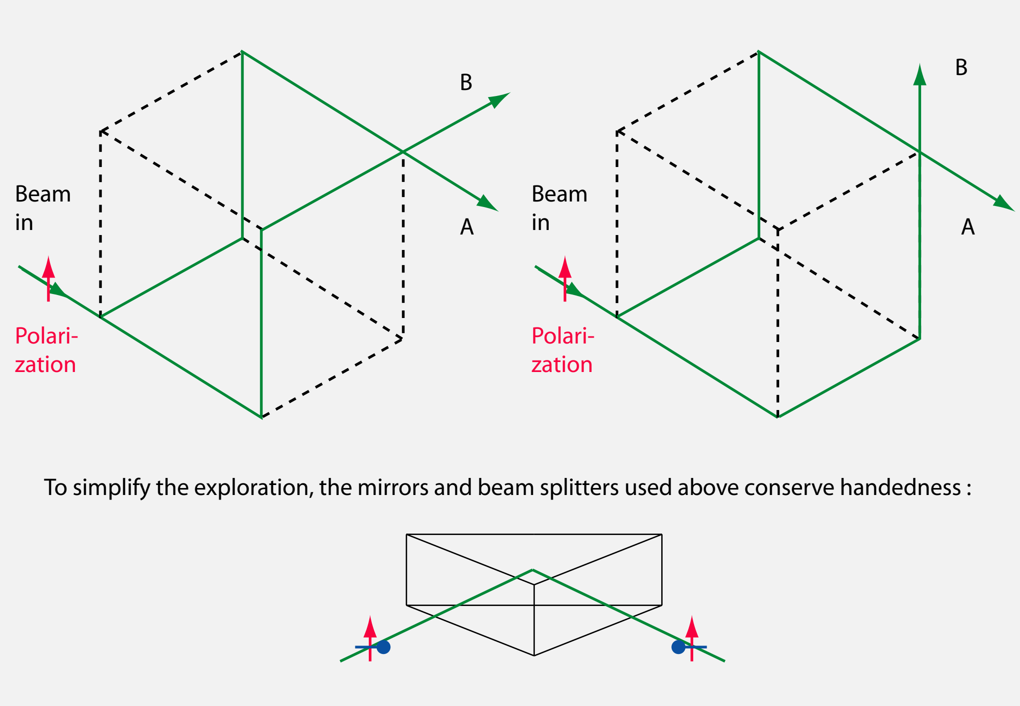


FIGURE 90 Two different three-dimensional interferometers, with all edges of equal lengths, the mirrors/beam splitters used, and their outputs A and B. Where does the light exit?

Ref. 101 Only in the 1990s people started asking what would happen in *three-dimensional* interferometers, such as the one shown in Figure 90. To clarify the situation, a few points are necessary. First, we need to specify the polarization of the light used, and recall that only light of the same polarization can interfere. Secondly, to simplify the discussion, we assume that the mirrors are of a special type (namely corner cubes based on total refraction) so that, in contrast to usual mirrors, they *conserve* polarization. Thirdly, we assume that all edges have equal length. Can you deduce which exits are bright in the two cases of Figure 90? Challenge 151 s

* *

It is possible to build a glass device that allows realizing the optical analog of the Stern-Gerlach experiment. The so-called *Fresnel triprism* separates a light beam into its left- and right-polarized components. To achieve this, three double refracting prisms of different handedness are glued together in a suitable geometric arrangement. Ref. 103

* *

In regions of destructive interference one finds so-called *phase singularities*. If the interfering light is white, such regions are not black but show, if the intensity is amplified, fascinating colour patterns. These colours, predicted in the 1970s, were found experimentally a few decades later. They follow an universal blue-orange pattern. Ref. 102

* *

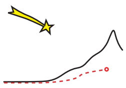
Maxwell's equations of the electromagnetic field are 150 years old. Is all about them known? Probably not. For example, only in the 1990s Antonio Rañada discovered that the equations have solutions with *knotted* field lines. The most spectacular solutions so far have been published by Arrayás and Trueba. More such surprising results are probably waiting to be found. Ref. 104

* *

Light can bleach hair. Many women turn their hair blond by using chemicals. Light can do this much better, if the wavelength is between 500 and 1100 nm, and if the pulse length is under 10 ps. Such short pulses, if powerful enough, destroy the melanin in the hair without destroying the keratin. In a not too distant future, we might see picosecond lasers at hair dressers.

SUMMARY ON LIGHT

Radio waves, infrared light, visible light, ultraviolet light, X-rays and gamma rays are electromagnetic waves. Their dispersion relation in vacuum is $\omega = ck$, where the phase velocity $c = 299\,792\,458$ m/s is a universal constant, an invariant. Electromagnetic waves carry energy, linear momentum and angular momentum. In vacuum, the phase velocity is also the group and the signal velocity. In addition, the speed of electromagnetic waves c is the (local) limit energy speed in nature: Electromagnetic waves in vacuum move faster than any material object.



CHAPTER 4

IMAGES AND THE EYE – OPTICS

Optics is the field that explores the production of images. In particular, ptics is the study and use of light *generation*, of light *transport*, and f light and image *detection*. With this definition of optics, we note directly that classical electrodynamics can describe only the transport of light. The generation and the detection of light are always quantum effects. Every lamp is a device based on quantum physics. Every detector of light, including the eye, is based on quantum physics. Therefore, in this chapter we mainly explore the motion of light and the way it forms images, and give only a short introduction into light sources and the eye. Light generation will be explored in more detail in the volumes on quantum physics.

Ref. 105

WAYS TO ACQUIRE IMAGES

Acquiring images is an important part of modern society. The quality of images depends on the smart use of optics, electronics, computers and materials science. Despite the long history of optics, there are still new results in the field. Images, i.e., two or three-dimensional reproductions of a physical situation, can be taken by at least six groups of techniques:

- *Photography* uses a light source, lenses and film – or another large area detector inside a camera. Photography can be used in reflection, in transmission, with phase-dependence, with various illuminations, and with light sources and detectors for various wavelengths.
- *Optical microscopy* uses a light source, magnifying lens systems and film (or another large area detector). If the illumination is through the sample, in transmission, one speaks of *bright-field microscopy*. (Variations use coloured or polarizing filters.) If the illumination is from the side, one speaks of *oblique microscopy*. If the illumination is confined to an outer ring of light, one speaks of *dark-field microscopy*. An even more elaborate illumination system, using plane waves, allows *phase-contrast microscopy*. (It was invented by Frits Zernike in the 1930s and earned him the Nobel Prize in Physics in 1953.) If one splits a polarized illumination beam into two components that pass the sample at close (but not identical) locations, and then recombines them afterwards, one speaks of *differential interference contrast microscopy*. If a sample is treated with a fluorescent dye, the illuminating light is filtered out, and only the fluorescence is observed, one speaks of *fluorescence microscopy*. The image quality of expensive microscopes can be further improved with the help of a computer, using deconvolution techniques.

Ref. 106

FIGURE 91 An X-ray photographic image of a ten-year old boy with polydactyly (© Drgnu23).

- *Telescopy* is used most of all in geodesy and astronomy. Since over a hundred years, telescopes are so powerful that, at large magnification, stars can be observed during the day. We will explore telescopes below. The most advanced astronomical telescopes can compensate star images for the effects of the turbulence of the atmosphere; they can also take images at various wavelengths, ranging from radio frequencies, infrared, visible, ultraviolet to X-rays. Simple telescopes are lens-based; high-performance telescopes are usually mirror-based. Telescopes also exists for non-visible wavelengths. Infrared telescopes can be ground-based, balloon-based, aeroplane-based or satellite-based. UV and X-ray telescopes have to be operated outside the atmosphere, to avoid absorption by air, for example on rockets, satellites or high-altitude balloons. They are all mirror based.

Page 163

- *Scanning techniques* acquire images point by point through the motion of the detector, the light source or both. There are numerous scanning microscopy techniques: *confocal laser scanning microscopy*, the fibre-based *near-field scanning optical microscopy*, and combinations of them with fluorescence techniques or various deconvolution techniques. Many of these scanning microscopy techniques allow resolutions much lower than the wavelength of light, a feat that is impossible with conventional microscopic techniques. Scanning techniques are also used in special fields of photography.
- *Tomography*, usually performed in transmission, uses a source and a detector that are rotated together around an object. This technique, effectively a specialized scanning technique, allows imaging cross sections of physical bodies. For example, light tomography is a promising technique, without any health risk, for breast cancer detection.
- *Holography* uses lasers and large area detectors and allows taking three-dimensional images of objects. Such images seem to float in space. Holography can be used in reflection or in transmission.

Each image acquisition method can be used with radio waves, infrared light, visible light, ultraviolet light, X-rays or with gamma rays. In fact, these techniques can even be used with electron beams; one then speaks of electron optics. In all imaging methods, the race is twofold: progress aims for images with the highest resolution possible and for

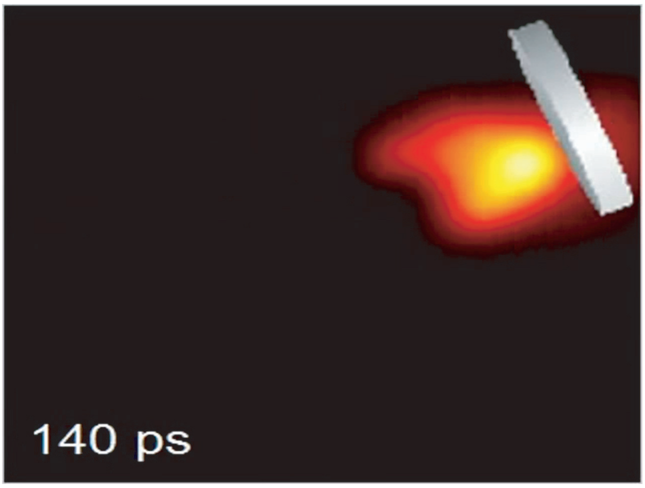


FIGURE 92 A film taken with a special ultrafast camera showing a short light pulse that bounces off a mirror (QuickTime film © Wang Lihong and Washington University at St. Louis).

images with the shortest shutter times possible. The shorter the shutter time, the more informative is the resulting film. An impressive example is the film of a moving light pulse shown in Figure 92. We start our overview of imaging techniques with the most important tool: light sources.

LIGHT SOURCES

Without radiation sources, there would be no images. All imaging techniques need sources of radiation. In the domain of visible light optics, the most common light sources of visible and infrared light are *hot* objects, such as candles, the Sun or flashlamps. Physically speaking, these light sources are approximations of black bodies. Let us see why they are used. *Cold* light sources, such as light emitting semiconductor diodes, fireflies or lasers, are explored later on.

WHY CAN WE SEE EACH OTHER? BLACK BODIES AND THE TEMPERATURE OF LIGHT

Physicists have a strange use of the term 'black'. A body that glows perfectly is called a *black body*. In this domain, 'perfect' means that the surface of the body has *no* effect on its colour.

▷ A ***black body*** is a body that absorbs all radiation impinging on it.

In other words, a black body is a body without reflection or transmission of radiation. Black bodies are an idealization; above all, they are only black at low temperature. With increasing temperature, black bodies glow or shine in black, brown, red, orange, yellow, white or light blue.

The essence of black bodies is that the colour they have, i.e., the light they radiate,

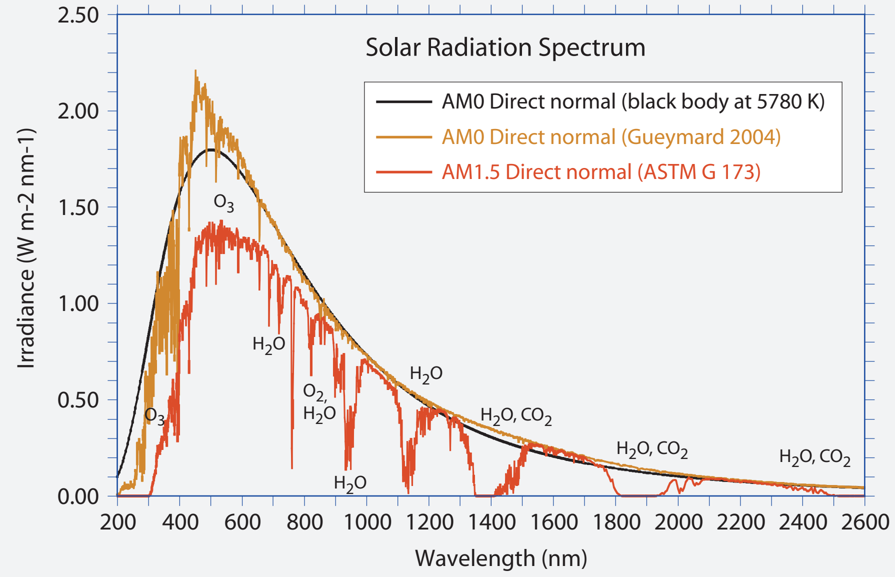


FIGURE 93 A black body spectrum at 5780 K, the solar spectrum *above* the atmosphere in direction of the Sun, with 1350 W/m^2, and the spectrum with 1.5 air masses, or atmospheric thicknesses, in between, with 844 W/m^2. The latter roughly describes the spectrum of a typical sunny day at sea level. The gases responsible for the absorption bands are also shown (© Chris Gueymard).

is independent of the surface. Black bodies are thus *ideal* in this sense. Real bodies, which do show surface effects, can be classified by their *emissivity*. The emissivity gives the degree to which a body approaches a black body. Mirrors have emissivities of around 0.02, whereas black soot can have values as high as 0.95. Practically all bodies at everyday temperature are not black bodies: their colour is not determined by emission, but mostly by the absorption and reflection of light at their surface.

Black bodies, as the section on quantum theory will show, have *smooth* light emission spectra. An example for a spectrum of a black body, and for a spectrum of a real body – in this case, the Sun – is shown in Figure 93.

Black bodies are also used to define the colour *white*. What we commonly call *pure white* is the colour emitted by the Sun. The sun is not a good black body, as Figure 93 shows (its effective temperature is 5780 K). Because of these problems, pure white is now defined as the colour of a black body of 6500 K, e.g. by the Commission Internationale
Ref. 107 d'Eclairage. As mentioned, hotter black bodies are bluish, colder ones are yellow, orange,
Vol. I, page 261 red, brown or black. The stars in the sky are classified in this way.

Black bodies are thus bodies that glow perfectly. Most real bodies are only rough approximations of black bodies, even at temperatures at which they shine yellow light. For example, the tungsten in incandescent light bulbs, at around 2000 K, has an emissivity of around 0.4 for most wavelengths, so that its spectrum is a corresponding fraction of that of black body. (However, the glass of the light bulb then absorbs much of the ultraviolet and infrared components, so that the final spectrum is not at all that of a black body.)

Black body radiation has two important properties: first, the emitted light power increases with the fourth power of the temperature. With this relation alone you can check the temperature of the Sun, mentioned above, simply by comparing the size of the Sun with the width of your thumb when your arm is stretched out in front of you. Are you Challenge 152 d able to do this? (Hint: use the excellent approximation that the Earth's average temper- Ref. 108 ature of about 14.0°C is due to the Sun's irradiation.)

The precise expression for the energy density u per frequency ν emitted a temperature T can be deduced from the radiation 'law' for black bodies discovered by Max Planck[*]

$$u(\nu, T) = \frac{8\pi h}{c^3} \frac{\nu^3}{\mathrm{e}^{h\nu/kT} - 1} . \qquad (76)$$

He made this important discovery, which we will discuss in more detail in the quantum part of our mountain ascent, simply by comparing this curve with experiment. The new constant h is called the *quantum of action* or *Planck's constant* and turns out to have the Vol. IV, page 17 value $6.6 \cdot 10^{-34}$ Js, and is central to all quantum theory, as we will find out. The other constant Planck introduced, the Boltzmann constant k, appears as a prefactor of temperature all over thermodynamics, as it acts as a conversion unit from temperature to energy.

Challenge 153 e The radiation 'law' gives for the total emitted energy density the expression

$$u(T) = T^4 \, \frac{8\pi^5 k^4}{15 c^3 h^3} . \qquad (77)$$

Below, we will deduce from it the expression for the intensity I of thermal radiation. That Challenge 154 ny Page 239 expression, equation (85), is deduced using $I = uc/4$. (Why?)

The second property of black body radiation is the value of the peak wavelength, i.e., the wavelength emitted with the highest intensity. This wavelength determines the colour Challenge 155 ny of a black body; it is deduced from equation (76) to be

$$\lambda_{\text{max}} = \frac{1}{T} \frac{hc}{4.956\,k} = \frac{2.90\,\text{mm K}}{T} \quad \text{but} \quad \hbar\nu_{\text{max}} = T \cdot 2.82\,k/h = T \cdot 5.9 \cdot 10^{10}\,\text{Hz/K} . \qquad (78)$$

Either of these expressions is called *Wien's colour displacement rule* after its discoverer.[**] The colour change with temperature is used in optical thermometers; this is also the way

---

[*] Max Planck (b. 1858 Kile, d. 1947 Göttingen), professor of physics in Berlin, was a central figure in thermodynamics. He discovered and named the *Boltzmann constant* k and the *quantum of action* h, often called Planck's constant. His introduction of the quantum hypothesis gave birth to quantum theory. He also made the works of Einstein known in the physical community, and later organized a job for him in Berlin. He received the Nobel Prize in Physics in 1918. He was an important figure in the German scientific establishment; he also was one of the very few who had the courage to tell Adolf Hitler *face to face* that it was a Ref. 109 bad idea to fire Jewish professors. (He got an outburst of anger as answer.) Famously modest, with many tragedies in his personal life, he was esteemed by everybody who knew him.

[**] Wilhelm Wien (b. 1864 Gaffken, d. 1928 Munich) received the Nobel Prize in Physics in 1911 for the discovery of this relation. The value of the constant appearing in Wien's rule can be uniquely calculated from equation (76), but cannot be expressed as a formula. Indeed, Wien's constant contains the solution of the equation $x = 5(1 - \mathrm{e}^{-x})$.

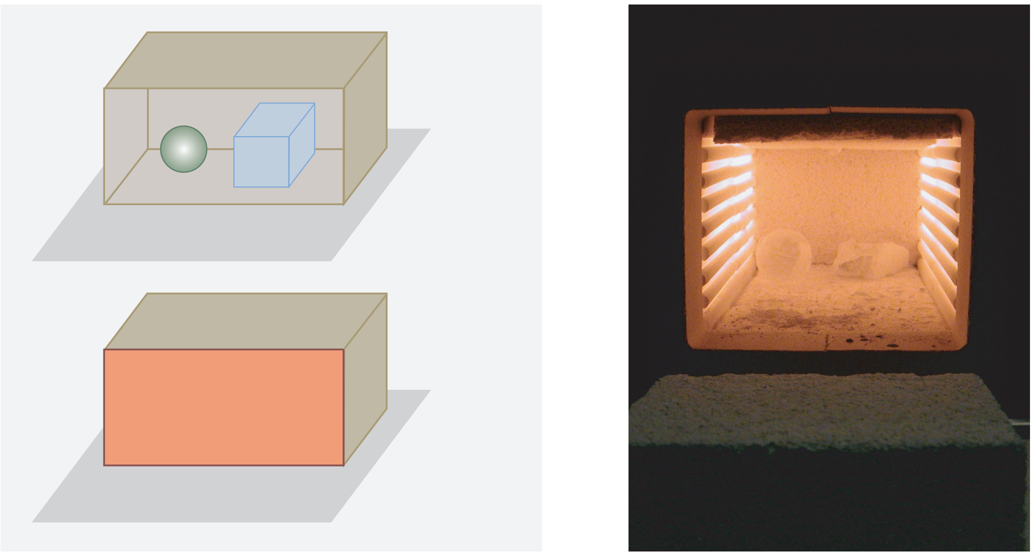


FIGURE 94 Bodies inside an oven at room temperature differ in colour, in contrast to bodies at high temperature (photo © Wolfgang Rueckner).

the temperatures of stars are measured. For 37°C, human body temperature, it gives a peak wavelength of 9.3 μm or 115 THz, which is therefore the colour of the bulk of the radiation emitted by every human being. (The peak wavelength does not correspond to the peak frequency. Why?) On the other hand, following the telecommunication laws of many countries, any radiation emitter needs a licence to operate; it follows that strictly in Germany only dead people are legal, and only if their bodies are at absolute zero temperature.

Challenge 156 s

We saw that a black body – or a star – can be blue, white, yellow, orange, red or brown. A black body is never green. Can you explain why?

Challenge 157 e

Above, we predicted that any material made of charges emits radiation. Are you able to find a simple argument showing whether heat radiation is or is not this classically predicted radiation?

Challenge 158 ny

But let us come back to the question in the section title. The existence of thermal radiation implies that any hot body will cool, even if it is left in the most insulating medium there is, namely in vacuum. More precisely, if the vacuum is surrounded by a wall, the temperature of a body in the vacuum will gradually approach that of the wall.

Interestingly, when the temperature of the wall and of the body inside have become the same, something strange happens. The effect is difficult to check at home, but impressive photographs exist in the literature.

Ref. 110

One arrangement in which walls and the objects inside them are at the same temperature is an *oven*. It turns out that it is *impossible* to see objects in an oven using the light coming from thermal radiation. For example, if an oven and all its contents are red hot, taking a picture of the inside of the oven (without a flash!) does not reveal anything; no contrast nor brightness changes exist that allow one to distinguish the objects from the walls or their surroundings. Can you explain the finding?

Challenge 159 s

In short, we are able to see each other only because the light sources we use are at

FIGURE 95 The last mirror of the solar furnace at Odeillo, in the French Pyrenees (© Gerhard Weinrebe).

a *different* temperature from us. We can see each other only because we do *not* live in thermal equilibrium with our environment.

Limits to the concentration of light

Light sources should be as bright as possible. Are there any limits? Interestingly, for black body radiation there is an important and instructive limitation.

If we build a large lens or a large curved mirror, we can collect the light of the Sun and focus it on a tiny spot. Everybody has used a converging lens as a child to burn black spots on newspapers – or ants – in this way. In Odeillo, in Spain, wealthier researchers have built a curved mirror as large as a house, in order to study solar energy use and material behaviour at high temperature. Essentially, the mirror provides a cheap way to fire an oven in its focus. (And 'focus' is the Latin word for 'hearth'.)

Kids find out quite rapidly that large lenses or mirrors allow them to burn things or paper more easily than small ones. The Odeillo site shown in Figure 95 is the record holder in the quest for the larges possible collection area. Interstingly, building a larger mirror does not make much sense. Whatever its size may be, the temperature in such a set-up is *limited*: Ref. 111

▷ The effective temperature of the light in a focus *cannot* exceed the temperature of the original light source.

In all practical situations, the temperature of the light source is much higher than in the focus. The surface temperature of the Sun is about 5780 K; the highest temperature reached so far in Odeillo is about 4000 K. Are you able to show that this limitation is equivalent to the second principle of thermodynamics, as Hemholtz, Clausius and Airy showed? Challenge 160 s

In short, nature provides a *limit* to the concentration of light energy. More precisely, we can say: thermodynamics limits what can be achieved through heating with thermal light sources.

The thermodynamic limit on heating with light does not prevent people to use light concentration to gather solar energy. Experimental power plants such as the one shown in Figure 96 are one promising way to supply energy to households when fossil fuel prices rise too much.

FIGURE 96 The solar power plant at Sanlucar la Mayor, near Seville, in Spain (© Wikimedia).

As we just saw, a beam of thermal light has entropy. In contrast, a laser beam only has a tiny entropy. We can also ascribe a temperature value to either beam: the temperature of a thermal beam is the temperature of the light source; the temperature of a laser beam is a 'negative' number. This makes some sense intuitively, because a laser beam is able to cool gases; more precisely, a laser beam is a non-equilibrium situation, and temperature is not defined for such cases.

In several countries, taxpayer's money is wasted in so-called *inertial confinement fusion* centres. In those centres, several powerful lasers are focused on a small sphere of material, typically, 1 mm in size; a target temperature of around 3 MK (or, equivalently, 300 eV) has been achieved. Why is this possible? Challenge 161 s

MEASURING LIGHT INTENSITY

Light sources differ in brightness. Measuring what we call 'dark' and 'bright' is somewhat involved, because light can be diffuse or directed. To achieve proper measurements, the *SI*, the international system of units, defines a specific base unit, the candela: Page 352

- 'The *candela* is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency $540 \cdot 10^{12}$ hertz and has a radiant intensity in that direction of (1/683) watt per steradian.'

The candela is thus a unit for light power per (solid) angle, usually called *luminous intens-*

TABLE 16 Some measured illuminance values.

Observation	Illuminance
Brightness of the human body	1 plx
Faint star	0.1 nlx
Sirius	10 μlx
phot (old illuminance unit)	10 μlx
Jupiter	20 μlx
Dark, moonless night	1 mlx
Full moon	0.01 to 0.24 lx
Street at night, low traffic, poor lighting	0.1 to 3 lx
Street at night, high traffic	10 to 30 lx
For reading	50 to 100 lx
Cinema screen	100 lx
Workplace	0.2 to 5 klx
Cloudy day	1 klx
Brightest lamps, used for surgery	120 klx
Sunny day	120 klx
Film in cinema projector	5 Mlx
Painful to the eye	100 Mlx

ity, except that it is corrected for the eye's sensitivity: the candela measures only *visible* power per angle. The definition of the candela simply says that 683 cd = 683 lm/sr corresponds to 1 W/sr. For example, a glow worm produces 0.01 cd, a candle indeed produces around 1 cd, a car light around 100 cd, and a lighthouse around 2 Mcd. Another way to look at the candela is the following: watching a source with 1 cd from a distance of 1 m is a just bit brighter than the full moon.

Total light power, irrespective of its direction, is measured in lumen. Therefore, 683 lm = 683 cd sr corresponds to 1 W. In other words, both the lumen and the watt measure power, or energy flux, but the lumen measures only the *visible* part of the power or energy flux. This difference is expressed by adding 'luminous' or 'radiant': thus, the lumen measures *luminous* flux, whereas the Watt measures *radiant* flux.

The factor 683 appearing in the definitions is historical. An ordinary candle emits a luminous intensity of about a candela. To put this into perspective: at night, a candle can be seen up to a distance of 10 or 20 kilometres. A 100 W incandescent light bulb produces 1700 lm, and the brightest commercial light emitting diodes about 20 lm, though laboratory devices exceed 1000 lm. Cinema projectors produce around 2 Mlm, and the brightest flashes, like lightning, 100 Mlm.

Challenge 162 e

The *irradiance* of sunlight is about 1300 W/m^2 on a sunny day; on the other hand, the *illuminance* is only 120 klm/m^2 = 120 klx or 170 W/m^2. A cloud-covered summer day or a clear winter day produces about 10 klx. These numbers show that most of the energy from the Sun that reaches the Earth is outside the visible spectrum.

Illuminance is essentially what we call 'brightness' in everyday life. On a glacier, near the sea shore, on the top of a mountain, or in particular weather condition the brightness can reach 150 klx. Museums are often kept dark because water-based paintings are degraded by light above 100 lx, and oil paintings by light above 200 lx. The eyes lose their ability to distinguish colours somewhere between 0.1 lx and 0.01 lx; the eye stops to work below 1 nlx. Technical devices to produce images in the dark, such as night goggles, start to work at 1 μlx. By the way, the human body itself *shines* with about 1 plx, a value too small to be detected with the eye, but easily measured with specialized apparatus. The origin of this emission is still a topic of research. Ref. 112

The highest achieved light intensities, produced with high-power lasers, are in excess of 10^{18} W/m^2, more than 15 orders of magnitude higher than the intensity of sunlight. (How much is that in lux?) Such intensities are produced by tight focusing of pulsed laser beams. The electric field in such light pulses is of the same order as the field inside atoms; such a laser beam therefore ionizes all matter it encounters, including the air. Challenge 163 e Ref. 113

The *luminous density* is a quantity often used by light technicians. Its unit is 1 cd/m^2, unofficially called 1 Nit and abbreviated 1 nt. Human eyes see using *rods* only from 0.1 μcd/m^2 to 1 mcd/m^2; they see with *cones* only above 5 cd/m^2. Eyes see best between 100 and 50 000 cd/m^2, and they get completely overloaded above 10 Mcd/m^2: a total range of 15 orders of magnitude. Very few technical detectors achieve this range.

Other light and radiation sources

Apart from black bodies, many other types of light sources exist. Cold sources of light range from glowing fish to high-power lasers. They range in size from an atom to a building, in cost from a fraction of an Euro to hundreds of millions of Euros, and in lifetime from a fraction of a second to hundreds of years.

Lasers are important light sources for industry, medicine and research. Lasers can emit visible, infrared and ultraviolet light, continuously or as light pulses, with various powers, polarizations and beam shapes; they are explored later on in our adventure. In the domain of imaging, lasers are used in many microscopy techniques, in scanning imaging systems, in tomography and in holography. Vol. V, page 108

Sources of radio waves are common in everyday life: mobile phones, radio transmitters, tv transmitters and walkie-talkies are all sources of radio waves. They are used for imaging in magnetic resonance imaging, which allows to image the interior of the human body, and in astronomy: Since many stars are radio emitters, one can image the sky at radio wavelengths. Nowadays, radio astronomy is an important part of modern astronomy and has led to many discoveries. Radio astronomy has also been an important tool for the precision testing and confirmation of general relativity. Vol. V, page 162 Vol. II, page 211

On the other end of the electromagnetic spectrum, light sources that emit X-rays and gamma rays are also common. They are routinely used in medicine and materials science, also for various imaging techniques.

All sources of electromagnetic radiation are potentially dangerous to humans, so that special care has to be taken when using them. This has also led to various unfortunate developments.

FIGURE 97 A modern picosecond pulse laser and an industrial X-ray source, both about 700 mm in size (© Time-Bandwidth, SPECS).

RADIATION AS WEAPON

High-intensity electromagnetic radiation is dangerous. In many countries, more money is available to study assault weapons than to increase the education and wealth of their citizen. Several types of assault weapons using electromagnetic radiation are being researched. Two are particularly advanced.

The first weapon using electromagnetic radiation is a truck with a movable parabolic antenna on its roof, about 1 m in size, that emits a high power – a few kW – microwave beam at 95 GHz. The beam, like all microwave beams, is invisible; depending on power and beam shape, it is painful or lethal, up to a distance of 100 m and more. This terrible device, officially called *active denial system*, with which the operator can make many victims even by mistake, was ready in 2006. Some extreme politicians want to give it to the police. (Who expects that a parabolic antenna is dangerous?) Efforts to ban it across the world are slowly gathering momentum.

The second weapon under development is the so-called *pulsed impulse kill laser*. The idea is to take a laser that emits radiation that is not absorbed by air, steam or similar obstacles. An example is a pulsed deuterium fluoride laser that emits at 3.5 μm. This laser burns every material it hits; in addition, the evaporation of the plasma produced by the burn produces a strong hit, so that people hit by such a laser are hurt and hit at

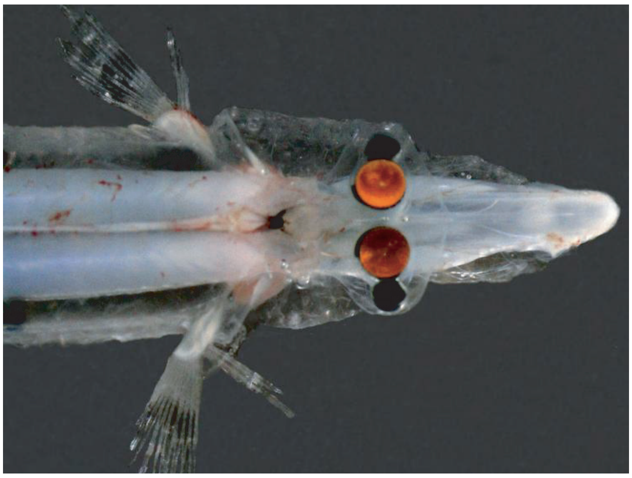


FIGURE 98 The spookfish *Dolichopteryx longipes* has orange mirrors that help him make sharp images also from the dim light coming upwards from bioluminescent lifeforms below it (© Tamara Frank).

the same time. Fortunately, it is still difficult to make such a device rugged enough for practical mobile use. Nevertheless, experts expect battle lasers, mounted on trucks, to appear soon – after a number of Potemkin's versions.

In short, it is probable that radiation weapons will appear in the coming years. What the men working on such developments tell their children when they come home in the evening is not clear, though.

IMAGES – TRANSPORTING LIGHT

Every image is formed by transporting light in a useful manner along known paths. The simplest possible path is the straight line.

Making images with mirrors

Since light moves in a straight line, a flat mirror produces an image of the same size than the original. Curved mirrors can be used to enlarge, reduce and distort images. For example, expensive bed room mirrors are often slightly curved, in order to make people appear thinner.

Most human-made mirrors are made of metal, usually evaporated onto a glass substrate; in contrast, living systems cannot produce pure metals. On the other hand, in living systems, mirrors abound: they are found as the *tapetum* in the eyes, on fish scales, on bugs, etc. How does nature produce mirrors, despite lacking the ability to use pure metals? It turns out that sandwiches of different thin transparent materials – one of which is typically crystalline guanine – can produce mirrors that are almost as good as metal mirrors. Such mirrors are based on interference effects and are called *dielectric mirrors*. Dielectric mirrors are also used to make laser mirrors.

Image-forming mirrors are used in large telescopes, in systems for X-rays, and in medical devices used by physicians. Interestingly, also some living beings use mirrors for imaging. The most famous example is the spookfish shown in Figure 98. It is able to

FIGURE 99 A Wolter-type grazing incidence collector for 13.5 nm radiation built with the help of concentric mirrors (© Media Lario Technologies).

look up and down at the same time, and does so using mirrors attached to his eyes.

Challenge 164 s By the way, why are mirrors frequently used in telescopes, but not in microscopes?

In illumination systems, mirrors are used for the shaping of light beams in cars, in pocket lamps and in LED lamps. It might be that some deep water creatures use mirrors for similar uses – but no example is known to the author.

The most involved mirror systems to date are used in the extreme ultraviolet mask lithography systems that will be used in the future production of integrated circuits. These systems use a wavelength of 13.5 nm, at which lenses are not available. Collimating an expanding beam thus requires many concentric mirrors, as shown in Figure 99. These optical systems are the very best that modern technology can provide; for example, the mirrors have a surface roughness below 0.4 nm. Similar optical mirror systems are also used in X-ray satellite telescopes.

Does light always travel in a straight line? – Refraction

Usually light moves in straight lines. A laser in a misty night shows this most clearly, as illustrated in Figure 100. But any laser pointer in the mist is equally fascinating. Indeed, we use light to *define* 'straightness', as we explained in the exploration of relativity. Page 15 However, there are a number of situations in which light does not travel in a straight line, Ref. 114 and every expert on motion should know them.

In diluted sugar syrup, light beams curve, as shown in Figure 101. The reason is that in such an experiment, the sugar concentration changes with depth. Are you able to explain Challenge 165 s the syrup effect?

More detailed observation show that a light beam is bent at every material change it encounters on its path. This effect, called *refraction*, is quite common. Refraction changes the appearance of the shape of our feet when we are in the bath tub; refraction also makes aquaria seem less deep than they actually are and produces effects such as those shown in Vol. I, page 263 Figure 102 and Figure 103. Refraction is a consequence of the change of the phase velocity

FIGURE 100 Light usually travels in a straight line. In the figure, a sodium frequency laser beam is used as laser guide star to provide a signal for adaptive optics in large telescopes. The laser illuminates a layer of sodium found in the atmosphere at around 90 km of altitude, thus providing an artificial star. The artificial star is used to improve the image quality of the telescope through adaptive optics. In the photograph, the images of the real stars are blurred because of the long exposure time of 3 min (photo by Paul Hirst).

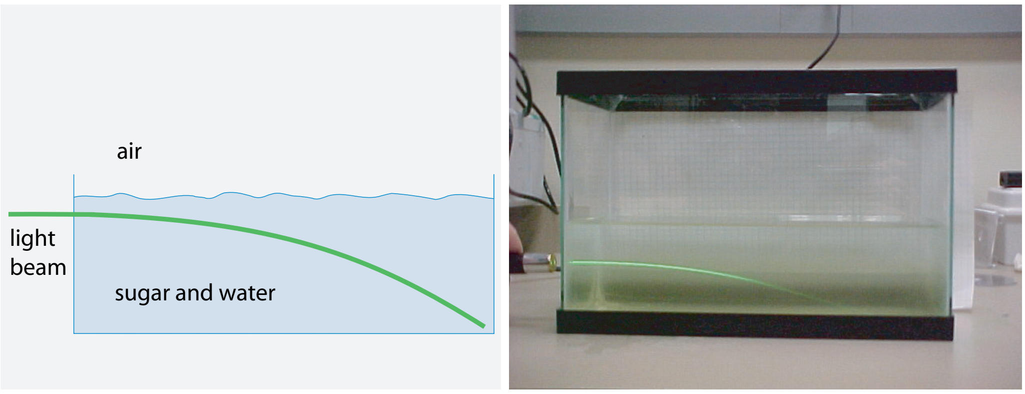


FIGURE 101 Diluted sugar syrup bends light (© Jennifer Nierer).

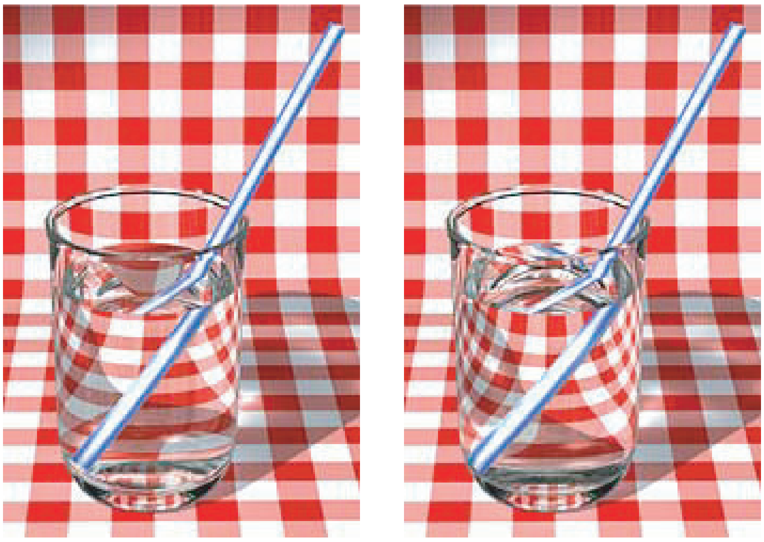
of light from material to material; all refraction effects are thus explained by Figure 104.

Refraction can also be seen to follow from the minimization principle for the motion of light:

▷ Light always takes the path that requires the *shortest* travel time.

For example, light moves more slowly in water than in air; that is the reason for the bend illustrated in Figure 105.

The speed ratio between air and water is called the *refractive index* of water. The refractive index, usually abbreviated n, is material-dependent. The value for water is about 1.3. This speed ratio, together with the minimum-time principle, leads to the 'law' of refraction, a simple relation between the sines of the two angles shown in Figure 105.Snell's
Challenge 167 s 'law' Can you deduce the relation? In fact, the exact definition of the refractive index of a material is with respect to vacuum, not to air. But the difference is negligible, because



Challenge 166 e

FIGURE 102 Realistic computer graphics showing the refraction in water and in diluted sugar syrup (graphics © Robin Wood). Can you tell which one is which?

FIGURE 103 A pretty effect of refraction at the water–air interface that you can repeat at home (© Maric Vladimir).

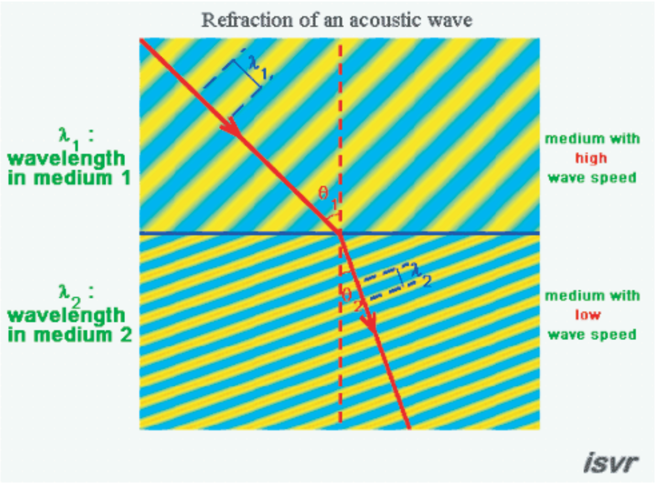


FIGURE 104 A visualisation of refraction (QuickTime film © ISVR, University of Southampton).

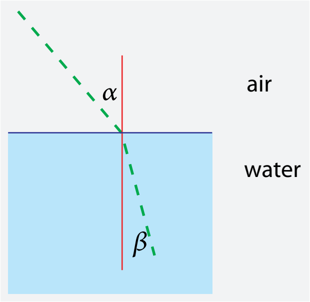


FIGURE 105 Refraction of light is due to travel-time optimization.

gases are mainly made of vacuum and their index of refraction is close to one.

In many fluids and solids, light signals move more slowly than in vacuum; also the (different) phase and group velocities of light inside materials are regularly *lower* than c, the light speed in vacuum. We discussed the difference between these speeds above. Page 133 For such 'normal' materials, the refractive index n, the ratio of c to the phase velocity inside the material, is larger than 1. The refractive index is an important material property for the description of optical effects. For example, the value for visible light in water is about 1.3, for glasses it is around 1.5, and for diamond 2.4. The high value is one reason for the sparkle of diamonds cut with the 57-face *brilliant* cut.

The refractive index also depends on wavelength; this effect, called *dispersion*, appears in most materials. Prisms make use of dispersion in glass to split white or other light into its constituent colours. Also diamond, and in particular the brilliant cut, works as a prism, and this is the second reason for their sparkle.

In contrast to 'normal' materials, various materials have refractive indices that are lower than 1, and thus phase velocities larger than c. For example, gold has a refractive index of around 0.2 for visible light, and thus a phase velocity of around $5c$ for such waves. In fact, almost all materials have refractive indices below 1 for *some* wave frequencies, including table salt. Ref. 115

In short, refraction of light, the change of the direction of light motion, is due to different phase velocities of light in different materials. *Material changes bend light paths.* Refraction is so common because it is extremely rare to have different adjacent materials with the same refractive index.

Gases have refractive indices close to the vacuum value 1. Nevertheless, also gases lead to refraction – including the air around us.

FROM ATMOSPHERIC REFRACTION TO MIRAGES

If light travels a long distance through air, the refraction can be considerable. For example, one we look at distant mountains, light does not follow a straight line; there is a deviation of several minutes of arc. This *terrestrial refraction* is a big problem for geodesy.

Light coming from the stars also gets refracted when it enters the terrestrial atmosphere. This *astronomic refraction* is about one minute of arc at an elevation of 45 degrees and usually 30 minutes of arc at the horizon. Therefore we can say that when we see the Sun touching the horizon, in reality it has already set! The exact value of the bending depends on the temperature gradients; they are often particularly strong at high latitudes.

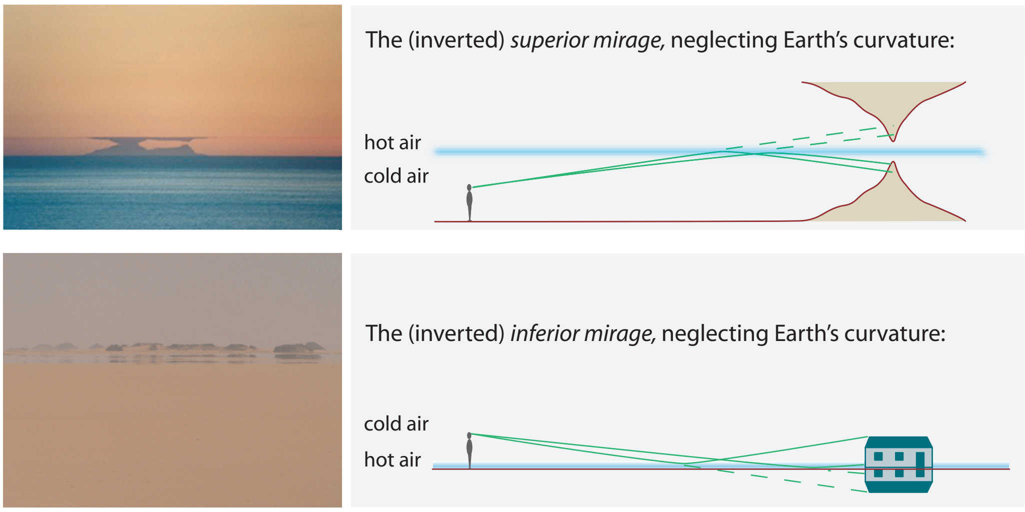


FIGURE 106 The basis of mirages is an effective reflection due to refraction in a hot air layer; it can lead to spectacular effects, such as the inverted superior mirage (top left and right) and the inferior image (bottom left and right) (photographs © Thomas Hogan and Andy Barson).

Sometimes the bending can be as high as 2 degrees; in these exceptional cases, the Sun can be visible when it should not be; this is now called the *Novaya Zemlya effect*.

Page 205

The refractive index of all gases depends on temperature; the temperature gradient is usually proportional to the density gradient. In air of varying temperature, terrestrial refraction leads to curved light paths and produces many effects.

The twinkling of the stars is due to the varying refraction induced by air turbulence. It was presented in the first volume.

Vol. I, page 88

Refraction by the atmosphere can make objects at the horizon float in the air – an effect called *looming* – or disappear below the horizon – an effect called *sinking*. If the images are stretched or compressed instead, the effects are called *towering* and *stooping*.

By far the most well-known effect due to refraction is the *mirage*. Mirages are – despite their name – due to the refraction of light rays in a horizontal layer of air that is warmer than the adjacent layers, as shown in Figure 106. Mirages always appear near the horizon, in a stripe narrower than the width of a finger at an arms's length.

Ref. 116

If the layer is below the observer, for example on the ground, an *inferior mirage* can appear, in which an additional inverted image appears *below* the direct image. Inferior mirages are regularly seen on hot highways. But they also appear in deserts, as shown in Figure 107, over snow and ice.

If the hotter layer is up in the air, one speaks of an *inversion layer.* If the observer is below the inversion layer, many kinds of mirages can appear: the superior mirage, in which a inverted mirror image is added *above* the direct image, or more complex mirages, in which several additional images appear. This latter mirage, but sometimes also any kind of mirage, is called *fata morgana*. All mirage types are due to refraction; their detailed appearance depends on the given temperature profile in the air, and the relative heights of the observer, the inversion layer and the observed object. Often, the curvature of the Earth also plays a role.

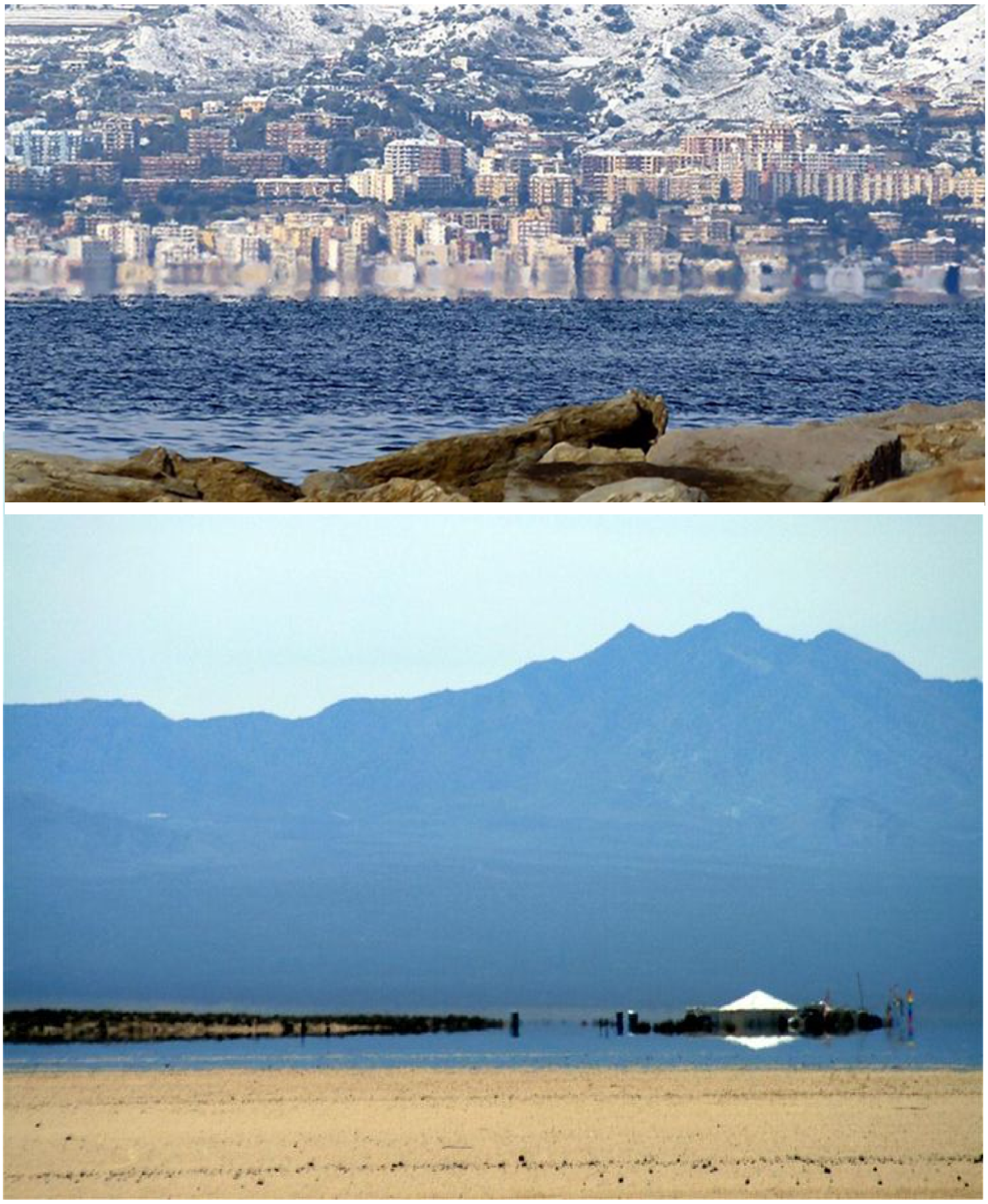


FIGURE 107 Two inferior mirages: one at the place were the term 'fata morgana' comes from, the Strait of Messina (top) and another in a desert (photographs © Nicola Petrolino and Mila Zinkova).

FROM REFRACTION TO LENSES

Above all, refraction is used in the design of *lenses*. With glass one can produce precisely curved surfaces that allow us to *focus* light. All focusing devices, such as lenses, can be used to produce images. The two main types of lenses, with their focal points and the images they produce, are shown in Figure 109; they are called *converging lenses* and *diver-*

FIGURE 108 Two inferior mirages producing looming (photographs © Olaf Schneider and Gerold Prenger).

gent lenses. When an object is more distant from a single converging lens than its focus, the lens produces a *real* image, i.e., an image that can be projected onto a screen. In all other cases single converging or diverging lenses produce so-called *virtual images*: such images can be seen with the eye but not be projected onto a screen. For example, when an object is put between a converging lens and its focus, the lens works as a *magnifying glass*. Figure 109 also allows one to deduce the *thin lens formula* that connects the lengths d_o, d_i and f. What is it? Challenge 168 s

Even though glasses and lenses have been known since antiquity, the Middle Ages had to pass by before two lenses were combined to make more elaborate optical instruments.

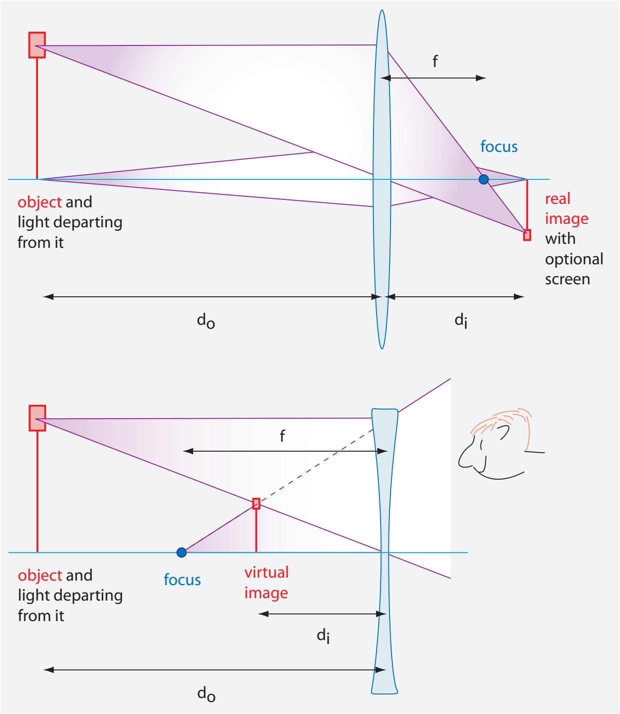


FIGURE 109 A real image produced by a converging lens (if used in the way shown) and the virtual image produced by a diverging lens.

Ref. 117 The various effects that can be observed with one or two lenses are shown in Figure 110. The *telescope* was invented – after a partial success in Italy by Giambattista della Porta – just before 1608 in the Netherlands. The most well-known of at least three simultaneous inventors was the lens grinder Johannes Lipperhey (b. *c.* 1570 Wesel, d. 1619 Middelburg) who made a fortune by selling his telescopes to the Dutch military. When Galileo heard about the discovery, he quickly took it over and improved it. Already in 1609, Galileo performed the first astronomical observations; they made him world-famous. The *Dutch telescope* design has a short tube yielding a bright and upright image, and its magnification is the ratio of the focal distances of the two lenses. It is still used today in opera glasses. Over the years, many other ways of building telescopes have been developed; nowadays, high-performance telescopes use mirrors instead of lenses; they are not as heavy and they allow the use of adaptive optics. Challenge 169 e Ref. 118

By the way, telescopes also exist in living beings. Most spiders have several types of eyes, and some spiders have up to 6 different pairs. In particular, the jumping spider genus *Portia (Salticidae)* has two especially large eyes, made to see distant objects, which have two lenses behind each other; the second lens and the retina behind it can be moved with muscles, so that such spiders can effectively point their telescope in different directions without moving their head. In order to process the input from all their eyes, jumping spiders need a large brain. In fact, about 50 % of the body mass of jumping spiders is brain mass.

Another way to combine two lenses leads to the *microscope*. Can you explain to a non-physicist how a microscope works? Werner Heisenberg almost failed his Ph.D. exam Challenge 170 s

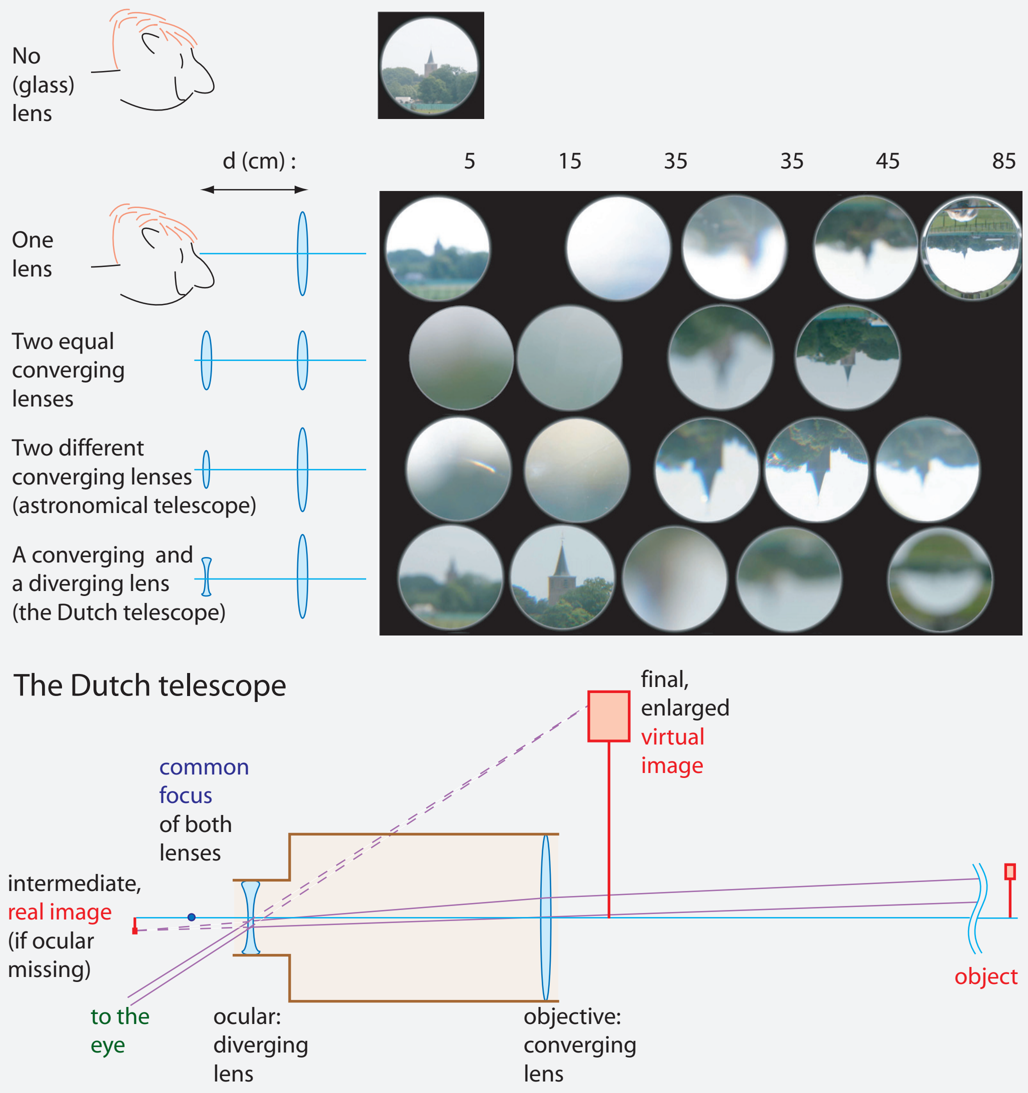


FIGURE 110 Lens refraction is the basis of the telescope: above, the experiments with lenses that lead to the development of the telescope: the object to watch compared with the images produced by a single converging lens, by two equal converging lenses, by two different converging lenses in the astronomical telescope, and by a diverging and a converging lens in the Dutch telescope, at various distances from the eye; below, the explanation of the Dutch telescope (photographs © Eric Kirchner).

because he could not. The problem is not difficult, though. Indeed, the inventor of the microscope was an autodidact of the seventeenth century: the technician Antoni van Leeuwenhoek (b. 1632 Delft, d. 1723 Delft) made a living by selling over five hundred of his microscopes to his contemporaries. (This is a somewhat nasty remark: Van Leeuwenhoek only used one lens, not two, as in the modern microscope.)

No ray tracing diagram, be it that of a simple lens, of a telescope or of a microscope, is really complete if the eye, with its lens and retina, is missing. Can you add it and convince yourself that these devices really work?

Challenge 171 ny

FIGURE 111 The glory produced by the droplets in a cloud (© Brocken Inaglory).

FIGURE 112 Watching this graphic at higher magnification shows the dispersion of the human eye: the letters float at different depths.

As mentioned, refraction is almost always colour-dependent; it shows dispersion. Because of dispersion, lenses produce chromatic aberrations; they are visible as coloured borders of images. To avoid this, microscopes or photographic cameras have *several* lenses made of *different* types of glass. (They also contain several lenses of the same glass type in order to compensate the geometric lens imaging errors called *Seidel aberrations* that are independent of colour.) The different glass types compensate dispersion and thus avoid the coloured image borders. The colour dependence of refraction in water droplets is also the basis of the rainbow, as shown below; the rainbow can be thought of as the coloured border of a white disk produced by the water droplets acting as lenses. Refraction in ice crystals – sometimes with dispersion and sometimes without – in the atmosphere is at the basis of the halos, the Sun pillars and the many other light patterns often seen around the Sun or the Moon in cold weather.

Page 125

Ref. 119

Also the human eye shows colour-dependent refraction, i.e., dispersion. Fortunately, the effect is small. Indeed, for the working of the eye, the curved *shape of the cornea* is more important than the refractive power of the lens, because the lens is embedded in

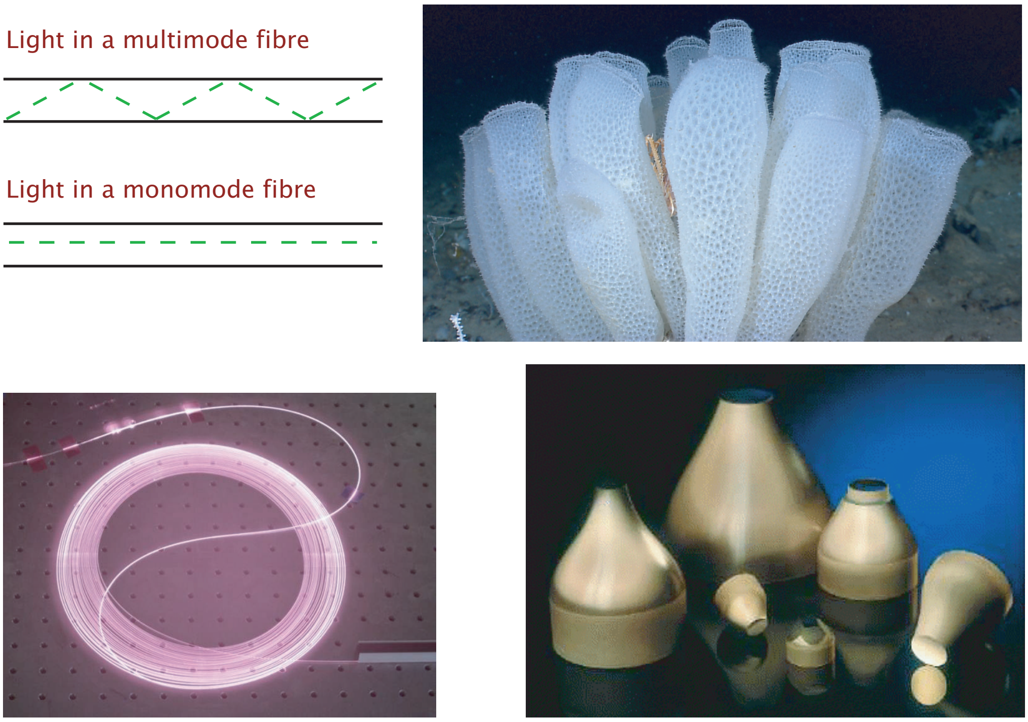


FIGURE 113 Optical fibres: the working principle of the two extreme fibre types, the astonishing marine sponge *Euplectella aspergillum* (height about 30 cm) that contains silica optical fibres with lenses at the end and synthesized at water temperature to help symbiotic algae, a modern fibre laser used in material processing and in medicine, and, glued together in large numbers, fibre tapers to change image sizes (maximum diameter about 20 cm) (© NOAA, Hochschule Mittweida, Schott).

a medium with nearly the same index of refraction, thus limiting the effects of refraction. The small effects of colour-dependent refraction is not corrected in the eye, but in the brain. Therefore, the dispersion of the eye lens can be noticed if this correction by the brain is prevented, for example when red or blue letters are printed on a black background, as shown in Figure 112. We get the impression that the red letters float in front of the blue letters. Can you explain how dispersion leads to this floating effect? Challenge 172 s

BENDING LIGHT WITH TUBES – FIBRE OPTICS

Another way to bend light, also based on refraction, is used by many animals and by many technical devices: the *optical fibre*. Optical fibres are based on total internal reflection; an overview of their uses is given in Figure 113.

In nature, optical fibres appear in at least three systems. In insect eyes, such as the eyes Page 195 of the house fly or the eye of a honey bee, the light for each image pixel is transported along a structure that works as a conical optical fibre. In certain sea animals, such as the Ref. 120 glass sponge *Euplectella aspergillum* and a number of other sponges, actual silica fibres are used to provide structural stability and to transport light signals to photodetectors. Finally, all vertebrate eyes, including the human eye, contain a large number of optical fibres above the retina, to avoid the image problems that might be caused by the blood

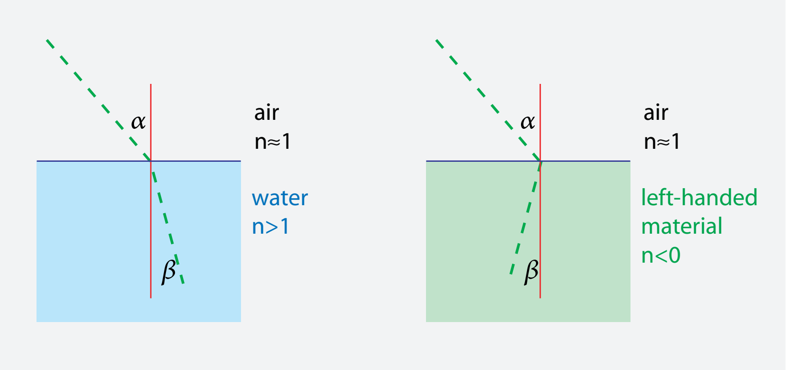


FIGURE 114 Positive and negative indices of refraction n.

Ref. 121 vessels, which lie *above* the retina in all vertebrate eyes. By the way, the frequently heard Ref. 122 claim that the white hair of polar bears works as optical fibres for UV light is *false*.

In technical applications, optical fibres are essential for the working of the telephone network and the internet, for signal distribution inside aeroplanes and cars, for the transport of laser light for medical uses, for high-power lasers and in many other settings. Hollow glass fibres are successfully used for the guiding of X-rays in X-ray imaging systems.

200 YEARS TOO LATE – NEGATIVE REFRACTION INDICES

In 1967 Victor Veselago made a strange prediction: the index of refraction could have *negative* values without invalidating any known 'law' of physics. A negative index means that a beam is refracted to the same side of the vertical, as shown in Figure 114. As a result, concave lenses made of such materials focus parallel beams and convex lenses disperse them, in contrast to usual lens materials.

In 1996, John Pendry and his group proposed ways of realizing such materials. In 2000, a first experimental confirmation for microwave refraction was published, but it Ref. 123 met with strong disbelief. In 2002 the debate was in full swing. It was argued that negative refraction indices imply speeds greater than that of light and are only possible for either phase velocity or group velocity, but not for the energy or true signal velocity. The conceptual problems would arise only because in some physical systems the refraction angle for phase motion and for energy motion differ.

In the meantime, the debate is over. Negative indices of refraction have indeed been observed frequently; the corresponding systems are being extensively explored all over Ref. 124 the world. Systems with negative index of refraction do exist. Following Veselago, the materials showing this property are called *left-handed*. The reason is that the vectors of the electric field, the magnetic field and the wave vector form a left-handed triplet, in contrast to vacuum and usual materials, where the triplet is right-handed. All left-handed materials have *negative* magnetic permeability μ_r and *negative* dielectric coeffi- Ref. 125 cient, i.e., negative permittivity ε_r. However, in actual systems, these properties are only realized for a narrow range of frequencies, usually in the microwave range.

Apart from the unusual refraction properties, left-handed materials have negative phase velocities, i.e., a phase velocity opposed to the energy velocity and show a reversed

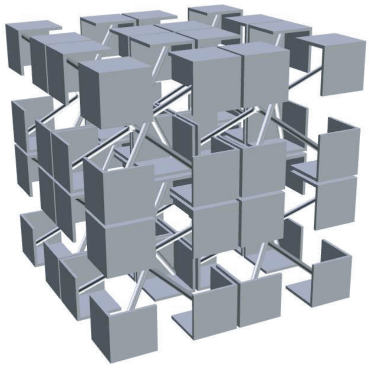


FIGURE 115 An example of an isotropic metamaterial (M. Zedler et al., © 2007 IEEE).

Doppler effect. These properties have been confirmed by experiment. Left-handed materials should also yield obtuse angles in the Vavilov–Çerenkov effect, thus emitting Vavilov–Çerenkov radiation in the backward instead of in the forward direction, they are predicted to have an inverted *Goos-Hänchen effect* and to show a repulsive Casimir effect. However, these predictions have not been verified yet.

Most intriguing, negative index materials are predicted to allow constructing lenses that are completely flat. In addition, in the year 2000, John Pendry gained the attention of the whole physics community world-wide by predicting that lenses made with such materials, in particular for a refractive index $n = -1$, would be *perfect*, thus beating the usual diffraction limit. This would happen because such a perfect lens would also image the *evanescent* parts of the waves – i.e., the exponentially decaying ones – by amplifying them accordingly. First experiments claim to confirm the prediction. Exploration of the topic is still in full swing. Ref. 126 Ref. 125

So far, left-handed materials have been realized only for microwave and terahertz frequencies. First claims in the visible domain have been published, but have to be taken with care. It should be mentioned that one type of negative refraction systems have been known since a long time: diffraction gratings. We could argue that left-handed materials are gratings that attempt to work in all spatial directions. And indeed, all left-handed materials realized so far are periodic arrangements of electromagnetic circuits.

Metamaterials

The simplest realization of left-handed systems are metamaterials. *Metamaterials* are engineered metal-insulator structures with a periodicity below the wavelength of the radiation for which they are designed, so that the structure behaves like a homogeneous material. Metamaterials have negative or otherwise unusual permittivity or permeability properties in a certain wavelength range, usually in the microwave domain; some metamaterials are left-handed.

Ref. 127 Currently, there are two basic approaches to realize metamaterials. The first is to build a metamaterial from a large array of compact resonant substructures, such as inductor-capacitor (LC-) circuits or dielectric spheres. The second approach is to build a metamaterial from transmission lines. The latter approach has lower losses and a wider spectral range; an example for this type is shown in Figure 115. Comparing and exploring differ- Ref. 128

ent realizations is subject of intense research.

Most metamaterials are conceived for microwaves or terahertz waves. Industrial applications of metamaterials are expected for antenna design; for example, an antenna dipole could be located just above a metamaterial and thus allowing to build flat directional antennas. Applications in terahertz technology might also arise.

Less serious workers in the field claim that *invisibility cloaks* can be realized with metamaterials. While this is a good marketing slogan to attract funding and get into newspapers, the dream is not realistic, due to inevitable signal losses in the materials, dispersion, refraction, finite cell size, the need for windows to observe the outside from inside and the impossibility to achieve invisibility for all wavelengths. So far, all aeroplanes that were claimed to be invisible even only for specific radar frequencies have turned out to be visible to radar after all. But sources of military funding are known to have only a distant relation to reality.

Metamaterials for sound and lower-frequency waves are also subject of research. Such acoustic or mechanical metamaterials have not found a technical application yet.

Light around corners – diffraction

Light goes around corners. This effect was called *diffraction* by Francesco Grimaldi, in his text *Physico-mathesis de lumine*, published in 1665. Grimaldi studied shadows very carefully. He found out what everybody now learns in secondary school: light goes around corners in the same way that sound does, and light diffraction is due to the wave nature of light. (Newton got interested in optics after he read Grimaldi; Newton then wrongly dismissed Grimaldi's conclusions.) Ref. 129

Because of diffraction, it is *impossible* to produce strictly parallel light beams. For example, every laser beam diverges by a certain minimum amount, called the *diffraction limit*. Maybe you know that the world's most expensive Cat's-eyes are on the Moon, where they have been deposited by the Lunokhod and the Apollo missions. Can you determine how wide a laser beam with minimum divergence has become when it arrives at the Moon and returns back to Earth, assuming that it was 1 m wide when it left Earth? How wide would it be on its return if it had been 1 mm wide at the start? In short, both diffraction and the impossibility of non-diverging beams confirm that light is a wave. Ref. 130 Challenge 173 s

Diffraction implies that there are no perfectly sharp images: there exists a *limit on resolution*. This is true for every optical instrument, including the eye. The resolution of the eye is between one and two minutes of arc, i.e., between 0.3 and 0.6 mrad. The limit is partly due to the finite size of the pupil. (That is why squinting helps to see more sharply.) In practice, the resolution of the eye is often limited by chromatic aberrations and shape imperfections of the cornea and lens. (Can you check the numbers and their interpretation by calculation? Is it true that the number of rods in the eye is tuned exactly to its resolution?) Therefore, for example, there is a maximum distance at which humans can distinguish the two headlights of a car. Can you estimate it? Challenge 174 d Challenge 175 s

Resolution limits also make it impossible to see the Great Wall in northern China from the Moon, contrary to what is often claimed. In the few parts that are not yet in ruins, the wall is about 6 metres wide, and even if it casts a wide shadow during the morning or the evening, the angle it subtends is way below a second of arc, so that it is completely invisible to the human eye. In fact, three different cosmonauts who travelled

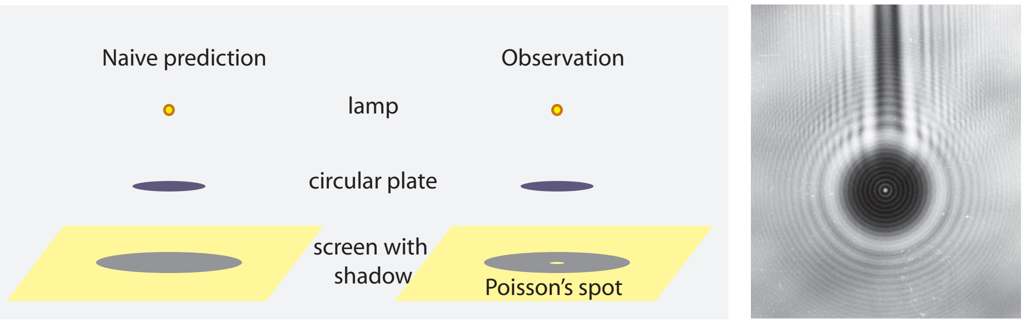


FIGURE 116 Shadows show that light is a wave: the naive expectation (left), neglecting the wave idea, and the actual observation (middle and right) of the shadow of a circular object (photo © Christopher Jones).

Ref. 131 to the Moon performed careful searches and confirmed that the claim is absurd. The story is one of the most tenacious urban legends. (Is it possible to see the Wall from the space shuttle?) Challenge 176 ny The largest human-made objects are the polders of reclaimed land in the Netherlands; they *are* visible from outer space. So are most large cities as well as the highways in Belgium at night; their bright illumination makes them stand out clearly from the dark side of the Earth.

Diffraction has a strange consequence. The shadow of a small illuminated ball from a ball bearing, shows, against expectations, a bright spot at its centre. The effect is illustrated in Figure 116. This 'hole' in the shadow was predicted in 1819 by Denis Poisson (b. 1781 Pithiviers, d. 1840 Paris) in order to show to what absurd consequences the wave theory of light would lead. He had just read the mathematical description of diffraction developed by Augustin Fresnel* on the basis of the wave description of light. But shortly afterwards, François Arago actually observed Poisson's spot, converting Poisson, making Fresnel famous and accelerating the general acceptance of the wave properties of light.

Diffraction can also be used, in certain special applications, to produce images. A few examples of the use of diffraction in optics are shown in Figure 117. Of these, acousto-optic modulators are used in many laser systems, for example in laser shows. Also holograms, to be discussed in detail below, Page 175 can be considered a special kind of diffractive images.

In summary, diffraction is sometimes used to form or to influence images; but above all, in every image, diffraction determines the resolution, i.e., the image quality.

Beating the diffraction limit

In all imaging methods, the race is for images with the highest resolution possible. The perfect lens mentioned above Page 168 has not been realized for visible light. However, other

---

* Augustin Jean Fresnel (b. 1788 Broglie, d. 1827 Ville d'Avray), engineer and part time physicist. The 's' in his name is silent. In 1818, he published his great paper on wave theory for which he got the prize of the French Academy of Sciences in 1819. To improve his finances, he worked in the commission responsible for lighthouses, for which he developed the well-known Fresnel lens. He died prematurely, partly of tuberculosis and partly of exhaustion due to overwork.

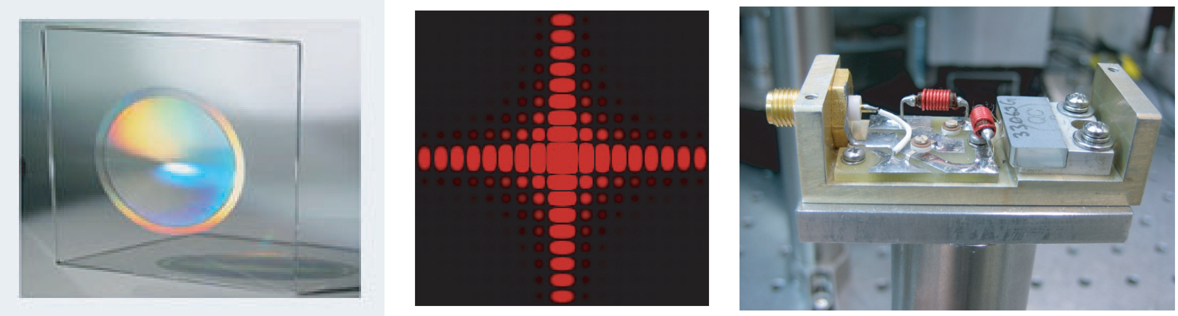


FIGURE 117 Examples of diffractive optics: a diffractive aspherical lens, the result shining a red laser through of a plastic sheet with a diffractive cross generator, and an acousto-optic modulator used to modulate laser beams that are transmitted through the built-in crystal (© Jenoptik, Wikimedia, Jeff Sherman).

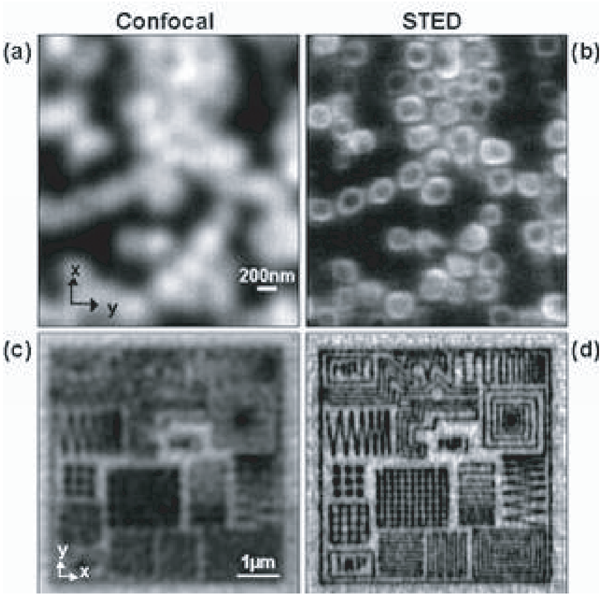


FIGURE 118 Sub-wavelength optical microscopy using stimulated emission depletion (right) compared to conventional confocal microscopy (left) (© MPI für biophysikalische Chemie/Stefan Hell).

techniques of producing images with resolutions *less* than the wavelength of light have made great progress in recent years.

Nowadays, extraordinary images can be produced with modified commercial light microscopes. The conventional diffraction limit for microscopes is

$$d \geqslant \frac{\lambda}{2n\ \sin\alpha} , \tag{79}$$

where λ is the wavelength, n the index of refraction and α is the angle of observation. There are three main ways to circumvent this limit. The first is to work in the 'near field', where the diffraction limit is not valid, the second way is to observe and measure the diffraction effects and then to use computers to reduce the effects via image processing, the third way is to use effects that produces light emission from the sample that is smaller than the wavelength of light, and the fourth way is to use resolution in time to increase resolution in time.

A well-known near-field technique is the near-field scanning optical microscope. Light is sent through a tapered glass fibre with a small transparent hole at the end, down to 15 nm; the tip is scanned over the sample, so that the image is acquired point by point. These microscopes achieve the highest resolution of all optical microscopes. However, it is hard to get a practical amount of light through the small hole at the end of the tip.

Many computational techniques can achieve images that achieve resolutions below the diffraction limit. The simpler types of these deconvolution microscopy techniques are already commercially available.

One of the first techniques that beat the diffraction limit by a substantial amount using a conventional microscope is *stimulated emission depletion microscopy*. Using a clever illumination system based on two laser beams, the technique allows spot sizes of almost molecular size. The new technique, a special type of fluorescence microscopy developed by Stefan Hell, uses an illuminating laser beam with a circular spot and a second laser beam with a ring-like shape. As a result of this combination, the techniques modifies the diffraction limit to

$$d \geqslant \frac{\lambda}{2n\ \sin\alpha\ \sqrt{I/I_{\text{sat}}}}\ , \tag{80}$$

so that a properly chosen saturation intensity I_{sat} allows one to reduce the diffraction limit to arbitrary low values. So far, light microscopy with a resolution of 16 nm has been performed. An example image is shown in Figure 118. This and similar techniques have galvanized the microscopy field; they are now commonplace in materials science, medicine and biology. In 2014, Stefan Hell received the Nobel Prize in Chemistry for his achievements. Ref. 132

Research in new microscopy techniques is still ongoing, also in the numerous attempts to transfer resolution in time to resolution in space. Another important domain of research is the development of microscopes that can be included in endoscopes, so that physicians can explore the human body without the need of large operations. Microscopy is still a field in full swing.

OTHER WAYS TO BEND LIGHT

Optical technology can be defined as the science of bending light. Reflection, refraction and diffraction are the most important methods to achieve this. But it makes sense to explore the question more generally: what other ways can be used to bend light beams?

Vol. I, page 201
Vol. II, page 161
A further way to bend light is gravity, as discussed already in the chapters on universal gravity and those on general relativity. Since the effect of gravity is weak, it is only of importance in astronomy. Gravitational lensing is used in various projects to measure the size, mass and distance of galaxies and galaxy groups. The usually negligible effect of gravity between two light beams was also discussed earlier on. Vol. II, page 260

In practice, there are thus no laboratory-scale methods to bend light beams apart from reflection, refraction and diffraction. All known methods are specialized cases of these three options.

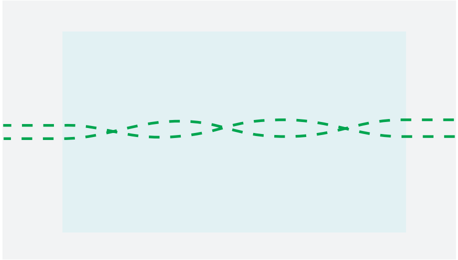


FIGURE 119 In certain materials, light beams can spiral around each other.

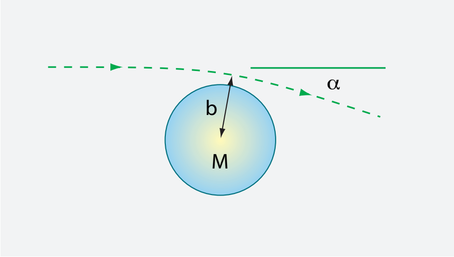


FIGURE 120 Masses bend light.

An important way in which materials can be used to bend light are *acousto-optic deflectors*. They work like acousto-optic modulators, i.e., a sound wave travelling through a crystal generates a diffraction grating that is used to deflect a laser beam. Such modulators thus use diffraction to bend light. Page 171

Additional electromagnetic fields usually do not influence light directly, since light has no charge and since Maxwell's equations are linear. But in some materials the effective equations are non-linear, and the story changes. For example, in certain photorefractive materials, two nearby light beams can even *twist* around each other, as was shown by Segev and coworkers in 1997. This is illustrated in Figure 119. This effect is thus a form of refraction. Ref. 133

Another common way to deflect light uses its polarization. Many materials, for example liquid crystals or electro-optic materials, bend light beams depending on their polarization. These materials can be used to steer or even to block laser beams. Liquid crystal modulators and electro-optic modulators are thus based in refraction.

Scattered light also changes direction. It is debatable whether it is appropriate to call this process an example of bending of light. In any case, scattering is important: without it, we would not see almost anything around us. After all, everyday seeing is detection of scattered light. And of course, scattering is a case of diffraction. Vol. IV, page 70

The next question is: what methods exist to *move* light beams? Even though photons have zero mass and electrons have non-zero mass, scanning electron beams is easily achieved with more than 1 GHz frequency, whereas scanning powerful light beams is hard for more than 10 kHz.

Moving light beams – and *laser beams* in particular – is important: solutions are the basis of a sizeable industry. Moving laser beams are used for laser treatments of the eye, for laser marking, for laser shows, for laser cutting, for barcode reading in supermarkets, for rapid prototyping, for laser sintering three-dimensional parts, for laser distance measurements, for lidar, for the mentioned microscopy techniques, and for various industrial processes in the production of electronic printed circuits, of semiconductor products, and of displays for mobile phones. Most laser scanners are based on moving mirrors, prisms or lenses, though acousto-optic scanners and electro-optic scanners, which achieve a few MHz scanning rate for low power beams, are also used in special applications. Many applications are eagerly waiting for inventions that allow faster laser scanning. Page 146

In summary, moving light beams requires to move matter, usually in the form of mir-

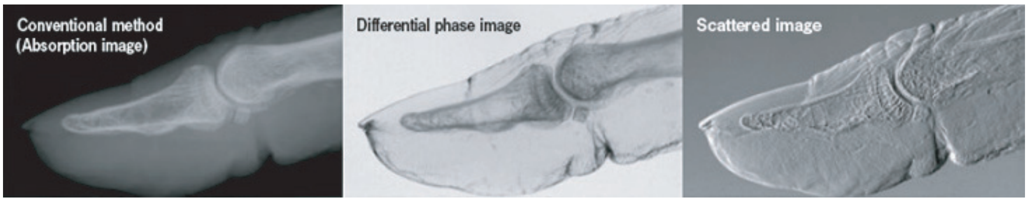


FIGURE 121 Three types of X-ray images of a thumb: the conventional image (left) and two images taken using interference effects (© Momose Atsushi).

ror or lenses. Light travels in straight line only if it travels *far from matter*. In everyday life, 'far' simply means more than a few millimetres, because electromagnetic effects are negligible at these distances, mainly due to light's truly supersonic speed. However, as we have seen, in some cases that involve gravitation, larger distances from matter are necessary to ensure undisturbed motion of light.

Using interference for imaging

Page 104 As we saw above for the case of the guitar, images produced by interference can be useful. Above all, interference effects can be used to measure the deformation and the motion of objects.

Interference can also be used to enhance images. Figure 121 show the improvement that is possible when a special case of interferometer, a so-called *Talbot-Lau interferometer*, is used with X-rays. In particular, the technique increases the sensitivity of X-rays for soft tissue. Ref. 134

Interference is also at the basis of holography, an important technique to produce three-dimensional images.

How does one make holograms and other three-dimensional images?

Our sense of sight gives us an image of the world around us that includes the impression of depth. We constantly experience our environment as *three*-dimensional. Stereopsis, the experience of depth, occurs because of three main effects. First, the two eyes see *different images*. Second, the images formed in each eye are *position dependent*: when we move the head, we observe parallax effects between the bodies near and far from us. Third, for different distances, our eyes needs to *focus differently* and to *converge* more or less strongly, depending on the position of the object.

A usual paper photograph does not capture any of these three-dimensional effects: a paper photograph corresponds to the picture taken by one eye, from one particular spot and at one particular focus. In fact, all photographic cameras are essentially copies of a single, static eye with fixed focus. Challenge 177 e

Any system trying to produce the perception of depth for the observer must include at least one of the three three-dimensional effects just mentioned. In fact, the third effect, varying focus with distance, is the weakest one, so that most systems concentrate on the other two effects, different images for the two eyes, and an image that depends on the position of the head.

FIGURE 122 The highest-quality holograms available in the world at present are produced by Yves Gentet and can be found on his website www.ultimate-holography.com. They are Denisjuk holograms. The viewer is tricked into thinking that there are real butterflies behind the glass pane. (© Yves Gentet).

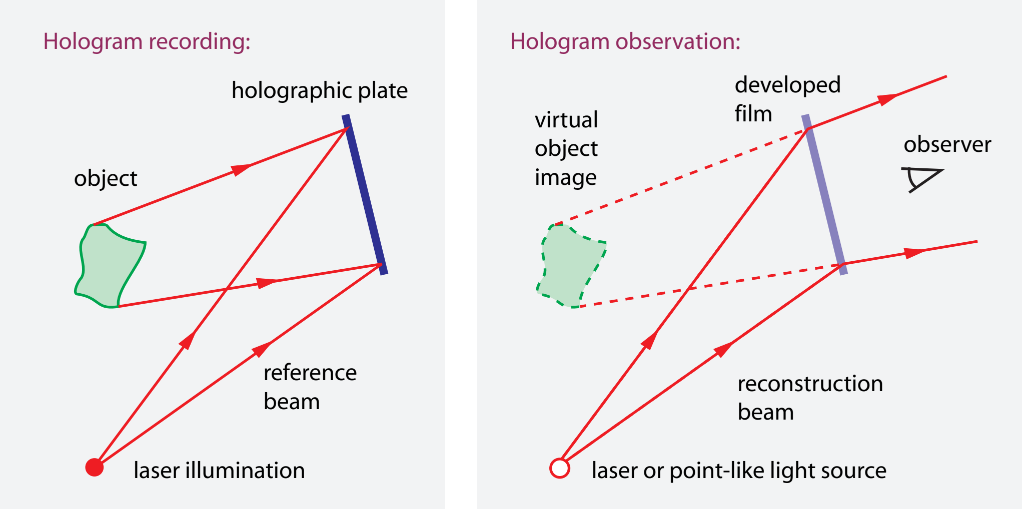


FIGURE 123 The recording (left) and the observation (right) of a monochromatic hologram (in this case, in transmission). True colour holograms use three lasers, for red, green and blue.

Stereo photography and stereo films extensively use the first effect, sending different images to different eyes, by various technical tricks. A common trick is to use coloured glasses. Also certain post cards and computer screens are covered by thin cylindrical lenses that allow sending two different images to the two eyes, thus generating an impression of depth. It is well known that at large object distances, the two images in the two human eyes do not differ any more. This limit distance is called the *stereoscopic radius* and lies somewhere between 200 m and 500 m.

But obviously the most spectacular depth effect is the second, obtained whenever head-position-dependent images can be created. Modern *virtual reality systems* take films using a number of cameras in all directions. The use up to 12 cameras, for example with two cameras at eye distance pointing along each coordinate axis. In this way they include also the first depth effect. Using a goggle with direction sensors attached to the head, these systems interpolate the taken film in the actual head direction of the viewer or generate a computer-calculated film that depends on the head orientation. Such virtual reality systems allow anybody to experience in a surprisingly realistic way a ride on the back of an eagle flying through the mountains or a dive among sharks in the deep sea.

So far, the only method that achieves all three depth effects is *holography*. The resulting images are called *holograms*. An example of a hologram is shown in Figure 122. Even though a hologram is only a film with a thickness of a fraction of a millimetre, the observer has the impression that there are objects behind it. Depending on the details of the geometry, objects can also seem to float in front of the film.

A hologram reproduces all data that is seen from any point of a region of space. A *hologram* is thus a stored set of position-dependent pictures of an object. In a first step, a hologram is captured by storing amplitude *and phase* of the light emitted or scattered by an object, as shown in Figure 123 and Figure 125. To achieve this storage of the whole

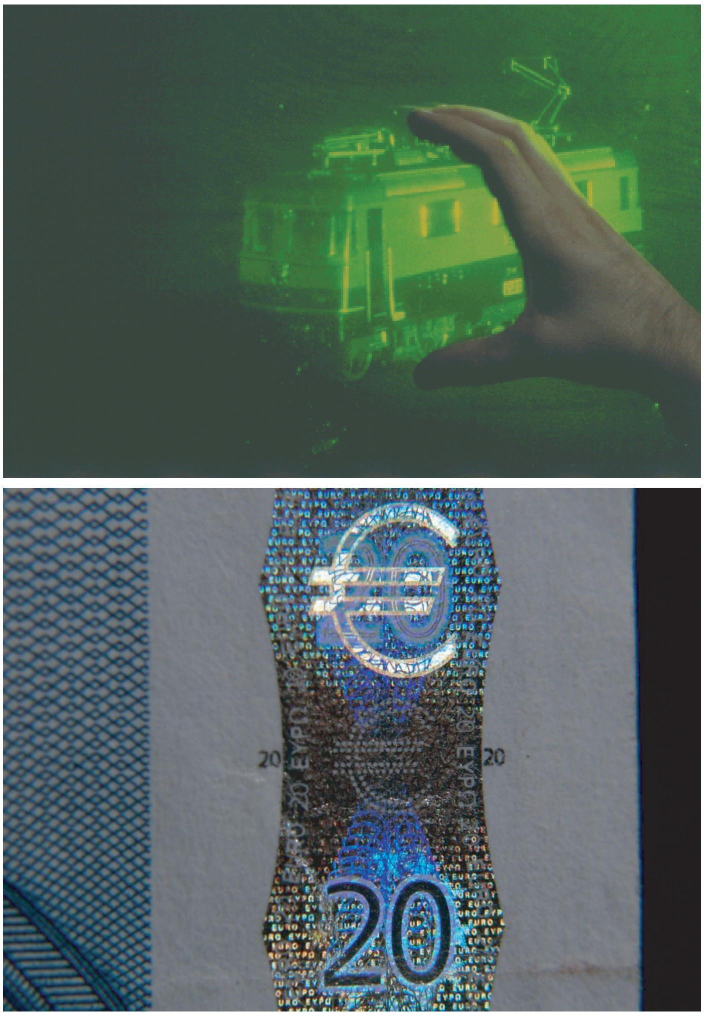


FIGURE 124 A hologram of a train and the reflection hologram on a Euro bill (© Anonymous, Hans-Ulrich Pötsch).

light field, the object is illuminated by *coherent* light,* such as light from a laser, and the interference pattern between the illumination and the scattered light is stored; usually it is stored in a photographic film. The procedure is shown schematically in Figure 123. In a second step, illuminating the developed film by coherent light – from a laser or a lamp that is as point-like as possible – allows one to see a full three-dimensional image. In particular, due to the reproduction of the situation, the image appears to float in free space.

A few examples of holograms are shown in Figure 124. Holograms were developed in 1947 by the famous physicist Dennis Gabor (b. 1900 Budapest, d. 1979 London), who

---

* Generally speaking, two light beams or two parts of one light beam – or other waves – are called *coherent* if they have constant phase difference and frequency. In practice, due to ubiquitous disturbances, this only happens over a certain finite volume, which is then called the *volume of coherence*. Coherence enables and is required for interference.

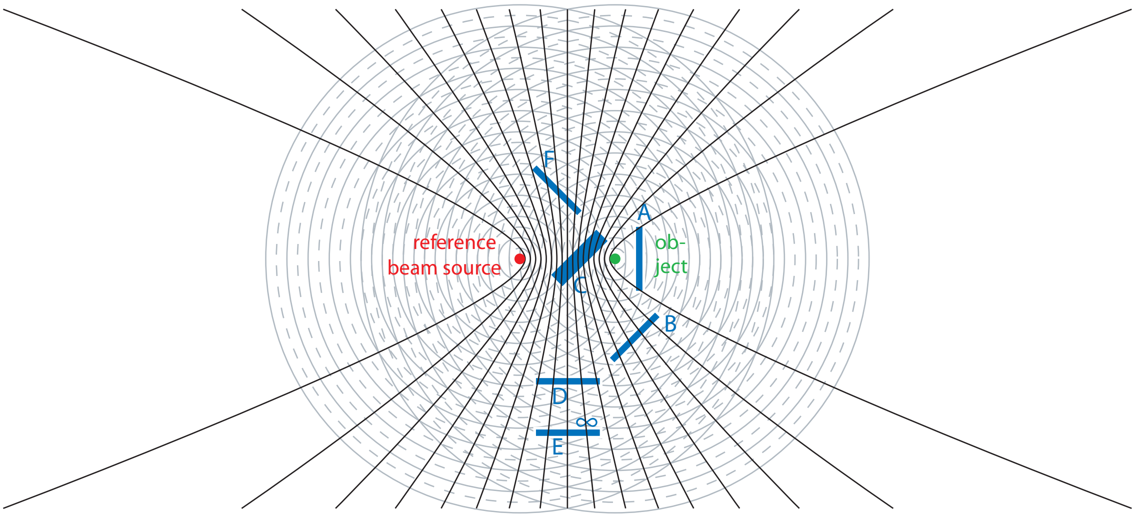


FIGURE 125 Different types of holograms arise through different relative position of object (green), holographic plate (blue) and reference beam (red). Situation A denotes a thin inline transmission hologram as proposed by Gabor, B a thin offline transmission hologram following Leith and Upatnieks, C a thick reflection hologram, or white light hologram, following Denisyuk, D a Fourier hologram at large distance, E a Fraunhofer hologram at infinite distance and F a two-dimensional hologram with inverted wave train (© DGH).

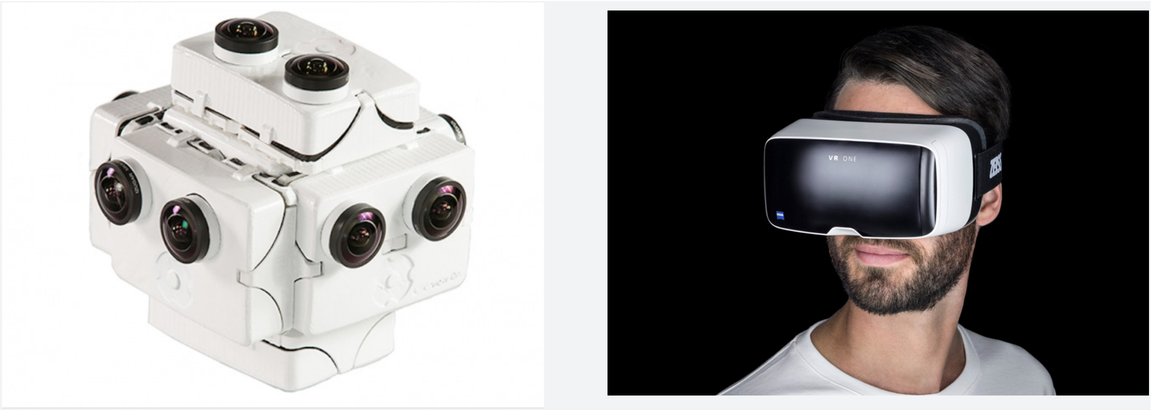


FIGURE 126 A virtual reality camera proposed for a trip to the International Space Station and a headset to experience the resulting videos (© SpaceVR and Zeiss)

received the 1971 Nobel Prize in Physics for this work. The beauty of Gabor's invention is that it was mainly theoretical, since lasers were not yet available at the time.

Holograms can be *transmission* holograms, like those in seen in museums, or *reflection* holograms, like those found on credit cards or currency bills. Holograms can be laser holograms and white light holograms. Most coloured holograms are rainbow holograms, showing false colours that are unrelated to the original objects. Real colour holograms, made and rendered with three different lasers, are possible but expensive.

Holograms are based on interference. Interference images can also be used in other ways. By a double illumination at two different times, one obtains a so-called *interfero-*

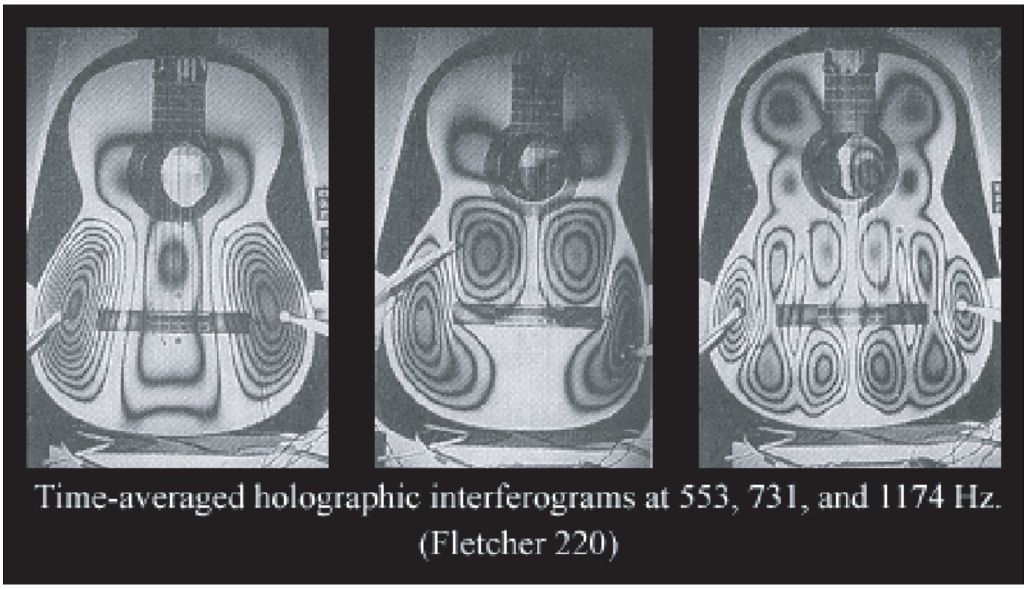


FIGURE 127 Interferograms of a guitar (© Wikimedia).

gram, which allows visualizing and measuring the deformation of an object. Interferograms are used to observe and measure deformation, oscillation or temperature effects.

Is it possible to make *moving* holograms? Yes; however, the technical set-ups are still subject of research. So far, such systems exist only in a few laboratories (for example, www.optics.arizona.edu/pstg/index.html) and are expensive. By the way, can you describe how you would distinguish a high quality moving hologram from a real body without touching it? Challenge 178 s

In the beginning of the computer industry, the aim of display makers was to produce *photo-realistic* displays, i.e., displays that could not be distinguished from a photograph. This aim has become reality. In 2012, a technology visionary proposed that the next aim of the industry should be to produce *window-realistic* displays, i.e., displays that cannot be distingusihed from a window. This should include the three-dimensionality of everything that is shown inside such a display. Will such a display ever be possible? Challenge 179 d

Not all three-dimensional images are holograms. Using rotating displays, rotating mirrors or rotating screens, it is possible to produce stunning three-dimensional images. An impressive example of such technology demonstrators is presented in Figure 128. Can you deduce why it was not a commercial success? Challenge 180 e

A well-known toy that make floating images with two stacked parabolic mirrors is shown in Figure 129. It is sometimes called a 'mirascope', but this awful term mixes latin and greek and like all such awful terms, including 'automobile', should never be used. Can you find out how the parabolic mirrors produce this astonishing effect? Challenge 181 e

IMAGES THROUGH SCANNING

When images are produced using lenses or mirrors, all the pixels of an image are produced in parallel. In contrast, in scanning techniques, images are constructed seri-

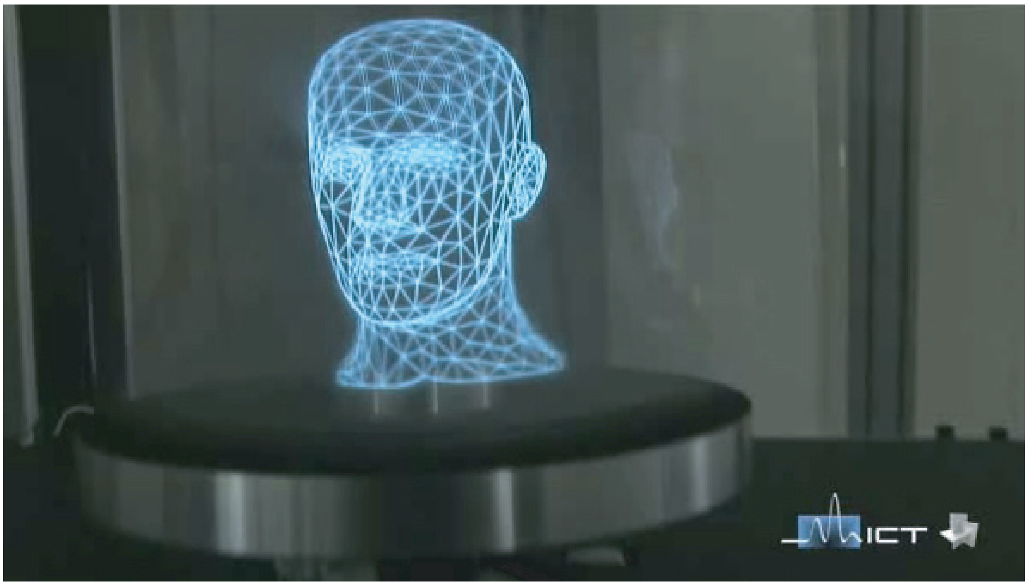


FIGURE 128 A three-dimensional image system based on a rotating mirror, from the University of Southern California, at gl.ict.usc.edu/Research/3DDisplay (© USC Stevens Institute for Innovation).

FIGURE 129 A floating Lego brick displayed by two stacked parabolic mirrors, the upper one with a hole. The right picture shows both the brick lying at the bottom and the floating image. (© Christoph Schiller).

ally, pixels after pixels. Even though scanning is intrinsically slower than any parallel technique, it has its own advantages: scanning allows imaging in three dimensions and achieving resolutions higher than the diffraction limit. Scanning techniques are mainly used in microscopy.

The most famous scanning technique does not use light rays, but electrons: the *scanning electron microscope*. As shown in Figure 131, such microscopes can produce stunning images. However, the images produced are two-dimensional. In special cases, *ion microscopes* are also used. All microscopes that use charged particles exist both as scanning and as transmission microscopes.

A typical example for a modern three-dimensional imaging technique based on light

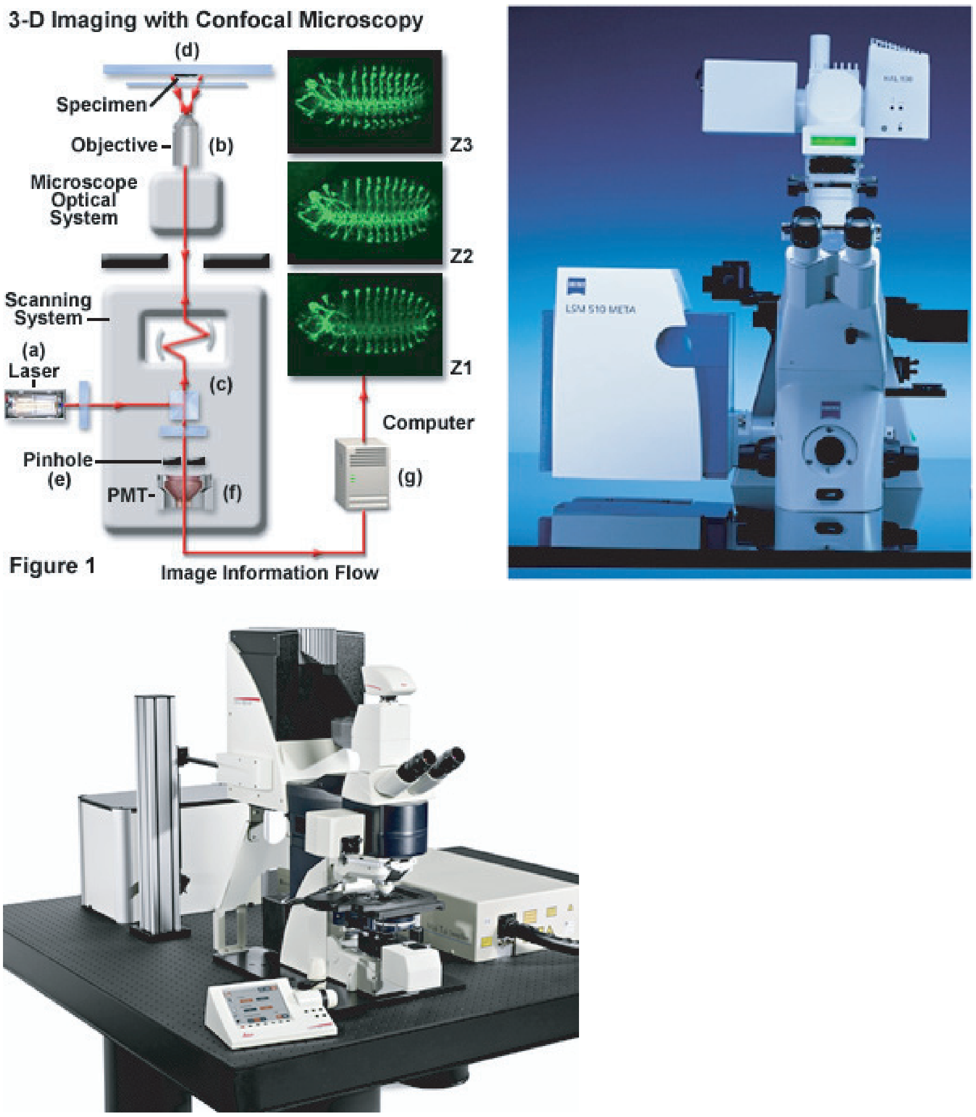


FIGURE 130 Two scanning imaging techniques: confocal laser scanning microscopy and multiphoton microscopy (© Nikon, Carl Zeiss).

FIGURE 131 A modern scanning electron microscope, and an image of pollen – field size about 0.3 mm – showing the resolution and the depth of field achievable with the technique (© Zeiss, Wikimedia).

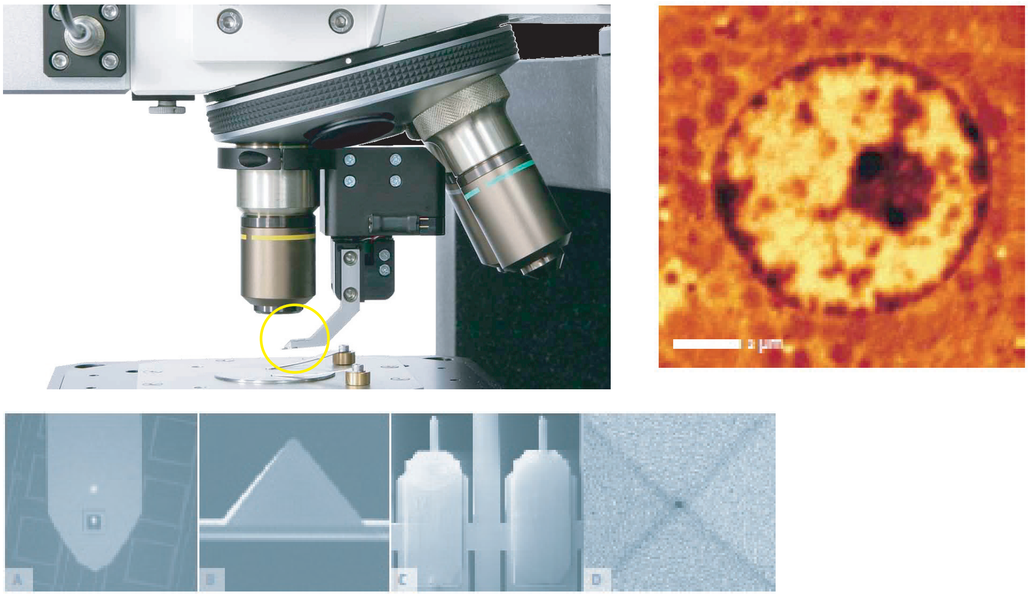


FIGURE 132 A scanning near-field optical microscope (SNOM) combined with an optical microscope, the details of the scanning probe, and an image of a liver cell nucleus produced with it (© WITec).

is *confocal laser scanning microscopy*. The technique is based on eliminating all light signals that are outside the focus of the microscope. The technique allows taking a picture of a more or less transparent specimen at a specified depth below its surface, up to a maximum depth of about 500 µm. Confocal microscopes are now available from various manufacturers.

An example of a technique for high-resolution is *multiphoton microscopy*. In this technique, the fluorescence of a specimen is excited using two or three photons of longer wavelengths. Like all fluorescence techniques, the image is produced from the fluorescent light emitted by certain chemical substances found in living organisms. In contrast to usual fluorescence microscopy, multiphoton imaging is based on a nonlinear effect, so that the emission region is extremely narrow and therefore high resolution is achieved.

For the highest possible optical resolution, *scanning near-field optical microscopy* is unsurpassed. Usually, a tiny optical probe is scanned across the surface, as shown in Figure 132. By working in the near field, the diffraction limit is circumvented, and resolution in the nanometre range becomes possible.

Another group of scanning microscopes also use electromagnetism to produce highest resolution images, though they do not use light. The most famous examples are the *scanning tunnelling microscope* or STM, the *atomic force microscope* or AFM and the *magnetic force microscope* or MFM. These instruments, though small and easy to build, have revolutionized material science in the last decades, because they allow to achieve atomic resolution in air on a normal laboratory table.

Vol. I, page 345

In summary, technological advances nowadays allow sophisticated imaging systems based on scanning, in particular in the field of microscopy. Since the field is still in flux, scanning techniques are expected to yield even more impressive results in the coming years. This progress in scanning techniques reminds one of the past progress of a fur-

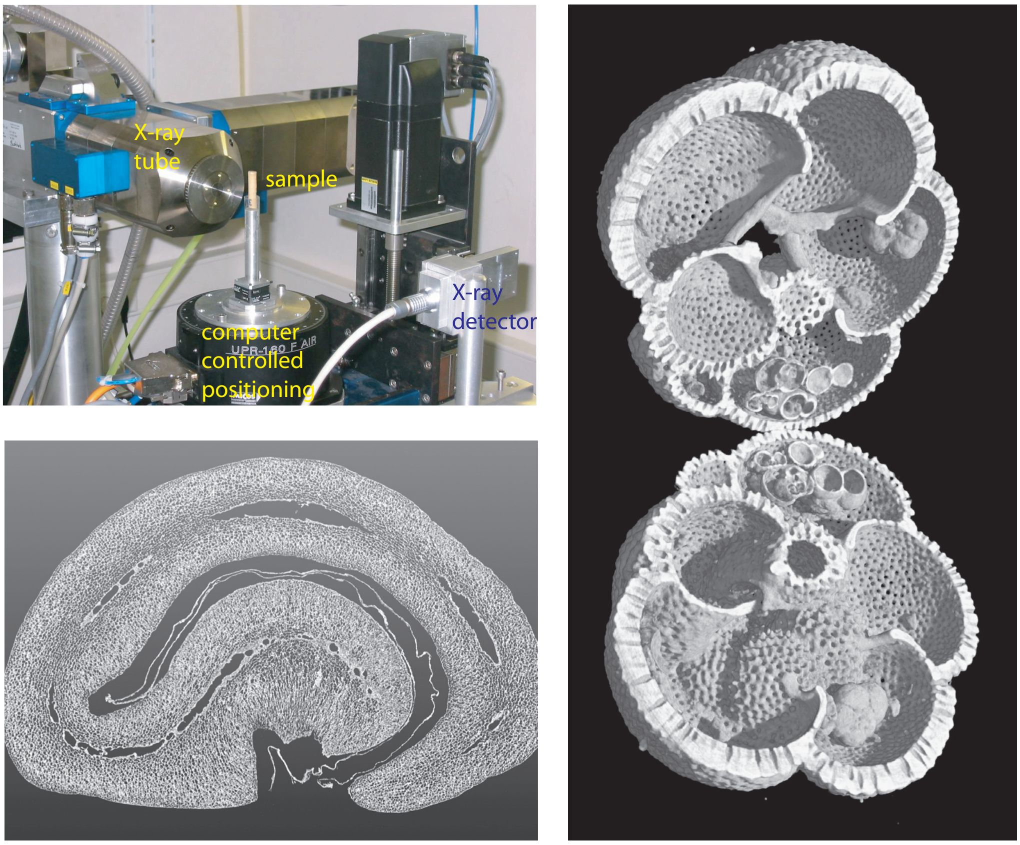


FIGURE 133 A set-up for high-resolution X-ray tomography, and two examples of images produced with it: a cross-section of a coffee bean (lower left) with a size of 8 mm, and a three-dimensional reconstruction of the exoskeleton of a foraminiferan, with a diameter of only 0.5 mm (© Manuel Dierick).

ther type of imaging principle that reconstructs images in an even more involved way: tomography.

TOMOGRAPHY

A spectacular type of imaging has become possible only after high-speed computers became cheap: *tomography*. In tomography, a radiation source rotates around the object to be imaged; the radiation that is scattered and/or transmitted is detected, and with sophisticated computer programming, a cross section of the object is reconstructed. Three-dimensional reconstructions are also possible. Tomography can be performed with any type of radiation that can be emitted in sufficiently well-defined beams, such as gamma rays, X-rays, light, radio waves, electron beams, neutron beams, sound and even earthquakes. X-ray tomography is a standard method in health care; visible light tomography, which has no side effects on humans, is being developed for breast tumour detection. Additional specialized techniques are electrical resistivity tomography, magnetic induction tomography and cryo-electron tomography.

In several types of tomography, the resolution achieved is breath-taking. An example

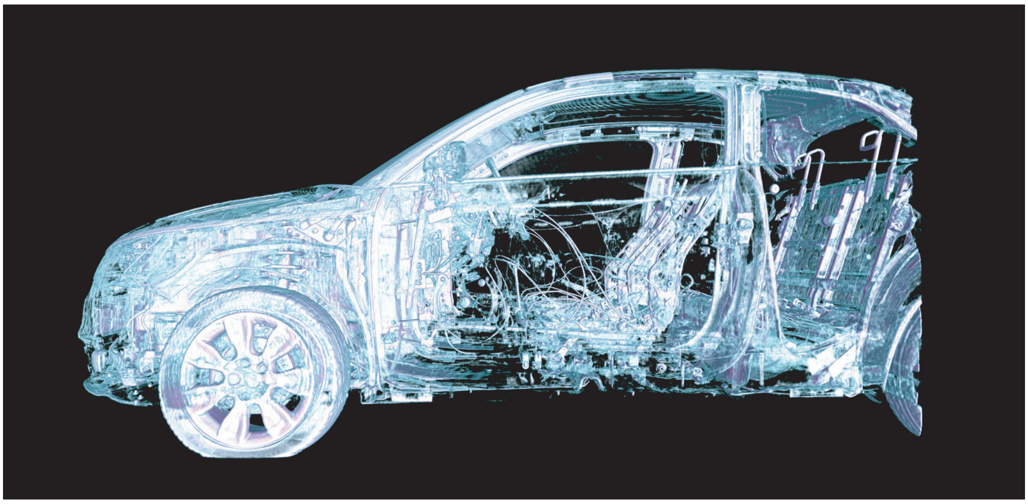


FIGURE 134 An X-ray CT image of a modern passenger car, with a resolution of less than 1 mm (© Fraunhofer IIS).

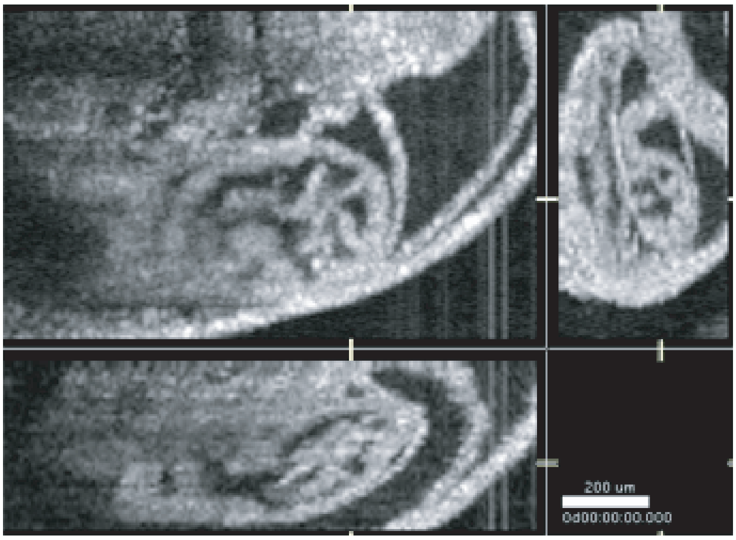


FIGURE 135 An OCT film of the heartbeat of a mouse embryo taken by Kyrill Larin. The three views correspond to the three coordinate axes. (QuickTime film © Kyrill Larin).

for modern high-resolution *X-ray tomography* of really small objects is shown in Figure 133. An example of X-ray tomography of a large object is shown in Figure 134. Building a set-up that produces such images is a large project and an impressive feat. Also magnetic resonance imaging, widely used in health care to image the interior of the human body, is a type of tomography, based on radio waves; it will be presented later on in our journey. Various types of tomographic systems – including *opto-acoustic tomography* based on sound produced by pulsed light, positron emission tomography, optical coherence tomography and the common sonography – also allow the production of film sequences.

Vol. V, page 162

An example of a technique that allows both three-dimensional imaging and high-resolution is *optical coherence tomography*. The technique is free of danger for the patient or specimen, allows a depth of a few millimetres in animal or human tissue, and allows resolutions down to 500 nm. Modern systems allow imaging of 10 GVoxel/s and more, so that films of biological processes can be produced in vivo, such as the blood flow in a human finger. Using the Doppler effect, the direction of the blood flow can also be determined. Another fascinating example is given in Figure 135. OCT is commonly used in ophthalmology; OCT is also being researched for applications in dermatology. *Endoscopic* OCT, i.e., performing OCT through a small catheter inserted into the human body, will become an important tool in oncology and cardiology in the near future. OCT is also being used in material research to image turbid media or to produce topographic profiles.

Page 196

An unusual imaging method is *muon tomography*, an imaging method that uses the muons in cosmic rays to detect heavy metals in boxes, luggage and trucks. This method is particularly interesting for searching for hidden heavy metals, such as plutonium, which scatter muons much more strongly than other materials such as iron.

THE EYE AND THE BRAIN: BIOLOGICAL IMAGE ACQUISITION AND PROCESSING

Image processing systems acquire images and then exctract information from them. In technical image processing systems, the acquisition occurs with a camera and the extraction is realized with software running on a computer. An interesting image processing system is built into each of us: the combination of eye and brain. The eye and the brain are involved devices. We start by exploring the construction and performance of our eyes.

Do we see what exists?

Sometimes we see *less* than there is. Close your left eye, look at the white spot in Figure 136, bring the page slowly towards your right eye, and pay attention to the middle lines. At a distance of about 15 to 20 cm the middle line will seem uninterrupted. Why?

Challenge 182 s

Look with one eye at a full computer screen that is blinking blue and black, at a rate of once or twice a second. Now look at the same blinking screen through a blue filter (a Balzers K45 or a Kodak BG12 filter). You will see a spot. Why?

Ref. 135

Challenge 183 s

Sometimes we see *more* than there is, as Figures 137 and 138 show. The first figure shows that parallel lines can look skewed, and the second show a so-called *Hermann lattice*, named after its discoverer.* The Hermann lattice of Figure 138, discovered by Elke Lingelbach in 1995, is especially striking. Variations of these lattices are now used to understand the mechanisms at the basis of human vision. For example, they can be used to determine how many light sensitive cells in the retina are united to one signal pathway towards the brain. The illusions are angle dependent because this number is also angle dependent.

Ref. 136

---

* Ludimar Hermann (b. 1838 Berlin, d. 1914 Königsberg) was an important physiologist. The lattices are

FIGURE 136 A limitation of the eye (see text).

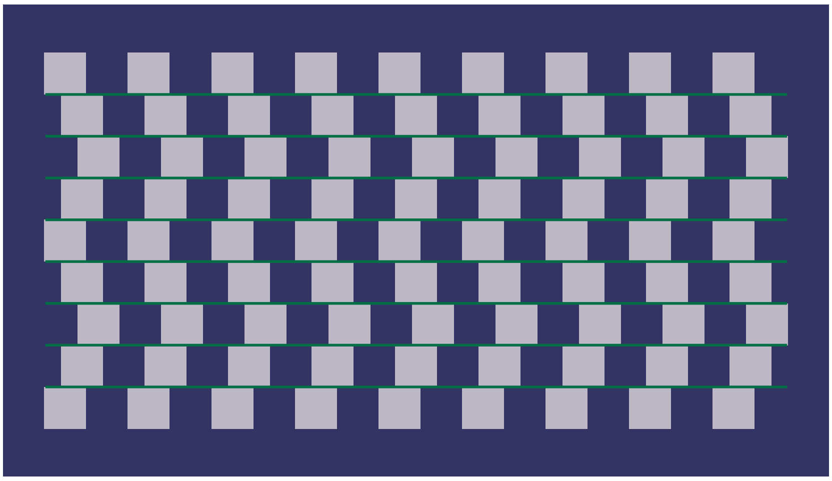


FIGURE 137 What is the angle between the thin lines between the squares?

Our eyes also 'see' things *differently:* the retina sees an *inverted* image of the world. There is a simple method to show this, due to Helmholtz.* You need only a needle and a piece of paper, e.g. this page of text. Use the needle to make two holes inside the two letters 'oo'. Then keep the page as close to your eye as possible, look through the two holes towards the wall, keeping the needle vertical, a few centimetres behind the paper. You will see two images of the needle. If you now cover the *left* hole with your finger, the

---

often falsely called 'Hering lattices' after the man who made Hermann's discovery famous.

* See Hermann von Helmholtz, *Handbuch der physiologischen Optik*, 1867. This famous classic is available in English as *Handbook of Physiological Optics*, Dover, 1962. Physician, physicist and science politician, born as Hermann Helmholtz (b. 1821 Potsdam, d. 1894 Charlottenburg), was famous for his works on optics, acoustics, electrodynamics, thermodynamics, epistemology and geometry. He founded several physics institutions across Germany. He was one of the first to propagate the idea of conservation of energy. His other important book, *Die Lehre von den Tonempfindungen*, published in 1863, describes the basis of acoustics and, like the Handbook, is still worth reading.

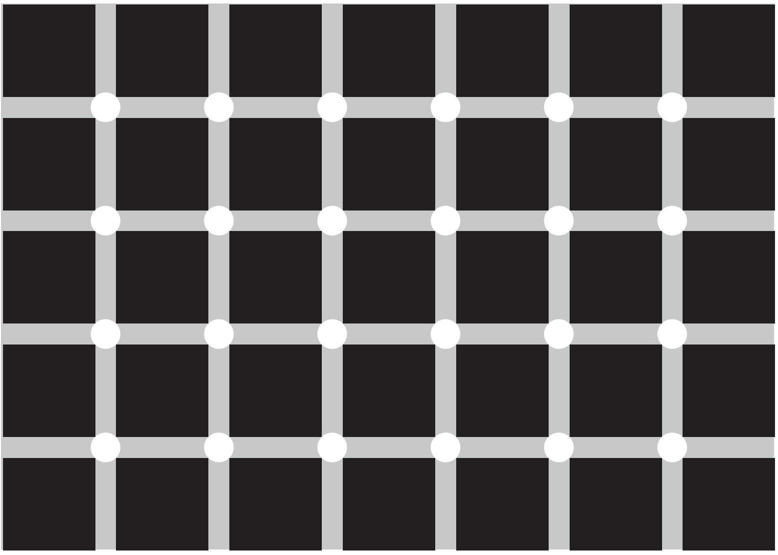


FIGURE 138 The Lingelbach lattice: do you see white, grey, or black dots at the crossings?

FIGURE 139 An example of an infrared photograph, slightly mixed with a colour image (© Serge Augustin).

right needle will disappear, and vice versa. This shows that the image inside the eye, on the retina, is inverted. Are you able to complete the proof? Challenge 184 ny

An urban legend, spread by many medical doctors and midwives to this day, claims that newborn babies see everything upside down. Can you explain why this idea is wrong? Challenge 185 s

Two additional experiments can show that retinas acquire inverted images. If you push very lightly on the *inside* of your eye (careful!), you will see a dark spot appear on the *outside* of your vision field. And if you stand in a dark room and ask a friend to look at a burning candle, explore his eye: you will see three reflections: two upright ones, reflected from the cornea and from the lens, and a dim third one, *upside-down*, reflected

FIGURE 140 How the appearance of a sunflower changes with wavelength: how it looks to the human eye, how it might look to a bird, and how it looks in the ultraviolet (© Andrew Davidhazy).

from the retina.

Our eyes do not produce a faithful image of nature: they have a limited wavelength sensitivity. This sensitivity peaks around 560 nm; outside the red and the violet, our eyes does not detect radiation. We thus see only part of nature. As a result, infrared photographs of nature, such as the one shown in Figure 139, are interesting because they show us something different from what we see usually. The same happens to ultraviolet photographs, as shown in Figure 140. Also images of the sky differ with wavelength; the website www.chromoscope.net shows this in detail. Ref. 137

The eye sees most sharply in the region of the fovea. But the highest light sensitivity is not in that region. As a result, we often do not see faint stars at night when we look directly at them, but see them when we look *next* to them. This effect is due to the peculiar distribution of rods, which has its peak density 20°away from the axis of sharpest vision.

Several other optical illusions are found throughout this text. In summary, we have to be careful whenever we maintain that seeing means observing. Our sense of vision is limited. Are there other limitations of our senses which are less evident? Our adventure will indeed uncover several of them. But let us now turn to see what the eye *can* do.

The human eye

The eye is the part that moves most frequently in the human body – more than the heart. It is estimated that the eye performs 200 million saccades every year. Therefore the motion and lubrication mechanisms of the eye are especially involved. Eye movements exist in various types: apart from saccades, the eye shows pursuit movements, motions that compensate head rotation, called the vestibulo-ocular reflex, and ocular microtremor.

The human eye is a so-called *camera eye*. Like a photographic camera, and in contrast to insect eyes and other compound eyes, the vertebrate camera eye works by producing an image of the outer world on a surface consisting of light sensors, the *retina*. The retina covers more than half of the inside of the eye ball, whose typical diameter in an adult is about 16.7 mm. The pupil has a diameter between 2 mm – below which one gets problems with diffraction – and 7 mm – for which lens aberrations are just acceptable.

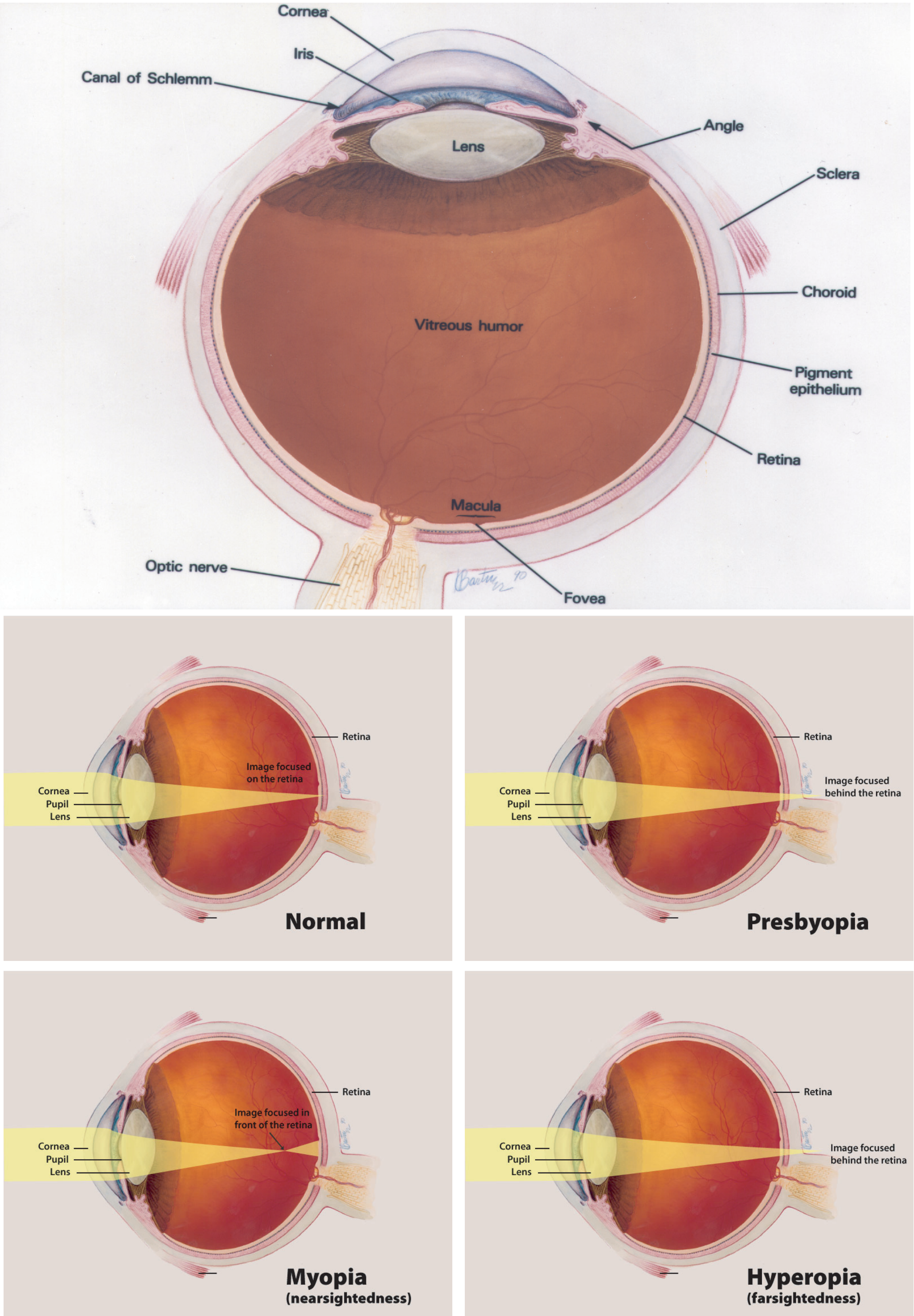


FIGURE 141 Top: a simplified cross section of the human eye; bottom: the comparison of the optical imaging for a healthy eye and for the most common eye problems, myopia, hyperopia and presbyopia (© NEI at NIH).

The image on the retina has low image distortion, low chromatic aberrations (about 1 dioptre between red and blue) and low coma; the eye achieves this performance by using an deformable aspheric gradient-index lens and a cornea whose shape is always near the ideal shape within 30 μm – an extremely good value for a deformable body. The eye, together with the brain, also has a powerful autofocus – still not fully understood – and an excellent motion compensation and image stabilization system built in. A section of this amazing device is shown in Figure 141.

The retina is an outgrowth of the brain. It contains 120 million *rods*, or black and white pixels, and 6 million *cones*, or colour pixels. Each pixel can detect around 300 to 500 intensity levels (9 bit). The eye works over an intensity range of 8 to 10 orders of magnitude; the involved mechanism is incredibly complex, takes place already inside the receptors, involves calcium ions, and is fully known only since a few years. The region of highest resolution, the *fovea*, has an angular size of about 1°. The resolution of the eye is about 1′. The integration time of the retina is about 100 ms – despite this vale, no artefacts are noticed during the saccades. The retina itself is 200 μm thick and is transparent: this means that all cables leading to the receptors are transparent as well.

The retina has very low energy consumption and uses a different type of neurons than usual nerves: the neurons in the retina use *electrotonic potentials*, not the action potentials or spikes used in most other nerves, which would generate interferences that would make seeing impossible. In the fovea, every pixel has a connection to the brain. At the borders of the retina, around 10 000 pixels are combined to one signal channel. (If all pixels were connected 1 to 1 to the brain, the brain would need to be as large as a typical classroom.) As a result, the signals of the fovea, whose area is only about 0.3 % of the retina, use about 50 % of the processing in the brain's cortex. To avoid chromatic aberrations, the fovea has no blue receptors. The retina is also a graphic preprocessor: it contains three neuronal layers that end up as 1.3 million channels to the cortex, where they feed 5 million axons that in turn connect to 500 million neurons.

The compression methods between the 125 million pixel in the retina and the 1.3 million channels to the cortex is still subject of research. It is known that the signals do not transport pixel data, but data streams processed in about a dozen different ways. The streams do not carry brightness values, but only contrasts, and they do not transmit RGB values, but colour differences. The streams carry motion signals in a compressed way and the spatial frequency data is simplified. Explorations have shown how the ganglions in the retina provide a navigational horizon, how they detect objects moving against the background of the visual field, and how they subtract the motion of the head. The coming years and decades will provide many additional results; several data channels between the eye and the brain are still unknown.

Ref. 138

Apart from rods and cones, human eyes also contain a third type of receptor. This receptor type, the *photosensitive ganglion cell* or *intrinsically photosensitive retinal ganglion cell*, has only been discovered in the early 1990s, sparking a whole new research field. Photosensitive ganglion cells are sensitive mainly to blue light, use melanopsin as photopigment and are extremely slow. They are connected to the suprachiasmatic nucleus in the brain, a small structure of the size of a grain of rice that controls our circadian hormone cycle. For this reason you should walk a lot outside, where a lot of blue light is available, in order to reset the body's clock and get rid of jet-lag. Photosensitive ganglion cells also produce the signals that control the diameter of the pupil.

Ref. 139

It is worth recalling that drawings such as the one of Figure 141 are *simplified*. They do not show the structures in the transparent part of the eye, the vitreous body, such as the hyaloid canal, which plays an important role during the growth of the eye in the embryo stage. In fact, the growth of the eye inside the womb is even more amazing than its actual function – but this story is outside the scope of this text.

HUMAN VERSUS OTHER EYES

The human eye and many other animal eyes are better devices than most modern photographic or video cameras. Not only does it have more pixels than most cameras, it is also insensitive to pixel errors, to the blood vessels in front of the sensors. No camera covers the same range of intensity variation. No human-made camera has a lens system of comparable quality or capabilities: the large viewing angle, the low field distortions – also due to the spherical shape of the retina – and the low chromatic aberrations. No technical autofocus system, image stabilizer or motion compensation system matches that of the eye.

One limitation of the eye is its speed. The human eye produces an effective number of 30 images per second and up to 120 images per second under the most ideal conditions; dogs and birds achieve twice the basic rate and insects about ten times as much. Modern video cameras can produce more than 10 000 images per second. When developing the eye, evolution has traded speed for resolution. To achieve high resolution, the eye continuously performs small movements, called *micronystagmus*. In detail, the eye continuously oscillates around the direction of vision with around 40 to 50 Hz; it constantly averages an image pixel over 30 to 50 receptors, but the exact sharpening mechanism is not clear yet. This motion increases the effective number of pixels, avoids issues with dead pixels and also allows the rods and cones to recharge.

All vertebrate eyes have *rods*, the pixel types that produce black and white images at night. Additionally, the retina of the human eye contains three types of cones, for the colours red, green and blue. As mentioned, much better eyes are found in birds, many reptiles and fish: they have four or more types of cones, built-in colour filters and an ultraviolet-transparent lens. The fourth type of cones and the special eye lens make the eyes of birds and reptiles sensitive to near-ultraviolet light; birds use their ultraviolet sense to find food and to distinguish males from females. Indeed, most birds whose males and females look the same to humans differ markedly in the ultraviolet.

Birds and reptiles also have coloured oil droplets built into the top of their cones, with each cone type containing a different oil colour. These droplets act as colour filters. In this way, the spectral resolution of their cones is much sharper than in mammals. The sense of colour in birds is much more evolved than in humans – it would be fascinating to watch the world with a bird's eye. Birds are the best colour seers overall. They have cone receptors for red, blue, green, ultraviolet, and, depending on the bird, for up to three more sets of colours.

Eagles and a number of other birds (but not many) also have a better eye resolution than humans. They achieve this in two ways. First, their photoreceptors are small; in other words, their pixel size is the smallest known with respect to the eye diameter, with only 1.6 μm. Secondly, the eye includes bones. These bones fix the relative position of lens and retina, like a rigid camera body. With these technical solutions, the eye of the eagle is clearly better than that of humans.

In the course of evolution, the eye of mammals lost two types of cones that were part of the vertebrate heritage, and were left with only two types of cones. The (Old-World) primates later regained one type, in order to distinguish more clearly tree fruit, which are so important as food for the primate brain, from the surrounding leaves. But despite this change, primates never reached the capability of the best bird's eyes. Thus, of all mammals, only primates can see full *colours* as human do. Bulls for example, don't; they cannot distinguish red from blue.

Usual humans are thus *trichromatic*: they have three types of cones that detect red, green and blue. However, around 1 % of women are (somewhat) *tetrachromatic*. This is possible because humans can have two different red pigments. The red pigment details are encoded on the X chromosome. Now, in some women, the two X chromosomes code for two different red pigments. In a part of these women, both pigments are found in the cones of their eyes. These women thus seem to have something like RR'GB eyes. Ref. 140 Tests showed that they can distinguish more red shades than men and than most other women.

Every expert of motion should also know that the highest sensitivity of the human eye Ref. 141 does *not* correspond to the brightest part of sunlight. This myth has been spread around the world by the numerous textbooks that have copied from each other. Depending on whether frequency or wavelength or wavelength logarithm is used, the solar spectrum peaks at 500 nm, 880 nm or 720 nm. The human eye's spectral sensitivity, like the completely different sensitivity of birds or frogs, is due to the chemicals used for detection. In short, the human eye can only be understood by a careful analysis of its particular evolutionary history.

Camera eyes are found in all vertebrates. Mammals have eyes similar to ours, with a flexible lens; in contrast, snakes have eyes with rigid lenses that are moved with respect to the retina in order to put images into focus. Camera eyes evolved independently several times in other animal groups. Most known are the cephalopods, such as the octopus, and indeed, the largest eyes known, up to 30 cm in diameter, are from animals of this group. Camera eyes are also found in some spiders, in snails and in a number of other groups.

By the way, the human eye–brain system processes colours mainly around the direction of gaze. This allows a fun trick: if a vision system follows the direction of your gaze, it can command a computer display to show colours only in the display region at which you are looking to, and to leave the rest of the picture in black and white. If the command system is fast enough, you get the impression that the whole picture is coloured, whereas every bystander sees that the picture is mainly black and white, and just shows colours in a spot that is constantly moving around.

The most common eyes in nature are not camera eyes, but *compound eyes*, as found in bees, dragonflies or house flies. Compound eyes have one lens for each axon. These units are usually hexagonal in shape are called *ommatidia* and typically contain a handful of photoreceptors that are connected to the outgoing axon. An ommatidium is a tiny eye; depending on the species, a compound eyes consist of at least a hundred and at most 30 000 ommatidia (for some dragonflies). Many compound eyes are also tetra- or pentachromatic. Compound eyes have low resolution – it is suspected that no insect can see the stars – but such eyes have a number of advantages. Compound eyes need no

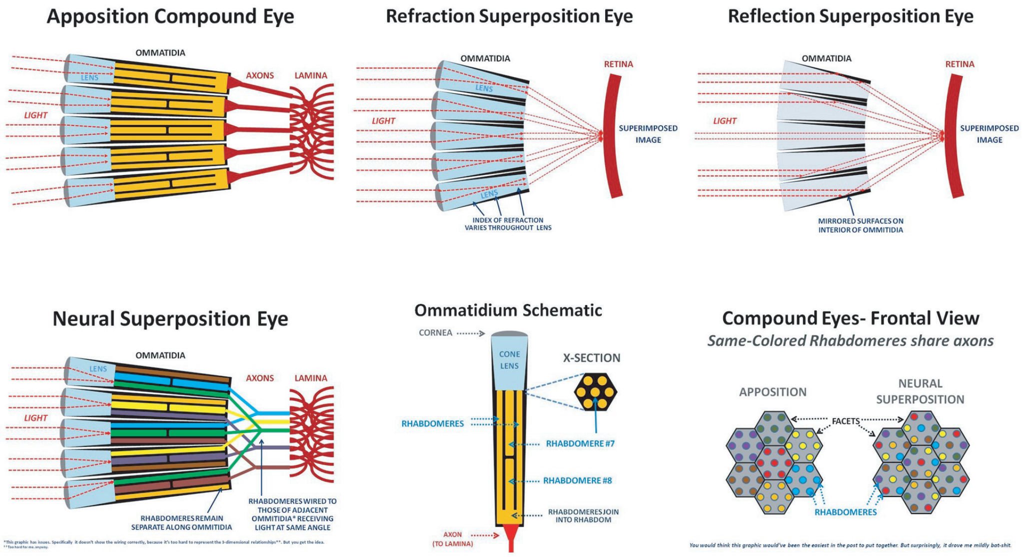


FIGURE 142 Compound eyes: the apposition compound eye found in bees and dragonflies, the refraction superposition eye of moths, the reflection superposition eye of lobsters, (not shown: the parabolic superposition eye of certain crabs) and the neural superposition eye of the house fly (© Watcher, from watchingtheworldwakeup.blogspot.com).

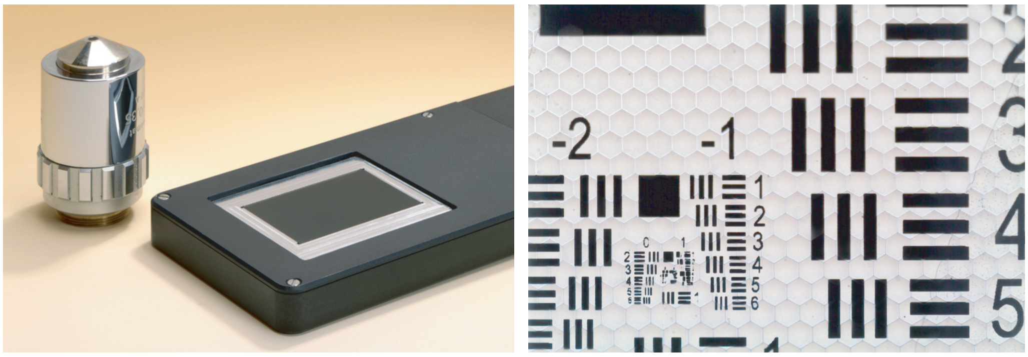


FIGURE 143 A flat microscope based on stacked microlens arrays – in front of a conventional objective – and an image it produces (© Frank Wippermann).

focussing mechanism, can cover a large field of view, and above all, they are extremely fast. These advantages are so interesting that compound-eye-style electronic cameras are also being explored as alternatives to usual, one-lens-plus-one-sensor cameras.

Using ideas from insect eyes is also interesting for other uses. For example, modern technology provides the possibilities to think anew how a microscope should look like. Figure 143 shows a microscope that is in fact an array of thousands of tiny microscopes. The lenses produce images on a CMOS imaging chip with 16 megapixel. Ref. 142

In summary, the microscopic structures inside the eye are important and fascinating. But here we face a question.

HOW CAN WE MAKE PICTURES OF THE INSIDE OF THE EYE?

When we look through a small hole towards a bright surface, we can see the blood vessels in our eye. In particular, we can see that the fovea has no blood vessels at all. But how can we observe other people's microscopic eye structure?

Imaging the details inside of a living eye is not easy. The retina is far away from the surface of the eye, so that a normal microscope cannot be used. In addition, the continuous motions of the lens and of the eye itself disturb any imaging system. Finally, two separate developments changed the situation in the 1990s.

Page 187 The first breakthrough in eye imaging was the technique, mentioned above, of *optical coherence tomography*. This imaging method uses a scanned low-power laser beam and allows imaging scattering media up to a depth of a few millimetres with a resolution of the order of a few µm. This microscopy technique, developed in the 1990s, allows observing in detail the retina of the human eye and the region below it; it also allows cross sections of the cornea and the lens. Through the detailed pictures it provides in a few milliseconds, shown in Figure 144, optical coherence tomography allows extremely precise diagnoses; it has profoundly changed modern ophthalmology. The fascinating pictures from the research group on optical coherence tomography at the University of Vienna are shown in Figure 145.

Optical coherence tomography also allows imaging the skin to a depth of about 8 mm; this is already improving skin cancer diagnosis. In the future, the technique will also simplify cancer diagnosis for gynaecologists and otolaryngologists. Endoscopic systems are also being developed. Optical coherence tomography is becoming standard also in various industrial applications.

The second breakthrough in eye imaging was the technique of *adaptive optics*, a technique, also used in astronomy, that continuously and quickly changes the shape of the imaging lens. The most beautiful pictures so far of a *living* human retina, such as that of Figure 146, were made by the group of David Williams and Austin Roorda at the University at Rochester in New York using this modern technique. They used adaptive optics in order to compensate for the shape variations of the lens in the eye of the patient. Ref. 143

The human eye produces the sensation of colour by averaging the intensity arriving at the red, blue and green sensitive cones. This explains the possibility, mentioned above, of getting the same impression of colour, e.g. yellow, either by a pure yellow laser beam, or by a suitable mixture of red and green light. Page 125

But if the light is focused on to one cone only, the eye makes mistakes. Using adaptive optics it is possible to focus a red laser beam such that it hits a green cone only. In this case, something strange happens: even though the light is *red*, the eye sees *green* colour!

Incidentally, Figure 146 is quite puzzling. In the human eye, as in all vertebrate eyes, the blood vessels are located *in front* of the cones. Why don't they appear in the picture? And why don't they disturb us in everyday life? (The picture does not show the other type of sensitive light cells, the *rods*, because the subject was in daylight; rods come to the front of the retina only in the dark, and then produce black and white pictures.) Challenge 186 s

In 2016, the technique of *optical coherence tomography* (OCT) allowed to make an even more astonishing measurement, shown in Figure 147. Observing the retina of a living human allows seeing what a person is watching. The exact details for this possibility are not yet understood; somehow, illuminated photoreceptors have a different optical Ref. 144

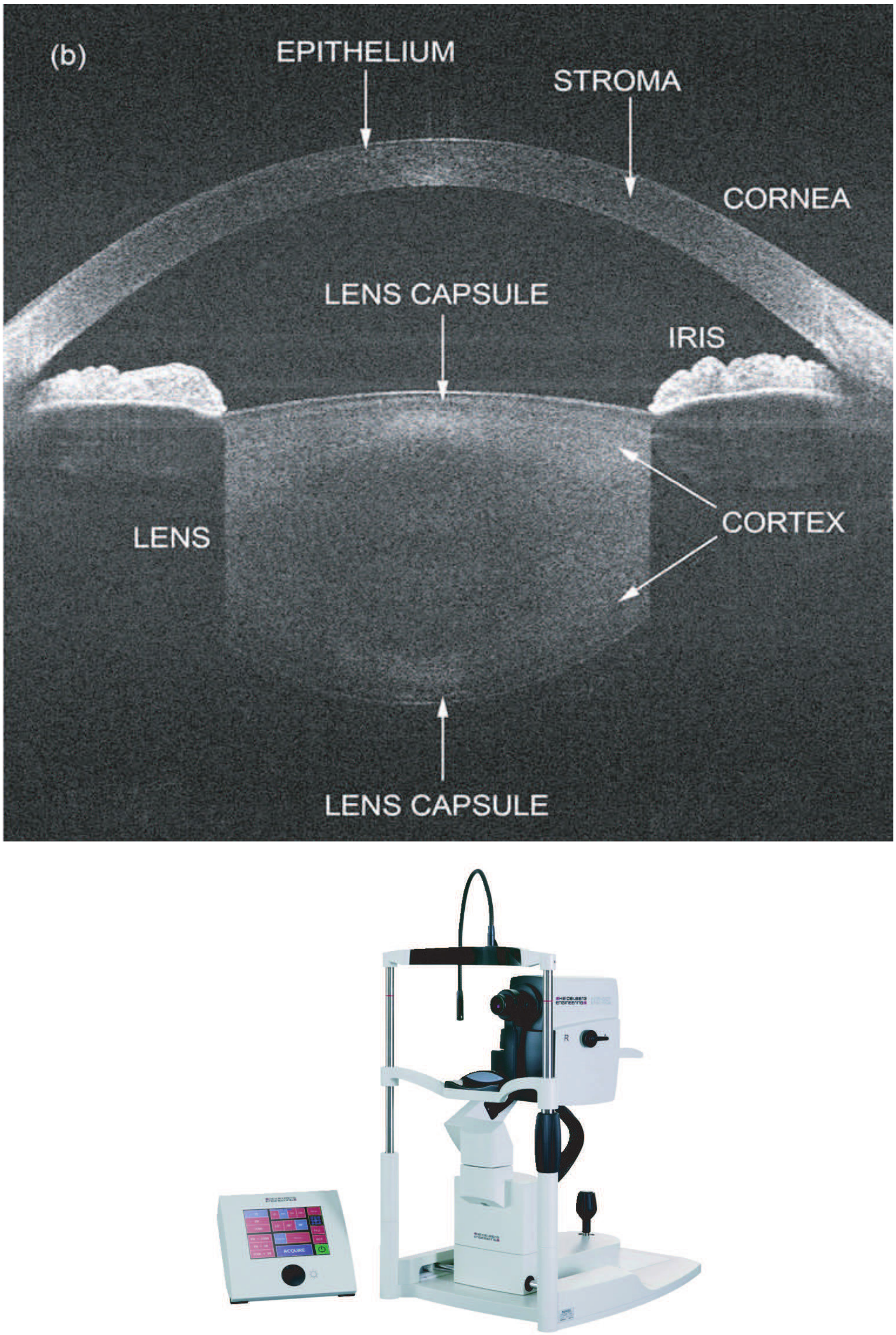


FIGURE 144 Top: an image of the front of the human eye acquired by optical coherence tomography, showing the cornea, the iris and the lens. Bottom: a typical apparatus used by ophthalmologues. (© www.zmpbmt.meduniwien.ac.at/forschung/optical-imaging/advanced-imaging-technologies/, Heidelberg Engineering)

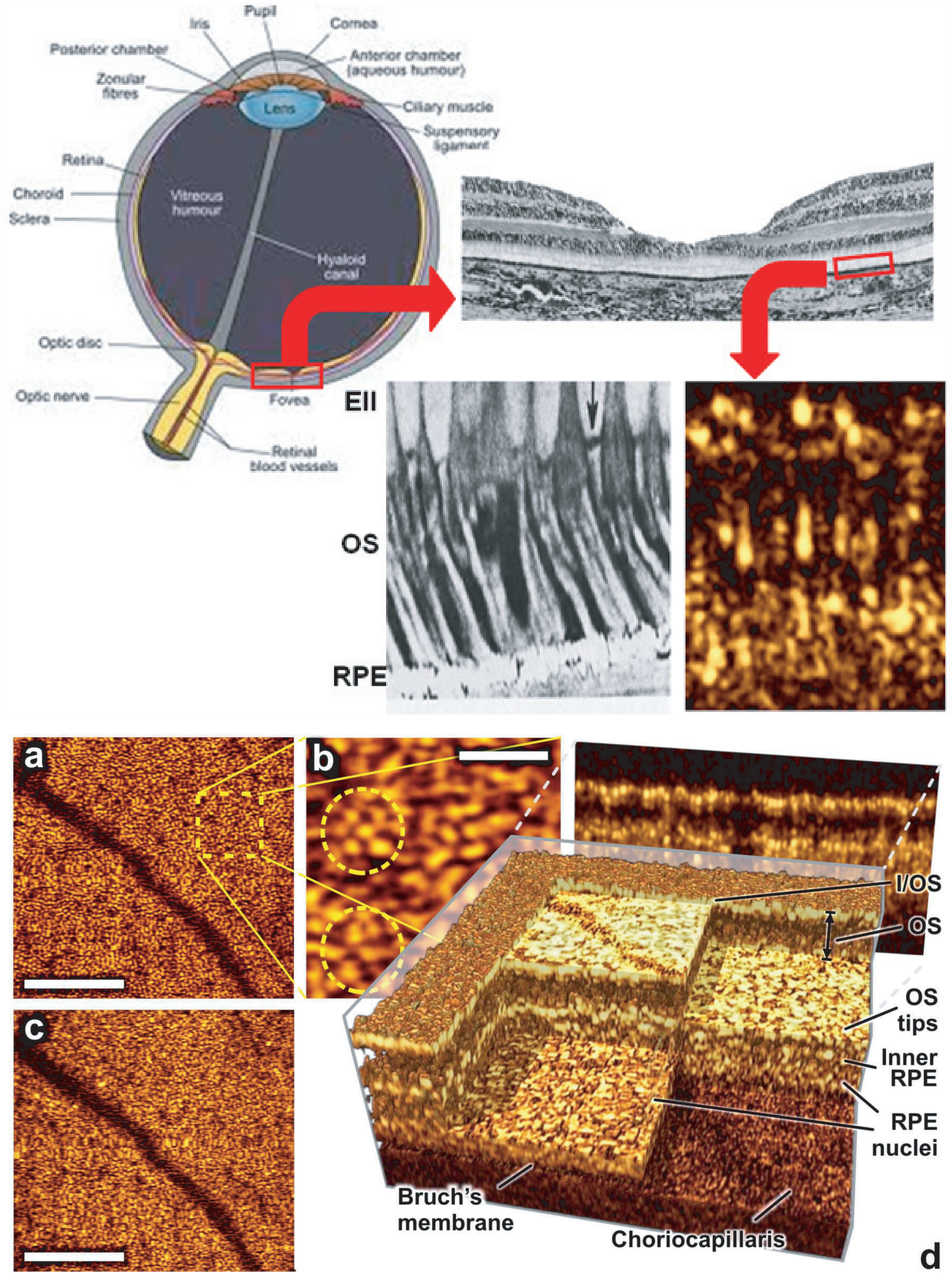


FIGURE 145 Images of the live human retina taken with adaptive optics optical coherence tomography. Top: Cross section of the human eye indicating a special region of the retina, the fovea, at the back of the eye; histology of this area indicating the outer segment (OS) of the photoreceptor cells; enlarged histology of the OS; in vivo cellular resolution OCT of living photoreceptor cells; Ell indicates the ellipsoid of photoreceptors; RPE the retinal pigment epithelium. Bottom: OCT tomograms of the inner/outer junction of human photoreecptors (a), their outer segment tips (c) with enlarged field of view (b). The bright spots in the dashed circles indicate single photoreceptors cells. The representation (d) at different depths reveals intraretinal microstructures at cellular resolution. (© www.zmpbmt.meduniwien.ac.at/forschung/optical-imaging/advanced-imaging-technologies/)

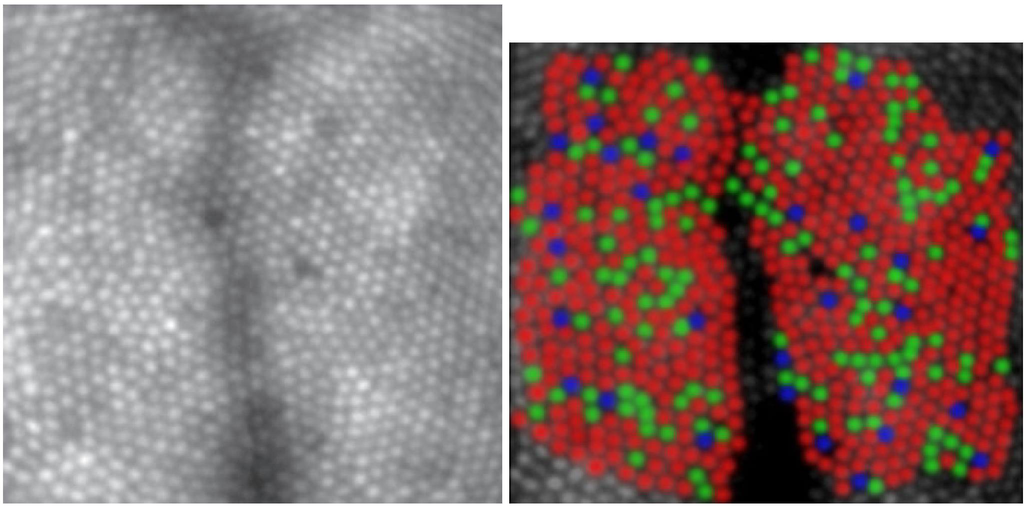


FIGURE 146 Left: a high quality photograph of a living human retina taken with adaptive optics; right: same image with a superimposed measured indication of the sensitivity of each cone cell (© Austin Roorda).

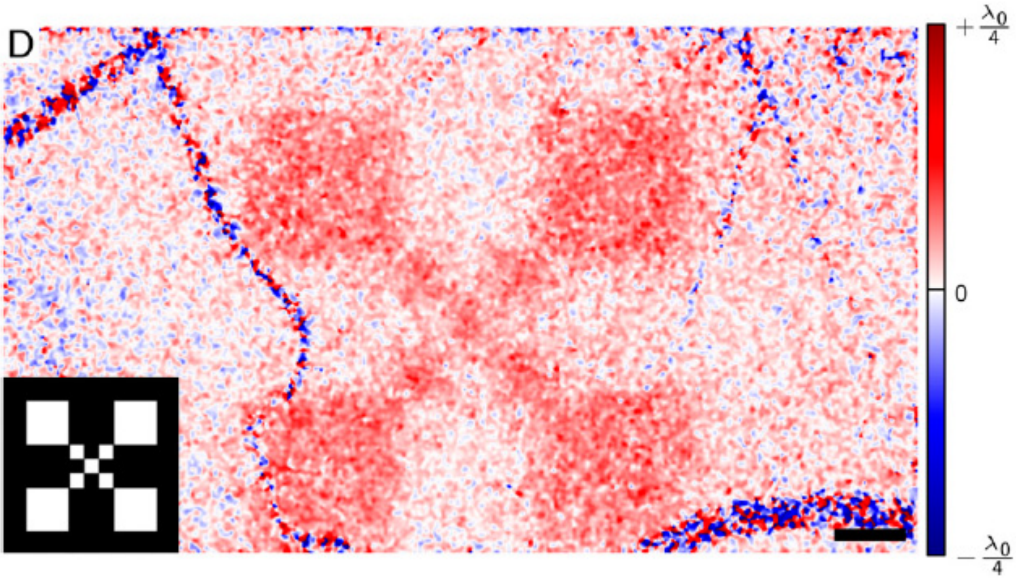


FIGURE 147 Using optical coherence tomography to image the live retina 247 ms after the eye stopped watching at the pattern in the lower left. The afterimage on the retina can be observed (© PNAS).

path length than receptors in the dark.

Ref. 145 In summary, evolution has provided us with an observations system that has amazing properties. Take good care of your eyes.

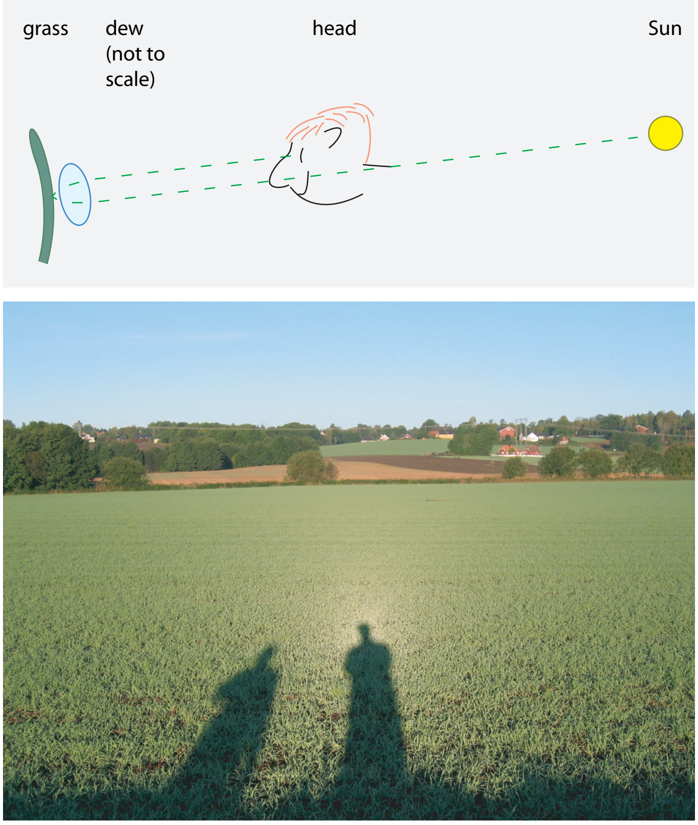


FIGURE 148 The path of light for the dew on grass that is responsible for the aureole or Heiligenschein, and a photo showing that it is seen only around one's own head (© Bernt Rostad).

HOW TO PROVE YOU'RE HOLY

Light reflection and refraction are responsible for many striking effects. The originally Indian symbol for a holy person, now used throughout most of the world, is the *aureole*, also called *halo* or *Heiligenschein*: a ring of light surrounding the head. You can easily observe it around your own head. You need only to get up early in the morning and look into the wet grass while turning your back to the Sun. You will see an aureole around your shadow. The effect is due to the morning dew on the grass, which reflects the light back predominantly in the direction of the light source, as shown in Figure 148. The fun Ref. 146 part is that if you do this in a group, you will see the aureole around only *your own* head.

Retroreflective paint works in the same way: it contains tiny glass spheres that play the role of the dew. A large surface of retroreflective paint, a traffic sign for example, can Ref. 147 also show your halo if the light source is sufficiently far away. Also the so-called 'glow' of the eyes of a cat at night is due to the same effect; it is visible only if you look at the Challenge 187 s cat with a light source behind you. By the way, do Cat's-eyes work like a cat's eyes?

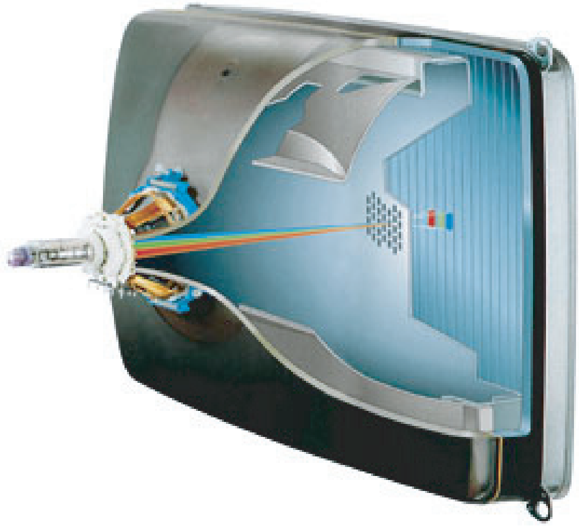


FIGURE 149 A cathode ray tube in older televisions: the first way – now obsolete – to produce changing colour images using electric signals. Television tubes emit an electron beam, deflect it, and generate light by electroluminescence on a coloured screen covered with patterned phosphors.

DISPLAYING IMAGES

Systems that display images are of importance in technical devices and, to a smaller degree, in nature. In nature, these displays are of two types: The first type is used by squids living in shallow water: they are able to produce moving colour patterns on their skin, and they use these patterns to confuse prey. The second type is found in the deep sea, where there is no ambient light: there, many living beings produce moving light displays to attract prey or to confuse predators.

In short, images can be generated by changing surface colours – passive displays – or by emitting light. Also human-made systems can divided into these two classes.

At present, the most common passive displays are liquid crystal displays – or LCDs – and electronic ink displays. The former are used in watches and mobile phones, the latter in electronic book readers.

The most common light emitting displays are the dated cathode ray tube, plasma displays, the light emitting diode displays and projection displays. These displays are used mostly in entertainment devices.

Hopping electrons and the biggest disappointment of the television industry

It is well known that when an electric field in a vacuum points along a glass surface, electrons can *hop* along the glass surface. The general effect is shown in Figure 150; usually, the effect is unwelcome. Among others, the hopping effect is responsible for sparks in vacuum systems that contain high voltage. To avoid the effect, the glass insulators on high voltage lines have complex shapes.

When this effect was studied in more detail, it turned out that reasonably low electric fields are sufficient to create sizeable electric hopping currents in hollow glass tubes with an internal diameter around a millimetre. The low elctric field can also lead elec-

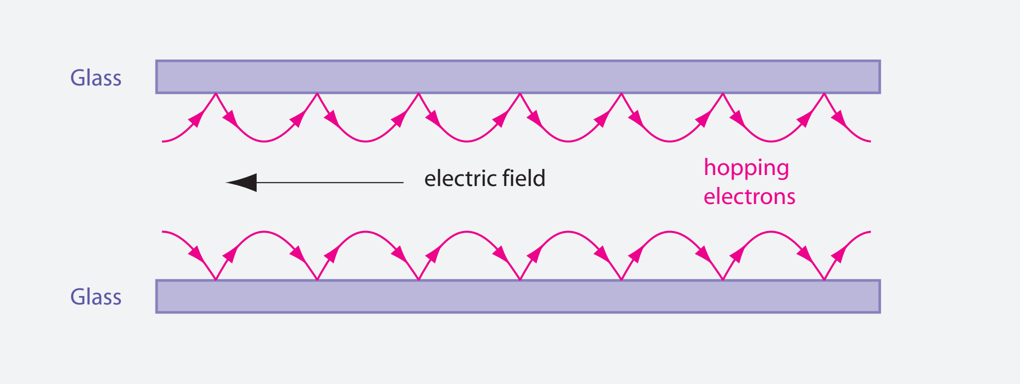


FIGURE 150 Free electrons can hop along a glass wall.

tron around bends and corners. Furthermore, electric switches that change the hopping direction can be constructed. In short, the hopping effect can be used to make extremely cheap flat television displays of high image quality. The idea is to put an array of electron sources – essentially sharp metal tips – at the start of many closeby glass channels. Each channel transports the emitted electrons along a line of the display. Making use of suitable switches at each pixel, the electrons are made to hit phosphorescent colour emitters. These are the same pixels that were used in the then common – bulky and heavy – television tubes and that are used in flat plasma displays. Since the hopping effect also works around bends and corners, and since it only needs glass and a bit of metal, the whole system can be made extremely thin, and lightweight; moreover, the machines are cheap, the yield is high and the production cost is low. Already in the early 1990s, the laboratory samples of the electron hopping displays were spectacularly good: the small displays were brighter, sharper and cheaper than liquid crystal displays, and the large ones brighter, sharper and cheaper than plasma displays. Affordable flat television was on the horizon. Ref. 148

Then came the disappointment. The lifetime of the displays was only of the order of a few hundred hours. The limitation was due to the necessity to use helium inside the display, which cannot be contained inside a vacuum system for a long time. Despite the most intense material research, achieving a higher lifetime turned out to be impossible. All tricks that were tried did not help. Despite all their fantastic properties, despite huge investments in the technology, despite the best material researchers working on the issue, electron hopping displays could not be brought to market. Not a single display was ever sold.

CHALLENGES AND FUN CURIOSITIES ABOUT IMAGES AND THE EYE

An image sensor does not need a lens. The temple viper (or Wagler's pit viper) has two infrared sensors – one is shown in Figure 151 – with a resolution of 40 times 40 pixels each, and it just has a hole instead of a lens. The pit viper uses these sensors to catch mice even in the dark. The working of this infrared sensor has been explored and simulated by several research groups. It is now known how the sensor acquires the data, how the snake brain reconstructs the image, and how it achieves the high resolution. Ref. 149

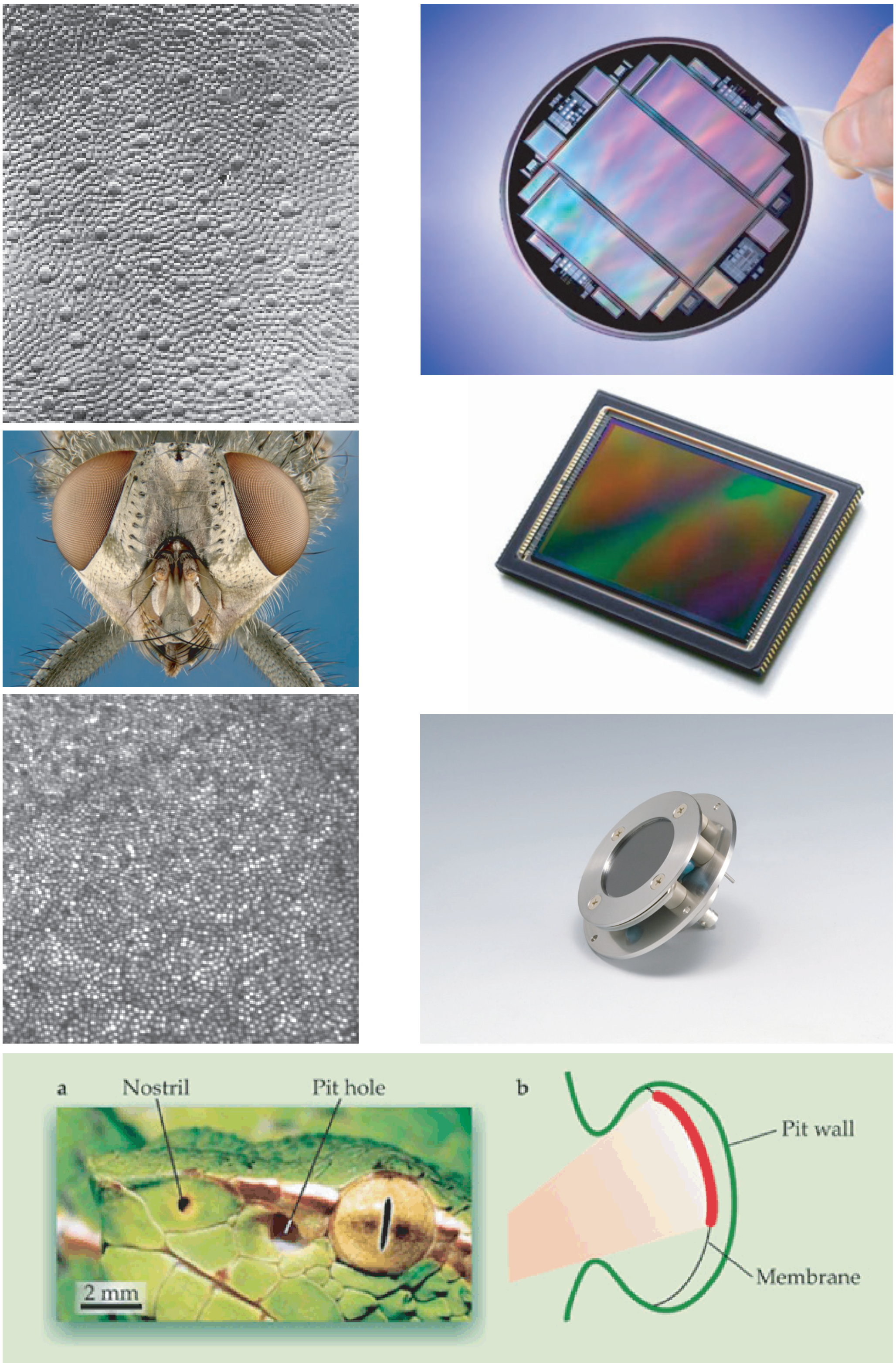


FIGURE 151 A collection of image sensors – thus of pixel systems: A cat's retina, a CCD sensor still on a wafer, the eye of a house fly, a CMOS sensor, a human retina, a multichannel plate, and a temple viper's infrared pit (© Wikimedia, Austin Roorda, Hamamatsu Photonics, Guido Westhoff/Leo van Hemmen).

* *

The simplest imaging system are eye glasses. A child that has no proper glasses misses an important experience: seeing the stars. Such a child will not understand the famous statement by Immanuel Kant: 'Two things fill the mind with ever new and increasing admiration and awe, the more often and persistently thought considers them: the starred sky above me and the moral law inside me.' Always be sure that children can see the stars.

Two lenses of 40 cents each are sufficient to change the life of a child or that of an adult. See the website www.onedollarglasses.org for an effective way to do it across the world.

* *

Among the most impressive nature photographs are those found on www.microsculpture.net; they show beetles to an extremely high resolution. Each beetle photograph is a composition of many thousands of usual high-resolution photographs. They provide a stunning sight – enjoy it. Challenge 188 e

* *

Challenge 189 ny How does the eye correct pixel (photoreceptor) failure? How many pixels are bad in a typical eye?

* *

Infrared light can be seen, if it is of sufficient intensity. (Never try this yourself!) People who observed such light sources – semiconductor lasers, for example – saw it as a white spot with some red borders. In other cases, it is also possible to see short infrared pulses through double-photon absorption in the retina; in this way, infrared of 1000 nm produces a green flash inside the eye.

* *

Among vertebrates, the largest eye is the eye of the blue whale; it has a diameter of 150 mm. (Only squids have larger eyes.) The smallest vertebrate eye seems to be that of juvenile *Brookesia micra*, a small chameleon whose head is half the size of the head of a match and whose eye is around 0.3 mm in diameter. The eye is a wonderful organ. To learn more about it, read the beautiful book Simon Ings, *The Eye – A Natural History*, Bloomsbury, 2007.

* *

In many applications, it is important to avoid reflections. Anti-reflection coatings are used on the glass of shop-windows and in lens systems that need to work in dim conditions, when light is scarce. Such coatings usually are interference coatings made of various layers of transparent materials deposited on the surface. Also living beings have anti-reflection coatings; the eyes of moths are famous for appearing black also in bright daylight. They are black because they do not reflect any light, and thus keep the moths hidden from their predators. However, moth eyes use a different effect to avoid reflections: their surface is covered with a hexagonal pattern of pillars of about 200 nm height. A similar effect is achieved by the glasswing butterfly, *Greta oto*, whose wings are as transparent as glass, but without any reflections. Various companies are trying to reproduce this so-called *moth-eye effect* in commercial applications, for example to improve photovoltaic cells.

* *

Ref. 150 Modern technology allows producing microscopes at low cost. For a fascinating example, see the 1 Euro microscope that can be folded from a sheet of paper, embedded with some additional devices, and shown in Figure 152. The device is used by holding it in front of the eye or by holding it in front of a lamp and observing the projected image on a screen.

* *

If a sufficient number of images is available, it is possible to identify the camera that produced them. Every camera has a specific image noise pattern; by extracting it through clever averaging, computer software that processes camera images is able to support police investigations.

* *

Ref. 151 Mirages often have surprising effects. In 1597, a group of sailors were stranded on Novaya Zemlya during the winter. On 24 January they saw the Sun – roughly two weeks before it should be visible there. Such an unusual sighting of the Sun is now called a *Novaya Zemlya effect*.

* *

Challenge 190 s It is possible to measure the width of a hair with a laser pointer. How?

* *

Ref. 152 Modern imaging techniques allow high sensitivity and high spatial resolution. As shown in Figure 153, using a Fresnel lens, a cooled CCD sensor and a laser as a light source, it is even possible to photograph the shadow of a single floating ion.

* *

An important device in medicine is the *endoscope*. An endoscope, shown in Figure 154, allows looking into a body cavity through a very small hole. It is a metal tube, typically with a diameter of around 5 mm and a length of 300 mm. How would you build one? Challenge 191 e (The device must resist at least 150 disinfection cycles in an autoclave; each cycle implies staying at 134°C and 3 bar for three hours.) Made of a sequence of carefully designed cylinder lenses, endoscopes allow surgeons to watch the inside of a human body through a tiny hole, thus avoiding large cuts and dangerous operations. Endoscopes have saved many lives, and their production and development employs a large industry.

* *

Challenge 192 s The Sun is visible to the naked eye only up to a distance of 50 light years. Is this true?

* *

Ref. 153 Challenge 193 s Grass is usually greener on the other side of the fence. Can you give an explanation based on observations for this statement?

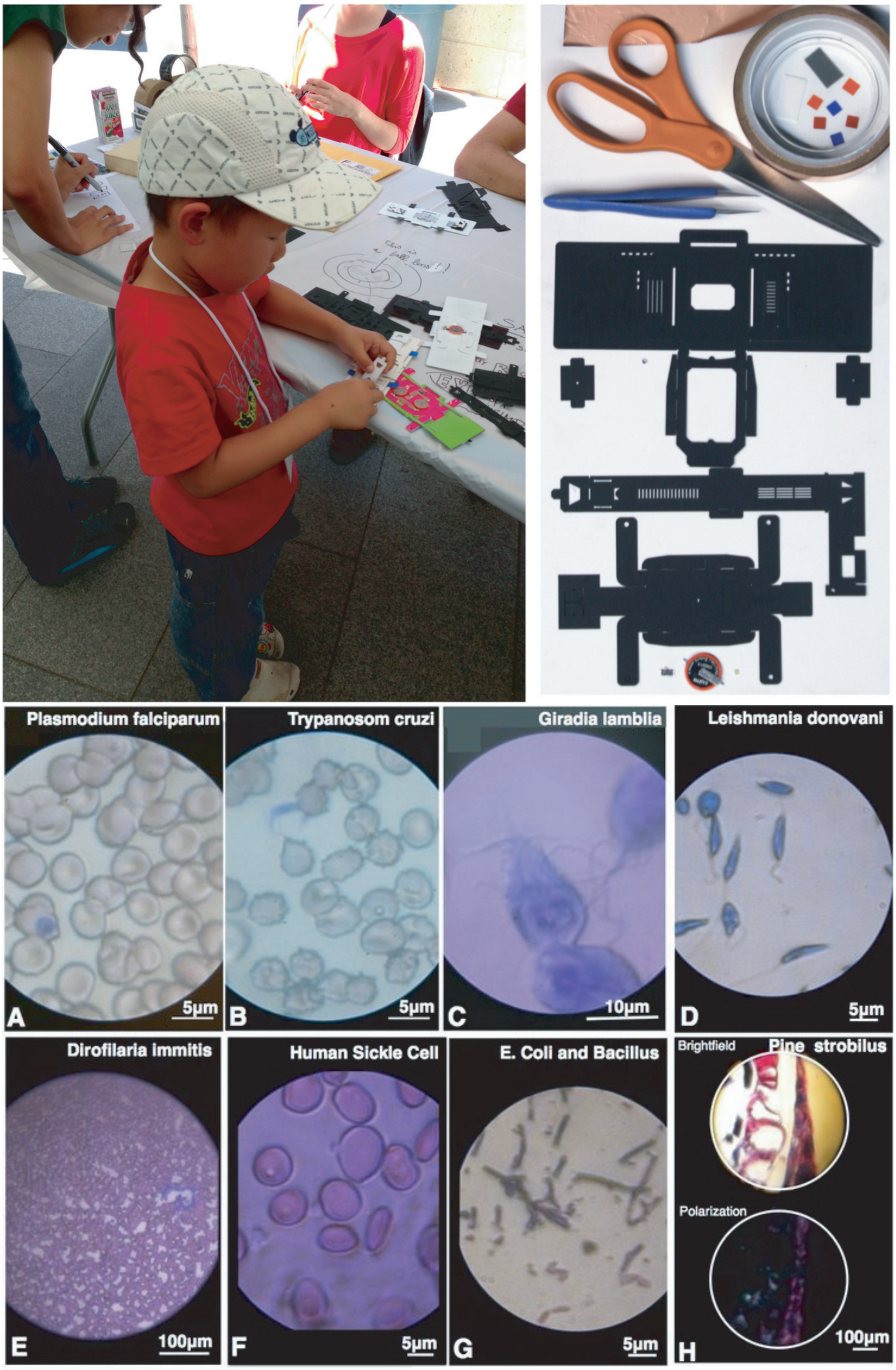


FIGURE 152 Top: the production and the parts of a flat microscope for medical use in developing countries made from sheet paper; bottom: the images it produces (© Foldscope team at www.foldscope.com).

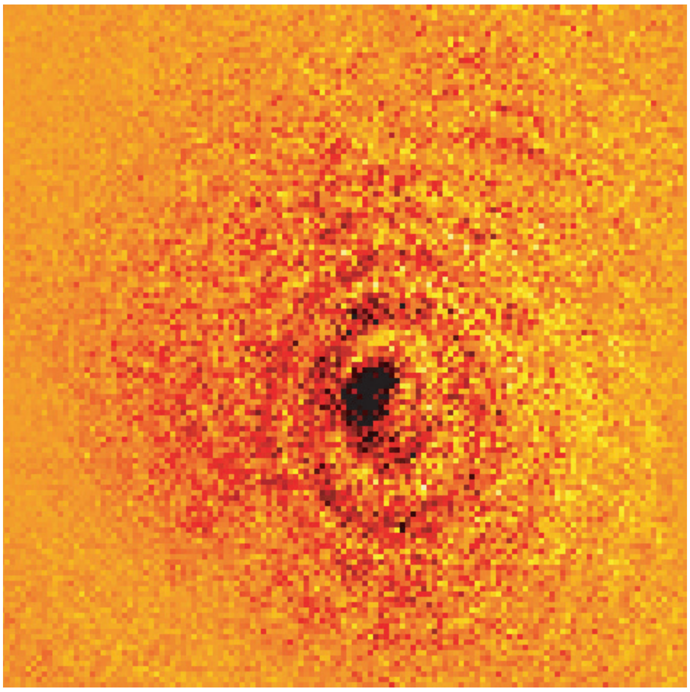


FIGURE 153 The shadow of a single ytterbium ion levitated in an ion trap and illuminated with a laser; picture size is about 16 μm in both directions (© Dave Kielpinski).

* *

It is said that astronomers have telescopes so powerful that they could see whether somebody would be lighting a match on the Moon. Can this be true? Challenge 194 s

* *

Total refraction is an interesting phenomenon it itself; but its details are even more fascinating. In 1943 Fritz Goos and Hilda Hänchen showed that the reflected beam is slightly shifted; in other words, the reflected beam is effectively reflected by a plane that lies slightly *behind* the material interface. This so-called *Goos-Hänchen shift* can be as large as a few wavelengths and is due to travelling evanescent waves in the thinner medium.

Ref. 154 In fact, recent research into this topic discovered something even more interesting. When reflection is explored with high precision, one discovers that no reflected light ray is exactly on the position one expects them: there is also a lateral shift, the *Imbert–Fedorov shift*, and even the angle of the reflected ray can deviate from the expected one. The fascinating details depend on the polarization of the beam, on the divergence of the beam and on the material properties of the reflecting layer. These observations can be seen as higher-order effects of quantum field theory; their details are still a topic of research.

* *

Materials that absorb light strongly also emit strongly. Why then does a door with dark paint in the sun get hotter than a door that is painted white? The reason is that the

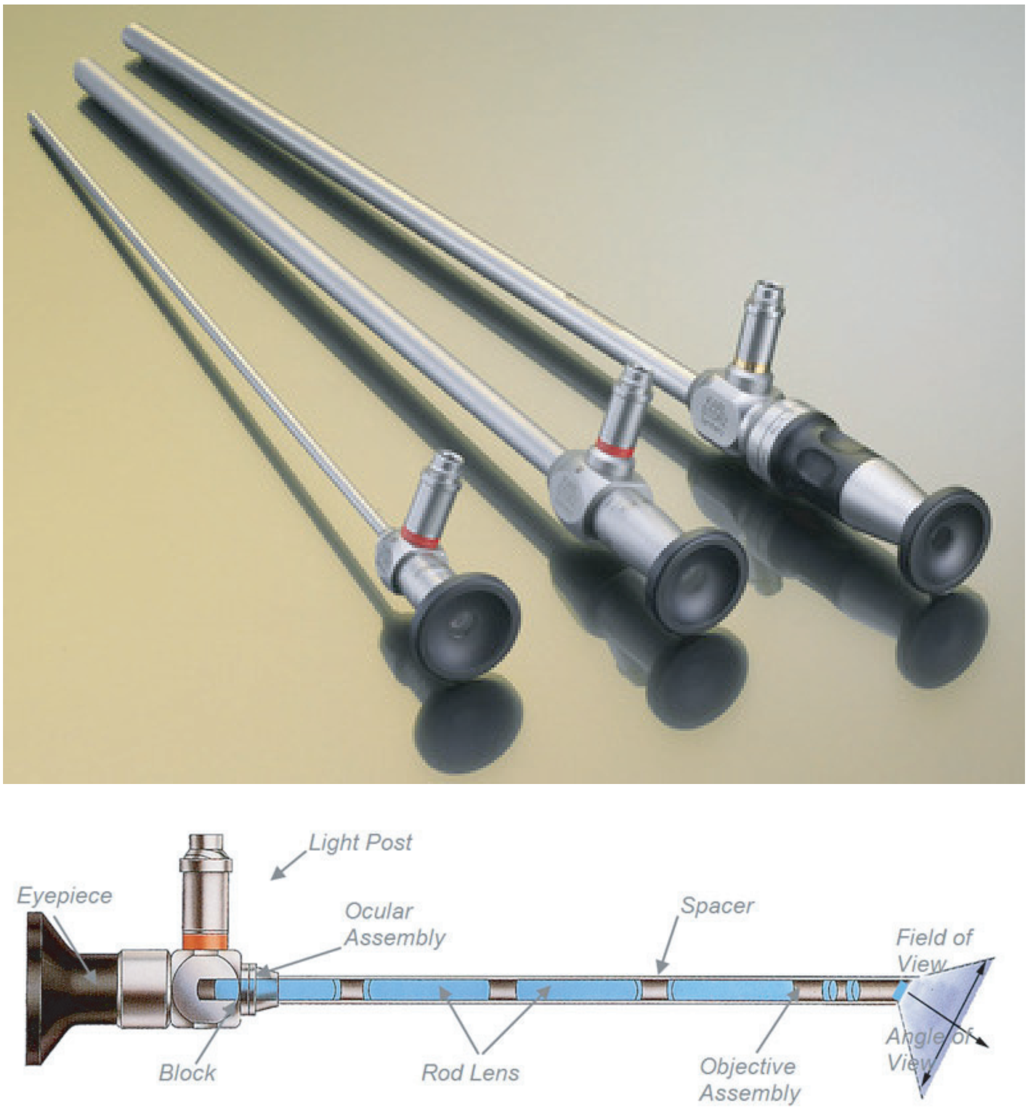


FIGURE 154 The endoscope invented by Hopkins, in which rod lenses allow large field of view and high brightness – the more so the higher the glass/air ratio is (© Karl Storz).

emission takes place at a much lower wavelength than that of visible light; for everyday situations and temperatures, emission is around 10 μm. And at that wavelength, almost all paints are effectively black, with emissivities of the order of 0.9, irrespective of their colour. And for the same reason, when you paint your home radiator, the colour is not important.

* *

Ref. 155 When two laser beams cross at a small angle, they can form light pulses that seem to
Challenge 195 s move faster than light. Does this contradict special relativity?

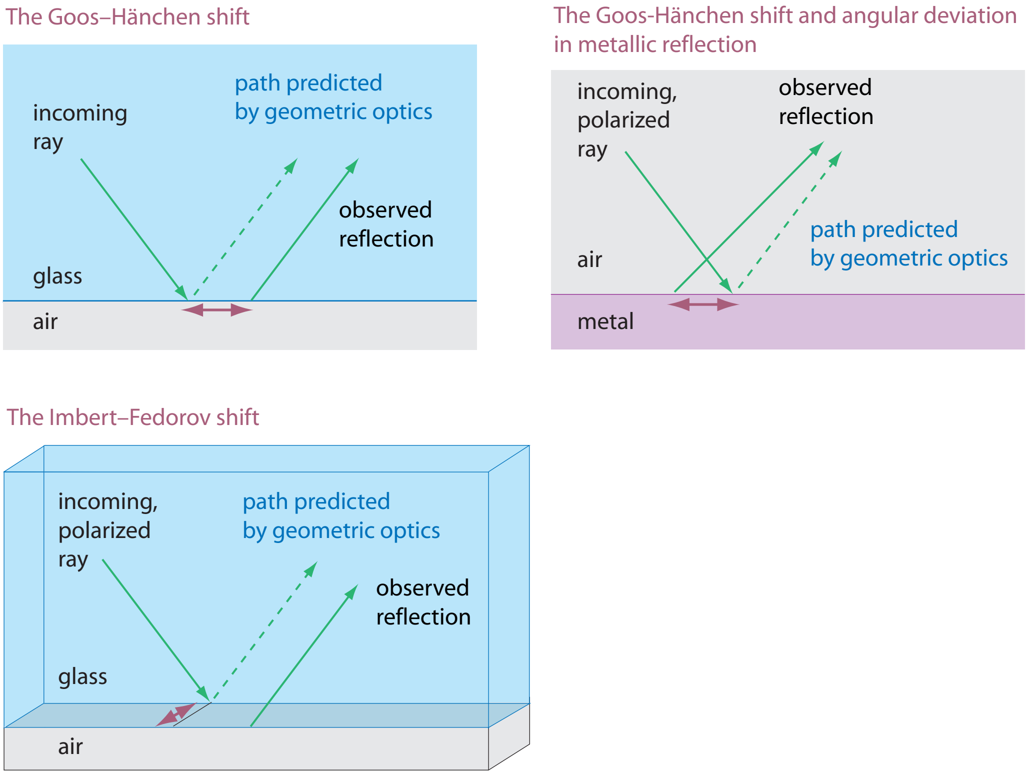


FIGURE 155 The Goos-Hänchen shift and other deviations from geometric reflection: in total reflection, the reflected light beam is slightly displaced from its naively expected position; in metallic reflection, even more deviations are observed.

FIGURE 156 How natural colours (top) change for three types of colour blind: deutan, protan and tritan (© Michael Douma).

* *

Colour blindness was discovered by the great scientist John Dalton (b. 1766 Eaglesfield, d. 1844 Manchester) – on himself. Can you imagine how he found out? It affects, in all its forms, one in 20 men. In many languages, a man who is colour blind is called *daltonic*. Women are almost never daltonic or colour blind, as the property is linked to defects on the X chromosome. If you are colour blind, you can check to which type you belong with the help of Figure 156. (The X chromosome is also at the origin of the rare tetrachromatic women mentioned above.)

Challenge 196 s

Ref. 156

Page 194

* *

Artificial colour blindness is induced by certain types of illumination. For example, violet light is used to reduce intravenous drug consumption, because violet light does not allow finding veins under the skin.

Artificial contrast enhancement with illumination is also useful. Pink light is used by beauticians to highlight blemishes, so that the skin can be cleaned as well as possible. In 2007, the police officer Mike Powis in Nottingham discovered that this 'acne light' could be used to reduce the crime rate; since acne is not fashionable, pink light deters youth from gathering in groups, and thus calms the environment where it is installed.

Yellowish light is used by by supermarkets to increase their sales of fruit and vegetables. In yellow light, tomatoes look redder and salad looks greener. Check by yourself: you will not find a single supermarket without these lights installed over fruit and vegetables.

Challenge 197 e

* *

Light beams, such as those emitted from lasers, are usually thought of as thin lines. However, light beams can also be *tubes*, with the light intensity lower in the centre than on the rim. Tubular laser beams, i.e., Bessel beams of high order, are used in modern research experiments to guide plasma channels and sparks.

* *

Is it possible to see stars from the bottom of a deep pit or of a well, even during the day, as is often stated?

Challenge 198 s

* *

Ref. 157

Humans are the *only* primates that have *white* eyes. All apes have *brown* eyes, so that it is impossible to see in which direction they are looking. Apes make extensive use of this impossibility: they often turn their head in one direction, pretending to look somewhere, but turn their eyes in another. In other words, brown eyes are useful for deception. The same effect is achieved in humans by wearing dark sunglasses. So if you see somebody with sunglasses in a situation where there is no sunlight, you know that he or she is behaving like an ape.

Apes use this type of deception to flirt with the opposite sex without their steady partner noticing. Sunglasses are tools for the unfaithful.

* *

Challenge 199 s

How can you measure the power of the Sun with your eyes closed?

FIGURE 157 Ames rooms in Paris and in San Francisco (© Sergio Davini, David Darling).

* *

Even in a dark, moonless and starless night, a forest is not dark. You can see luminescent mushrooms (of which there are over 70 different species), luminescent moulds, you can see sparks when you take off your pullover or when your friend bites a mint bonbon or when you unroll a roll of adhesive tape or open a letter.

* *

Challenge 200 d How do you produce X-rays with a roll of adhesive tape?

* *

The number of optical illusions is enormous, and there are many time-wasting websites devoted to the topic. Films often use the so-called *Ames room* to transform actors into dwarfs. It is shown in Figure 157.

* *

The brain is important in many aspects of vision. It happens that the brain, together with the eye, makes colours *disappear*, as shown in Figure 158. (The effect only works with a colour version of the figure.) The example is taken from the beautiful collection of visual illusions at www.psy.ritsumei.ac.jp/~akitaoka/color9e.html. Several related illusions, based on this one, use moving coloured dots.

The brain is also able to correct, in a matter of minutes, *deformations* of the field of view, such as those generated by glasses, for example. Even more impressive is the ability of the brain to compensate *cyclotorsion*; cyclotorsion is the rotation of the eyes along the front-back axis; when we lie down, this rotation has a value between 2 and 14 degrees, compared to the orientation while standing. The value of the angle depends on age and stress; it rotates each eye into opposite directions.

* *

If you want to experience how essential the brain is for stereopsis, build and then look through a so-called *pseudoscope*. It uses 4 mirrors or two prisms to switch the images between the left and the right eyes. An example is shown in Figure 159. You will see concave things as convex, and your sense of depth gets utterly confused. Enjoy it. Challenge 201 e

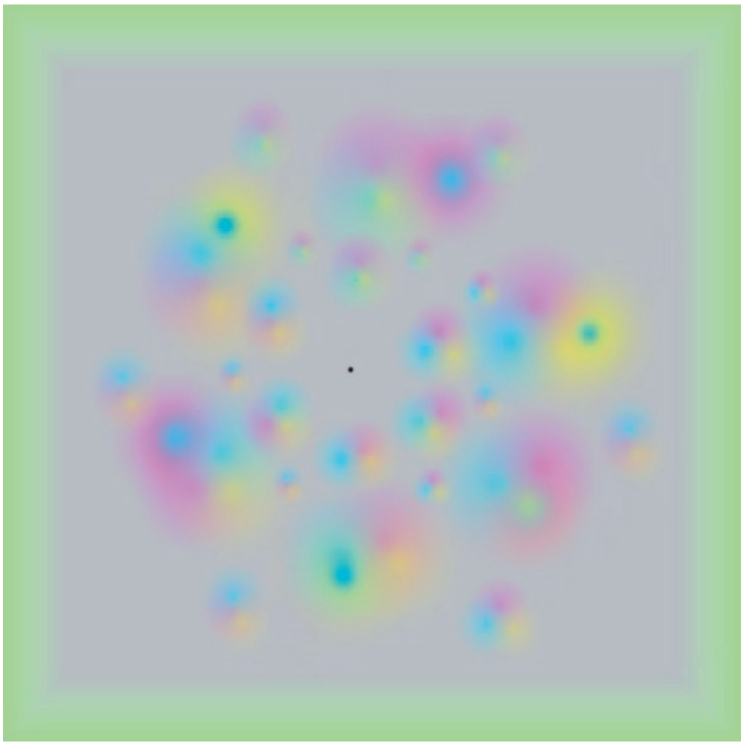


FIGURE 158 Look at the central dot for twenty seconds: the colours will disappear (© Kitaoka Akiyoshi).

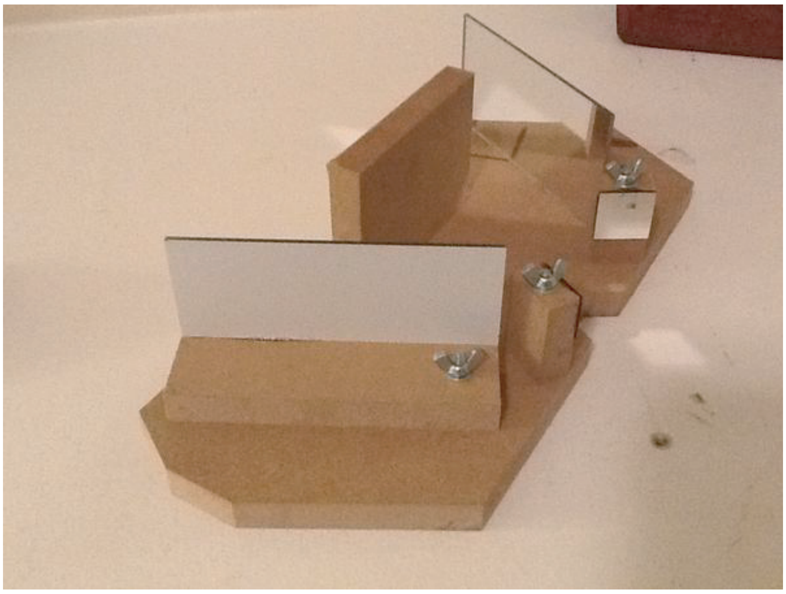


FIGURE 159 Looking through a pseudoscope changes our perception of depth (© Joshua Foer).

∗ ∗

Even more astonishing are devices that turn upside down all what you see. They can be made with mirrors or with two Dove prisms. Interestingly, after wearing them for a

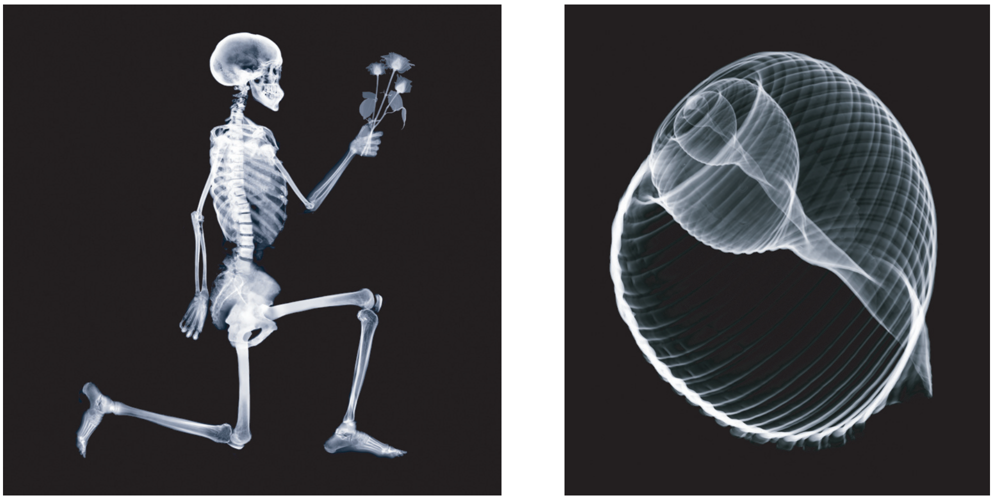


FIGURE 160 The beauty of X-rays: X-ray images of a person (taken with a corpse) and of a sea shell (© Nick Veasey).

while, the brain switches the images back to the correct orientation.

* *

X-ray imaging is so impressive that it has become a form of art. One of the foremost X-ray artists is Nick Veasey, and two of his works are shown in Figure 160. Among many examples, he has even taken X-ray images of complete buses and aeroplanes.

* *

Lenses are important components in most optical systems. Approximately, the distance of the lens focus f, the distance of the object to be imaged o, and the distance of its image i are related by the *thin lens formula*

$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i} . \qquad (81)$$

Challenge 202 e It is not hard to deduce it with the help of raytracing.

If you ever are in the situation to design a lens, you will want to know the relation between the shape of a lens and its focal distance. It turns out that there are two types of lenses: The first type are *spherical* lenses which are easy and thus cheap to make, but whose images are not perfect. The second lens type are *aspherical* lenses, which are hard to fabricate, more expensive, but provide much better image quality. High-quality optical systems always contain aspherical lenses.

For historical reasons, most books on optics teach readers the approximate relation between the geometric radii of a thin spherical lens, its refractive index n and its focal distance: Challenge 203 e

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) . \qquad (82)$$

This is called the *lensmaker formula*. Most aspherical lenses are approximately spherical, so that the formula helps as a rough first estimate also in these cases.

* *

Ref. 111 Imaging is an important part of modern industry. Without laser printers, photocopying machines, CD players, DVD players, microscopes, digital photographic cameras, film and video cameras, lithography machines for integrated circuit production, telescopes, film projectors, our world would look much more boring. Nowadays, designing optical systems is done with the help of dedicated software packages. They allow to calculate image quality, temperature effects and mechanical tolerances with high precision. Despite the beauty of optical design, there is a shortage of experts on this fascinating field, across the world.

* *

Additional types of videos cameras are still being developed. Examples are time-of-flight cameras, laser scanning cameras, ultraviolet video cameras, video cameras that measure polarization and infrared video cameras. The latter cameras will soon appear in cars, in order to recognize people and animals from the heat radiation they emit and help avoiding accidents.

* *

Challenge 204 e What are the best colour images one can produce today? At present, affordable images on paper have about 400 dots/mm, or dots of about 2.5 μm. What is the theoretical maximum? You will find that several unserious research groups claim to have produced colour images with a resolution that is higher than the theoretical maximum.

* *

Vol. I, page 313 Challenge 205 e Ultrasound imaging is regularly used in medical applications. As mentioned earlier on, unfortunately it is not safe for imaging pregnancies. Is *ultrasound imaging*, though not an optical imaging method, a type of tomography?

* *

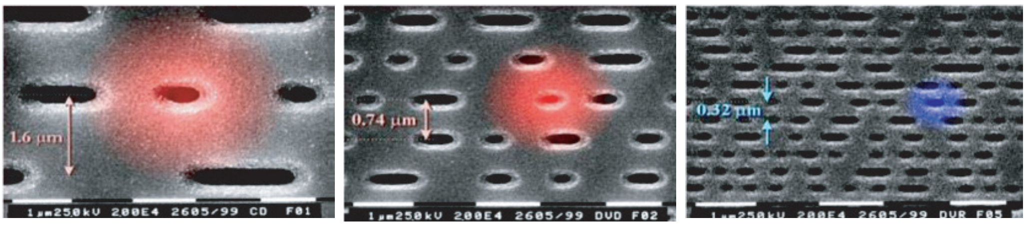
CMOS cameras, batteries and radio transmitters have become so small that they can be made into a package with the size of a pill. Such a camera can be swallowed, and with electrodes attached to the belly of a person, one can record movies of the intestine while the person is continuing its daily activities.

* *

The most *common* optical systems are those found inside CD and DVD drives. If you ever have the opportunity to take one apart, do it. They are fascinating pieces of technology, in which every cubic millimetre has been optimized by hundreds of engineers. Can you imagine how a CD or DVD player works, starting from the photographs of Figure 161?

* *

The most *expensive* optical systems are not those found on espionage satellites – which can read the headlines of a newspaper from space – but those found in wafer steppers.



CD
track pitch 1.6 µm
minimum pit length 0.8 µm

DVD
track pitch 0.74 µm
minimum pit length 0.4 µm

Blue Ray Disk
track pitch 0.32 µm
minimum pit length 0.15µm

FIGURE 161 Composed image of the tracks and the laser spot in a drive reading a CD, a DVD and a blue ray disc (© Wikimedia).

FIGURE 162 One of the many kinds of Benham's wheels. Rotating it with a top, a CD player or a drill is the simplest way to produce Fechner colours, i.e., false colours that appear from intermittent black and white patterns.

Wafer steppers are machines used for the production of electronic integrated circuits. In such steppers, a metal mask is imaged, using light from a UV laser at 193 nm, onto a photo-resist covered silicon wafer. The optical systems used have the size of an average human, are precise within a few nanometres, and cost more than six million Euro a piece. Objectives for extreme UV will be at least ten times more expensive. EUV steppers are probably the most daring industrial systems ever conceived.

* *

You can buy transparent window panes that can be switched to translucent and back – thus from a clear glass to milk glass and back – by toggling an electrical switch. How do they work? Challenge 206 e

* *

A rotating wheel coloured in a specific black and white pattern, such as *Benham's wheel*, will produce false colour effects in the eye. Unfortunately, a video of the effect does not work inside a pdf file such as the one of this book; instead, have a look at Kenneth Brecher's website at lite.bu.edu/vision/applets/Color/Benham/Benham.html or lite.bu.edu/vision-flash10/applets/Color/Benham/Benham.html. False colours can also be induced by flickering monochromatic images on computer screens. All these false colours are mainly due to the different response times of red, green and blue cones.

* *

Ref. 158 The size of the eye in mammals depends on their maximum running speed. This dependence has been verified for 50 different species. Interestingly, the correlation does not hold for the flying speed of birds.

* *

Children swimming a lot under water can learn to see *sharply* in about 10 sessions – in contrast to adults. The children of the Moken people in Thailand were studied for this feat. The study confirmed that all children have this ability, but most children do not spend enough time in the sea.

* *

Challenge 207 e Did you ever see a shadow on a mirror or on a flat water surface? Why not?

* *

Can you use lasers to produce images floating in mid-air? Yes, and there are at least three ways.

With a laser tuned to the orange sodium resonance, one can write a simulated star in the sky, at a height of about 80 km. If you would move that laser, you could write a text in the night sky. Alas, the brightness of a few watts of light at that distance is not visible with the naked eye. And there are no lasers with more power at present.

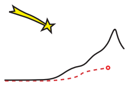
During the day, a laser with short pulses (nanoseconds) is able to write simple moving shapes at a height of a few metres. Demonstrations can be found on the internet, e.g., at www.burton-jp.com/en/.

During the day, a laser with ultrashort pulses (picoseconds or femtoseconds) of sufficient power, together with a fast scanning system, is able to write moving three-dimensional shapes of a few cubic centimetres with high resolution. There is a race across the world to be the first to demonstrate this.

Challenge 208 ny Will you be the first to show one of these systems?

SUMMARY ON APPLIED OPTICS

The art and science of making images is central to modern health care, industry, science, entertainment and telecommunications. Acquiring images is in large part the result of bending light beams in predefined ways and then detecting them. All image acquisition systems, biological or human-made, are based on reflection, refraction or diffraction, combined with pixel detectors. All imaging systems that acquire or display high-quality images – biological or human-made – use clever combinations of materials science, sensors, actuators and signal processing. This fascinating field is still evolving rapidly.



CHAPTER 5

ELECTROMAGNETIC EFFECTS

Looking carefully, the atmosphere is full of electrical effects. The most impressive, lightning, is now reasonably well understood. However, it took decades and a large number of researchers to discover and put together all the parts of the puzzle. Also below our feet there is something important going on: the hot magma below the continental crust produces the magnetic field of the Earth and other planets. Strong magnetic fields are fascinating for a third reason: they can be used for levitation. We first explore these three topics, then give an overview about the many effects that electromagnetic fields produce and conclude with some curiosities and challenges about electric charge.

Ref. 159

IS LIGHTNING A DISCHARGE? – ELECTRICITY IN THE ATMOSPHERE

Inside thunderstorm clouds, especially inside tall *cumulonimbus* clouds,[**] charges are separated by collision between the large 'graupel' ice crystals falling due to their weight and the small 'hail' ice crystallites rising due to thermal upwinds. Since the collision takes part in an electric field, charges are separated in a way similar to the mechanism in the Kelvin generator. Discharge takes place when the electric field becomes too high, taking a strange path influenced by ions created in the air by cosmic rays. (There are however, at least ten other competing explanations for charge separation in clouds.) It seems that cosmic rays are at least partly responsible for the zigzag shape of lightning. For a striking example, see Figure 163.

Ref. 161 Page 20 Ref. 162

A lightning flash typically transports 20 to 30 C of charge, with a peak current of up to 20 kA. But lightning flashes have also strange properties. First, lightnings appear at fields around 200 kV/m (at low altitude) instead of the 2 MV/m of normal sparks. Second, *lightning emits radio pulses.* Third, lightning emits X-rays and gamma rays. Russian researchers, from 1992 onwards explained all three effects by a newly discovered discharge mechanism. At length scales of 50 m and more, cosmic rays can trigger the appearance of

Ref. 163

[**] Clouds have Latin names. They were introduced in 1802 by the explorer Luke Howard (b. 1772 London, d. 1864 Tottenham), who found that all clouds could be seen as variations of three types, which he called *cirrus*, *cumulus* and *stratus*. He called the combination of all three, the rain cloud, *nimbus* (from the Latin 'big cloud'). Today's internationally agreed system has been slightly adjusted and distinguishes clouds by the height of their lower edge. The clouds starting above a height of 6 km are the cirrus, the cirrocumulus and the cirrostratus; those starting at heights of between 2 and 4 km are the altocumulus, the altostratus and the nimbostratus; clouds starting below a height of 2 km are the stratocumulus, the stratus and the cumulus. The rain or thunder cloud, which crosses all heights, is today called *cumulonimbus*. For beautiful views of clouds, see the www.goes.noaa.gov and www.osei.noaa.gov websites.

Ref. 160

FIGURE 163 A rare photograph of a lightning stroke hitting a tree (© Niklas Montonen).

FIGURE 164 Cumulonimbus clouds from ground and from space (NASA).

lightning; the relativistic energy of these rays allows for a discharge mechanism that does not exist for low energy electrons. At relativistic energy, so-called runaway breakdown leads to discharges at much lower fields than usual laboratory sparks. The multiplication

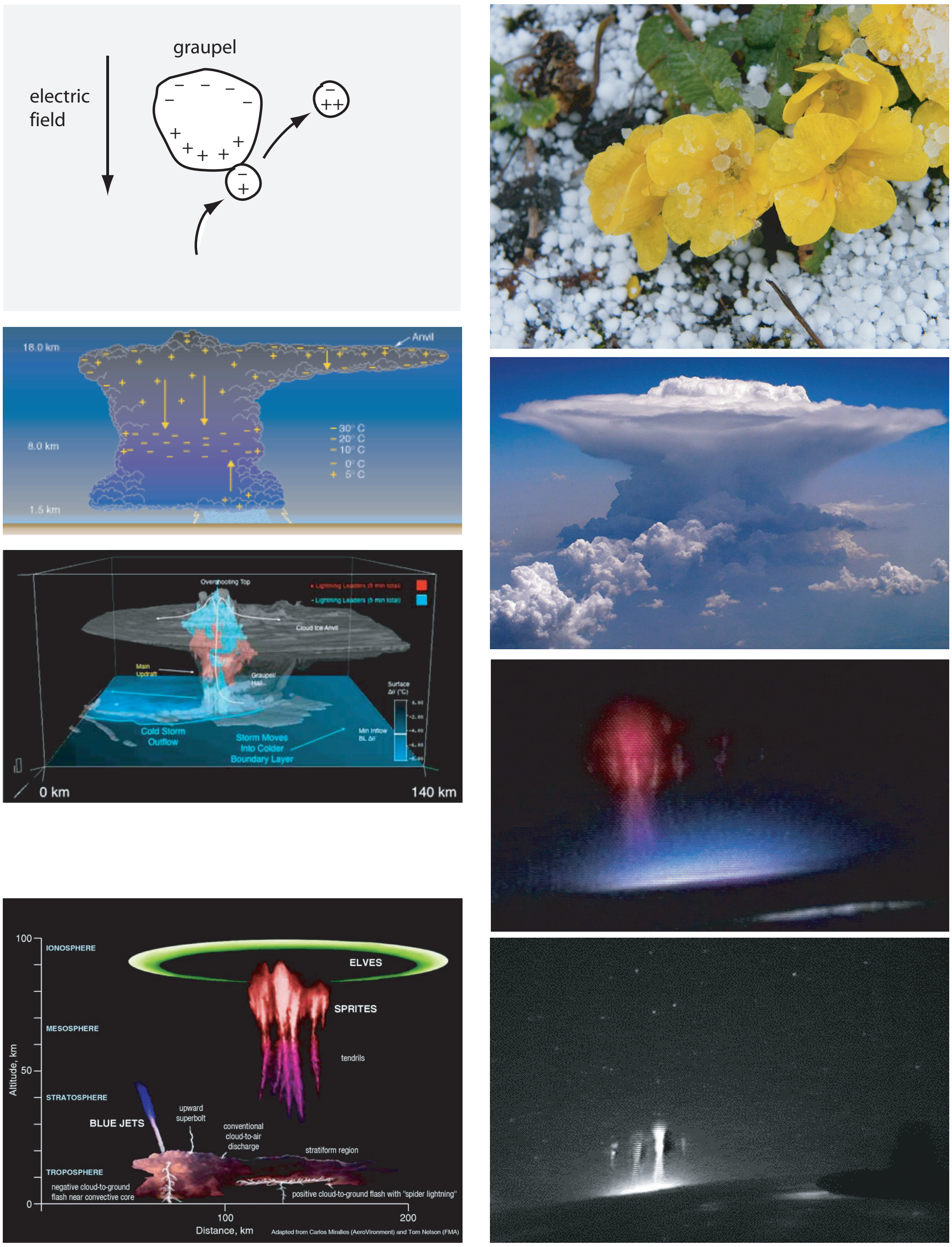


FIGURE 165 The charging and discharging of clouds: the most probable microscopic mechanism, namely charging of graupel particles by collision with ice particles, the cloud charge distribution, the three-dimensional structure and the large scale processes discovered in the past decades from aeroplanes (© nordique, NASA, NOAA).

of these relativistic electrons also leads to the observed radio and gamma ray emissions.

In the 1990s more electrical details about thunderstorms became known. Airline pilots and passengers sometime see weak and coloured light emissions spreading from the top of thunderclouds. There are various types of such emissions: blue *jets* and mostly red *sprites* and *elves*, which are somehow due to electric fields between the cloud top and the ionosphere. The details are still under investigation, and the mechanisms are not yet clear.*

Ref. 166 The emission of X-rays by lightning dates from the early twentieth century. The experimental confirmation was not easy though; it is necessary to put a detector near the lightning flash. To achieve this, the lightning has to be directed into a given region, where the detector is located. This is possible using a missile pulling a metal wire, the other end of which is attached to the ground. These experimental results are now being collated into a new description of lightning which also explains the red-blue sprites above thunderclouds. In particular, the processes also imply that inside clouds, electrons can be accelerated up to energies of a few MeV. Thunderclouds are electron accelerators. Ref. 167

Incidentally, you have a 75 % chance of survival after being hit by lightning, especially if you are completely wet, as in that case the current will mainly flow outside the skin. Usually, wet people who are hit lose all their clothes, as the evaporating water tears them off. Rapid resuscitation is essential to help somebody to recover after a hit. If you are ever hit by lightning and survive, go to the hospital! Many people died three days later having failed to do so. A lightning strike often leads to coagulation effects in the blood. These substances block the kidneys, and one can die three days later because of kidney failure. The simply remedy is to have dialysis treatment.

As a note, you might know how to measure the distance of a lightning by counting the seconds between the lightning and the thunder and multiplying this by the speed of sound, 340 m/s; it is less well known that one can estimate the *length* of the lightning bolt by measuring the *duration* of the thunder, and multiplying it by the same factor.

Lightning is part of the electrical circuit around the Earth. This fascinating part of geophysics would lead us too far from the aim of our adventure. But every physicist should know that there is a vertical electric field of between 100 and 300 V/m on a clear day, as discovered already in 1752. (Can you guess why it is not noticeable in everyday life? And why despite its value it cannot be used to extract large amounts of energy?) Challenge 209 s The field is directed from the ionosphere down towards the ground; in fact the Earth is permanently negatively charged, and in clear weather current flows downwards (electrons flow upwards) through the clear atmosphere, trying to *discharge* our planet. The current of about 1 to 2 kA is spread over the whole planet; it is possibly due to the ions formed by cosmic radiation. (The resistance between the ground and the ionosphere is about 200 Ω, so the total voltage drop is about 200 kV.) At the same time, the Earth is constantly being *charged* by several effects: there is a dynamo effect due to the tides of the atmosphere and there are currents induced by the magnetosphere. But the most important charging effect is lightning.

In other words, contrary to what one may think, lightning does not discharge the Ref. 164 ground, it actually charges it up! Indeed, the Earth is charged to about −1 MC. Can you

---

* For images, have a look at the interesting elf.gi.alaska.edu/, www.fma-research.com/spriteres.htm and pasko.ee.psu.edu/Nature websites.

Challenge 210 s confirm this? Of course, lightning does discharge the cloud to ground potential difference; but by doing so, it actually sends (usually) a negative charge down to the Earth as a whole. Thunderclouds are batteries; the energy from the batteries comes from the thermal uplifts mentioned above, which transport charge *against* the global ambient electrical field.

Using a few electrical measurement stations that measure the variations of the electrical field of the Earth it is possible to locate the position of all the lightning that comes down towards the Earth at a given moment. Distributed around the world, there are Ref. 165 *about a hundred lightning flashes per second.* Present research also aims at measuring the activity of the related electrical sprites and elves in this way.

The ions in air play a role in the charging of thunderclouds via the charging of ice crystals and rain drops. In general, all small particles in the air are electrically charged. When aeroplanes and helicopters fly, they usually hit more particles of one charge than of the other. As a result, aeroplanes and helicopters are charged up during flight. When a helicopter is used to rescue people from a raft in high seas, the rope pulling the people upwards must first be earthed by hanging it in the water; if this is not done, the people on the raft could die from an electrical shock when they touch the rope, as has happened a few times in the past.

Why are sparks and lightning blue? This turns out to be a material property: the colour comes from the material that happens to be excited by the energy of the discharge, usually air. This excitation is due to the temperature of 30 kK inside the channel of a typical lightning flash. For everyday sparks, the temperature is much lower. Depending on the situation, the colour may arise from the gas between the two electrodes, such as oxygen or nitrogen, or it may due to the material evaporated from the electrodes by the discharge. For an explanation of such colours, as for the explanation of all colours due to materials, we need to wait for the next part of our walk, on quantum theory.

DOES BALL LIGHTNING EXIST?

For hundreds of years, people have reported sightings of so-called *ball lightning*. The sightings are rare but recurrent. Usually ball lightning was reported during a thunder- Ref. 168 storm, often after a usual lightning had struck. With a few exceptions, nobody took these reports seriously, because no reproducible data existed.

When microwave ovens become popular, several methods to produce ball-shaped discharges became known. To observe one, just stick a toothpick into a candle, light the toothpick, and put it into (somebody else's) microwave oven at maximum power. This set-up produces a beautiful ball-like discharge. However, humans do not live in a microwave oven; therefore, this mechanism is not related to ball lightning.

The experimental situation changed completely in the years 1999 to 2001. In those Ref. 169 years the Russian physicists Anton Egorov and Gennady Shabanov discovered a way to produce plasma clouds, or *plasmoids*, floating in air, using three main ingredients: water, metal and high voltage. If high voltage is applied to submerged metal electrodes of the proper shape and make, plasma clouds emerge from the water, about 10 to 20 cm in size, float above the surface, and disappear after about half a second. Two examples can be seen in Figure 166. The phenomenon of floating plasmoids is still being explored. There are variations in shape, colour, size and lifetime. The spectrum of observations

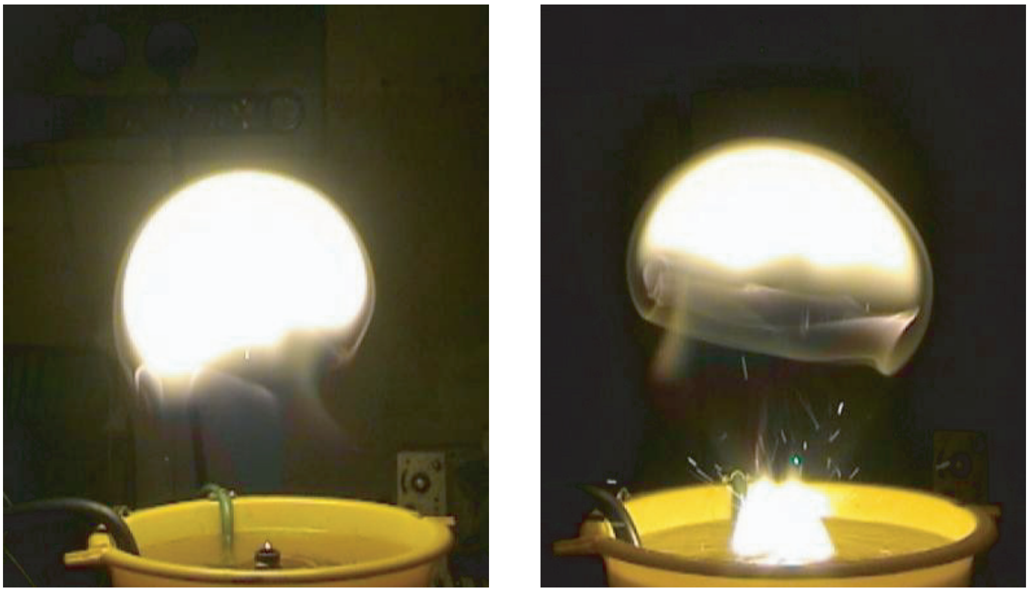


FIGURE 166 A floating plasma cloud produced in the laboratory (© Sergei Emelin and Alexei Pirozerski).

and techniques will surely evolve in the coming years.

An even more astonishing effect was published in 2007. A Brazilian research team found a way to make golf-ball sized discharges that seem to roll along the floor for as long as 8 s. Their method was beautifully simple: with the help of a 25 V power supply, they passed a current of 140 A through an arc at the surface of a silicon wafer. They discovered that small silicon particles detach and move away, while being surrounded by a luminous glow. These luminous clouds can wander around the table and floor of the laboratory, until they extinguish. Ref. 170

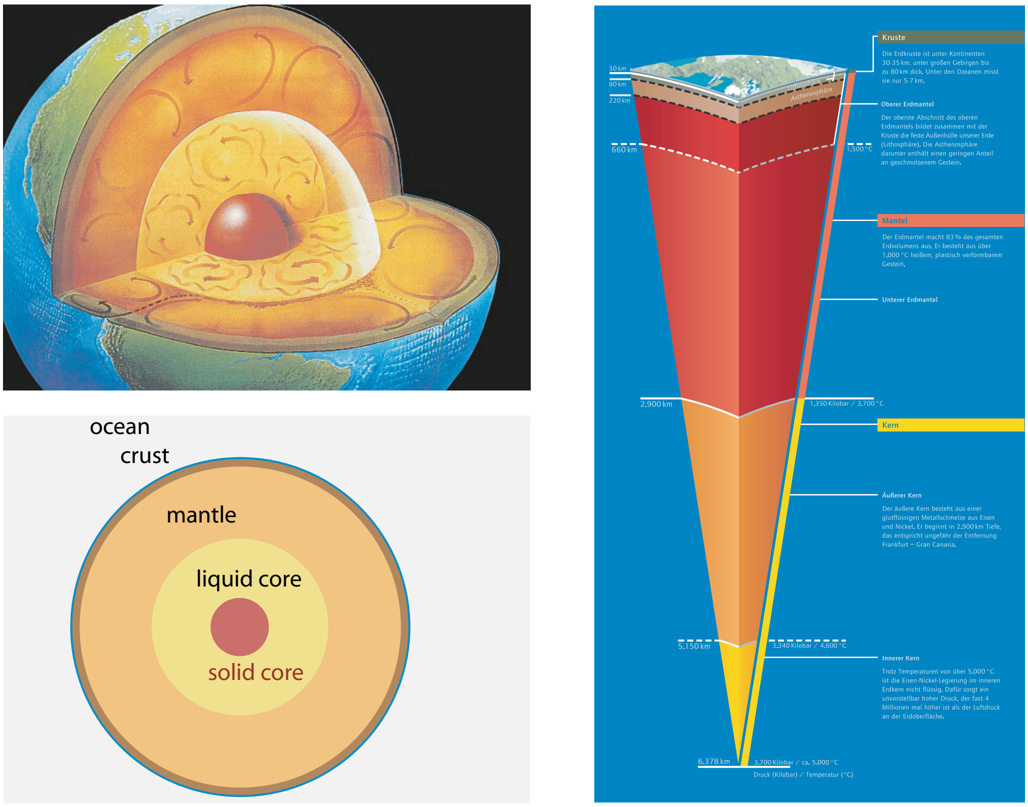
It seems that these phenomena could explain a number of ball lightning observations. But it is equally possible that additional effects will be discovered in the future.

PLANETARY MAGNETIC FIELDS

The classical description of electrodynamics is consistent and complete; nevertheless there are still many subjects of research. A fascinating example is the origin of the magnetic fields of the Earth, the other planets, the Sun and the galaxies.

The magnetic field on the Earth that determines the direction of a compass has eight sources:

1. The *main* component of the magnetic field is the geodynamo in the fluid core of the Earth.
2. A further component, the *lithospheric* field, is due to the magnetisation of the rocks.
3. The *tidal* fields are due to the induction by the main field via the moving, electrically conductive ocean currents.
4. The *Sq* fields are due to the solar irradiation of the ionosphere.
5. The magnetospheric fields are due to the distribution and drift of the charged particles it contains.



Die Erdkruste ist unter Kontinenten 30-35 km, unter großen Gebirgen bis zu 80 km dick. Unter den Ozeanen misst sie nur 5-7 km.

Der oberste Abschnitt des oberen Erdmantels bildet zusammen mit der Kruste die feste Außenhülle unserer Erde (Lithosphäre). Die Asthenosphäre darunter enthält einen geringen Anteil an geschmolzenem Gestein.

Der Erdmantel macht 83 % des gesamten Erdvolumens aus. Er besteht aus über 1.000 °C heißem, plastisch verformbarem Gestein.

Der äußere Kern besteht aus einer glutflüssigen Metallschmelze aus Eisen und Nickel. Er beginnt in 2.900 km Tiefe, das entspricht ungefähr der Entfernung Frankfurt – Gran Canaria.

Trotz Temperaturen von über 5.000 °C ist die Eisen-Nickel-Legierung im inneren Erdkern nicht flüssig. Dafür sorgt ein unvorstellbar hoher Druck, der fast 4 Millionen mal höher ist als der Luftdruck an der Erdoberfläche.

FIGURE 167 The structure of our planet (© MPI-Chemie, Mainz/GEO).

6. Polar and equatorial electrojets are induced by specific ionospheric conductivity distributions.
7. Magnetic storms are induced by the solar wind.
8. Human sources of all kinds.

The main magnetic field is due to the convection of the liquid outer core deep inside the Earth, which is made mainly of liquid iron. The convection is mainly due to the radial gradient of composition in the outer core – but also to the temperature gradient – and leads to motions of the liquid iron with speeds of up to 30 km/a. The Coriolis force strongly influences these motions. The motion of the conductive iron in the already existing magnetic field in turn generates, like in a dynamo, an additional magnetic field. The mechanism at the basis of the geodynamo is not easy to picture, as it is intrinsically three-dimensional. An impression is given by Figure 168. The influences of turbulence, non-linearities and chaos make this a surprisingly complex phenomenon. Similar processes occur inside the other planets and the stars.

The details of the generation of the magnetic field of the Earth, usually called the *geodynamo*, began to appear only in the second half of the twentieth century, when the knowledge of the Earth's interior reached a sufficient level. The Earth's interior starts below the Earth's crust. The *crust* is typically 30 to 40 km thick (under the continents), though it is thicker under high mountains and thinner near volcanoes or under Ref. 33

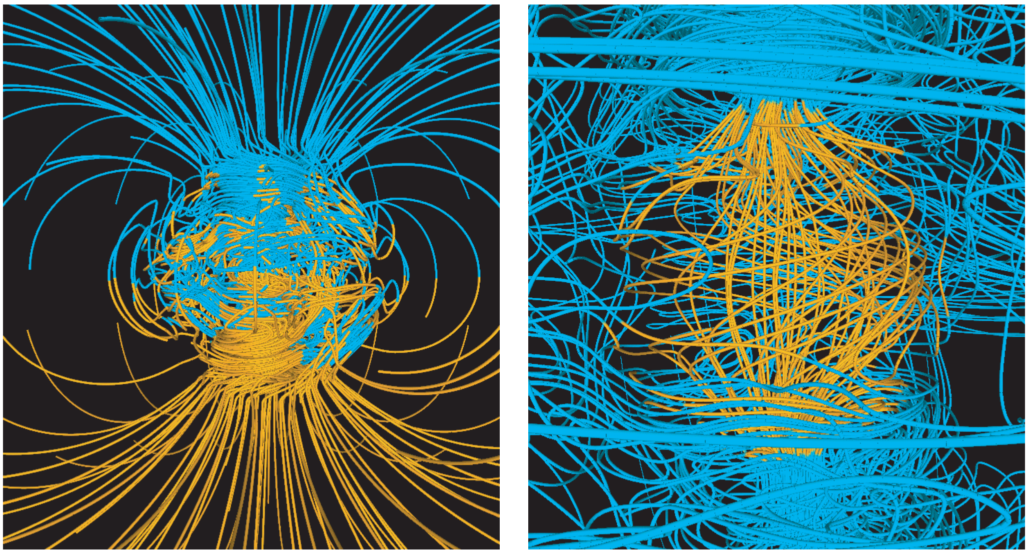


FIGURE 168 *Left:* an impression of the magnetic field lines inside and outside the rotating Earth, up to a distance of two Earth radii, calculated with a computer simulation. North is up, south is down. Field lines directed inwards are blue, directed outwards are yellow. Inside the fluid core, the field is complex and strong. Outside the Earth's core, it is a much weaker, smooth and mainly dipolar field. *Right:* the field lines inside the solid inner core of the Earth (yellow) and the liquid outer core (blue); the relative rotation between the two is central for the geodynamo. The computer model was developed and run by Gary A. Glatzmaier (University of California, Santa Cruz) and Paul H. Roberts (University of California, Los Angeles) (© Gary Glatzmaier)

the oceans. As already mentioned, the crust consists of large segments, the *plates*, that float on magma and move with respect to one another. The Earth's interior is divided into the *mantle* – the first 2900 km from the surface – and the *core*. The core is made up of a *liquid outer core*, 2210 km thick, and a *solid inner core* of 1280 km radius. (The temperature of the core is not well known; it is believed to be in the range of 6 ± 1 kK. Can you find a way to determine it? The temperature might even have decreased a few hundred kelvin during the last 3000 million years.)

Challenge 211 d

The Earth's core consists mainly of iron that has been collected from the asteroids that collided with the Earth during its youth. The liquid and electrically conducting outer core acts as a dynamo that keeps the magnetic field going. This is possible because the liquid core not only rotates, but also *convects* from deep inside the Earth to more shallow depths. As mentioned, the convection is driven by the radial gradient of its composition and, probably a bit less, by the temperature gradient between the hot inner core and the cooler mantle. Due to the convection, rotation and the Coriolis effect, the average fluid motion near the inner core is helical. Huge electric currents flow in complex ways through the liquid. The liquid motion, maintained by friction, creates the magnetic field. At present, the surface magnetic field has an intensity between 20 and 70 μT, depending on the location; inside the core, the values are about 50 times higher.

The magnetic energy of the Earth thus comes from the kinetic energy of the liquid outer core, which in turn is due to buoyancy. The convection is due to what happens in the core, which is finally due to the radioactive decays that keeps the core hot. (The

Vol. V, page 184 radioactive processes are explained later on.) The detailed story is fascinating. The liquid in the outer core rotates with respect to the Earth's surface; but this motion cannot be measured. Geodynamo simulations by Gary Glatzmaier and his team predicted in 1995 that as a consequence, the solid inner core of the Earth is dragged along by the liquid outer core and thus should also rotate faster than the Earth's crust. Experimental evidence for this effect appeared from 1996 onwards. In 2005, it has been definitely reported Ref. 173 that the inner core of the Earth rotates faster than the Earth's crust by up to half a degree per year.

The magnetic field of the Earth switches orientation at irregular intervals of between a few tens of thousands and a few million years. Understanding this process is one of the central subjects of research. This is not easy; experiments are not yet possible, 150 years of measurements is a short time when compared with the last transition – about 730 000 years ago – and computer simulations are extremely involved. In fact, since the field measurements started, the dipole moment of the magnetic field has steadily diminished, presently by 5 % a year, and the quadrupole moment has steadily increased. Maybe we are heading towards a surprise.

Also in stars, the magnetic field is due to convection. The moving fluid is the plasma. Because of its low viscosity and the lack of solid material, the processes and motions in the solar dynamo differ from those in the geodynamo. For example, the rotation period of the solar surface depends on the latitude; it is 24.5 days at the equator and 38 days at the poles. Due to the low viscosity of the plasma, the solar magnetic field switches polarity rapidly and regularly, every 11 years. The switch has important effects on the number of sunspots and on the intensity of the solar wind that arrives on Earth. The typical surface solar magnetic field is 0.1 to 0.2 mT, a few times that of the Earth; in sunspots it can be as high as 0.3 T.

The study of *galactic* magnetic fields is even more complex, and still in its infancy. Many measurements are available, showing typical intensities of a few nT. The origin of the galactic fields is not yet understood.

Levitation

We have seen that it is possible to move certain objects without touching them, using a magnetic or electric field or, of course, using gravity. Is it also possible, without touching an object, to keep it fixed, floating in mid-air? Does this type of rest exist?

It turns out that there are several methods of levitating objects. These are commonly Ref. 174 divided into two groups: levitation methods that *consume energy* and those who *do not*. Among the methods that consume energy is the floating of objects on a jet of air or of water, the floating of objects through sound waves, e.g. on top of a siren, or through a laser beam coming from below, and the floating of conducting material, even of liquids, in strong radio-frequency fields. Presently, levitation of liquids or solids by strong ul- Ref. 175 trasound waves is becoming popular in research laboratories. All these methods give *stationary* levitation. (Self-propelled objects like drones do not count as example of levitation.)

Another group of energy-consuming levitation methods sense the way a body is falling and kick it up again in the right way via a feedback loop; these methods are *non-stationary* and usually use magnetic fields to keep the objects from falling. The magnetic

Ref. 176 train being built in Shanghai by a German consortium is levitated this way. The whole train, including the passengers, is levitated and then moved forward using electromagnets. It is thus possible, using magnets, to levitate many tens of tonnes of material.

For levitation methods that do *not* consume energy – all such methods are necessarily stationary – a well-known limitation can be found by studying Coulomb's 'law' of electrostatics:

▷ No static arrangement of electric fields can levitate a *charged* object in free space or in air.

The same result is valid for gravitational fields and *massive* objects:*

▷ No static arrangement of masses can levitate a *massive* object.

In other words, we cannot produce a local minimum of potential energy in the middle of a box using electric or gravitational fields. This impossibility is called *Earnshaw's* Ref. 177 *theorem*. Speaking mathematically, the solutions of the Laplace equation $\Delta\varphi = 0$, the so-called *harmonic functions*, have minima or maxima only at the border, and never inside the domain of definition. (You proved this yourself on page 188 in volume I.) Earnshaw's theorem can also be proved by noting that given a potential minimum in free space, Gauss' theorem for a sphere around that minimum requires that a source of the field be present inside, which is in contradiction with the original assumption.

We can deduce that it is also impossible to use electric fields to levitate an electrically *neutral* body in air: the potential energy U of such a body, with volume V and dielectric constant ε, in an environment of dielectric constant ε_0, is given by

$$\frac{U}{V} = -\frac{1}{2}(\varepsilon - \varepsilon_0)\,E^2 \;. \qquad (83)$$

Challenge 212 ny Since the electric field E never has a maximum in the absence of space charge, and since for all materials $\varepsilon > \varepsilon_0$, there cannot be a minimum of potential energy in free space for a neutral body.**

To sum up, using *static electric* or *static gravitational* fields it is impossible to keep an object from falling; neither quantum mechanics, which incorporates phenomena such as antimatter, nor general relativity, including phenomena such as black holes, change this basic result.

For *static magnetic* fields, the discussion is analogous to electrical fields: the potential energy U of a magnetizable body of volume V and permeability μ in a medium with

---

Vol. I, page 106 * To the disappointment of many science-fiction addicts, this would even be true if a negative mass existed. And even though gravity is not really due to a field, but to space-time curvature, the result still holds in general relativity.

Ref. 178 ** It is possible, however, to 'levitate' gas bubbles in liquids – 'trap' them to prevent them from rising would be a better expression – because in such a case the dielectric constant of the environment is higher than that of the gas. Can you find a liquid–gas combination where bubbles fall instead of rise? Challenge 213 ny

Challenge 214 ny permeability μ_0 containing no current is given by

$$\frac{U}{V} = -\frac{1}{2}\left(\frac{1}{\mu} - \frac{1}{\mu_0}\right)B^2 \; . \tag{84}$$

Due to the inequality $\Delta B^2 \geqslant 0$ for the magnetic field, isolated maxima of a static magnetic field B are not possible, only isolated minima. Therefore, it is impossible to levitate paramagnetic ($\mu > \mu_o$) or ferromagnetic ($\mu \gg \mu_0$) materials such as steel, including bar magnets, which are all attracted, and not repelled to magnetic field maxima. Challenge 215 e

Two ways to realize magnetic levitation are possible: *levitating a diamagnet* or using a *time-dependent magnetic field*.

Page 39 Diamagnetic materials ($\mu < \mu_0$, or $\mu_r = \mu/\mu_0 < 1$) were discovered shortly after Earnshaw published his theorem, and allow circumventing it. Indeed, diamagnetic materials, such as graphite or water, can be levitated by static magnetic fields because they are attracted to magnetic field minima. In fact, it is possible to levitate magnets if one uses a combination containing diamagnets. A few cases that can easily be replicated on a kitchen table – together with a few other ones – are shown in Figure 169. Ref. 180 Ref. 179

Another well-known example of diamagnetic levitation is the levitation of superconductors. Indeed, superconductors, at least those of type I, are perfects diamagnets ($\mu = 0$). In some cases, superconductors can even be *suspended* in mid-air, below a magnet. Also single atoms with a magnetic moment are diamagnets; they are routinely levitated this way and have also been photographed in this state. Single neutrons, which have a magnetic dipole moment, have been kept in magnetic bottles through magnetic levitation, until they decay. Ref. 181

Diamagnets levitate if $\nabla B^2 > 2\mu_0 \rho g/\chi$, where ρ is the mass density of the object and $\chi = 1 - \mu/\mu_0$ its magnetic susceptibility. Since χ is typically about 10^{-5} and ρ of order $1000\,\mathrm{kg/m^3}$, field gradients of about $1000\,\mathrm{T^2/m}$ are needed. In other words, levitation requires fields changes of 10 T over 10 cm, which is nowadays common for high field laboratory magnets. Challenge 216 ny

Recently, scientists have levitated pieces of wood and of plastic, strawberries, water droplets, liquid helium droplets as large as 2 cm, grasshoppers, fish and frogs (all alive and without any harm) using magnetic levitation. Indeed, animals, like humans, are all made of diamagnetic material. Humans themselves have not yet been levitated, but the feat, expected to require 40 T and large amounts of electrical power, is being planned and worked on. In fact, a similar feat has already been achieved: diamagnetic levitation is being explored for the levitation of passenger trains, especially in Japan, though with little commercial success. Ref. 182 Ref. 176

Time-dependent electrical or magnetic fields, e.g. periodic fields, can lead to levitation in many different ways without any consumption of energy. This is one of the methods used in the magnetic bearings of turbomolecular vacuum pumps. Also single charged particles, such as ions and electrons, are now regularly levitated with Paul traps and Penning traps. The mechanical analogy is shown in Figure 170. Ref. 174 Ref. 174

Ref. 183 Figure 171 shows a toy that allows you to personally levitate a spinning top or a spinning magnetic sphere in mid-air above a ring magnet, a quite impressive demonstration of levitation for anybody looking at it. The photo shows that is not hard to build such a