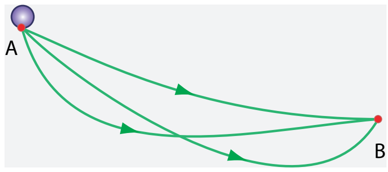


FIGURE 91 What shape of rail allows the black stone to glide most rapidly from point A to the lower point B?

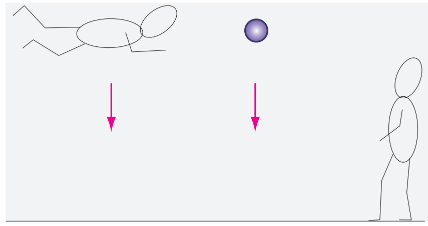


FIGURE 92 Can motion be described in a manner common to all observers?

3. GLOBAL DESCRIPTIONS OF MOTION – THE SIMPLICITY OF COMPLEXITY

> “ Πλεῖν ἀνάγκε, ζῆν οὐκ ἀνάγκη.*
> Pompeius ”

All over the Earth – even in Australia – people observe that stones fall ‘down’. This ancient observation led to the discovery of the universal ‘law’ of gravity. To find it, all that was necessary was to look for a description of gravity that was valid globally. The only additional observation that needs to be recognized in order to deduce the result $a = GM/r^2$ is the variation of gravity with height.

In short, thinking *globally* helps us to make our description of motion more precise. How can we describe motion as globally as possible? It turns out that there are six approaches to this question, each of which will be helpful on our way to the top of Motion Mountain. We will start with an overview, and then explore the details of each approach.

— The first global approach to motion arises from a limitation of what we have learned so far. When we predict the motion of a particle from its current acceleration, we are using the most *local* description of motion possible. For example, whenever we use an evolution equation we use the acceleration of a particle at a certain place and time to determine its position and motion *just after* that moment and *in the immediate neighbourhood* of that place.

Evolution equations thus have a mental ’horizon’ of radius zero.

The opposite approach is illustrated in the famous problem of Figure 91. The challenge is to find the path that allows the fastest possible gliding motion from a high point to a distant low point. To solve this we need to consider the motion as a whole, for all times and positions. The global approach required by questions such as this one will lead us to a description of motion which is simple, precise and fascinating: the so-called principle of cosmic laziness, also known as the principle of least action.

Challenge 343 d

— The second global approach to motion emerges when we compare the various descriptions of the same system produced by different observers. For example, the observations by somebody falling from a cliff, a passenger in a roller coaster, and an observer

* Navigare necesse, vivere non necesse. ‘To navigate is necessary, to live is not.’ Gnaeus Pompeius Magnus (106–48 bce), as cited by Plutarchus (*c.* 45 to *c.* 125).

Ref. 145

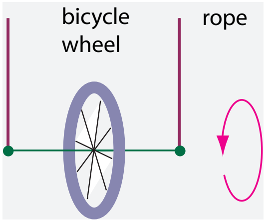


FIGURE 93 What happens when one rope is cut?

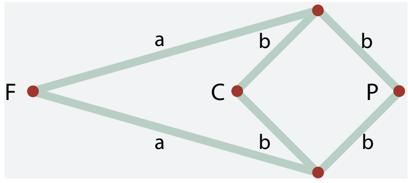


FIGURE 94 How to draw a straight line with a compass: fix point F, put a pencil into joint P and move C with a compass along a circle

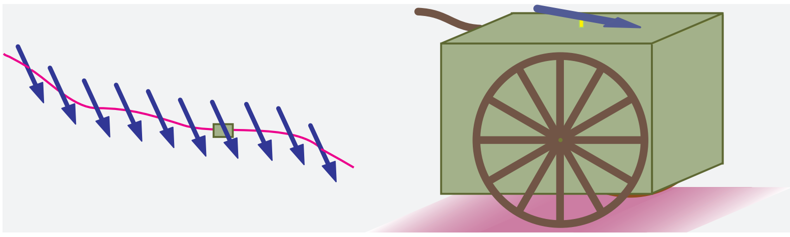


FIGURE 95 A south-pointing carriage

on the ground will usually differ. The relationships between these observations lead us to a global description, valid for everybody. This approach leads us to the theory of relativity.

— The third global approach to motion is to exploring the motion of *extended and rigid* bodies, rather than mass points. The counter-intuitive result of the experiment in Figure 93 shows why this is worthwhile.

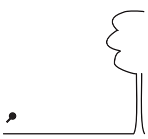
In order to design machines, it is essential to understand how a group of rigid bodies interact with one another. As an example, the mechanism in Figure 94 connects the motion of points C and P. It implicitly defines a circle such that one always has the relation $r_C = 1/r_P$ between the distances of C and P from its centre. Can you find that circle? Challenge 344 ny

Ref. 146 Another famous challenge is to devise a wooden carriage, with gearwheels that connect the wheels to an arrow in such a way that whatever path the carriage takes, the arrow always points south (see Figure 95). The solution to this is useful in helping us to understand general relativity, as we will see. Challenge 345 d

Another interesting example of rigid motion is the way that human movements, such as the general motions of an arm, are composed from a small number of basic motions. All these examples are from the fascinating field of engineering; unfortunately, we will have little time to explore this topic in our hike. Ref. 147

— The fourth global approach to motion is the description of *non-rigid extended bodies*. For example, *fluid mechanics* studies the flow of fluids (like honey, water or air) around solid bodies (like spoons, ships, sails or wings). Fluid mechanics thus seeks to explain how insects, birds and aeroplanes fly,* why sailing-boats can sail against the wind, what

* The mechanisms of insect flight are still a subject of active research. Traditionally, fluid dynamics has



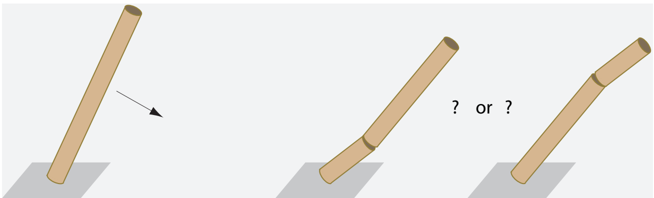


FIGURE 96 How and where does a falling brick chimney break?

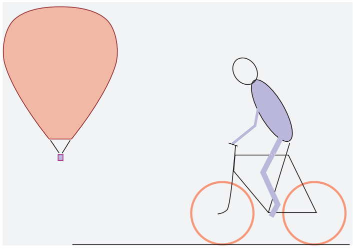


FIGURE 97 Why do hot-air balloons stay inflated? How can you measure the weight of a bicycle rider using only a ruler?

FIGURE 98 What determines the number of petals in a daisy?

Ref. 148
Challenge 346 n

happens when a hard-boiled egg is made to spin on a thin layer of water, or how a bottle full of wine can be emptied in the fastest way possible.

As well as fluids, we can study the behaviour of deformable *solids*. This area of research is called *continuum mechanics*. It deals with deformations and oscillations of extended structures. It seeks to explain, for example, why bells are made in particular shapes; how large bodies – such as falling chimneys – break when under stress; and how cats can turn themselves the right way up as they fall. During the course of our journey we will repeatedly encounter issues from this field, which impinges even upon general relativity and the world of elementary particles.

Challenge 347 n

— The fifth global approach to motion is the study of the motion of huge numbers of particles. This is called *statistical mechanics*. The concepts needed to describe gases, such as temperature and pressure (see Figure 97), will be our first steps towards the understanding of black holes.
— The sixth global approach to motion involves all of the above-mentioned viewpoints *at the same time*. Such an approach is needed to understand everyday experience, and *life* itself. Why does a flower form a specific number of petals? How does an embryo differentiate in the womb? What makes our hearts beat? How do mountains ridges and cloud patterns emerge? How do stars and galaxies evolve? How are sea waves formed by the wind?

concentrated on large systems, like boats, ships and aeroplanes. Indeed, the smallest human-made object that can fly in a controlled way – say, a radio-controlled plane or helicopter – is much larger and heavier than many flying objects that evolution has engineered. It turns out that controlling the flight of small things requires more knowledge and more tricks than controlling the flight of large things. There is more about this topic on page 991.

All these are examples of *self-organization*; life scientists simply speak of *growth*. Whatever we call these processes, they are characterized by the spontaneous appearance of patterns, shapes and cycles. Such processes are a common research theme across many disciplines, including biology, chemistry, medicine, geology and engineering.

We will now give a short introduction to these six global approaches to motion. We will begin with the first approach, namely, the global description of moving point-like objects. The beautiful method described below was the result of several centuries of collective effort, and is the highlight of mechanics. It also provides the basis for all the further descriptions of motion that we will meet later on.

Measuring change with action

Motion can be described by numbers. For a single particle, the relations between the spatial and temporal coordinates describe the motion. The realization that expressions like $(x(t), y(t), z(t))$ could be used to describe the path of a moving particle was a milestone in the development of modern physics.

We can go further. Motion is a type of change. And this change can itself be usefully described by numbers. In fact, change can be measured by a single number. This realization was the next important milestone. Physicists took almost two centuries of attempts to uncover the way to describe change. As a result, the quantity that measures change has a strange name: it is called *(physical) action*.* To remember the connection of 'action' with change, just think about a Hollywood film: a lot of action means a large amount of change.

Imagine taking two snapshots of a system at different times. How could you define the amount of change that occurred in between? When do things change a lot, and when do they change only a little? First of all, a system with a lot of motion shows a lot of change. So it makes sense that the action of a system composed of independent subsystems should be the sum of the actions of these subsystems.

Secondly, change often – but not always – builds up over time; in other cases, recent change can compensate for previous change. Change can thus increase or decrease with time.

Thirdly, for a system in which motion is stored, transformed or shifted from one subsystem to another, the change is smaller than for a system where this is not the case.

* Note that this 'action' is not the same as the 'action' appearing in statements such as 'every action has an equal and opposite reaction'. This last usage, coined by Newton, has not stuck; therefore the term has been recycled. After Newton, the term 'action' was first used with an intermediate meaning, before it was finally given the modern meaning used here. This last meaning is the only meaning used in this text.

Another term that has been recycled is the 'principle of least action'. In old books it used to have a different meaning from the one in this chapter. Nowadays, it refers to what used to be called *Hamilton's principle* in the Anglo-Saxon world, even though it is (mostly) due to others, especially Leibniz. The old names and meanings are falling into disuse and are not continued here.

Behind these shifts in terminology is the story of an intense two-centuries-long attempt to describe motion with so-called *extremal* or *variational principles*: the objective was to complete and improve the work initiated by Leibniz. These principles are only of historical interest today, because all are special cases of the
Ref. 149 principle of least action described here.

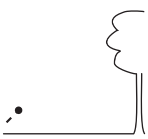


TABLE 22 Some action values for changes either observed or imagined

Change	Approximate action value
Smallest measurable change	$0.5 \cdot 10^{-34}$ Js
Exposure of photographic film	$1.1 \cdot 10^{-34}$ Js to 10^{-9} Js
Wing beat of a fruit fly	*c.* 1 pJs
Flower opening in the morning	*c.* 1 nJs
Getting a red face	*c.* 10 mJs
Held versus dropped glass	0.8 Js
Tree bent by the wind from one side to the other	500 Js
Making a white rabbit vanish by 'real' magic	100 PJs
Hiding a white rabbit	*c.* 0.1 Js
Maximum brain change in a minute	*c.* 5 Js
Levitating yourself within a minute by 1 m	*c.* 40 kJs
Car crash	*c.* 2 kJs
Birth	*c.* 2 kJs
Change due to a human life	*c.* 1 EJs
Driving car stops within the blink of an eye	20 kJs
Large earthquake	*c.* 1 PJs
Driving car disappears within the blink of an eye	1 ZJs
Sunrise	*c.* 0.1 ZJs
Gamma ray burster before and after explosion	*c.* 10^{46} Js
Universe after one second has elapsed	undefined and undefinable

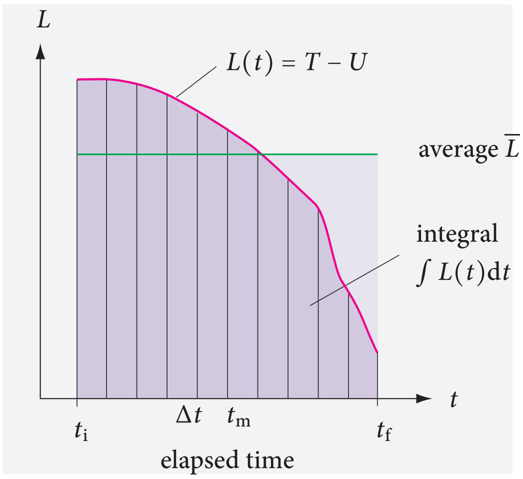
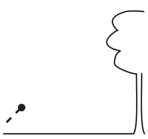


FIGURE 99 Defining a total effect as an accumulation (addition, or integral) of small effects over time



Joseph Lagrange

The mentioned properties imply that the natural measure of change is the average difference between kinetic and potential energy multiplied by the elapsed time. This quantity has all the right properties: it is (usually) the sum of the corresponding quantities for all subsystems if these are independent; it generally increases with time (unless the evolution compensates for something that happened earlier); and it decreases if the system transforms motion into potential energy.

Challenge 348 e

Thus the (physical) *action* S, measuring the change in a system, is defined as

$$S = \overline{L} \cdot (t_f - t_i) = \overline{T - U} \cdot (t_f - t_i) = \int_{t_i}^{t_f} (T - U)\, \mathrm{d}t = \int_{t_i}^{t_f} L\, \mathrm{d}t \,, \qquad (57)$$

Page 123

where T is the kinetic energy, U the potential energy we already know, L is the difference between these, and the overbar indicates a time average. The quantity L is called the *Lagrangian (function)* of the system,* describes what is being added over time, whenever things change. The sign $\int$ is a stretched 'S', for 'sum', and is pronounced 'integral of'. In intuitive terms it designates the operation (called *integration*) of adding up the values of a varying quantity in infinitesimal time steps $\mathrm{d}t$. The initial and the final times are written below and above the integration sign, respectively. Figure 99 illustrates the idea: the integral is simply the size of the dark area below the curve $L(t)$.

Challenge 349 e

Mathematically, the integral of the curve $L(t)$ is defined as

$$\int_{t_i}^{t_f} L(t)\, \mathrm{d}t = \lim_{\Delta t \to 0} \sum_{\mathrm{m=i}}^{\mathrm{f}} L(t_\mathrm{m}) \Delta t = \overline{L} \cdot (t_f - t_i) \,. \qquad (58)$$

In other words, the integral is the limit, as the time slices get smaller, of the sum of the areas of the individual rectangular strips that approximate the function.** Since the $\sum$ sign also means a sum, and since an infinitesimal Δt is written $\mathrm{d}t$, we can understand the notation used for integration. Integration is a sum over slices. The notation was developed by Gottfried Leibniz to make exactly this point. Physically speaking, the integral of the Lagrangian measures the *effect* that L builds up over time. Indeed, action is called 'effect' in some languages, such as German.

In short, then, action is the integral of the Lagrangian over time. The unit of action, and thus of physical change, is the unit of energy (the Joule), times the unit of time (the second). Thus change is measured in Js. A large value means a big change. Table 22 shows some approximate values of actions.

* It is named after Giuseppe Lodovico Lagrangia (b. 1736 Torino, d. 1813 Paris), better known as Joseph Louis Lagrange. He was the most important mathematician of his time; he started his career in Turin, then worked for 20 years in Berlin, and finally for 26 years in Paris. Among other things he worked on number theory and analytical mechanics, where he developed most of the mathematical tools used nowadays for calculations in classical mechanics and classical gravitation. He applied them successfully to many motions in the solar system.

** For more details on integration see Appendix D.

FIGURE 100 The minimum of a curve has vanishing slope

To understand the definition of action in more detail, we will start with the simplest case: a system for which the potential energy is zero, such as a particle moving freely. Obviously, a large kinetic energy means a lot of change. If we observe the particle at two instants, the more distant they are the larger the change. Furthermore, the observed change is larger if the particle moves more rapidly, as its kinetic energy is larger. This is not surprising.

Next, we explore a single particle moving in a potential. For example, a falling stone loses potential energy in exchange for a gain in kinetic energy. The more energy is exchanged, the more change there is. Hence the minus sign in the definition of L. If we explore a particle that is first thrown up in the air and then falls, the curve for $L(t)$ first is below the times axis, then above. We note that the definition of integration makes us count the grey surface *below* the time axis *negatively*. Change can thus be negative, and be compensated by subsequent change, as expected.

To measure change for a system made of several independent components, we simply add all the kinetic energies and subtract all the potential energies. This technique allows us to define actions for gases, liquids and solid matter. Even if the components interact, we still get a sensible result. In short, action is an *additive* quantity.

Physical action thus measures, in a single number, the change observed in a system between two instants of time. The observation may be anything at all: an explosion, a caress or a colour change. We will discover later that this idea is also applicable in relativity and quantum theory. Any change going on in any system of nature can be measured with a single number.

The principle of least action

We now have a precise measure of change, which, as it turns out, allows a simple and powerful description of motion. In nature, the change happening between two instants is always the *smallest* possible. *In nature, action is minimal.** Of all possible motions, nature always chooses for which the change is *minimal*. Let us study a few examples.

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

Challenge 350 e

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

When gravity is present, a thrown stone flies along a parabola (or more precisely, along an ellipse) because any other path, say one in which the stone makes a loop in the air, Challenge 351 e would imply a *larger* action. Again you might want to verify this for yourself.

All observations support this simple and basic statement: things always move in a way that produces the smallest possible value for the action. This statement applies to the full path and to any of its segments. Betrand Russell called it the 'law of cosmic laziness'.

It is customary to express the idea of minimal change in a different way. The action varies when the path is varied. The actual path is the one with the smallest action. You will recall from school that at a minimum the derivative of a quantity vanishes: a minimum has a horizontal slope. In the present case, we do not vary a quantity, but a complete path; hence we do not speak of a derivative or slope, but of a variation. It is customary to write the variation of action as δS. The *principle of least action* thus states:

$$\triangleright \textit{ The actual trajectory between specified end points satisfies } \delta S = 0. \qquad (59)$$

Mathematicians call this a *variational principle*. Note that the end points have to be specified: we have to compare motions with the *same* initial and final situations.

Before discussing the principle further, we can check that it is equivalent to the evolution equation.* To do this, we can use a standard procedure, part of the so-called *calculus*

* For those interested, here are a few comments on the equivalence of Lagrangians and evolution equations. Page 152 First of all, Lagrangians do not exist for non-conservative, or *dissipative* systems. We saw that there is no potential for any motion involving *friction* (and more than one dimension); therefore there is no action in these cases. One approach to overcome this limitation is to use a generalized formulation of the principle of least action. Whenever there is no potential, we can express the *work* variation δW between different trajectories x_i as

$$\delta W = \sum_i m_i \ddot{x}_i \delta x_i \,. \qquad (60)$$

Motion is then described in the following way:

$$\triangleright \textit{ The actual trajectory satifies } \int_{t_\mathrm{i}}^{t_\mathrm{f}} (\delta T + \delta W)\mathrm{d}t = 0 \quad \textit{provided} \quad \delta x(t_\mathrm{i}) = \delta x(t_\mathrm{f}) = 0\,. \qquad (61)$$

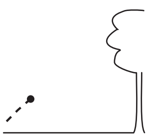
The quantity being varied has no name; it represents a generalized notion of change. You might want to Challenge 352 ny check that it leads to the correct evolution equations. Thus, although *proper* Lagrangian descriptions exist only for *conservative* systems, for dissipative systems the principle can be generalized and remains useful.

Many physicists will prefer another approach. What a mathematician calls a generalization is a special case for a physicist: the principle (61) hides the fact that *all* friction results from the usual principle of minimal action, if we include the complete microscopic details. There is no friction in the microscopic domain. Friction is an approximate, macroscopic concept.

Nevertheless, more mathematical viewpoints are useful. For example, they lead to interesting limitations for the use of Lagrangians. These limitations, which apply only if the world is viewed as purely classical – which it isn't – were discovered about a hundred years ago. In those times computers where not available, and the exploration of new calculation techniques was important. Here is a summary.

The coordinates used in connection with Lagrangians are not necessarily the Cartesian ones. *Generalized* coordinates are especially useful when there are *constraints* on the motion. This is the case for a pendulum, where the weight always has to be at the same distance from the suspension, or for an ice skater, where the Ref. 150 skate has to move in the direction in which it is pointing. Generalized coordinates may even be mixtures of positions and momenta. They can be divided into a few general types.

Generalized coordinates are called *holonomic–scleronomic* if they are related to Cartesian coordinates in a fixed way, independently of time: physical systems described by such coordinates include the pendulum and a particle in a potential. Coordinates are called *holonomic–rheonomic* if the dependence involves time.



of variations. The condition $\delta S = 0$ implies that the action, i.e. the area under the curve in Figure 99, is a minimum. A little bit of thinking shows that if the Lagrangian is of the form $L(x_n, v_n) = T(v_n) - U(x_n)$, then

Challenge 353 ny

$$\frac{d}{dt}\left(\frac{\partial T}{\partial v_n}\right) = \frac{\partial U}{\partial x_n} \tag{62}$$

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

Challenge 354 e

$$m\mathbf{a} = \nabla U \ . \tag{64}$$

This is the evolution equation: it says that the force on a particle is the gradient of the potential energy U. The principle of least action thus implies the equation of motion. (Can you show the converse?)

Challenge 355 n

In other words, *all systems evolve in such a way that the change is as small as possible.* Nature is economical. Nature is thus the opposite of a Hollywood thriller, in which the action is maximized; nature is more like a wise old man who keeps his actions to a minimum.

The principle of minimal action also states that the actual trajectory is the one for which the *average* of the Lagrangian over the whole trajectory is minimal (see Figure 99). Nature is a Dr. Dolittle. Can you verify this? This viewpoint allows one to deduce Lagrange's equations (62) directly.

Challenge 356 ny

The principle of least action distinguishes the actual trajectory from all other imaginable ones. This observation lead Leibniz to his famous interpretation that the actual world is the 'best of all possible worlds.'** We may dismiss this as metaphysical speculation, but

An example of a rheonomic systems would be a pendulum whose length depends on time. The two terms rheonomic and scleronomic are due to Ludwig Boltzmann. These two cases, which concern systems that are only described by their geometry, are grouped together as *holonomic systems*. The term is due to Heinrich Hertz.

Page 249

Page 570

The more general situation is called *anholonomic*, or *nonholonomic*. Lagrangians work well only for holonomic systems. Unfortunately, the meaning of the term 'nonholonomic' has changed. Nowadays, the term is also used for certain rheonomic systems. The modern use calls nonholonomic any system which involves velocities. Therefore, an ice skater or a rolling disc is often called a nonholonomic system. Care is thus necessary to decide what is meant by nonholonomic in any particular context.

Even though the use of Lagrangians, and of action, has its limitations, these need not bother us at microscopic level, since microscopic systems are always conservative, holonomic and scleronomic. At the fundamental level, evolution equations and Lagrangians are indeed equivalent.

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

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_n}\right) = \frac{\partial L}{\partial q_n} \ . \tag{63}$$

In order to deduce these equations, we also need the relation $\delta\dot{q} = d/dt(\delta q)$. This relation is valid only for *holonomic* coordinates introduced in the previous footnote and explains their importance.

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

Ref. 151

By the way, the letter q for position and p for momentum were introduced in physics by the mathematician Carl Jacobi (b. 1804 Potsdam, d. 1851 Berlin).

** This idea was ridiculed by the French philosopher Voltaire (1694–1778) in his lucid writings, notably in we should still be able to feel the fascination of the issue. Leibniz was so excited about the principle of least action because it was the first time that actual observations were distinguished from all other imaginable possibilities. For the first time, the search for reasons why things are the way they are became a part of physical investigation. Could the world be different from what it is? In the principle of least action, we have a hint of a negative answer. (What do you think?) The final answer will emerge only in the last part of our adventure.

Challenge 357 n

As a way to describe motion, the Lagrangian has several advantages over the evolution equation. First of all, the Lagrangian is usually more *compact* than writing the corresponding evolution equations. For example, only *one* Lagrangian is needed for one system, however many particles it includes. One makes fewer mistakes, especially sign mistakes, as one rapidly learns when performing calculations. Just try to write down the evolution equations for a chain of masses connected by springs; then compare the effort with a derivation using a Lagrangian. (The system behaves like a chain of atoms.) We will encounter another example shortly: David Hilbert took only a few weeks to deduce the equations of motion of general relativity using a Lagrangian, whereas Albert Einstein had worked for ten years searching for them directly.

Challenge 358 ny

In addition, the description with a Lagrangian is valid with *any* set of coordinates describing the objects of investigation. The coordinates do not have to be Cartesian; they can be chosen as one prefers: cylindrical, spherical, hyperbolic, etc. These so-called *generalized coordinates* allow one to rapidly calculate the behaviour of many mechanical systems that are in practice too complicated to be described with Cartesian coordinates. For example, for programming the motion of robot arms, the angles of the joints provide a clearer description than Cartesian coordinates of the ends of the arms. Angles are non-Cartesian coordinates. They simplify calculations considerably: the task of finding the most economical way to move the hand of a robot from one point to another can be solved much more easily with angular variables.

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

Page 197

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

As the principle of least action became well known, people applied it to an ever-increasing number of problems. Today, Lagrangians are used in everything from the study of elementary particle collisions to the programming of robot motion in artificial intelligence. However, we should not forget that despite its remarkable simplicity and usefulness, the Lagrangian formulation is *equivalent* to the evolution equations. It is neither more general nor more specific. In particular, it is *not an explanation* for any type of motion, but only a view of it. In fact, the search of a new physical 'law' of motion is *just* the search for a new Lagrangian. This makes sense, as the description of nature always requires the description of change. Change in nature is always described by actions and Lagrangians.

Ref. 149

Challenge 359 n

the brilliant book *Candide*, written in 1759, and still widely available.

The principle of least action states that the action is minimal when the end point of Ref. 152 the motion, and in particular the time between them, are fixed. It is less well known that the reciprocal principle also holds: if the action is kept fixed, the elapsed time is maximal. Challenge 360 ny Can you show this?

Even though the principle of least action is not an explanation of motion, it somehow calls for one. We need some patience, though. *Why* nature follows the principle of least action, and *how* it does so, will become clear when we explore quantum theory.

> Never confuse movement with action.
> Ernest Hemingway

Ref. 153

Why is motion so often bounded?

> The optimist thinks this is the best of all possible worlds, and the pessimist knows it.
> Robert Oppenheimer

Looking around ourselves on Earth and in the sky, we find that matter is not evenly distributed. Matter tends to be near other matter: it is lumped together in *aggregates*. Some Ref. 154 major examples of aggregates are given in Figure 101 and Table 23. In the mass–size diagram of Figure 101, both scales are logarithmic. One notes three straight lines: a line $m \sim l$ extending from the Planck mass* upwards, via black holes, to the universe itself; a line $m \sim 1/l$ extending from the Planck mass downwards, to the lightest possible aggregate; and the usual matter line with $m \sim l^3$, extending from atoms upwards, via the Earth and the Sun. The first of the lines, the black hole limit, is explained by general relativity; the last two, the aggregate limit and the common matter line, by quantum theory.**

The aggregates outside the common matter line also show that the stronger the interaction that keeps the components together, the smaller the aggregate. But why is matter mainly found in lumps?

First of all, aggregates form because of the existence of *attractive* interactions between objects. Secondly, they form because of *friction*: when two components approach, an aggregate can only be formed if the released energy can be changed into heat. Thirdly, aggregates have a finite size because of *repulsive* effects that prevent the components from collapsing completely. Together, these three factors ensure that bound motion is much more common than unbound, 'free' motion.

Only three types of attraction lead to aggregates: gravity, the attraction of electric charges, and the strong nuclear interaction. Similarly, only three types of repulsion are observed: rotation, pressure, and the Pauli exclusion principle (which we will encounter Page 806 later on). Of the nine possible combinations of attraction and repulsion, not all appear in Challenge 361 n nature. Can you find out which ones are missing from Figure 101 and Table 23, and why?

Together, attraction, friction and repulsion imply that change and action are minimized when objects come and stay together. The principle of least action thus implies the stability of aggregates. By the way, formation history also explains why so many aggreg-Challenge 362 ny ates *rotate*. Can you tell why?

Page 1173 * The Planck mass is given by $m_{Pl} = \sqrt{\hbar c/G} = 21.767(16)\ \mu g$.

** Figure 101 suggests that domains beyond physics exist; we will discover later on that this is not the case, as mass and size are not definable in those domains.

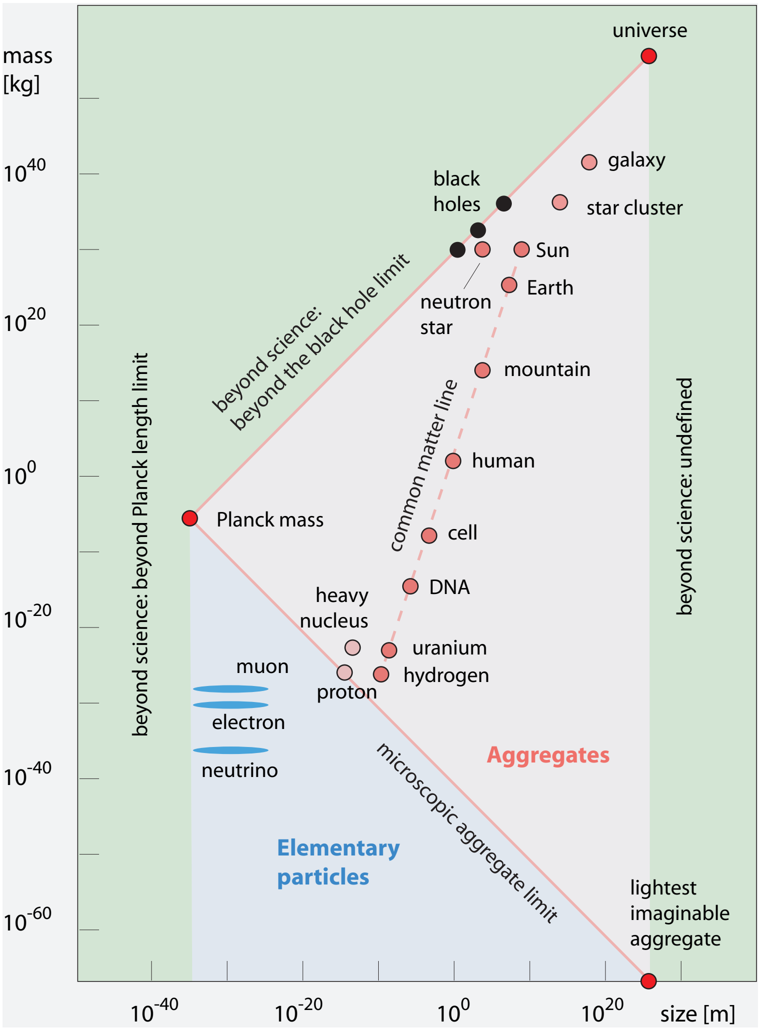
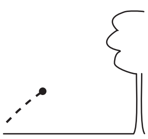


FIGURE 101 Aggregates in nature

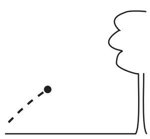
But why does friction exist at all? And why do attractive and repulsive interactions exist? And why is it – as it would appear from the above – that in some distant past matter was *not* found in lumps? In order to answer these questions, we must first study another global property of motion: symmetry.

TABLE 23 Some major aggregates observed in nature

Aggregate	Size (diameter)	Obs. num.	Constituents
gravitationally bound aggregates			
matter across universe	*c.* 100 Ym	1	superclusters of galaxies, hydrogen and helium atoms
quasar	10^{12} to 10^{14} m	$20 \cdot 10^{6}$	baryons and leptons



Aggregate	Size (diameter)	Obs. num.	Constituents
supercluster of galaxies	*c.* 3 Ym	10^7	galaxy groups and clusters
galaxy cluster	*c.* 60 Zm	$25 \cdot 10^9$	10 to 50 galaxies
galaxy group or cluster	*c.* 240 Zm		50 to over 2000 galaxies
our local galaxy group	50 Zm	1	*c.* 40 galaxies
general galaxy	0.5 to 2 Zm	$3.5 \cdot 10^{12}$	10^{10} to $3 \cdot 10^{11}$ stars, dust and gas clouds, probably solar systems
our galaxy	1.0(0.1) Zm	1	10^{11} stars, dust and gas clouds, solar systems
interstellar clouds	up to 15 Em	$\gg 10^5$	hydrogen, ice and dust
solar system [a]	unknown	> 100	star, planets
our solar system	30 Pm	1	Sun, planets (Pluto's orbit's diameter: 11.8 Tm), moons, planetoids, comets, asteroids, dust, gas
Oort cloud	6 to 30 Pm	1	comets, dust
Kuiper belt	60 Tm	1	planetoids, comets, dust
star [b]	10 km to 100 Gm	$10^{22\pm1}$	ionized gas: protons, neutrons, electrons, neutrinos, photons
our star	1.39 Gm		
planet [a] (Jupiter, Earth)	143 Mm, 12.8 Mm	9+*c.* 100	solids, liquids, gases; in particular, heavy atoms
planetoids (Varuna, etc)	50 to 1 000 km	*c.* 10 (est. 10^9)	solids
moons	10 to 1 000 km	*c.* 50	solids
neutron stars	10 km	*c.* 1000	mainly neutrons
electromagnetically bound aggregates [c]			
asteroids, mountains [d]	1 m to 930 km	>26 000	(10^9 estimated) solids, usually monolithic
comets	10 cm to 50 km	$> 10^6$	ice and dust
planetoids, solids, liquids, gases, cheese	1 nm to > 100 km	n.a.	molecules, atoms
animals, plants, kefir	5 μm to 1 km	$10^{26\pm2}$	organs, cells
brain	0.15 m	10^{10}	neurons and other cell types
cells:		$10^{31\pm1}$	organelles, membranes, molecules
smallest (*Nanoarchaeum equitans*)	*c.* 400 nm		molecules
amoeba	600 μm		molecules
largest (whale nerve, single-celled plants)	*c.* 30 m		molecules
molecules:		*c.* $10^{78\pm2}$	atoms
H_2	*c.* 50 pm	$10^{72\pm2}$	atoms
DNA (human)	2 m (total per cell)	10^{21}	atoms



Aggregate	Size (diameter)	Obs. num.	Constituents
atoms, ions	30 pm to 300 pm	$10^{80\pm2}$	electrons and nuclei
aggregates bound by the weak interaction [c]			
none			
aggregates bound by the strong interaction [c]			
nucleus	$> 10^{-15}$ m	$10^{79\pm2}$	nucleons
nucleon (proton, neutron)	*c.* 10^{-15} m	$10^{80\pm2}$	quarks
mesons	*c.* 10^{-15} m	n.a.	quarks
neutron stars: see above			

a. Only in 1994 was the first evidence found for objects circling stars other than our Sun; of over 100 *extrasolar planets* found so far, most are found around F, G and K stars, including neutron stars. For example, three
Ref. 155 objects circle the pulsar PSR 1257+12, and a matter ring circles the star β Pictoris. The objects seem to be dark stars, brown dwarfs or large gas planets like Jupiter. Due to the limitations of observation systems, none of the systems found so far form solar systems of the type we live in. In fact, only a few Earth-like planets have been found so far.

b. The Sun is among the brightest 7 % of stars. Of all stars, 80 %, are red M dwarfs, 8 % are orange K dwarfs, and 5 % are white D dwarfs: these are all faint. Almost all stars visible in the night sky belong to the bright 7 %. Some of these are from the rare blue O class or blue B class (such as Spica, Regulus and Riga); 0.7 % consist of the bright, white A class (such as Sirius, Vega and Altair); 2 % are of the yellow–white F class (such as Canopus, Procyon and Polaris); 3.5 % are of the yellow G class (like Alpha Centauri, Capella or the Sun). Exceptions include the few visible K giants, such as Arcturus and Aldebaran, and the rare M supergiants,
Page 474 such as Betelgeuse and Antares. More on stars later on.

c. For more details on *microscopic* aggregates, see the table of composites in Appendix C.

d. It is estimated that there are about 10^9 asteroids (or planetoids) larger than 1 km and about 10^{20} that are
Ref. 156 heavier than 100 kg. By the way, no asteroids between Mercury and the Sun – the hypothetical *Vulcanoids* – have been found so far.

Curiosities and fun challenges about Lagrangians

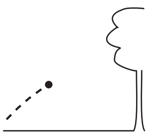
When Lagrange published his book *Mécanique analytique*, in 1788, it formed one of the high points in the history of mechanics. He was proud of having written a systematic exposition of mechanics without a single figure. Obviously the book was difficult to read and was not a sales success. Therefore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

* *

Given that action is the basic quantity describing motion, we can define energy as action per unit time, and momentum as action per unit distance. The *energy* of a system thus describes how much it changes over time, and the *momentum* how much it changes over
Challenge 363 n distance. What are angular momentum and rotational energy?

* *

'In nature, effects of telekinesis or prayer are impossible, as in most cases the change inside



the brain is much smaller than the change claimed in the outside world.' Is this argument
Challenge 364 n correct?

* *

In Galilean physics, the Lagrangian is the difference between kinetic and potential energy. Later on, this definition will be generalized in a way that sharpens our understanding of this distinction: the Lagrangian becomes the difference between a term for free particles and a term due to their interactions. In other words, particle motion is a continuous compromise between what the particle would do if it were free and what other particles want it to do. In this respect, particles behave a lot like humans beings.

* *

Challenge 365 ny Explain: why is $T + U$ constant, whereas $T - U$ is minimal?

* *

In nature, the sum $T + U$ of kinetic and potential energy is *constant* during motion (for closed systems), whereas the average of the difference $T - U$ is *minimal*. Is it possible to deduce, by combining these two facts, that systems tend to a state with minimum potential
Challenge 366 ny energy?

* *

There is a principle of *least effort* describing the growth of trees. When a tree – a *monopodal phanerophyte* – grows and produces leaves, between 40% and 60% of the mass it consists of, namely the water and the minerals, has to be lifted upwards from the ground.* Therefore, a tree gets as many branches as high up in the air as possible using the smallest amount of energy. This is the reason why not all leaves are at the very top of a tree. Can
Challenge 367 ny you deduce more details about trees from this principle?

* *

Another minimization principle can be used to understand the construction of animal
Ref. 157 bodies, especially their size and the proportions of their inner structures. For example, the heart pulse and breathing frequency both vary with animal mass m as $m^{-1/4}$, and the dissipated power varies as $m^{3/4}$. It turns out that such exponents result from three properties of living beings. First, they transport energy and material through the organism via a branched network of vessels: a few large ones, and increasingly many smaller ones. Secondly, the vessels all have the same minimum size. And thirdly, the networks are optimized in order to minimize the energy needed for transport. Together, these relations explain many additional scaling rules; they might also explain why animal lifespan scales as $m^{-1/4}$, or why most mammals have roughly the same number of heart beats in a lifetime.

A competing explanation, using a different minimization principle, states that quarter powers arise in any network built in order that the flow arrives to the destination by the
Ref. 158 most direct path.

* The rest of the mass comes from the CO_2 in the air.

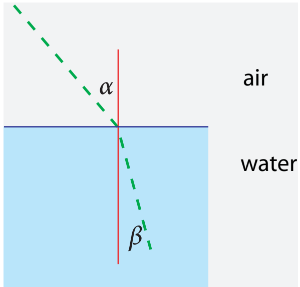


FIGURE 102 Refraction of light is due to travel-time optimization

* *

The minimization principle for the motion of light is even more beautiful: light always takes the path that requires the shortest travel time. It was known long ago that this idea describes exactly how light changes direction when it moves from air to water. In water, light moves more slowly; the speed ratio between air and water is called the *refractive index* of water. The refractive index, usually abbreviated n, is material-dependent. The value for water is about 1.3. This speed ratio, together with the minimum-time principle, leads to the 'law' of refraction, a simple relation between the sines of the two angles. Can you deduce it? (In fact, the exact definition of the refractive index is with respect to vacuum, not to air. But the difference is negligible: can you imagine why?)

Challenge 368 n

Challenge 369 n

For diamond, the refractive index is 2.4. The high value is one reason for the sparkle of diamonds cut with the 57-face *brilliant* cut. Can you think of some other reasons?

Challenge 370 n

* *

Can you confirm that each of these minimization principles is a special case of the principle of least action? In fact, this is the case for *all* known minimization principles in nature. Each of them, like the principle of least action, is a principle of least change.

Challenge 371 n

* *

In Galilean physics, the value of the action depends on the speed of the observer, but not on his position or orientation. But the action, when properly defined, should *not* depend on the observer. All observers should agree on the value of the observed change. Only special relativity will fulfil the requirement that action be independent of the observer's speed. How will the relativistic action be defined?

Challenge 372 n

* *

Measuring all the change that is going on in the universe presupposes that the universe is a physical system. Is this the case?

Challenge 373 n

* *

One motion for which action is particularly well minimized in nature is dear to us: walking. Extensive research efforts try to design robots which copy the energy saving functioning and control of human legs. For an example, see the website by Tao Geng at http://

Ref. 159

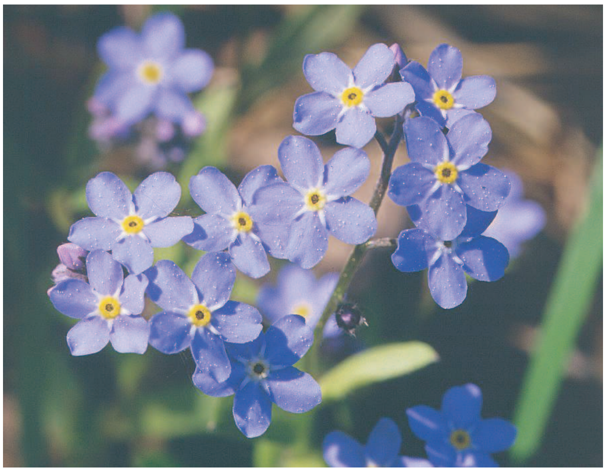


FIGURE 103 Forget-me-not, also called *Myosotis* (Boraginaceae) (© Markku Savela)

www.cn.stir.ac.uk/~tgeng/research.html.

Motion and symmetry

The second way to describe motion globally is to describe it in such a way that *all* observers agree. An object under observation is called *symmetric* if it looks the same when seen from different points of view. For example, a forget-me-not flower, shown in Figure 103, is symmetrical because it looks the same after turning around it by 72 degrees; many fruit tree flowers have the same symmetry. One also says that under change of viewpoint the flower has an *invariant property*, namely its shape. If many such viewpoints are possible, one talks about a *high* symmetry, otherwise a *low* symmetry. For example, a four-leaf clover has a higher symmetry than a usual, three-leaf one. Different points of view imply different observers; in physics, the viewpoints are often called *frames of reference* and are described mathematically by coordinate systems.

High symmetry means many agreeing observers. At first sight, not many objects or observations in nature seem to be symmetrical. But this is a mistake. On the contrary, we can deduce that nature as a whole is symmetric from the simple fact that we have the ability to talk about it! Moreover, the symmetry of nature is considerably higher than that of a forget-me-not. We will discover that this high symmetry is at the basis of the famous expression $E_0 = mc^2$.

Challenge 374 n

Why can we think and talk?

> “The hidden harmony is stronger than the apparent.
> Heraclitus of Ephesos, about 500 BCE”

Ref. 160

Why can we understand somebody when he is talking about the world, even though we are not in his shoes? We can for two reasons: because most things look *similar* from different viewpoints, and because most of us have already had similar experiences *beforehand*.

'Similar' means that what *we* and what *others* observe somehow correspond. In other words, many aspects of observations do not depend on viewpoint. For example, the number of petals of a flower has the same value for all observers. We can therefore say that this quantity has the highest possible symmetry. We will see below that mass is another such example. Observables with the highest possible symmetry are called *scalars* in physics. Other aspects change from observer to observer. For example, the apparent size varies with the distance of observation. However, the actual size is observer-independent. In general terms, any type of *viewpoint-independence* is a form of symmetry, and the observation that two people looking at the same thing from different viewpoints can understand each other proves that nature is symmetric. We start to explore the details of this symmetry in this section and we will continue during most of the rest of our hike.

In the world around us, we note another general property: not only does the same phenomenon look similar to different observers, but *different* phenomena look similar to the *same* observer. For example, we know that if fire burns the finger in the kitchen, it will do so outside the house as well, and also in other places and at other times. Nature shows *reproducibility*. Nature shows no surprises. In fact, our memory and our thinking are only possible because of this basic property of nature. (Can you confirm this?) As we will see, reproducibility leads to additional strong restrictions on the description of nature.

Challenge 375 n

Without viewpoint-independence and reproducibility, talking to others or to oneself would be impossible. Even more importantly, we will discover that viewpoint-independence and reproducibility do more than determine the possibility of talking to each other: they also fix the *content* of what we can say to each other. In other words, we will see that our description of nature follows logically, almost without choice, from the simple fact that we can talk about nature to our friends.

Viewpoints

> Tolerance ... is the suspicion that the other might be right.
> Kurt Tucholski (1890–1935), German writer

> Tolerance – a strength one mainly wishes to political opponents.
> Wolfram Weidner (b. 1925) German journalist

When a young human starts to meet other people in childhood, it quickly finds out that certain experiences are shared, while others, such as dreams, are not. Learning to make this distinction is one of the adventures of human life. In these pages, we concentrate on a section of the first type of experiences: *physical* observations. However, even among these, distinctions are to be made. In daily life we are used to assuming that weights, volumes, lengths and time intervals are independent of the viewpoint of the observer. We can talk about these observed quantities to anybody, and there are no disagreements over their values, provided they have been measured correctly. However, other quantities do depend on the observer. Imagine talking to a friend after he jumped from one of the trees along our path, while he is still falling downwards. He will say that the forest floor is approaching with high speed, whereas the observer below will maintain that the floor is stationary. Obviously, the difference between the statements is due to their different viewpoints. The velocity of an object (in this example that of the forest floor or of the friend himself) is thus a less symmetric property than weight or size. Not all observers agree on its value.

In the case of viewpoint-dependent observations, understanding is still possible with the help of a little effort: each observer can *imagine* observing from the point of view of the other, and *check* whether the imagined result agrees with the statement of the other.* If the statement thus imagined and the actual statement of the other observer agree, the observations are consistent, and the difference in statements is due only to the different viewpoints; otherwise, the difference is fundamental, and they cannot agree or talk. Using this approach, you can even argue whether human feelings, judgements, or tastes arise from fundamental differences or not. Challenge 376 n

The distinction between viewpoint-independent (invariant) and viewpoint-dependent quantities is an essential one. Invariant quantities, such as mass or shape, describe *intrinsic* properties, and quantities depending on the observer make up the *state* of the system. Therefore, we must answer the following questions in order to find a *complete* description of the state of a physical system:

— Which viewpoints are possible?
— How are descriptions transformed from one viewpoint to another?
— Which observables do these symmetries admit?
— What do these results tell us about motion?

In the discussion so far, we have studied viewpoints differing in location, in orientation, in time and, most importantly, in motion. With respect to each other, observers can be at rest, move with constant speed, or accelerate. These 'concrete' changes of viewpoint are those we will study first. In this case the requirement of consistency of observations made by different observers is called the *principle of relativity*. The symmetries associated with this type of invariance are also called *external* symmetries. They are listed in Table 25. Page 87 Page 202

A second class of fundamental changes of viewpoint concerns 'abstract' changes. Viewpoints can differ by the mathematical description used: such changes are called *changes of gauge*. They will be introduced first in the section on electrodynamics. Again, it is required that all statements be consistent across different mathematical descriptions. This requirement of consistency is called the *principle of gauge invariance*. The associated symmetries are called *internal* symmetries.

The third class of changes, whose importance may not be evident from everyday life, is that of the behaviour of a system under exchange of its parts. The associated invariance is called *permutation symmetry*. It is a *discrete* symmetry, and we will encounter it in the second part of our adventure.

The three consistency requirements described above are called 'principles' because these basic statements are so strong that they almost completely determine the 'laws' of physics, as we will see shortly. Later on we will discover that looking for a complete description of the state of objects will also yield a complete description of their *intrinsic* properties. But enough of introduction: let us come to the heart of the topic.

* Humans develop the ability to imagine that others can be in situations *different* from their own at the age of about four years. Therefore, before the age of four, humans are unable to conceive special relativity; afterwards, they can. Ref. 161

Symmetries and groups

Since we are looking for a complete description of motion, we need to understand and describe the full set of symmetries of nature. A system is said to be symmetric or to possess a *symmetry* if it appears identical when observed from different viewpoints. We also say that the system possesses an *invariance* under change from one viewpoint to the other. Viewpoint changes are called *symmetry operations* or *transformations*. A symmetry is thus a transformation, or more generally, a set of transformations. However, it is more than that: the successive application of two symmetry operations is another symmetry operation. To be more precise, a symmetry is a set $G = \{a, b, c, ...\}$ of elements, the transformations, together with a binary operation $\circ$ called *concatenation* or *multiplication* and pronounced 'after' or 'times', in which the following properties hold for all elements a, b and c:

$$\begin{aligned} \textit{associativity}\text{, i.e.} \quad & (a \circ b) \circ c = a \circ (b \circ c) \\ \text{a } \textit{neutral element } e \text{ exists such that} \quad & e \circ a = a \circ e = a \\ \text{an } \textit{inverse element } a^{-1} \text{ exists such that} \quad & a^{-1} \circ a = a \circ a^{-1} = e \quad . \end{aligned} \tag{65}$$

Any set that fulfils these three defining properties, or axioms, is called a *(mathematical) group*. Historically, the notion of group was the first example of a mathematical structure which was defined in a completely abstract manner.* Can you give an example of a group taken from daily life? Groups appear frequently in physics and mathematics, because symmetries are almost everywhere, as we will see.** Can you list the symmetry operations of the pattern of Figure 104?

Challenge 377 n
Ref. 162
Challenge 378 n

Representations

Looking at a symmetric and composed system such as the one shown in Figure 104, we notice that each of its parts, for example each red patch, belongs to a set of similar objects, usually called a *multiplet*. Taken as a whole, the multiplet has (at least) the symmetry properties of the whole system. For some of the coloured patches in Figure 104 we need four objects to make up a full multiplet, whereas for others we need two, or only one, as in the case of the central star. In fact, in any symmetric system each part can be classified according to what type of multiplet it belongs to. Throughout our mountain ascent we will perform the same classification with every part of nature, with ever-increasing precision.

Challenge 379 e

A *multiplet* is a set of parts that transform into each other under all symmetry transformations. Mathematicians often call abstract multiplets *representations*. By specifying

* The term is due to Evariste Galois (1811–1832), the structure to Augustin-Louis Cauchy (1789–1857) and the axiomatic definition to Arthur Cayley (1821–1895).

** In principle, mathematical groups need not be symmetry groups; but it can be proven that all groups can be seen as transformation groups on some suitably defined mathematical space, so that in mathematics we can use the terms 'symmetry group' and 'group' interchangeably.

A group is called *Abelian* if its concatenation operation is commutative, i.e. if $a \circ b = b \circ a$ for all pairs of elements a and b. In this case the concatenation is sometimes called *addition*. Do rotations form an abelian group?

A subset $G_1 \subset G$ of a group G can itself be a group; one then calls it a *subgroup* and often says sloppily that G is *larger* than G_1 or that G is a *higher* symmetry group than G_1.

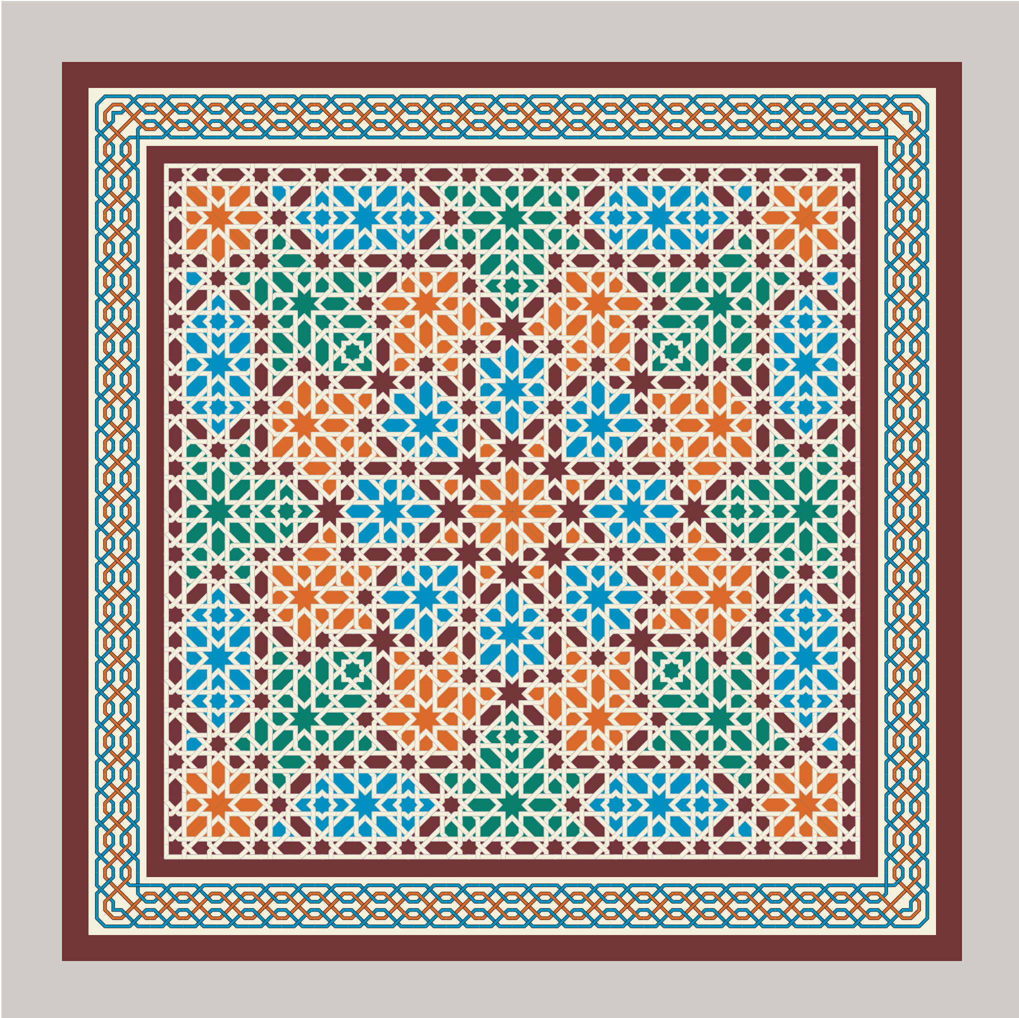


FIGURE 104 A Hispano–Arabic ornament from the Governor's Palace in Sevilla (© Christoph Schiller)

to which multiplet a component belongs, we describe in which way the component is part of the whole system. Let us see how this classification is achieved.

In mathematical language, symmetry transformations are often described by matrices. For example, in the plane, a reflection along the first diagonal is represented by the matrix

$$D(\text{refl}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad (66)$$

since every point (x, y) becomes transformed to (y, x) when multiplied by the matrix $D(\text{refl})$. Therefore, for a mathematician a *representation* of a symmetry group G is an assignment of a matrix $D(a)$ to each group element a such that the representation of the concatenation of two elements a and b is the product of the representations D of the

Challenge 380 e

elements:

$$D(a \circ b) = D(a)D(b) . \tag{67}$$

For example, the matrix of equation (66), together with the corresponding matrices for all the other symmetry operations, have this property.*

For every symmetry group, the construction and classification of all possible representations is an important task. It corresponds to the classification of all possible multiplets a symmetric system can be made of. In this way, understanding the classification of all multiplets and parts which can appear in Figure 104 will teach us how to classify all possible parts of which an object or an example of motion can be composed!

A representation D is called *unitary* if all matrices $D(a)$ are unitary.** Almost all representations appearing in physics, with only a handful of exceptions, are unitary: this term is the most restrictive, since it specifies that the corresponding transformations are one-to-one and invertible, which means that one observer never sees more or less than another. Obviously, if an observer can talk to a second one, the second one can also talk to the first.

The final important property of a multiplet, or representation, concerns its structure. If a multiplet can be seen as composed of sub-multiplets, it is called *reducible*, else *irreducible*; the same is said about representations. The irreducible representations obviously cannot be decomposed any further. For example, the (approximate) symmetry group of Figure 104, commonly called D_4, has eight elements. It has the general, faithful, unitary and irreducible matrix representation

Challenge 381 e

$$\begin{pmatrix} \cos n\pi/2 & -\sin n\pi/2 \\ \sin n\pi/2 & \cos n\pi/2 \end{pmatrix} n = 0..3, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}. \tag{69}$$

* There are some obvious, but important, side conditions for a representation: the matrices $D(a)$ must be invertible, or non-singular, and the identity operation of G must be mapped to the unit matrix. In even more compact language one says that a representation is a *homomorphism* from G into the group of non-singular or invertible matrices. A matrix D is invertible if its determinant $\det D$ is not zero.

In general, if a mapping f from a group G to another G' satisfies

$$f(a \circ_G b) = f(a) \circ_{G'} f(b) , \tag{68}$$

the mapping f is called an *homomorphism*. A homomorphism f that is one-to-one (injective) and onto (surjective) is called a *isomorphism*. If a representation is also injective, it is called *faithful*, *true* or *proper*.

In the same way as groups, more complex mathematical structures such as rings, fields and associative algebras may also be represented by suitable classes of matrices. A representation of the field of complex numbers is given in Appendix D.

** The *transpose* A^T of a matrix A is defined element-by-element by $(A^T)_{ik} = A_{ki}$. The *complex conjugate* A^* of a matrix A is defined by $(A^*)_{ik} = (A_{ik})^*$. The *adjoint* $A^\dagger$ of a matrix A is defined by $A^\dagger = (A^T)^*$. A matrix is called *symmetric* if $A^T = A$, *orthogonal* if $A^T = A^{-1}$, *Hermitean* or *self-adjoint* (the two are synonymous in all physical applications) if $A^\dagger = A$ (Hermitean matrices have real eigenvalues), and *unitary* if $A^\dagger = A^{-1}$. Unitary matrices have eigenvalues of norm one. Multiplication by a unitary matrix is a one-to-one mapping; since the time evolution of physical systems is a mapping from one time to another, evolution is always described by a unitary matrix. A *real* matrix obeys $A^* = A$, an *antisymmetric* or *skew-symmetric* matrix is defined by $A^T = -A$, an *anti-Hermitean* matrix by $A^\dagger = -A$ and an *anti-unitary* matrix by $A^\dagger = -A^{-1}$. All the mappings described by these special types of matrices are one-to-one. A matrix is *singular*, i.e. not one-to-one, if $\det A = 0$.

TABLE 24 Correspondences between the symmetries of an ornament, a flower and nature as a whole

System	Hispano–Arabic pattern	Flower	Motion
Structure and components	set of ribbons and patches	set of petals, stem	motion path and observables
System symmetry	pattern symmetry	flower symmetry	symmetry of Lagrangian
Mathematical description of the symmetry group	D_4	C_5	in Galilean relativity: position, orientation, instant and velocity changes
Invariants	number of multiplet elements	petal number	number of coordinates, magnitude of scalars, vectors and tensors
Representations of the components	multiplet types of elements	multiplet types of components	tensors, including scalars and vectors
Most symmetric representation	singlet	part with circular symmetry	scalar
Simplest faithful representation	quartet	quintet	vector
Least symmetric representation	quartet	quintet	no limit (tensor of infinite rank)

The representation is an *octet*. The complete list of possible irreducible representations of the group D_4 is given by *singlets*, *doublets* and *quartets*. Can you find them all? These representations allow the classification of all the white and black ribbons that appear in the figure, as well as all the coloured patches. The most symmetric elements are singlets, the least symmetric ones are members of the quartets. The complete system is always a singlet as well.

Challenge 382 ny

With these concepts we are ready to talk about motion with improved precision.

Symmetries, motion and Galilean physics

Every day we experience that we are able to talk to each other about motion. It must therefore be possible to find an *invariant* quantity describing it. We already know it: it is the *action*. Lighting a match is a change. It is the same whether it is lit here or there, in one direction or another, today or tomorrow. Indeed, the (Galilean) action is a number whose value is the same for each observer *at rest*, independent of his orientation or the time at which he makes his observation.

In the case of the Arabic pattern of Figure 104, the symmetry allows us to deduce the list of multiplets, or representations, that can be its building blocks. This approach must be possible for motion as well. We deduced the classification of the ribbons in the Arabic pattern into singlets, doublets, etc. from the various possible observation viewpoints. For a moving system, the building blocks, corresponding to the ribbons, are the *observables*. Since we observe that nature is symmetric under many different changes of viewpoint, we can classify all observables. To do so, we need to take the list of all viewpoint transformations and deduce the list of all their representations.

Our everyday life shows that the world stays unchanged after changes in position, orientation and instant of observation. One also speaks of space translation invariance, rotation invariance and time translation invariance. These transformations are different from those of the Arabic pattern in two respects: they are *continuous* and they are *unbounded*. As a result, their representations will generally be continuously variable and without bounds: they will be *quantities* or *magnitudes*. In other words, observables will be constructed with *numbers*. In this way we have deduced why numbers are *necessary* for any description of motion.*

Since observers can differ in orientation, most representations will be objects possessing a direction. To cut a long story short, the symmetry under change of observation position, orientation or instant leads to the result that all observables are either 'scalars', 'vectors' or higher-order 'tensors.'**

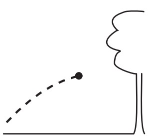
A *scalar* is an observable quantity which stays the same for all observers: it corresponds to a singlet. Examples are the mass or the charge of an object, the distance between two points, the distance of the horizon, and many others. Their possible values are (usually) continuous, unbounded and without direction. Other examples of scalars are the potential at a point and the temperature at a point. Velocity is obviously not a scalar; nor is the
Challenge 384 n coordinate of a point. Can you find more examples and counter-examples?

Energy is a puzzling observable. It is a scalar if only changes of place, orientation and instant of observation are considered. But energy is not a scalar if changes of observer speed are included. Nobody ever searched for a generalization of energy that is a scalar also for moving observers. Only Albert Einstein discovered it, completely by accident. More about this issue shortly.

Any quantity which has a magnitude and a direction and which 'stays the same' with respect to the environment when changing viewpoint is a *vector*. For example, the arrow between two fixed points on the floor is a vector. Its length is the same for all observers; its direction changes from observer to observer, but not with respect to its environment. On the other hand, the arrow between a tree and the place where a rainbow touches the Earth is *not* a vector, since that place does not stay fixed with respect to the environment, when the observer changes.

Mathematicians say that vectors are directed entities staying invariant under coordinate transformations. Velocities of objects, accelerations and field strength are examples of
Challenge 385 e vectors. (Can you confirm this?) The magnitude of a vector is a scalar: it is the same for any observer. By the way, a famous and baffling result of nineteenth-century experiments is that the velocity of light is *not* a vector for Galilean transformations. This mystery will be solved shortly.

* Only scalars, in contrast to vectors and higher-order tensors, may also be quantities which only take a
Challenge 383 e discrete set of values, such as +1 or −1 only. In short, only scalars may be *discrete* observables.
** Later on, *spinors* will be added to, and complete, this list.



Tensors are generalized vectors. As an example, take the moment of inertia of an object. Page 90 It specifies the dependence of the angular momentum on the angular velocity. For any object, doubling the magnitude of angular velocity doubles the magnitude of angular momentum; however, the two vectors are not parallel to each other if the object is not a Page 113 sphere. In general, if any two vector quantities are proportional, in the sense that doubling the magnitude of one vector doubles the magnitude of the other, but without the two vectors being parallel to each other, then the proportionality 'factor' is a (second order) *tensor*. Like all proportionality factors, tensors have a magnitude. In addition, tensors have a direction and a *shape*: they describe the connection between the vectors they relate. Just as vectors are the simplest quantities with a magnitude and a direction, so tensors are the simplest quantities with a magnitude and with a direction depending on a second, chosen direction. Vectors can be visualized as oriented arrows; tensors can be visualized Challenge 387 n as oriented ellipsoids.* Can you name another example of tensor?

Let us get back to the description of motion. Table 24 shows that in physical systems we always have to distinguish between the symmetry of the whole Lagrangian – corresponding to the symmetry of the complete pattern – and the representation of the observables – corresponding to the ribbon multiplets. Since the action must be a scalar, and since all observables must be tensors, Lagrangians contain sums and products of tensors only in combinations forming scalars. Lagrangians thus contain only scalar products or generalizations thereof. In short, Lagrangians always look like

$$L = \alpha\, a_i b^i + \beta\, c_{jk} d^{jk} + \gamma\, e_{lmn} f^{lmn} + \ldots \qquad (70)$$

where the indices attached to the variables a, b, c etc. always come in matching pairs to be summed over. (Therefore summation signs are usually simply left out.) The Greek letters represent constants. For example, the action of a free point particle in Galilean physics was given as

$$S = \int L\, \mathrm{d}t = \frac{m}{2} \int v^2\, \mathrm{d}t \qquad (71)$$

which is indeed of the form just mentioned. We will encounter many other cases during our study of motion.**

Galileo already understood that motion is also invariant under change of viewpoints Page 87 with different velocity. However, the action just given does not reflect this. It took some

* A rank-n tensor is the proportionality factor between a rank-1 tensor, i.e. between a vector, and an rank-$(n-1)$ tensor. Vectors and scalars are rank 1 and rank 0 tensors. Scalars can be pictured as spheres, vectors as arrows, and rank-2 tensors as ellipsoids. Tensors of higher rank correspond to more and more complex shapes.

A vector has the same length and direction for every observer; a tensor (of rank 2) has the same determinant, the same trace, and the same sum of diagonal subdeterminants for all observers.

A vector is described mathematically by a *list* of components; a tensor (of rank 2) is described by a *matrix* of components. The rank or order of a tensor thus gives the number of indices the observable has. Can you Challenge 386 e show this?

** By the way, is the usual list of possible observation viewpoints – namely different positions, different observation instants, different orientations, and different velocities – also *complete* for the action (71)? Sur- Ref. 163 prisingly, the answer is no. One of the first who noted this fact was Niederer, in 1972. Studying the quantum theory of point particles, he found that even the action of a Galilean free point particle is invariant under years to find out the correct generalization: it is given by the theory of special relativity. But before we study it, we need to finish the present topic.

Reproducibility, conservation and Noether's theorem

> I will leave my mass, charge and momentum to science.
>
> Graffito

The reproducibility of observations, i.e. the symmetry under change of instant of time or 'time translation invariance', is a case of viewpoint-independence. (That is not obvious; can you find its irreducible representations?) The connection has several important consequences. We have seen that symmetry implies invariance. It turns out that for *continuous* symmetries, such as time translation symmetry, this statement can be made more precise: for any continuous symmetry of the Lagrangian there is an associated conserved constant of motion and vice versa. The exact formulation of this connection is the theorem of Emmy Noether.* She found the result in 1915 when helping Albert Einstein and David Hilbert, who were both struggling and competing at constructing general relativity. However, the result applies to any type of Lagrangian.

Challenge 389 ny

Ref. 164

Noether investigated continuous symmetries depending on a continuous parameter b. A viewpoint transformation is a symmetry if the action S does not depend on the value of b. For example, changing position as

$$x \mapsto x + b \tag{74}$$

leaves the action

$$S_0 = \int T(v) - U(x)\,\mathrm{d}t \tag{75}$$

some additional transformations. If the two observers use the coordinates $(t, \mathbf{x})$ and (τ, ξ), the action (71) is invariant under the transformations

Challenge 388 ny

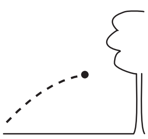
$$\xi = \frac{\mathbf{R}\mathbf{x} + \mathbf{x}_0 + \mathbf{v}t}{\gamma t + \delta} \quad \text{and} \quad \tau = \frac{\alpha t + \beta}{\gamma t + \delta} \quad \text{with} \quad \mathbf{R}^T\mathbf{R} = \mathbf{1} \quad \text{and} \quad \alpha\delta - \beta\gamma = 1\,. \tag{72}$$

where $\mathbf{R}$ describes the rotation from the orientation of one observer to the other, $\mathbf{v}$ the velocity between the two observers, and $\mathbf{x}_0$ the vector between the two origins at time zero. This group contains two important special cases of transformations:

$$\text{The connected, static Galilei group } \xi = \mathbf{R}\mathbf{x} + \mathbf{x}_0 + \mathbf{v}t \quad \text{and} \quad \tau = t$$
$$\text{The transformation group SL(2,R) } \xi = \frac{\mathbf{x}}{\gamma t + \delta} \quad \text{and} \quad \tau = \frac{\alpha t + \beta}{\gamma t + \delta} \tag{73}$$

The latter, three-parameter group includes *spatial inversion*, *dilations*, *time translation* and a set of time-dependent transformations such as $\xi = \mathbf{x}/t$, $\tau = 1/t$ called *expansions*. Dilations and expansions are rarely mentioned, as they are symmetries of point particles only, and do not apply to everyday objects and systems. They will return to be of importance later on, however.

* Emmy Noether (b. 1882 Erlangen, d. 1935 Bryn Mayr), German mathematician. The theorem is only a sideline in her career which she dedicated mostly to number theory. The theorem also applies to gauge symmetries, where it states that to every gauge symmetry corresponds an identity of the equation of motion, and vice versa.



invariant, since $S(b) = S_0$. This situation implies that

$$\frac{\partial T}{\partial v} = p = \text{const} \quad ; \tag{76}$$

in short, symmetry under change of position implies conservation of momentum. The converse is also true.

Challenge 390 ny

In the case of symmetry under shift of observation instant, we find

$$T + U = \text{const} \quad ; \tag{77}$$

in other words, time translation invariance implies constant energy. Again, the converse is also correct. One also says that energy and momentum are the *generators* of time and space translations.

The conserved quantity for a continuous symmetry is sometimes called the *Noether charge*, because the term *charge* is used in theoretical physics to designate conserved extensive observables. So, energy and momentum are Noether charges. 'Electric charge', 'gravitational charge' (i.e. mass) and 'topological charge' are other common examples. What is the conserved charge for rotation invariance? Challenge 391 n

We note that the expression 'energy is conserved' has several meanings. First of all, it means that the energy of a *single* free particle is constant in time. Secondly, it means that the total energy of any number of independent particles is constant. Finally, it means that the energy of a *system* of particles, i.e. including their interactions, is constant in time. Collisions are examples of the latter case. Noether's theorem makes all of these points at the same time, as you can verify using the corresponding Lagrangians. Challenge 392 e

But Noether's theorem also makes, or rather repeats, an even stronger statement: if energy were not conserved, time could not be defined. The whole description of nature requires the existence of conserved quantities, as we noticed when we introduced the concepts of object, state and environment. For example, we defined objects as *permanent* entities, that is, as entities characterized by conserved quantities. We also saw that the introduction of time is possible only because in nature there are 'no surprises'. Noether's theorem describes exactly what such a 'surprise' would have to be: the non-conservation of energy. However, energy jumps have never been observed – not even at the quantum level.

Page 39
Page 157

Since symmetries are so important for the description of nature, Table 25 gives an overview of all the symmetries of nature we will encounter. Their main properties are also listed. Except for those marked as 'approximate' or 'speculative', an experimental proof of incorrectness of any of them would be a big surprise indeed.

TABLE 25 The symmetries of relativity and quantum theory with their properties; also the complete list of logical *inductions* used in the two fields

Symmetry	Type [number of parameters]	Space of action	Group topology	Possible representations	Conserved quantity	Vacuum/ matter is symmetric	Main effect
Geometric or space-time, external, symmetries							
Time and space translation	$R \times R^3$ [4 par.]	space, time	not compact	scalars, vectors,	momentum and energy	yes/yes	allow everyday
Rotation	SO(3) [3 par.]	space	S^2	tensors	angular momentum	yes/yes	communi-cation
Galilei boost	R^3 [3 par.]	space, time	not compact	scalars, vectors, tensors	velocity of centre of mass	yes/for low speeds	relativity of motion
Lorentz	homogen-eous Lie SO(3,1) [6 par.]	space-time	not compact	tensors, spinors	energy-momentum $T^{\mu\nu}$	yes/yes	constant light speed
Poincaré ISL(2,C)	inhomo-geneous Lie [10 par.]	space-time	not compact	tensors, spinors	energy-momentum $T^{\mu\nu}$	yes/yes	
Dilation invariance	R^+ [1 par.]	space-time	ray	n-dimen. continuum	none	yes/no	massless particles
Special conformal invariance	R^4 [4 par.]	space-time	R^4	n-dimen. continuum	none	yes/no	massless particles
Conformal invariance	[15 par.]	space-time	involved	massless tensors, spinors	none	yes/no	light cone invariance
Dynamic, interaction-dependent symmetries: gravity							
$1/r^2$ gravity	SO(4) [6 par.]	config. space	as SO(4)	vector pair	perihelion direction	yes/yes	closed orbits
Diffeomorphism invariance	[∞ par.]	space-time	involved	space-times	local energy–momentum	yes/no	perihelion shift
Dynamic, classical and quantum-mechanical motion symmetries							

Symmetry	Type [number of parameters]	Space of action	Group topology	Possible representations	Conserved quantity	Vacuum/matter is symmetric	Main effect
Motion('time') inversion T	discrete	Hilbert or phase space	discrete	even, odd	T-parity	yes/no	reversibility
Parity('spatial') inversion P	discrete	Hilbert or phase space	discrete	even, odd	P-parity	yes/no	mirror world exists
Charge conjugation C	global, antilinear, anti-Hermitean	Hilbert or phase space	discrete	even, odd	C-parity	yes/no	anti-particles exist
CPT	discrete	Hilbert or phase space	discrete	even	CPT-parity	yes/yes	makes field theory possible
Dynamic, interaction-dependent, gauge symmetries							
Electromagnetic classical gauge invariance	[∞ par.]	space of fields	un- important	un-important	electric charge	yes/yes	massless light
Electromagnetic q.m. gauge inv.	abelian Lie U(1) [1 par.]	Hilbert space	circle S^1	fields	electric charge	yes/yes	massless photon
Electromagnetic duality	abelian Lie U(1) [1 par.]	space of fields	circle S^1	abstract	abstract	yes/no	none
Weak gauge	non-abelian Lie SU(2) [3 par.]	Hilbert space	as $SU(3)$	particles	weak charge	no/ approx.	
Colour gauge	non-abelian Lie SU(3) [8 par.]	Hilbert space	as $SU(3)$	coloured quarks	colour	yes/yes	massless gluons
Chiral symmetry	discrete	fermions	discrete	left, right	helicity	approximately	'massless' fermions[a]
Permutation symmetries							

Symmetry	Type [number of parameters]	Space of action	Group topology	Possible representations	Conserved quantity	Vacuum/matter is symmetric	Main effect
Particle exchange	discrete	Fock space etc.	discrete	fermions and bosons	none	n.a./yes	Gibbs' paradox
Selected *speculative* symmetries of nature							
GUT	E_8, SO(10)	Hilbert	from Lie group	particles	from Lie group	yes/no	coupling constant convergence
N-supersymmetry[b]	global	Hilbert		particles, sparticles	T_{mn} and N spinors [c] Q_{imn}	no/no	'massless'[a] particles
R-parity	discrete	Hilbert	discrete	+1, -1	R-parity	yes/yes sfermions, gauginos	
Braid symmetry	discrete	own space	discrete	unclear	unclear	yes/maybe	unclear
Space-time duality	discrete	all	discrete	vacuum	unclear	yes/maybe	fixes particle masses
Event symmetry	discrete	space-time	discrete	nature	none	yes/no	unclear

For details about the connection between symmetry and induction, see page 690. The explanation of the terms in the table will be completed in the rest of the walk. The real numbers are denoted as *R*.
a. Only approximate; 'massless' means that $m \ll m_{Pl}$, i.e. that $m \ll 22\,\mu g$.
b. $N = 1$ supersymmetry, but not $N = 1$ supergravity, is probably a good approximation for nature at everyday energies.
c. $i = 1 .. N$.

In summary, since we can *talk* about nature we can deduce several of its symmetries, in particular its symmetry under time and space translations. From nature's symmetries, using Noether's theorem, we can deduce the conserved charges, such as energy or linear and angular momentum. In other words, the definition of mass, space and time, together with their symmetry properties, is *equivalent* to the conservation of energy and momentum. Conservation and symmetry are two ways to express the same property of nature. To put it simply, our ability to talk about nature means that energy and momentum are conserved.

In general, the most elegant way to uncover the 'laws' of nature is to search for nature's symmetries. In many historical cases, once this connection had been understood, physics made rapid progress. For example, Albert Einstein discovered the theory of relativity in this way, and Paul Dirac started off quantum electrodynamics. We will use the same method throughout our walk; in its third part we will uncover some symmetries which are even more mind-boggling than those of relativity. Now, though, we will move on to the next approach to a global description of motion.

Curiosities and fun challenges about motion symmetry

Challenge 393 ny
What is the path followed by four turtles starting on the four angles of a square, if each of them continuously walks at the same speed towards the next one?

* *

Challenge 394 n
What is the symmetry of a simple oscillation? And of a wave?

* *

Challenge 395 n
For what systems is motion reversal a symmetry transformation?

* *

Challenge 396 ny
What is the symmetry of a continuous rotation?

* *

Challenge 397 ny
A sphere has a tensor for the moment of inertia that is diagonal with three equal numbers. The same is true for a cube. Can you distinguish spheres and cubes by their rotation behaviour?

* *

Challenge 398 ny
Is there a motion in nature whose symmetry is perfect?

Simple motions of extended bodies – oscillations and waves

We defined action, and thus change, as the integral of the Lagrangian, and the Lagrangian as the difference between kinetic and potential energy. One of the simplest systems in nature is a mass m attached to a spring. Its Lagrangian is given by

$$L = \frac{1}{2} m v^2 - k x^2 , \qquad (78)$$

Challenge 399 e
where k is a quantity characterizing the spring, the so-called spring constant. The Lagrangian is due to Robert Hooke, in the seventeenth century. Can you confirm it?

The motion that results from this Lagrangian is periodic, as shown in Figure 105. The Lagrangian describes the oscillation of the spring length. The motion is exactly the same as that of a long pendulum. It is called *harmonic motion*, because an object vibrating rapidly in this way produces a completely pure – or harmonic – musical sound. (The musical instrument producing the purest harmonic waves is the transverse flute. This instrument thus gives the best idea of how harmonic motion 'sounds'.) The graph of a

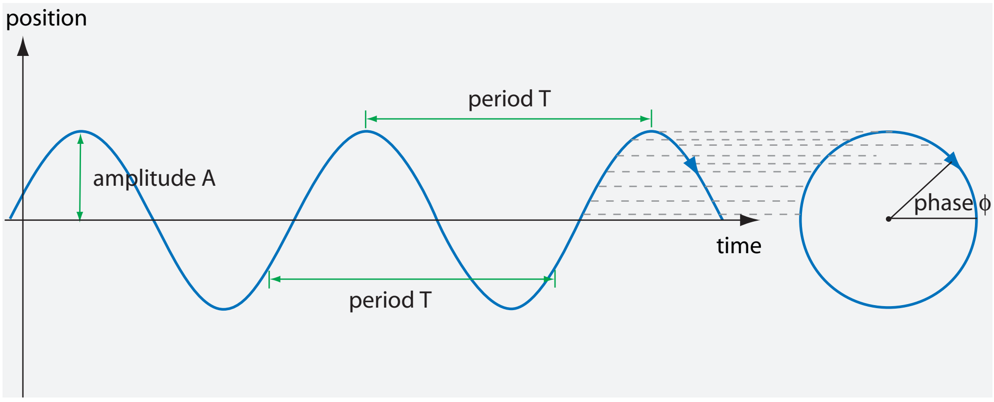


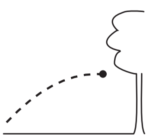
FIGURE 105 The simplest oscillation

TABLE 26 Some mechanical frequency values found in nature

Observation	Frequency
Sound frequencies in gas emitted by black holes	*c.* 1 fHz
Precision in measured vibration frequencies of the Sun	down to 2 nHz
Vibration frequencies of the Sun	down to *c.* 300 nHz
Vibration frequencies that disturb gravitational radiation detection	down to 3 μHz
Lowest vibration frequency of the Earth Ref. 165	309 μHz
Resonance frequency of stomach and internal organs (giving the 'sound in the belly' experience)	1 to 10 Hz
Wing beat of tiny fly	*c.* 1000 Hz
Sound audible to young humans	20 Hz to 20 kHz
Sonar used by bats	up to over 100 kHz
Sonar used by dolphins	up to 150 kHz
Sound frequency used in ultrasound imaging	up to 15 MHz
Phonon (sound) frequencies measured in single crystals	up to 20 THz and more

harmonic or linear oscillation, shown in Figure 105, is called a *sine curve*; it can be seen as the basic building block of all oscillations. All other, non-harmonic oscillations in nature can be composed from sine curves, as we shall see shortly. Page 207

Every oscillating motion continuously transforms kinetic energy into potential energy and vice versa. This is the case for the tides, the pendulum, or any radio receiver. But many oscillations also diminish in time: they are damped. Systems with large damping, such as the shock absorbers in cars, are used to avoid oscillations. Systems with *small* damping are useful for making precise and long-running clocks. The simplest measure of damping is the number of oscillations a system takes to reduce its amplitude to $1/e \approx 1/2.718$ times the original value. This characteristic number is the so-called *Q-factor*, named after the abbreviation of 'quality factor'. A poor Q-factor is 1 or less, an extremely good one is 100 000 or more. (Can you write down a simple Lagrangian for a damped oscillation with



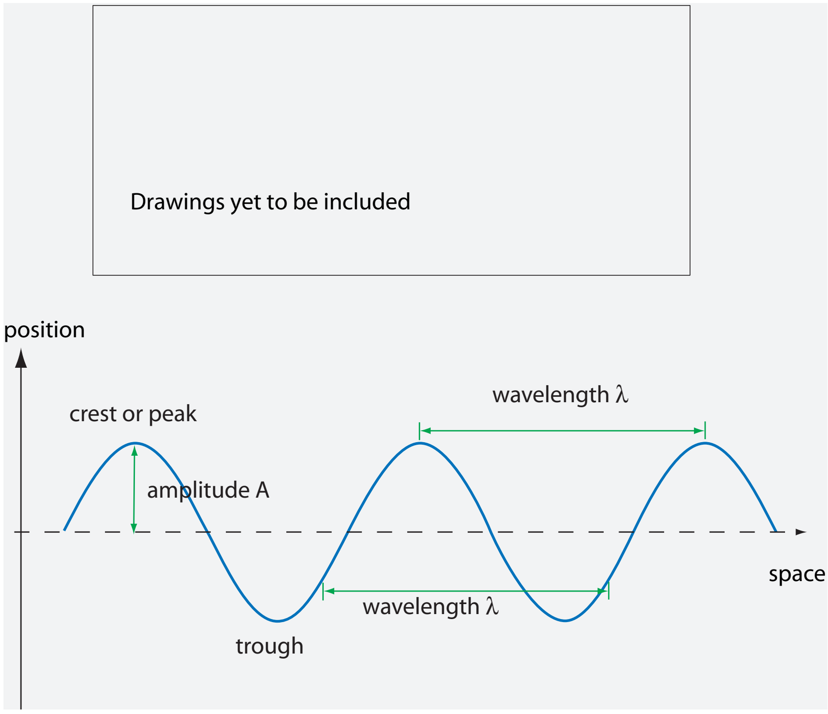


FIGURE 106 Decomposing a general wave or signal into harmonic waves

Challenge 400 ny

a given Q-factor?) In nature, damped oscillations do not usually keep constant frequency; however, for the simple pendulum this remains the case to a high degree of accuracy. The reason is that for a pendulum, the frequency does not depend significantly on the amplitude (as long as the amplitude is smaller than about 20°). This is one reason why pendulums are used as oscillators in mechanical clocks.

Obviously, for a good clock, the driving oscillation must not only show small damping, but must also be independent of temperature and be insensitive to other external influences. An important development of the twentieth century was the introduction of quartz crystals as oscillators. Technical quartzes are crystals of the size of a few grains of sand; they can be made to oscillate by applying an electric signal. They have little temperature dependence and a large Q-factor, and therefore low energy consumption, so that precise clocks can now run on small batteries.

Every harmonic oscillation is described by three quantities: the *amplitude*, the *period* (the inverse of the frequency) and the *phase*. The phase distinguishes oscillations of the same amplitude and period; it defines at what time the oscillation starts. Figure 105 shows how a harmonic oscillation is related to an imaginary rotation. As a result, the phase is best described by an angle between 0 and 2π.

All systems that oscillate also emit waves. In fact, oscillations only appear in extended systems, and oscillations are only the simplest of motions of extended systems. The general repetitive motion of an extended system is the wave.

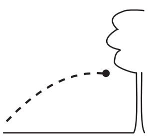


TABLE 27 Some wave velocities

Wave	Velocity
Tsunami	around 200 m/s
Sound in most gases	0.3 km/s
Sound in air at 273 K	331 m/s
Sound in air at 293 K	343 m/s
Sound in helium at 293 K	1.1 km/s
Sound in most liquids	1.1 km/s
Sound in water at 273 K	1.402 km/s
Sound in water at 293 K	1.482 km/s
Sound in gold	4.5 km/s
Sound in steel	5.790 km/s
Sound in granite	5.8 km/s
Sound in glass	5.9 km/s
Sound in beryllium	12.8 km/s
Sound in boron	up to 15 km/s
Sound in diamond	up to 18 km/s
Sound in fullerene (C_{60})	up to 26 km/s
Plasma wave velocity in InGaAs	600 km/s
Light in vacuum	$2.998 \cdot 10^8$ m/s

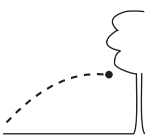
Waves and their motion

Waves are travelling imbalances, or, equivalently, travelling oscillations. Waves move, even though the substrate does not move. Every wave can be seen as a superposition of *harmonic* waves. Can you describe the difference in wave shape between a pure harmonic tone, a musical sound, a noise and an explosion? (Challenge 401 e) Every sound effect can be thought of as being composed of harmonic waves. Harmonic waves, also called *sine waves* or *linear waves*, are the building blocks of which all internal motions of an extended body are constructed.

Every harmonic wave is characterized by an oscillation *frequency*, a propagation *velocity*, a *wavelength*, and a *phase*, as can be deduced from Figure 106. Low-amplitude water waves show this most clearly. In a harmonic wave, every position performs a harmonic oscillation. The phase of a wave specifies the position of the wave (or a crest) at a given time. It is an angle between 0 and 2π. How are frequency and wavelength related in a wave? (Challenge 402 e)

Waves appear inside all *extended* bodies, be they solids, liquids, gases or plasmas. Inside fluid bodies, waves are *longitudinal*, meaning that the wave motion is in the same direction as the wave oscillation. Sound in air is an example of a longitudinal wave. Inside solid bodies, waves can also be *transverse*; in that case the wave oscillation is perpendicular to the travelling direction.

Waves appear also on *interfaces* between bodies: water–air interfaces are a well-known case. Even a saltwater–freshwater interface, so-called *dead water*, shows waves: they



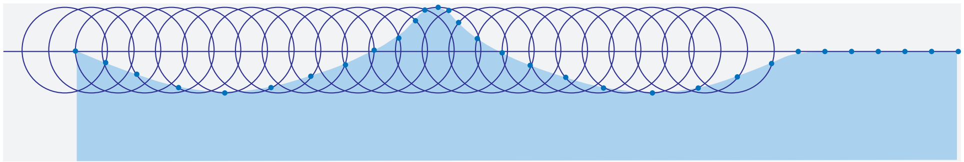


FIGURE 107 The formation of gravity waves on water

can appear even if the upper surface of the water is immobile. Any flight in an aeroplane provides an opportunity to study the regular cloud arrangements on the interface between warm and cold air layers in the atmosphere. Seismic waves travelling along the boundary between the sea floor and the sea water are also well-known. General surface waves are usually neither longitudinal nor transverse, but of a mixed type.

On water surfaces, one classifies waves according to the force that restores the plane surface. The first type, *surface tension waves*, plays a role on scales up to a few centimetres. At longer scales, gravity takes over as the main restoring force and one speaks of *gravity waves*. This is the type we focus on here. Gravity waves in water, in contrast to surface tension waves, are not sinusoidal. This is because of the special way the water moves in such a wave. As shown in Figure 107, the surface water moves in circles; this leads to the typical, asymmetrical wave shape with short sharp crests and long shallow troughs. (As long as there is no wind and the floor below the water is horizontal, the waves are also symmetric under front-to-back reflection.)

For water gravity waves, as for many other waves, the speed depends on the wavelength. Indeed, the speed c of water waves depends on the wavelength λ and on the depth of the water d in the following way:

$$c = \sqrt{\frac{g\lambda}{2\pi} \tanh \frac{2\pi d}{\lambda}} \;, \qquad (79)$$

where g is the acceleration due to gravity (and an amplitude much smaller than the wavelength is assumed). The formula shows two limiting regimes. First, short or deep waves appear when the water depth is larger than half the wavelength; for *deep waves*, the phase velocity is $c \approx \sqrt{g\lambda/2\pi}$, thus wavelength dependent, and the group velocity is about half the phase velocity. Shorter deep waves are thus slower. Secondly, shallow or *long waves* appear when the depth is less than 5% of the wavelength; in this case, $c \approx \sqrt{gd}$, there is no dispersion, and the group velocity is about the same as the phase velocity. The most impressive shallow waves are tsunamis, the large waves triggered by submarine earthquakes. (The Japanese name is composed of *tsu*, meaning harbour, and *nami*, meaning wave.) Since tsunamis are shallow waves, they show little dispersion and thus travel over long distances; they can go round the Earth several times. Typical oscillation times are between 6 and 60 minutes, giving wavelengths between 70 and 700 km and speeds in
Challenge 403 e
the open sea of 200 to 250 m/s, similar to that of a jet plane. Their amplitude on the open sea is often of the order of 10 cm; however, the amplitude scales with depth d as $1/d^4$ and heights up to 40 m have been measured at the shore. This was the order of magnitude of the large and disastrous tsunami observed in the Indian Ocean on 26 December 2004.

Waves can also exist in empty space. Both light and gravity waves are examples. The exploration of electromagnetism and relativity will tell us more about their properties.

Any study of motion must include the study of wave motion. We know from experience that waves can hit or even damage targets; thus every wave carries energy and momentum, even though (on average) no matter moves along the wave propagation direction. The *energy* E of a wave is the sum of its kinetic and potential energy. The kinetic energy (density) depends on the temporal change of the displacement u at a given spot: rapidly changing waves carry a larger kinetic energy. The potential energy (density) depends on the gradient of the displacement, i.e. on its spatial change: steep waves carry a larger potential energy than shallow ones. (Can you explain why the potential energy does not depend on the displacement itself?) For harmonic waves propagating along the direction z, each type of energy is proportional to the square of its respective displacement change:

Challenge 404 n

Ref. 166

$$E \sim \left(\frac{\partial u}{\partial t}\right)^2 + v^2 \left(\frac{\partial u}{\partial z}\right)^2 . \tag{80}$$

Challenge 405 ny

How is the energy density related to the frequency?

The *momentum* of a wave is directed along the direction of wave propagation. The momentum value depends on both the temporal and the spatial change of displacement u. For harmonic waves, the momentum (density) P is proportional to the product of these two quantities:

$$P_z \sim \frac{\partial u}{\partial t} \frac{\partial u}{\partial z} . \tag{81}$$

When two linear wave trains collide or interfere, the total momentum is conserved throughout the collision. An important consequence of momentum conservation is that waves that are reflected by an obstacle do so with an outgoing angle equal to minus the infalling angle. What happens to the phase?

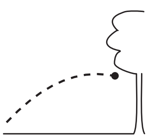
Challenge 406 n

Waves, like moving bodies, carry energy and momentum. In simple terms, if you shout against a wall, the wall is hit. This hit, for example, can start avalanches on snowy mountain slopes. In the same way, waves, like bodies, can carry also angular momentum. (What type of wave is necessary for this to be possible?) However, we can distinguish six main properties that set the motion of waves apart from the motion of bodies.

Challenge 407 ny

— Waves can add up or cancel each other out; thus they can interpenetrate each other. These effects, called *superposition* and *interference*, are strongly tied to the linearity of most waves.
— Transverse waves in three dimensions can oscillate in different directions: they show *polarization*.
— Waves, such as sound, can go around corners. This is called *diffraction*.
— Waves change direction when they change medium. This is called *refraction*.
— Waves can have a frequency-dependent propagation speed. This is called *dispersion*.
— Often, the wave amplitude decreases over time: waves show *damping*.

Material bodies in everyday life do not behave in these ways when they move. These six wave effects appear because wave motion is the motion of *extended* entities. The famous debate whether electrons or light are waves or particles thus requires us to check whether



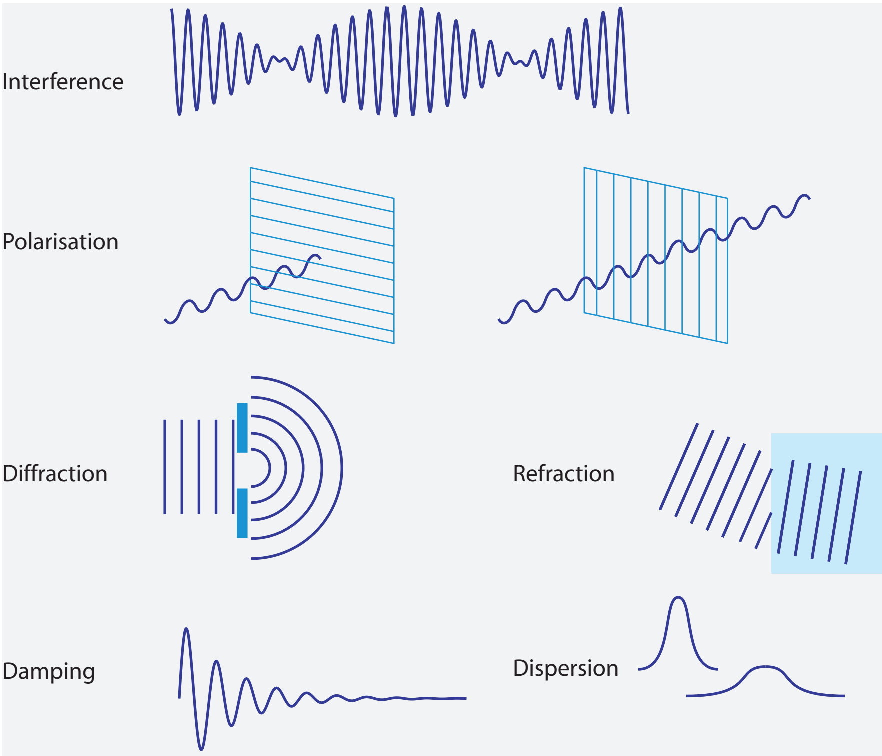


FIGURE 108 The six main properties of the motion of waves

these effects specific to waves can be observed or not. This is one topic of quantum theory. Before we study it, can you give an example of an observation that implies that a motion surely cannot be a wave? Challenge 408 n

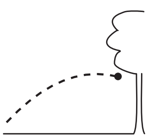
As a result of having a frequency f and a propagation velocity v, all sine waves are characterized by the distance λ between two neighbouring wave crests: this distance is called the wavelength. All waves obey the basic relation

$$\lambda f = v . \tag{82}$$

In many cases the wave velocity v depends on the wavelength of the wave. For example, this is the case for water waves. This change of speed with wavelength is called *dispersion*. In contrast, the speed of sound in air does not depend on the wavelength (to a high degree of accuracy). Sound in air shows almost no dispersion. Indeed, if there were dispersion for sound, we could not understand each other's speech at larger distances.

In everyday life we do not experience light as a wave, because the wavelength is only around one two-thousandth of a millimetre. But light shows all six effects typical of wave motion. A rainbow, for example, can only be understood fully when the last five wave effects are taken into account. Diffraction and interference can even be observed with your fingers only. Can you tell how? Page 568 Challenge 409 n

Like every anharmonic oscillation, every anharmonic wave can be decomposed into sine waves. Figure 106 gives examples. If the various sine waves contained in a disturbance



propagate differently, the original wave will change in shape while it travels. That is the reason why an echo does not sound exactly like the original sound; for the same reason, a nearby thunder and a far-away one sound different.

All systems which oscillate also emit waves. Any radio or TV receiver contains oscillators. As a result, any such receiver is also a (weak) transmitter; indeed, in some countries the authorities search for people who listen to radio without permission listening to the radio waves emitted by these devices. Also, inside the human ear, numerous tiny structures, the hair cells, oscillate. As a result, the ear must also emit sound. This prediction, made in 1948 by Tommy Gold, was confirmed only in 1979 by David Kemp. These so-called *otoacoustic emissions* can be detected with sensitive microphones; they are presently being studied in order to unravel the still unknown workings of the ear and in order to diagnose various ear illnesses without the need for surgery. Ref. 167

Since any travelling disturbance can be decomposed into sine waves, the term 'wave' is used by physicists for all travelling disturbances, whether they look like sine waves or not. In fact, the disturbances do not even have to be travelling. Take a standing wave: is it a wave or an oscillation? Standing waves do not travel; they are oscillations. But a standing wave can be seen as the superposition of two waves travelling in opposite directions. Since all oscillations are standing waves (can you confirm this?), we can say that all oscillations are special forms of waves. Challenge 410 ny

The most important travelling disturbances are those that are localized. Figure 106 shows an example of a localized wave group or pulse, together with its decomposition into harmonic waves. Wave groups are extensively used to talk and as signals for communication.

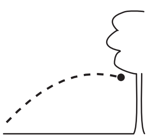
Why can we talk to each other? – Huygens' principle

The properties of our environment often disclose their full importance only when we ask simple questions. Why can we use the radio? Why can we talk on mobile phones? Why can we listen to each other? It turns out that a central part of the answer to these questions is that the space we live has an *odd* numbers of dimensions.

In spaces of *even* dimension, it is impossible to talk, because messages do not stop. This is an important result which is easily checked by throwing a stone into a lake: even after the stone has disappeared, waves are still emitted from the point at which it entered the water. Yet, when we stop talking, no waves are emitted any more. Waves in two and three dimensions thus behave differently.

In three dimensions, it is possible to say that the propagation of a wave happens in the following way: Every point on a wave front (of light or of sound) can be regarded as the source of secondary waves; the surface that is formed by the envelope of all the secondary waves determines the future position of the wave front. The idea is illustrated in Figure 109. It can be used to describe, without mathematics, the propagation of waves, their reflection, their refraction, and, with an extension due to Augustin Fresnel, their diffraction. (Try!) Challenge 411 e

This idea was first proposed by Christiaan Huygens in 1678 and is called *Huygens' principle*. Almost two hundred years later, Gustav Kirchoff showed that the principle is a consequence of the wave equation in three dimensions, and thus, in the case of light, a consequence of Maxwell's field equations.



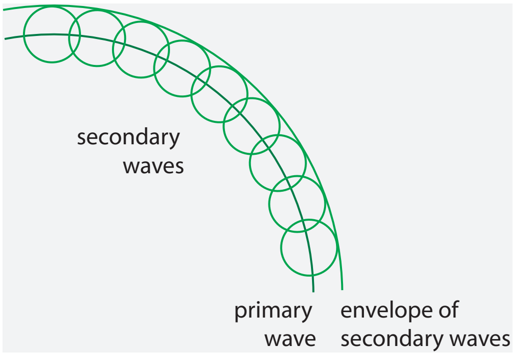


FIGURE 109 Wave propagation as a consequence of Huygens' principle

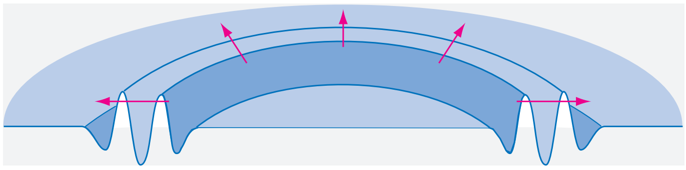


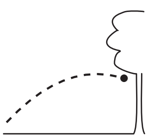
FIGURE 110 An impossible water wave: the centre is never flat

But the description of wave fronts as envelopes of secondary waves has an important limitation. It is not correct in two dimensions (even though Figure 109 is two-dimensional!). In particular, it does not apply to water waves. Water wave propagation *cannot* be calculated in this way in an exact manner. (It is only possible if the situation is limited to a waves of a single frequency.) It turns out that for water waves, secondary waves do not only depend on the wave front of the primary waves, but depend also on their interior. The reason is that in two (and other even) dimensions, waves of different frequency necessarily have different speeds. And a stone falling into water generates waves of many frequencies. In contrast, in three (and larger odd) dimensions, waves of all frequencies have the same speed.

We can also say that Huygens' principle holds if the wave equation is solved by a circular wave leaving no amplitude behind it. Mathematicians translate this by requiring that the evolving delta function $\delta(c^2t^2 - r^2)$ satisfies the wave equation, i.e. that $\partial_t^2\delta = c^2\Delta\delta$. The delta function is that strange 'function' which is zero everywhere except at the origin, where it is infinite. A few more properties describe the precise way in which this happens.* It turns out that the delta function is a solution of the wave equation only if the space dimension is odd and at least three. In other words, while a spherical wave pulse is possible, a circular pulse is not: there is no way to keep the centre of an expanding wave quiet. (See Figure 110.) That is exactly what the stone experiment shows. You can try to produce a circular pulse (a wave that has only a few crests) the next time you are in the bathroom or near a lake: you will not succeed.

In summary, the reason a room gets dark when we switch off the light, is that we live in a space with a number of dimensions which is odd and larger than one.

* The main property is $\int \delta x \mathrm{d}x = 1$. In mathematically precise terms, the delta 'function' is a distribution.



Signals

A signal is the transport of information. Every signal is motion of energy. Signals can be either objects or waves. A thrown stone can be a signal, as can a whistle. Waves are a more practical form of communication because they do not require transport of matter: it is easier to use electricity in a telephone wire to transport a statement than to send a messenger. Indeed, most modern technological advances can be traced to the separation between signal and matter transport. Instead of transporting an orchestra to transmit music, we can send radio signals. Instead of sending paper letters we write email messages. Instead of going to the library we browse the internet.

The greatest advances in communication have resulted from the use of signals to transport large amounts of energy. That is what electric cables do: they transport energy without transporting any (noticeable) matter. We do not need to attach our kitchen machines to the power station: we can get the energy via a copper wire.

For all these reasons, the term 'signal' is often meant to imply waves only. Voice, sound, electric signals, radio and light signals are the most common examples of wave signals.

Signals are characterized by their speed and their information content. Both quantities turn out to be limited. The limit on speed is the central topic of the theory of special relativity.

Page 283

A simple limit on information content can be expressed when noting that the information flow is given by the detailed shape of the signal. The shape is characterized by a frequency (or wavelength) and a position in time (or space). For every signal – and every wave – there is a relation between the time-of-arrival error Δt and the angular frequency error $\Delta\omega$:

$$\Delta t \; \Delta\omega \geqslant \frac{1}{2} \; . \qquad (83)$$

This time–frequency indeterminacy relation expresses that, in a signal, it is impossible to specify both the time of arrival and the frequency with full precision. The two errors are (within a numerical factor) the inverse of each other. (One also says that the time-bandwidth product is always larger than $1/4\pi$.) The limitation appears because on one hand one needs a wave as similar as possible to a sine wave in order to precisely determine the frequency, but on the other hand one needs a signal as narrow as possible to precisely determine its time of arrival. The contrast in the two requirements leads to the limit. The indeterminacy relation is thus a feature of every wave phenomenon. You might want to test this relation with any wave in your environment.

Challenge 412 e

Similarly, there is a relation between the position error Δx and the wave vector error $\Delta k = 2\pi/\Delta\lambda$ of a signal:

$$\Delta x \; \Delta k \geqslant \frac{1}{2} \; . \qquad (84)$$

Like the previous case, also this indeterminacy relation expresses that it is impossible to specify both the position of a signal and its wavelength with full precision. Also this position–wave-vector indeterminacy relation is a feature of any wave phenomenon.

Every indeterminacy relation is the consequence of a smallest entity. In the case of waves, the smallest entity of the phenomenon is the period (or cycle, as it used to be called). Whenever there is a smallest unit in a natural phenomenon, an indeterminacy

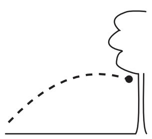


Figure to be included

FIGURE 111 The electrical signals measured in a nerve

relation results. We will encounter other indeterminacy relations both in relativity and in quantum theory. As we will find out, they are due to smallest entities as well.

Whenever signals are sent, their content can be lost. Each of the six characteristics of waves listed on page 210 can lead to content degradation. Can you provide an example for each case? The energy, the momentum and all other conserved properties of signals are never lost, of course. The disappearance of signals is akin to the disappearance of motion. When motion disappears by friction, it only seems to disappear, and is in fact transformed into heat. Similarly, when a signal disappears, it only seems to disappear, and is in fact transformed into noise. *(Physical) noise* is a collection of numerous disordered signals, in the same way that heat is a collection of numerous disordered movements.

Challenge 413 ny

All signal propagation is described by a wave equation. A famous example is the equation found by Hodgkin and Huxley. It is a realistic approximation for the behaviour of electrical potential in nerves. Using facts about the behaviour of potassium and sodium ions, they found an elaborate equation that describes the voltage V in nerves, and thus the way the signals are propagated. The equation accurately describes the characteristic voltage spikes measured in nerves, shown in Figure 111. The figure clearly shows that these waves differ from sine waves: they are not harmonic. Anharmonicity is one result of nonlinearity. But nonlinearity can lead to even stronger effects.

Ref. 168

Solitary waves and solitons

In August 1834, the Scottish engineer John Scott Russell (1808–1882) recorded a strange observation in a water canal in the countryside near Edinburgh. When a boat pulled through the channel was suddenly stopped, a strange water wave departed from it. It consisted of a *single* crest, about 10 m long and 0.5 m high, moving at about 4 m/s. He followed that crest, shown in a reconstruction in Figure 112, with his horse for several kilometres: the wave died out only very slowly. Russell did not observe any dispersion, as is usual in water waves: the width of the crest remained constant. Russell then started producing such waves in his laboratory, and extensively studied their properties. He showed that the speed depended on the amplitude, in contrast to linear, harmonic waves. He also found that the depth d of the water canal was an important parameter. In fact, the speed v, the amplitude A and the width L of these single-crested waves are related by

Ref. 169

$$v = \sqrt{gd}\left(1 + \frac{A}{2d}\right) \quad \text{and} \quad L = \sqrt{\frac{4d^3}{3A}} \,. \qquad (85)$$

FIGURE 112 A solitary water wave followed by a motor boat, reconstructing the discovery by Scott Russel (© Dugald Duncan)

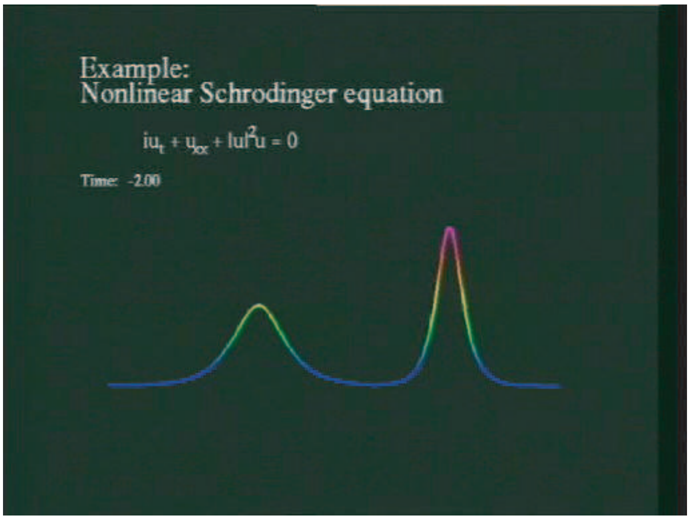
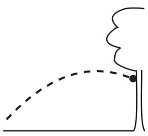


FIGURE 113 Solitons are stable against encounters (mpeg © Jarmo Hietarinta)

As shown by these expressions, and noted by Russell, high waves are narrow and fast, whereas shallow waves are slow and wide. The shape of the waves is fixed during their motion. Today, these and all other stable waves with a single crest are called *solitary waves*. They appear only where the dispersion and the nonlinearity of the system exactly compensate for each other. Russell also noted that the solitary waves in water channels can cross each other unchanged, even when travelling in opposite directions; solitary waves with this property are called *solitons*. Solitons are stable against encounters, as shown in Figure 113, whereas solitary waves in general are not.



Only sixty years later, in 1895, Korteweg and de Vries found out that solitary waves in water channels have a shape described by

$$u(x,t) = A \operatorname{sech}^2 \frac{x - vt}{L} \quad \text{where} \quad \operatorname{sech} x = \frac{2}{e^x + e^{-x}} , \tag{86}$$

and that the relation found by Russell was due to the wave equation

$$\frac{1}{\sqrt{gd}} \frac{\partial u}{\partial t} + \left(1 + \frac{3}{2d} u\right) \frac{\partial u}{\partial x} + \frac{d^2}{6} \frac{\partial^3 u}{\partial x^3} = 0 . \tag{87}$$

This equation for the elongation u is called the *Korteweg–de Vries equation* in their honour.* The surprising stability of the solitary solutions is due to the opposite effect of the two terms that distinguish the equation from linear wave equations: for the solitary solutions, the nonlinear term precisely compensates for the dispersion induced by the third-derivative term.

For many decades such solitary waves were seen as mathematical and physical curiosities. But almost a hundred years later it became clear that the Korteweg–de Vries equation is a universal model for weakly nonlinear waves in the weak dispersion regime, and thus of basic importance. This conclusion was triggered by Kruskal and Zabusky, who in 1965 proved mathematically that the solutions (86) are unchanged in collisions. This discovery prompted them to introduce the term *soliton*. These solutions do indeed interpenetrate one another without changing velocity or shape: a collision only produces a small positional shift for each pulse. Ref. 170

Solitary waves play a role in many examples of fluid flows. They are found in ocean currents; and even the red spot on Jupiter, which was a steady feature of Jupiter photographs for many centuries, is an example.

Solitary waves also appear when extremely high-intensity sound is generated in solids. In these cases, they can lead to sound pulses of only a few nanometres in length. Solitary light pulses are also used inside certain optical communication fibres, where they provide (almost) lossless signal transmission. Ref. 171 Ref. 169

Towards the end of the twentieth century a second wave of interest in the mathematics of solitons arose, when quantum theorists became interested in them. The reason is simple but deep: a soliton is a 'middle thing' between a particle and a wave; it has features of both concepts. For this reason, solitons are now an essential part of any description of elementary particles, as we will find out later on.

Curiosities and fun challenges about waves and extended bodies

> Society is a wave. The wave moves onward, but the water of which it is composed does not.
> Ralph Waldo Emerson, *Self-Reliance*.

* The equation can be simplified by transforming the variable u; most concisely, it can be rewritten as $u_t + u_{xxx} = 6uu_x$. As long as the solutions are sech functions, this and other transformed versions of the equation are known by the same name.

When the frequency of a tone is doubled, one says that the tone is higher by an octave. Two tones that differ by an octave, when played together, sound pleasant to the ear. Two other agreeable frequency ratios – or 'intervals', as musicians say – are quarts and quints. Challenge 414 e What are the corresponding frequency ratios? (Note: the answer was one of the oldest discoveries in physics; it is attributed to Pythagoras, around 500 bce.)

* *

An orchestra is playing music in a large hall. At a distance of 30 m, somebody is listening to the music. At a distance of 3000 km, another person is listening to the music via the Challenge 415 n radio. Who hears the music first?

* *

What is the period of a simple pendulum, i.e. a mass *m* attached to a massless string of Challenge 416 ny length *l*? What is the period if the string is much longer than the radius of the Earth?

* *

What path is followed by a body moving in a plane, but attached by a spring to a fixed Challenge 417 n point on the plane?

* *

A device that shows how rotation and oscillation are linked is the alarm *siren*. Find out Challenge 418 e how it works, and build one yourself.

* *

Light is a wave, as we will discover later on. As a result, light reaching the Earth from space is refracted when it enters the atmosphere. Can you confirm that as a result, stars Challenge 419 e appear somewhat higher in the night sky than they really are?

* *

What are the highest sea waves? This question has been researched systematically only Ref. 172 recently, using satellites. The surprising result is that sea waves with a height of 25 m and more are *common*: there are a few such waves on the oceans at any given time. This result confirms the rare stories of experienced ship captains and explains many otherwise ship sinkings.

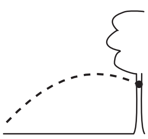
Surfers may thus get many chances to ride 30 m waves. (The record is just below this size.) But maybe the most impressive waves to surf are those of the Pororoca, a series of 4 m waves that move from the sea into the Amazon River every spring, against the flow of the river. These waves can be surfed for tens of kilometres.

* *

All waves are damped, eventually. This effect is often frequency-dependent. Can you Challenge 420 n provide a confirmation of this dependence in the case of sound in air?

* *

When you make a hole with a needle in black paper, the hole can be used as a magnifying Challenge 421 e lens. (Try it.) Diffraction is responsible for the lens effect. By the way, the diffraction of



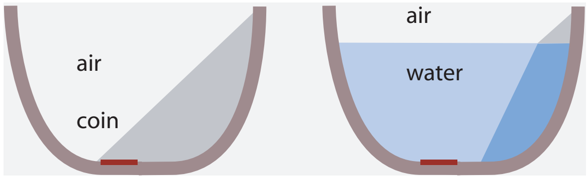


FIGURE 114 Shadows and refraction

light by holes was noted by Francesco Grimaldi in the seventeenth century; he deduced that light is a wave. His observations were later discussed by Newton, who wrongly dismissed them.

* *

Put an empty cup near a lamp, in such a way that the bottom of the cup remains in the shadow. When you fill the cup with water, some of the bottom will be lit, because of the refraction of the light from the lamp. The same effect allows us to build lenses. The same Page 581 effect is at the basis of instruments such as the telescope.

* *

Challenge 422 n Are water waves transverse or longitudinal?

* *

The speed of water waves limits the speeds of ships. A surface ship cannot travel (much) faster than about $v_{crit} = \sqrt{0.16gl}$, where $g = 9.8\,\mathrm{m/s^2}$, l is its length, and 0.16 is a number determined experimentally, called the critical Froude number. This relation is valid for all vessels, from large tankers ($l = 100\,\mathrm{m}$ gives $v_{crit} = 13\,\mathrm{m/s}$) down to ducks ($l = 0.3\,\mathrm{m}$ gives $v_{crit} = 0.7\,\mathrm{m/s}$). The critical speed is that of a wave with the same wavelength as the ship. In fact, moving at higher speeds than the critical value is possible, but requires much more energy. (A higher speed is also possible if the ship surfs on a wave.) Therefore all water animals and ships are faster when they swim below the surface – where the limit due to surface waves does not exist – than when they swim on the surface. For example, ducks can swim three times as fast under water than on the surface.

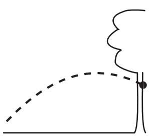
Challenge 423 n How far away is the *olympic swimming* record from the critical value?

* *

The group velocity of water waves (in deep water) is less than the velocity of the individual waves. As a result, when a group of wave crests travels, within the group the crests move from the back to the front, appearing at the back, travelling forward and then dying out at the front.

* *

One can hear the distant sea or a distant highway more clearly in the evening than in the morning. This is an effect of refraction. Sound speed decreases with temperature. In the evening, the ground cools more quickly than the air above. As a result, sound leaving the ground and travelling upwards is refracted downwards, leading to the long hearing



distance. In the morning, usually the air is cold above and warm below. Sound is refracted upwards, and distant sound does not reach a listener on the ground. Refraction thus implies that mornings are quiet, and that one can hear more distant sounds in the evenings. Elephants use the sound situation during evenings to communicate over distances of more than 10 km. (They also use sound waves in the ground to communicate, but that is another story.)

* *

Refraction also implies that there is a sound channel in the ocean, and in the atmosphere. Sound speed decreases with temperature, and increases with pressure. At an ocean depth of 1 km, or at an atmospheric height of 13 to 17 km (that is at the top of the tallest cumulonimbus clouds or equivalently, at the middle of the ozone layer) sound has minimal speed. As a result, sound that starts from that level and tries to leave is channelled back to it. Whales use the sound channel to communicate with each other with beautiful songs; one can find recordings of these songs on the internet. The military successfully uses microphones placed at the sound channel in the ocean to locate submarines, and microphones on balloons in the atmospheric channel to listen for nuclear explosions. (In fact, sound experiments conducted by the military are the main reason why whales are deafened and lose their orientation, stranding on the shores. Similar experiments in the air with high-altitude balloons are often mistaken for flying saucers, as in the famous Roswell incident.)

Challenge 424 e

Ref. 173

* *

Ref. 174

Also much smaller animals communicate by sound waves. In 2003, it was found that herring communicate using noises they produce when farting. When they pass wind, the gas creates a ticking sound whose frequency spectrum reaches up to 20 kHz. One can even listen to recordings of this sound on the internet. The details of the communication, such as the differences between males and females, are still being investigated. It is possible that the sounds may also be used by predators to detect herring, and they might even by used by future fishing vessels.

* *

On windy seas, the white wave crests have several important effects. The noise stems from tiny exploding and imploding water bubbles. The noise of waves on the open sea is thus the superposition of many small explosions. At the same time, white crests are the events where the seas absorb carbon dioxide from the atmosphere, and thus reduce global warming.

* *

Challenge 425 n

Why are there many small holes in the ceilings of many office buildings?

* *

Which quantity determines the wavelength of water waves emitted when a stone is thrown into a pond?

Challenge 426 ny

* *

Ref. 3

Yakov Perelman lists the following four problems in his delightful physics problem book.

(1) A stone falling into a lake produces circular waves. What is the shape of waves produced by a stone falling into a river, where the water flows in one direction?
Challenge 427 n

(2) It is possible to build a lens for sound, in the same way as it is possible to build lenses for light. What would such a lens look like?
Challenge 428 n

Challenge 429 ny
(3) What is the sound heard inside a shell?

(4) Light takes about eight minutes to travel from the Sun to the Earth. What consequence does this have for a sunrise?
Challenge 430 n

* *

Can you describe how a Rubik's Cube is built? And its generalizations to higher numbers of segments? Is there a limit to the number of segments? These puzzles are even tougher than the search for a rearrangement of the cube. Similar puzzles can be found in the study of many mechanisms, from robots to textile machines.
Challenge 431 n

* *

Typically, sound produces a pressure variation of 10^{-8} bar on the ear. How is this determined?
Challenge 432 ny

The ear is indeed a sensitive device. It is now known that most cases of sea mammals, like whales, swimming onto the shore are due to ear problems: usually some military device (either sonar signals or explosions) has destroyed their ear so that they became deaf and lose orientation.

* *

Infrasound, inaudible sound below 20 Hz, is a modern topic of research. In nature, infrasound is emitted by earthquakes, volcanic eruptions, wind, thunder, waterfalls, falling meteorites and the surf. Glacier motion, seaquakes, avalanches and geomagnetic storms also emit infrasound. Human sources include missile launches, traffic, fuel engines and air compressors.
Ref. 175

It is known that high intensities of infrasound lead to vomiting or disturbances of the sense of equilibrium (140 dB or more for 2 minutes), and even to death (170 dB for 10 minutes). The effects of lower intensities on human health are not yet known.

Infrasound can travel several times around the world before dying down, as the explosion of the Krakatoa volcano showed in 1883. With modern infrasound detectors, sea surf can be detected hundreds of kilometres away. Sea surf leads to a constant 'hum' of the Earth's crust at frequencies between 3 and 7 mHz. The *Global infrasound Network* uses infrasound to detect nuclear weapon tests, earthquakes and volcanic eruptions, and can count meteorites. Only very rarely can meteorites be heard with the human ear.

* *

The method used to deduce the sine waves contained in a signal, as shown in Figure 106, is called the Fourier transformation. It is of importance throughout science and technology. In the 1980s, an interesting generalization became popular, called the *wavelet transformation*. In contrast to Fourier transformations, wavelet transformations allow us to localize signals in time. Wavelet transformations are used to compress digitally stored images in an efficient way, to diagnose aeroplane turbine problems, and in many other applications.
Ref. 176

* *

If you like engineering challenges, here is one that is still open. How can one make a
Challenge 433 r robust and efficient system that transforms the energy of sea waves into electricity?

* *

If you are interested in ocean waves, you might also enjoy the science of *oceanography*. For an introduction, see the open source textbooks at http://oceanworld.tamu.edu/.

* *

In our description of extended bodies, we assumed that each spot of a body can be followed separately throughout its motion. Is this assumption justified? What would happen
Challenge 434 r if it were not?

* *

A special type of waves appears in explosions and supersonic flight: *shock waves*. In a shock wave, the density or pressure of a gas changes abruptly, on distances of a few micrometers. Studying shock waves is a research field in itself; shock waves determine the flight of bullets, the snapping of whips and the effects of detonations.

* *

Bats fly at night using *echolocation*. Dolphins also use it. Sonar, used by fishing vessels to look for fish, copies the system of dolphins. Less well known is that humans have the
Ref. 177 same ability. Have you ever tried to echolocate a wall in a completely dark room? You will be surprised at how easily this is possible. Just make a loud hissing or whistling noise that
Challenge 435 e stops abruptly, and listen to the echo. You will be able to locate walls reliably.

* *

Birds sing. If you want to explore how this happens, look at the X-ray film found at the http://www.indiana.edu/~songbird/multi/cineradiography_index.html website.

* *

Every soliton is a one-dimensional structure. Do two-dimensional analogues exist? This
Ref. 178 issue was open for many years. Finally, in 1988, Boiti, Leon, Martina and Fumagalli found that a certain evolution equation, the so-called Davey–Stewartson equation, can have solutions that are localized in two dimensions. These results were generalized by Fokas and Santini and further generalized by Hietarinta and Hirota. Such a solution is today called a *dromion*. Dromions are bumps that are localized in two dimensions and can move, without disappearing through diffusion, in non-linear systems. An example is shown in Figure 115. However, so far, no such solution has be observed in experiments; this is one of the most important experimental challenges left open in non-linear science.

Do extended bodies exist?

We have just discussed the motion of extended bodies in some detail. We have seen that extended bodies show wave motion. But are extended bodies found in nature? Strangely enough, this question has been one of the most intensely discussed questions in physics.

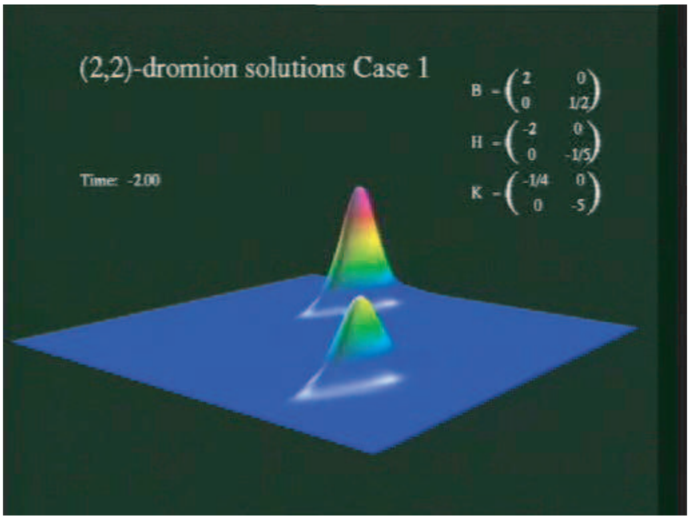


FIGURE 115 The calculated motion of a dromion across a two-dimensional substrate (mpeg © Jarmo Hietarinta)

Over the centuries, it has reappeared again and again, at each improvement of the description of motion; the answer has alternated between the affirmative and the negative. Many thinkers have been imprisoned, and many still are being persecuted, for giving answers that are not politically correct! In fact, the issue already arises in everyday life.

Mountains and fractals

Whenever we climb a mountain, we follow the outline of its shape. We usually describe this outline as a curved two-dimensional surface. In everyday life we find that this is a good approximation. But there are alternative possibilities. The most popular is the idea that mountains are fractal surfaces. A *fractal* was defined by Benoit Mandelbrot as a set that is self-similar under a countable but infinite number of magnification values.*
Page 53 We have already encountered fractal lines. An example of an algorithm for building a (random) fractal *surface* is shown on the right side of Figure 116. It produces shapes which
Ref. 179 look remarkably similar to real mountains. The results are so realistic that they are used in Hollywood films. If this description were correct, mountains would be extended, but not continuous.

But mountains could also be fractals of a different sort, as shown in the left side of Figure 116. Mountain surfaces could have an infinity of small and smaller holes. In fact, one could also imagine that mountains are described as three-dimensional versions of the left side of the figure. Mountains would then be some sort of mathematical Swiss cheese. Can you devise an experiment to decide whether fractals provide the correct description
Challenge 436 n for mountains? To settle the issue, a chocolate bar can help.

* For a definition of uncountability, see page 665.

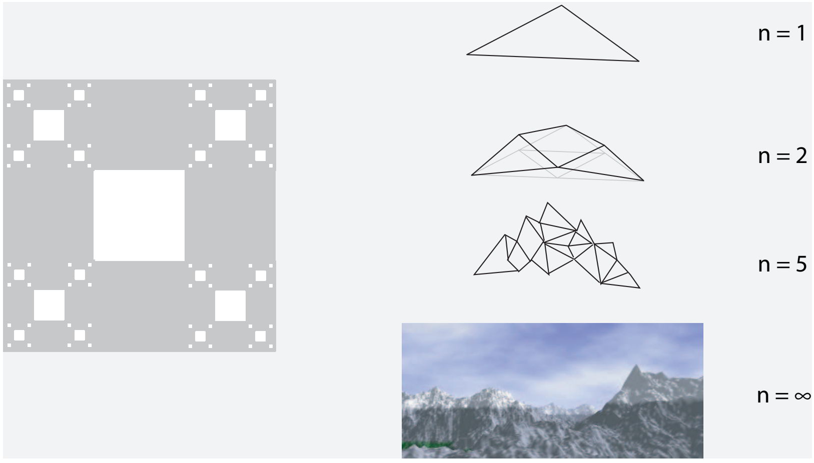


FIGURE 116 Floors and mountains as fractals (photograph © Paul Martz)

Can a chocolate bar last forever?

> From a drop of water a logician could predict an Atlantic or a Niagara.
>
> Arthur Conan Doyle, *A Study in Scarlet*

Any child knows how to make a chocolate bar last forever: eat half the remainder every day. However, this method only works if matter is scale-invariant. In other words, the method only works if matter is either *fractal*, as it then would be scale-invariant for a discrete set of zoom factors, or *continuous*, in which case it would be scale-invariant for any zoom factor. Which case, if either, applies to nature?

Page 53 We have already encountered a fact making continuity a questionable assumption: continuity would allow us, as Banach and Tarski showed, to multiply food and any other matter by clever cutting and reassembling. Continuity would allow children to eat the *same* amount of chocolate every day, without ever buying a new bar. Matter is thus not continuous. Now, fractal chocolate is not ruled out in this way; but other experiments settle the question. Indeed, we note that melted materials do not take up much smaller volumes than solid ones. We also find that even under the highest pressures, materials do not shrink. Thus matter is not a fractal. What then is its structure?

To get an idea of the structure of matter we can take fluid chocolate, or even just some oil – which is the main ingredient of chocolate anyway – and spread it out over a large surface. For example, we can spread a drop of oil onto a pond on a day without rain or wind; it is not difficult to observe which parts of the water are covered by the oil and which are not. A small droplet of oil cannot cover a surface larger than – can you guess the value? Challenge 437 n Trying to spread the film further inevitably rips it apart. The child's method of prolonging chocolate thus does not work for ever: it comes to a sudden end. The oil experiment shows that there is a *minimum* thickness of oil films, with a value of about 2 nm. This simple experiment can even be conducted at home; it shows that there is a smallest size in matter. Matter is made of tiny components. This confirms the observations made by Joseph Loschmidt* in 1865, who was the first person to measure the size of the components of matter.** In 1865, it was not a surprise that matter was made of small components, as the existence of a smallest size – but not its value – had already been deduced by Galileo, when studying some other simple questions.***

How high can animals jump?

Fleas can jump to heights a hundred times their size, humans only to heights about their own size. In fact, biological studies yield a simple observation: most animals, regardless of their size, achieve about the same jumping height of between 0.8 and 2.2 m, whether they are humans, cats, grasshoppers, apes, horses or leopards. The explanation of this fact takes only two lines. Can you find it?

Ref. 180

Challenge 438 n

The above observation seems to be an example of scale invariance. But there are some interesting exceptions at both ends of the mass range. At the small end, mites and other small insects do not achieve such heights because, like all small objects, they encounter the problem of air resistance. At the large end, elephants do not jump that high, because doing so would break their bones. But why do bones break at all?

* Joseph Loschmidt (b. 1821 Putschirn, d. 1895 Vienna) Austrian chemist and physicist. The oil experiment was popularized a few decades later, by Kelvin. It is often claimed that Benjamin Franklin was the first to conduct the oil experiment; that is wrong. Franklin did not measure the thickness, and did not even consider the question of the thickness. He did pour oil on water, but missed the most important conclusion that could be drawn from it. Even geniuses do not discover everything.

** Loschmidt knew that the (dynamic) viscosity of a gas was given by $\eta = \rho l v/3$, where ρ is the density of the gas, v the average speed of the components and l their mean free path. With Avogadro's prediction (made in 1811 without specifying any value) that a volume V of any gas always contains the same number N of components, one also has $l = V/\sqrt{2\pi N\sigma^2}$, where σ is the cross-section of the components. (The cross-section is the area of the shadow of an object.) Loschmidt then assumed that when the gas is liquefied, the volume of the liquid is the sum of the volumes of the particles. He then measured all the involved quantities and determined N. The modern value of N, called *Avogadro's number* or *Loschmidt's number*, is $6.02 \cdot 10^{23}$ particles in 22.4 l of any gas at standard conditions (today called 1 mol).

*** Galileo was brought to trial because of his ideas about atoms, not about the motion of the Earth, as is often claimed. To get a clear view of the matters of dispute in the case of Galileo, especially those of interest to physicists, the best text is the excellent book by Pietro Redondi, *Galileo eretico*, Einaudi, 1983, translated into English as *Galileo Heretic*, Princeton University Press, 1987. It is also available in many other languages. Redondi, a renowned historical scholar and colleague of Pierre Costabel, tells the story of the dispute between Galileo and the reactionary parts of the Catholic Church. He discovered a document of that time – the anonymous denunciation which started the trial – that allowed him to show that the condemnation of Galileo to life imprisonment for his views on the Earth's motion was organized by his friend the Pope to *protect* him from a sure condemnation to death over a different issue.

The reasons for his arrest, as shown by the denunciation, were not his ideas on astronomy and on the motion of the Earth, but his statements on matter. Galileo defended the view that since matter is not scale invariant, it must be made of 'atoms' or, as he called them, *piccolissimi quanti* – smallest quanta. This was and still is a heresy. A true Catholic is still not allowed to believe in atoms. Indeed, the theory of atoms is not compatible with the change of bread and wine into human flesh and blood, called *transsubstantiation*, which is a central tenet of the Catholic faith. In Galileo's days, church tribunals punished heresy, i.e. deviating personal opinions, by the death sentence. Despite being condemned to prison in his trial, Galileo published his last book, written as an old man under house arrest, on the scaling issue. Today, the Catholic Church still refuses to publish the proceedings and other documents of the trial. Its officials carefully avoid the subject of atoms, as any statement on this subject would make the Catholic Church into a laughing stock. Indeed, quantum theory, named after the term used by Galileo, has become the most precise description of nature yet.

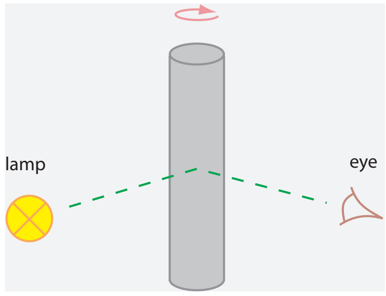


FIGURE 117 Atoms exist: rotating an aluminium rod leads to brightness oscillations

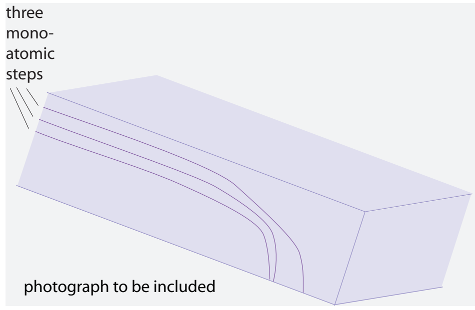


FIGURE 118 Atomic steps in broken gallium arsenide crystals can be seen under a light microscope

Why are all humans of about the same size? Why are there no giant adults with a height of ten metres? Why aren't there any land animals larger than elephants? The answer yields the key to understanding the structure of matter. In fact, the materials of which we are made would not allow such changes of scale, as the bones of giants would collapse under the weight they have to sustain. Bones have a finite strength because their constituents stick to each other with a finite attraction. Continuous matter – which exists only in cartoons – could not break at all, and fractal matter would be infinitely fragile. Matter breaks under finite loads because it is composed of small basic constituents.

Felling trees

The gentle lower slopes of Motion Mountain are covered by trees. Trees are fascinating structures. Take their size. Why do trees have limited size? Already in the sixteenth century, Galileo knew that it is not possible to increase tree height without limits: at some point a tree would not have the strength to support its own weight. He estimated the maximum height to be around 90 m; the actual record, unknown to him at the time, seems to be 150 m, for the Australian tree *Eucalyptus regnans*. But why does a limit exist at all? The answer is the same as for bones: wood has a finite strength because it is not scale invariant; and it is not scale invariant because it is made of small constituents, namely atoms.*

In fact, the derivation of the precise value of the height limit is more involved. Trees must not break under strong winds. Wind resistance limits the height-to-thickness ratio h/d to about 50 for normal-sized trees (for $0.2\,\mathrm{m} < d < 2\,\mathrm{m}$). Can you say why? Thinner trees are limited in height to less than 10 m by the requirement that they return to the vertical after being bent by the wind.

Challenge 439 ny

Ref. 182

Such studies of natural constraints also answer the question of why trees are made from wood and not, for example, from steel. You could check for yourself that the maximum height of a column of a given mass is determined by the ratio E/ρ^2 between the elastic module and the square of the mass density. Wood is actually the material for which this ratio is highest. Only recently have material scientists managed to engineer slightly better ratios with fibre composites.

Challenge 440 n

Ref. 183

Ref. 181

* There is another important limiting factor: the water columns inside trees must not break. Both factors seem to yield similar limiting heights.

Why do materials break at all? All observations yield the same answer and confirm Galileo's reasoning: because there is a smallest size in materials. For example, bodies under stress are torn apart at the position at which their strength is minimal. If a body were completely homogeneous, it could not be torn apart; a crack could not start anywhere. If a body had a fractal Swiss-cheese structure, cracks would have places to start, but they would need only an infinitesimal shock to do so.

A simple experiment that shows that solids have a smallest size is shown in Figure 117. A cylindrical rod of pure, single crystal aluminium shows a surprising behaviour when it is illuminated from the side: its brightness depends on how the rod is oriented, even though it is completely round. This angular dependence is due to the atomic arrangement of the aluminium atoms in the rod.

It is not difficult to confirm experimentally the existence of smallest size in solids. It is sufficient to break a single crystal, such as a gallium arsenide wafer, in two. The breaking surface is either completely flat or shows extremely small steps, as shown in Figure 118.

Challenge 441 ny These steps are visible under a normal light microscope. (Why?) It turns out that all the step heights are multiples of a smallest height: its value is about 0.2 nm. The existence of a smallest height, corresponding to the height of an atom, contradicts all possibilities of scale invariance in matter.

The sound of silence

Climbing the slopes of Motion Mountain, we arrive in a region of the forest covered with deep snow. We stop for a minute and look around. It is dark; all the animals are asleep; there is no wind and there are no sources of sound. We stand still, without breathing, and listen to the silence. (You can have this experience also in a sound studio such as those used for musical recordings, or in a quiet bedroom at night.) In situations of complete silence, the ear automatically becomes more sensitive*; we then have a strange experience. We hear two noises, a lower- and a higher-pitched one, which are obviously generated inside the ear. Experiments show that the higher note is due to the activity of the nerve cells in the inner ear. The lower note is due to pulsating blood streaming through the head. But why do we hear a noise at all?

Many similar experiments confirm that whatever we do, we can never eliminate noise from measurements. This unavoidable type of noise is called *shot noise* in physics. The statistical properties of this type of noise actually correspond precisely to what would be expected if flows, instead of being motions of continuous matter, were transportation of a large number of equal, small and discrete entities. Thus, simply listening to noise proves that electric current is made of electrons, that air and liquids are made of molecules, and that light is made of photons. In a sense, the sound of silence is the sound of atoms. Shot noise would not exist in continuous systems.

Little hard balls

> “I prefer knowing the cause of a single thing to being king of Persia.
>
> Democritus ”

* The human ear can detect pressure variations at least as small as 20 μPa.

Precise observations show that matter is neither continuous nor a fractal: matter is made of smallest basic particles. Galileo, who deduced their existence by thinking about giants and trees, called them 'smallest quanta.' Today they are called 'atoms', in honour of a famous argument of the ancient Greeks. Indeed, 2500 years ago, the Greeks asked the following question. If motion and matter are conserved, how can change and transformation exist? The philosophical school of Leucippus and Democritus of Abdera* studied two particular observations in special detail. They noted that salt dissolves in water. They also noted that fish can swim in water. In the first case, the volume of water does not increase when the salt is dissolved. In the second case, when fish advance, they must push water aside. Leucippus and Democritus deduced that there is only one possible explanation that satisfies observations and also reconciles conservation and transformation: nature is made of void and of small, indivisible and conserved particles.** In this way any example of motion, change or transformation is due to rearrangements of these particles; change and conservation are thus reconciled.

In short, since matter is hard, has a shape and is divisible, Leucippus and Democritus imagined it as being made of atoms. Atoms are particles which are hard, have a shape, but are indivisible. In other words, the Greeks imagined nature as a big Lego set. Lego pieces are first of all hard or *impenetrable*, i.e. repulsive at very small distances. They are *attractive* at small distances: they remain stuck together. Finally, they have *no interaction* at large distances. Atoms behave in the same way. (Actually, what the Greeks called 'atoms' partly corresponds to what today we call 'molecules'. The latter term was invented by Amadeo Avogadro in 1811 in order to clarify the distinction. But we can forget this detail for the moment.)

Since atoms are invisible, it took many years before all scientists were convinced by the experiments showing their existence. In the nineteenth century, the idea of atoms was Page 248 beautifully verified by the discovery of the 'laws' of chemistry and those of gas behaviour. Later on, the noise effects were discovered.

Nowadays, with advances in technology, single atoms can be seen, photographed, holo- Ref. 184, Ref. 185 grammed, counted, touched, moved, lifted, levitated, and thrown around. And indeed,

* Leucippus of Elea (Λευκιππος) (*c.* 490 to *c.* 430 bce), Greek philosopher; Elea was a small town south of Naples. It lies in Italy, but used to belong to the Magna Graecia. Democritus (Δεμοκριτος) of Abdera (*c.* 460 to *c.* 356 or 370 bce), also a Greek philosopher, was arguably the greatest philosopher who ever lived. Together with his teacher Leucippus, he was the founder of the atomic theory; Democritus was a much admired thinker, and a contemporary of Socrates. The vain Plato never even mentions him, as Democritus was a danger to his own fame. Democritus wrote many books which all have been lost; they were not copied during the Middle Ages because of his scientific and rational world view, which was felt to be a danger by religious zealots who had the monopoly on the copying industry. Nowadays, it has become common to claim – incorrectly – that Democritus had no proof for the existence of atoms. That is a typical example of disinformation with the aim of making us feel superior to the ancients.

** The story is told by Lucretius, in full Titus Lucretius Carus, in his famous text *De rerum natura*, around 60 bce. (An English translation can be found on http://perseus.uchicago.edu/hopper/text.jsp?doc=Perseus:text:1999.02.0131.) Lucretius relates many other proofs; in Book 1, he shows that there is vacuum in solids – as proven by porosity and by density differences – and in gases – as proven by wind. He shows that smells are Challenge 442 ny due to particles, and that so is evaporation. (Can you find more proofs?) He also explains that the particles cannot be seen due to their small size, but that their effects can be felt and that they allow to consistently explain all observations.

Especially if we the imagine particles as little balls, we cannot avoid calling this a typically male idea. Challenge 443 d (What would be the female approach?)

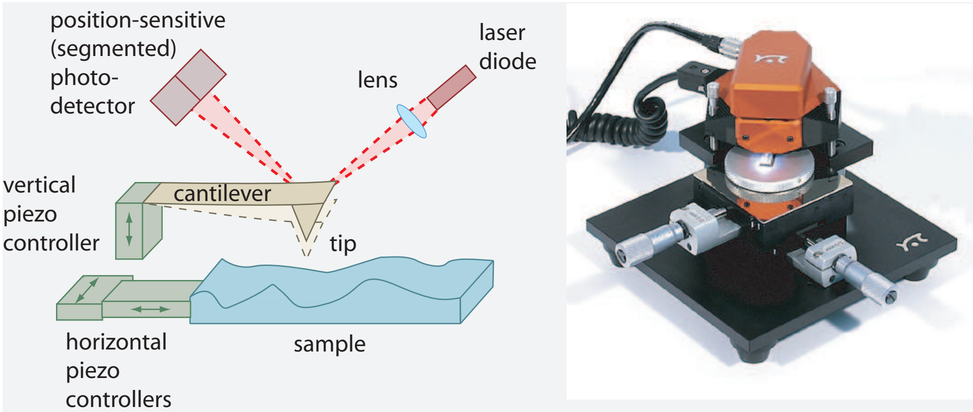


FIGURE 119 The principle and a realization of an atomic force microscope (photograph © Nanosurf)

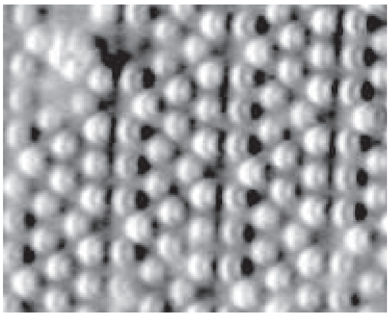


FIGURE 120 The atoms on the surface of a silicon crystal mapped with an atomic force microscope (© Universität Augsburg)

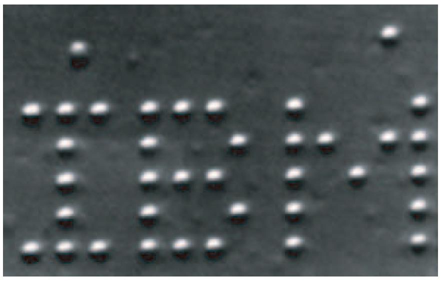


FIGURE 121 The result of moving helium atoms on a metallic surface (© IBM)

like everyday matter, atoms have mass, size, shape and colour. Single atoms have even been used as lamps and lasers. Ref. 186

Modern researchers in several fields have fun playing with atoms in the same way that children play with Lego. Maybe the most beautiful demonstration of these possibilities is provided by the many applications of the atomic force microscope. Ref. 187 If you ever have the opportunity to see one, do not miss it!* It is a simple device which follows the surface of an object with an atomically sharp needle; such needles, usually of tungsten, are easily manufactured with a simple etching method. Ref. 188 The changes in the height of the needle along its path over the surface are recorded with the help of a deflected light ray. With a little care, the atoms of the object can be felt and made visible on a computer screen. With special types of such microscopes, the needle can be used to move atoms one by one to specified places on the surface. It is also possible to scan a surface, pick up a given atom and throw it towards a mass spectrometer to determine what sort of atom it is. Ref. 189

* A cheap version costs only a few thousand euro, and will allow you to study the difference between a silicon wafer – crystalline – a flour wafer – granular-amorphous – and a consecrated wafer.

Figures to be inserted

FIGURE 122 Some examples of fluid motion

Incidentally, the construction of atomic force microscopes is only a small improvement on what nature is building already by the millions; when we use our ears to listen, we are actually detecting changes in eardrum position of about 1 nm. In other words, we all have two 'atomic force microscopes' built into our heads.

In summary, matter is not scale invariant: in particular, it is neither smooth nor fractal. Matter is made of atoms. Different types of atoms, as well as their various combinations, produce different types of substances. Pictures from atomic force microscopes show that the size and arrangement of atoms produce the *shape* and the *extension* of objects, confirming the Lego model of matter.* As a result, the description of the motion of extended objects can be reduced to the description of the motion of their atoms. Atomic motion will be a major theme in the following pages. One of its consequences is especially important: heat. Before we study it, we have a look at fluids.

The motion of fluids

Fluids can be liquids or gases. Their motion can be exceedingly complex, as Figure 122 shows. Such complex motions are often examples of self-organization or chaos; these are Page 263 discussed below.

Like all motion, fluid motion obeys energy conservation. In the case that no energy is transformed into heat, the conservation of energy is particularly simple. Motion that does not generate heat implies the lack of vortices; such fluid motion is called *laminar*. If the speed of the fluid does not depend on time at all positions, it is called *stationary*. For motion that is both laminar and stationary, energy conservation can be expressed with speed v and pressure p:

$$\frac{1}{2}\rho v^2 + p + \rho g z = \text{const} \qquad (88)$$

where z is the height above ground. This is called *Bernoulli's equation*. In this equation, the first term is the kinetic energy (per volume) of the fluid, and the other two terms are

* Studying matter in even more detail yields the now well-known idea that matter, at higher and higher magnifications, is made of molecules, atoms, nuclei, protons and neutrons, and finally, quarks. Atoms also contain electrons. A final type of matter, neutrinos, is observed coming from the Sun and from certain types of radioactive materials. Even though the fundamental bricks have become smaller with time, the basic idea remains: matter is made of smallest entities, nowadays called elementary particles. In the second part of our Page 943 mountain ascent we will explore this idea in detail. Appendix C lists the measured properties of all known elementary particles.

potential energies (per volume). The last term is only important if the fluid rises against ground. The second term is the potential energy (per volume) resulting from the compression of the fluid. Indeed, pressure is a potential energy per volume. Challenge 444 e

Energy conservation implies that the lower the pressure is, the larger the speed of a fluid becomes. One can use this relation to measure the speed of a stationary water flow in a tube. One just has to narrow the tube somewhat at one location along the tube, and measure the pressure difference before and at the tube restriction. One finds that the sped v is given as $v = k\sqrt{p_1 - p_2}$. (What is the constant k?) A device using this method is called a Venturi gauge. Challenge 445 n

If the geometry of a system is kept fixed and the fluid speed is increased, at a certain speed one observes a transition: the liquid loses its clarity, the flow is not stationary any more. This is seen whenever a water tap is opened. The flow has changed from laminar to turbulent. At this point, Bernoulli's equation is not valid any more.

The description of turbulence might be the toughest of all problems in physics. When the young Werner Heisenberg was asked to continue research on turbulence, he refused – rightly so – saying it was too difficult; he turned to something easier and discovered and developed quantum mechanics instead. Turbulence is such a vast topic, with many of its concepts still not settled, that despite the number and importance of its applications, only now, at the beginning of the twenty-first century, are its secrets beginning to be unravelled. It is thought that the equations of motion describing fluids, the so-called *Navier–Stokes equations*, are sufficient to understand turbulence.* But the mathematics behind them is mind-boggling. There is even a prize of one million dollars offered by the Clay Mathematics Institute for the completion of certain steps on the way to solving the equations. Ref. 190

Important systems which show laminar flow, vortices and turbulence at the same time are wings and sails. All wings work best in laminar mode. The essence of a wing is that it imparts air a downward velocity with as little turbulence as possible. (The aim to minimize tubulence is the reason that wings are curved. If the engine is very powerful, a flat wing at an angle also works. Strong turbulence is also of andavntage for landing safely.) The downward velocity of the trailing air leads to a centrifugal force acting on the air that passes above the wing. This leads to a lower pressure, and thus to lift. (Wings thus do *not* rely on the Bernoulli equation, where lower pressure *along* the flow leads to higher air speed, as unfortunately, many books used to say. Above a wing, the higher speed is related to lower pressure *across* the flow.) Ref. 191

The different speeds of the air above and below the wing lead to vortices at the end of every wing. These vortices are especially important for the takeoff of any insect, bird and aeroplane. More details on wings are discussed later on. Page 991

Curiosities and fun challenges about fluids

How much water is necessary to moisten the air in a room in winter? At 0°C, the vapour pressure of water is 6 mbar, 20°C it is 23 mbar. As a result, heating air in the winter gives at most a humidity of 25%. To increase the humidity by 50%, one thus needs about 1 litre Challenge 446 e

* They are named after Claude Navier (b. 1785 Dijon, d. 1836 Paris), important French engineer and bridge builder, and Georges Gabriel Stokes (b. 1819 Skreen, d. 1903 Cambridge), important Irish physicist and mathematician.

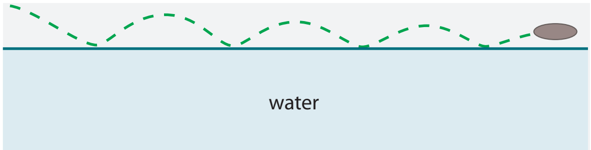


FIGURE 123 What is your personal stone-skipping record?

of water per 100 m^3.

* *

Challenge 447 n You are in a boat on a pond with a stone, a bucket of water and a piece of wood. What happens to the water level of the pond after you throw the stone in it? After you throw the water into the pond? After you throw the piece of wood?

* *

Challenge 448 n A ship leaves a river and enter the sea. What happens?

* *

Challenge 449 e Put a rubber air balloon over the end of a bottle and let it hang inside the bottle. How much can you blow up the balloon inside the bottle?

* *

Challenge 450 e Put a small paper ball into the neck of a horizontal bottle and try to blow it into the bottle. The paper will fly *towards* you. Why?

* *

Challenge 451 e It is possible to blow an egg from one egg-cup to a second one just behind it. Can you perform this trick?

* *

Challenge 452 n In the seventeenth century, engineers who needed to pump water faced a challenge. To pump water from mine shafts to the surface, no water pump managed more than 10 m of height difference. For twice that height, one always needed two pumps in series, connected by an intermediate reservoir. Why? How then do trees manage to pump water upwards for larger heights?

* *

Challenge 453 n When hydrogen and oxygen are combined to form water, the amount of hydrogen needed is exactly twice the amount of oxygen, if no gas is to be left over after the reaction. How does this observation confirm the existence of atoms?

* *

Challenge 454 n How are alcohol-filled chocolate pralines made? Note that the alcohol is not injected into them afterwards, because there would be no way to keep the result tight enough.

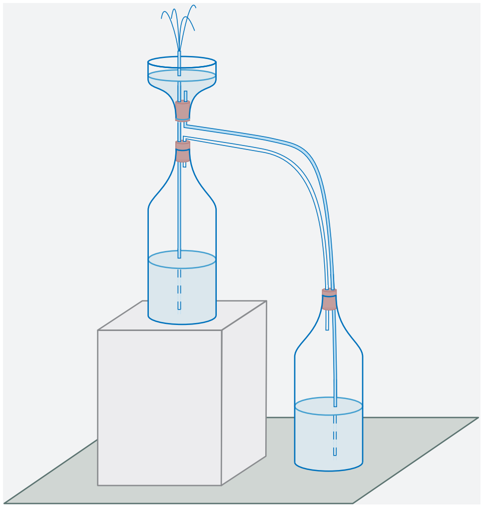


FIGURE 124 Heron's fountain

* *

How often can a stone jump when it is thrown over the surface of water? The present world record was achieved in 2002: 40 jumps. More information is known about the previous world record, achieved in 1992: a palm-sized, triangular and flat stone was thrown with a speed of 12 m/s (others say 20 m/s) and a rotation speed of about 14 revolutions per second along a river, covering about 100 m with 38 jumps. (The sequence was filmed with a video recorder from a bridge.) Ref. 192

What would be necessary to increase the number of jumps? Can you build a machine that is a better thrower than yourself? Challenge 455 r

* *

The biggest component of air is nitrogen (about 78 %). The second biggest component is oxygen (about 21 %). What is the third biggest one? Challenge 456 n

* *

Water can flow uphill: Heron's fountain shows this most clearly. Heron of Alexandria (*c.* 10 to *c.* 70) described it 2000 years ago; it is easily built at home, using some plastic bottles and a little tubing. How does it work? Challenge 457 n

* *

A light bulb is placed, underwater, in a stable steel cylinder with a diameter of 16 cm. A Fiat Cinquecento (500 kg) is placed on a piston pushing onto the water surface. Will the bulb resist? Challenge 458 n

* *

Challenge 459 ny What is the most dense gas? The most dense vapour?

* *

Every year, the Institute of Maritime Systems of the University of Rostock organizes a contest. The challenge is to build a paper boat with the highest carrying capacity. The paper boat must weigh at most 10 g; the carrying capacity is measured by pouring lead small shot onto it, until the boat sinks. The 2002 record stands at 2.6 kg. Can you achieve this value? (For more information, see the http://www.paperboat.de website.) Challenge 460 e

* *

A modern version of an old question – already posed by Daniel Colladon (1802–1893) – is the following. A ship of mass m in a river is pulled by horses walking along the riverbank attached by ropes. If the river is of superfluid helium, meaning that there is no friction between ship and river, what energy is necessary to pull the ship upstream along the river until a height h has been gained? Challenge 461 n

* *

The Swiss professor Auguste Piccard (1884–1962) was a famous explorer of the stratosphere. He reached a height of 16 km in his *aerostat*. Inside the airtight cabin hanging under his balloon, he had normal air pressure. However, he needed to introduce several ropes attached at the balloon into the cabin, in order to be able to pull them, as they controlled his balloon. How did he get the ropes into the cabin while preventing air from leaving the cabin? Challenge 462 n

* *

A human cannot breathe at any depth under water, even if he has a tube going to the surface. At a few metres of depth, trying to do so is inevitably fatal! Even at a depth of 60 cm only, the human body can only breathe in this way for a few minutes. Why? Challenge 463 n

* *

A human in air falls with a limiting speed of about 180 km/h, depending on clothing. How long does it take to fall from a plane at 3000 m down to a height of 200 m? Challenge 464 ny

* *

Several humans have survived free falls from aeroplanes for a thousand metres or more, even though they had no parachute. How was this possible? Challenge 465 n

* *

Liquid pressure depends on height. If the average human blood pressure at the height of the heart is 13.3 kPa, can you guess what it is inside the feet when standing? Challenge 466 n

* *

The human heart pumps blood at a rate of about 0.1 l/s. A capillary has the diameter of a red blood cell, around 7 μm, and in it the blood moves at a speed of half a millimetre per second. How many capillaries are there in a human? Challenge 467 n

* *

A few drops of tea usually flow along the underside of the spout of a teapot (or fall onto the table). This phenomenon has even been simulated using supercomputer simulations of

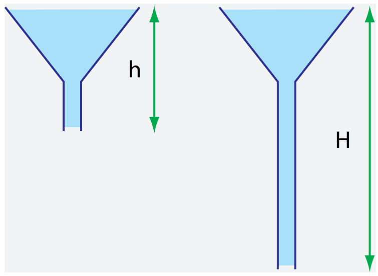


FIGURE 125 Which funnel is faster?

Ref. 193 the motion of liquids, by Kistler and Scriven, using the Navier–Stokes equations. Teapots are still shedding drops, though.

* *

The best giant soap bubbles can be made by mixing 1.5 l of water, 200 ml of corn syrup and 450 ml of washing-up liquid. Mix everything together and then let it rest for four hours. You can then make the largest bubbles by dipping a metal ring of up to 100 mm diameter into the mixture. But why do soap bubbles burst? Challenge 468 n

* *

A drop of water that falls into a pan containing hot oil dances on the surface for a considerable time, if the oil is above 220 °C. Cooks test the temperature of oil in this way. Why does this so-called Leidenfrost effect* take place? Challenge 469 ny

* *

Challenge 470 n Why don't air molecules fall towards the bottom of the container and stay there?

* *

Challenge 471 n Which of the two water funnels in Figure 125 is emptied more rapidly? Apply energy conservation to the fluid's motion (also called Bernoulli's 'law') to find the answer. Ref. 194

* *

As we have seen, fast flow generates an underpressure. How do fish prevent their eyes from popping when they swim rapidly? Challenge 472 n

* *

Golf balls have dimples for the same reasons that tennis balls are hairy and that shark and dolphin skin is not flat: deviations from flatness reduce the flow resistance because many small eddies produce less friction than a few large ones. Why? Challenge 473 ny

* *

Glass is a solid. Nevertheless, many textbooks say that glass is a liquid. This error has been propagated for about a hundred years, probably originating from a mistranslation of a

* It is named after Johann Gottlieb Leidenfrost (1715–1794), German physician.

sentence in a German textbook published in 1933 by Gustav Tamman, *Der Glaszustand*.
Challenge 474 n Can you give at least three reasons why glass is a solid and not a liquid?

* *

The recognized record height reached by a helicopter is 12 442 m above sea level, though 12 954 m has also been claimed. (The first height was reached in 1972, the second in 2002, both by French pilots in French helicopters.) Why, then, do people still continue to use their legs in order to reach the top of Mount Sagarmatha, the highest mountain in the world?
Challenge 475 n

* *

A loosely knotted sewing thread lies on the surface of a bowl filled with water. Putting a bit of washing-up liquid into the area surrounded by the thread makes it immediately become circular. Why?
Challenge 476 e

* *

Challenge 477 n How can you put a handkerchief under water using a glass, while keeping it dry?

* *

Are you able to blow a ping pong ball out of a funnel? What happens if you blow through a funnel towards a burning candle?

* *

The fall of a leaf, with its complex path, is still a topic of investigation. We are far from being able to predict the time a leaf will take to reach the ground; the motion of the air around a leaf is not easy to describe. One of the simplest phenomena of hydrodynamics remains one of its most difficult problems.

* *

Fluids exhibit many interesting effects. Soap bubbles in air are made of a thin spherical film of liquid with air on both sides. In 1932, anti-bubbles, thin spherical films of air with liquid on both sides, were first observed. In 2004, the Belgian physicist Stéphane Dorbolo and his team showed that it is possible to produce them in simple experiments, and in particular, in Belgian beer.
Ref. 196

* *

Have you ever dropped a Mentos candy into a Diet Coca Cola bottle? You will get an interesting effect. (Do it at your own risk...) Is it possible to build a rocket in this way?
Challenge 478 e

* *

Challenge 479 e A needle can swim on water, if you put it there carefully. Just try, using a fork.

* *

Fluids exhibit many complex motions. To see an overview, have a look at the beautiful collection on the web site http://serve.me.nus.edu.sg/limtt. One of the most famous examples of fluid motion is the leapfrogging of vortex rings, shown in Figure 126. Lim Tee
Challenge 480 e

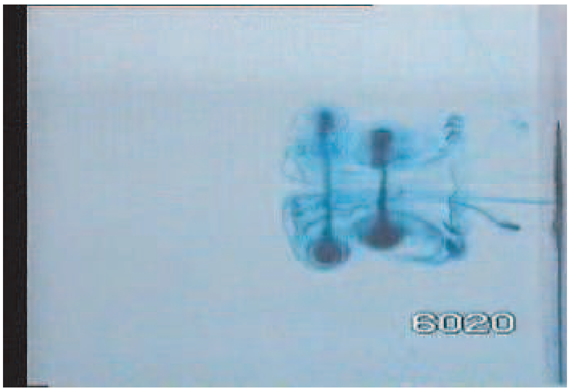


FIGURE 126 Two leapfrogging vortex rings (mpeg © Lim Tee Tai)

Tai explains that more two leapfrogs are extremely hard to achieve, because the slightest vortex ring misalignment leads to the collapse of the system. Ref. 200

Curiosities and fun challenges about solids

What is the maximum length of a vertically hanging wire? Could a wire be lowered from a suspended geostationary satellite down to the Earth? This would mean we could realize a space 'lift'. How long would the cable have to be? How heavy would it be? How would you build such a system? What dangers would it face? Challenge 481 n

* *

Matter is made of atoms. Over the centuries the stubborn resistance of many people to this idea has lead to the loss of many treasures. For over a thousand years, people thought that genuine pearls could be distinguished from false ones by hitting them with a hammer: only false pearls would break. However, *all* pearls break. (Also diamonds break in this situation.) As a result, all the most beautiful pearls in the world have been smashed to pieces.

* *

Comic books have difficulties with the concept of atoms. Could Asterix really throw Romans into the air using his fist? Are Lucky Luke's precise revolver shots possible? Can Spiderman's silk support him in his swings from building to building? Can the Roadrunner stop running in three steps? Can the Sun be made to stop in the sky by command? Can space-ships hover using fuel? Take any comic-book hero and ask yourself whether matter made of atoms would allow him the feats he seems capable of. You will find that most cartoons are comic precisely because they assume that matter is not made of atoms, but continuous! In a sense, atoms make life a serious adventure. Challenge 482 e

* *

Can humans start earthquakes? What would happen if 1000 million Indians were to jump

Challenge 483 n at the same time from the kitchen table to the floor?

In fact, several strong earthquakes *have* been triggered by humans. This has happened when water dams have been filled, or when water has been injected into drilling holes. It has been suggested that the extraction of deep underground water also causes earthquakes. If this is confirmed, a sizeable proportion of all earthquakes could be human-triggered.

* *

Challenge 484 n How can a tip of a stalactite be distinguished from a tip of a stalagmite? Does the difference exist also for icicles?

* *

How much more weight would your bathroom scales show if you stood on them in a vacuum? Challenge 485 n

* *

One of the most complex extended bodies is the human body. In modern simulations of the behaviour of humans in car accidents, the most advanced models include ribs, vertebrae, all other bones and the various organs. For each part, its specific deformation properties are taken into account. With such models and simulations, the protection of passengers and drivers in cars can be optimized.

* *

The deepest hole ever drilled into the Earth is 12 km deep. In 2003, somebody proposed to enlarge such a hole and then to pour millions of tons of liquid iron into it. He claims that the iron would sink towards the centre of the Earth. If a measurement device communication were dropped into the iron, it could send its observations to the surface using sound waves. Can you give some reasons why this would not work? Challenge 486 n

* *

The economic power of a nation has long been associated with its capacity to produce high-quality steel. Indeed, the Industrial Revolution started with the mass production of steel. Every scientist should know the basics facts about steel. *Steel* is a combination of iron and carbon to which other elements, mostly metals, may be added as well. One can distinguish three main types of steel, depending on the crystalline structure. *Ferritic steels* have a body-centred cubic structure, *austenitic steels* have a face-centred cubic structure, and *martensitic steels* have a body-centred tetragonal structure. Table 28 gives further details.

* *

A simple phenomenon which requires a complex explanation is the cracking of a whip. Since the experimental work of Peter Krehl it has been known that the whip cracks when the tip reaches a velocity of *twice* the speed of sound. Can you imagine why? Ref. 195 Challenge 487 ny

* *

A bicycle chain is an extended object with no stiffness. However, if it is made to ro-

TABLE 28 Steel types, properties and uses

Ferritic steel	Austenitic steel	Martensitic steel
'usual' steel	'soft' steel	hardened steel, brittle
body centred cubic (bcc)	face centred cubic (fcc)	body centred tetragonal (bct)
iron and carbon	iron, chromium, nickel, manganese, carbon	carbon steel and alloys
Examples		
construction steel	most stainless (18/8 Cr/Ni) steels	knife edges
car sheet steel	kitchenware	drill surfaces
ship steel	food industry	spring steel, crankshafts
12 % Cr stainless ferrite	Cr/V steels for nuclear reactors	
Properties		
phases described by the iron-carbon phase diagram	phases described by the Schaeffler diagram	phases described by the iron-carbon diagram and the TTT (time–temperature transformation) diagram
in equilibrium at RT	some alloys in equilibrium at RT	not in equilibrium at RT, but stable
mechanical properties and grain size depend on heat treatment	mechanical properties and grain size depend on thermo-mechanical pre-treatment	mechanical properties and grain size strongly depend on heat treatment
hardened by reducing grain size, by forging, by increasing carbon content or by nitration	hardened by cold working only	hard anyway – made by laser irradiation, induction heating, etc.
grains of ferrite and paerlite, with cementite (Fe_3C)	grains of austenite	grains of martensite
ferromagnetic	not magnetic or weakly magnetic	ferromagnetic

tate rapidly, it acquires dynamical stiffness, and can roll down an inclined plane or along the floor. This surprising effect can be watched at http://www.iwf.de/NR/rdonlyres/EEFA7FDC-DDDC-490C-9C49-4537A925EFE6/718/C14825.asx or http://www.iwf.de/NR/rdonlyres/EEFA7FDC-DDDC-490C-9C49-4537A925EFE6/793/C148292.smil.

* *

Mechanical devices are not covered in this text. There is a lot of progress in the area even at present. For example, people have built robots that are able to ride a unicycle. But even the physics of human unicycling is not simple.

Ref. 197
Ref. 198

* *

TABLE 29 Extensive quantities in nature, i.e. quantities that *flow* and *accumulate*

Domain	Extensive quantity (energy carrier)	Current (flow intensity)	Intensive quantity (driving strength)	Energy flow (power)	Resistance to transport (intensity of entropy generation)
Rivers	mass m	mass flow m/t	height difference gh	$P = gh\, m/t$	$R_{m} = ght/m$ $[m^2/s\,kg]$
Gases	volume V	volume flow V/t	pressure p	$P = pV/t$	$R_{V} = pt/V$ $[kg/s\,m^5]$
Mechanics	momentum $\mathbf{p}$	force $\mathbf{F} = d\mathbf{p}/dt$	velocity $\mathbf{v}$	$P = \mathbf{v}\,\mathbf{F}$	$R_{p} = t/m$ $[s/kg]$
	angular momentum $\mathbf{L}$	torque $\mathbf{M} = d\mathbf{L}/dt$	angular velocity $\boldsymbol{\omega}$	$P = \boldsymbol{\omega}\,\mathbf{M}$	$R_{L} = t/mr^2$ $[s/kg\,m^2]$
Chemistry	amount of substance n	substance flow $I_n = dn/dt$	chemical potential μ	$P = \mu\, I_n$	$R_n = \mu t/n$ $[Js/mol^2]$
Thermodynamics	entropy S	entropy flow $I_S = dS/dt$	temperature T	$P = T\, I_S$	$R_S = Tt/S$ $[K^2/W]$
Light	like all massless radiation, it can flow but cannot accumulate				
Electricity	charge q	electrical current $I = dq/dt$	electrical potential U	$P = U\, I$	$R = U/I$ $[\Omega]$
Magnetism	no accumulable magnetic sources are found in nature				
Nuclear physics	extensive quantities exist, but do not appear in everyday life				
Gravitation	empty space can move and flow, but the motion is not observed in everyday life				

There are many arguments against the existence of atoms as hard balls. Thomson-Kelvin put it in writing: "the monstruous assumption of infinitely strong and infinitely rigid pieces of matter." Even though Thomson was right in his comment, atoms do exist. Why?

Ref. 199
Challenge 488 n

What can move in nature?

Before we continue to the next way to describe motion globally, we will have a look at the possibilities of motion in everyday life. One overview is given in Table 29. The domains that belong to everyday life – motion of fluids, of matter, of matter types, of heat, of light and of charge – are the domains of continuum physics.

Within continuum physics, there are three domains we have not yet studied: the motion of charge and light, called electrodynamics, the motion of heat, called thermodynamics, and the motion of the vacuum. Once we have explored these domains, we will have completed the first step of our description of motion: continuum physics. In continuum physics, motion and moving entities are described with continuous quantities that can take any value, including arbitrarily small or arbitrarily large values.

But nature is *not* continuous. We have already seen that matter cannot be indefinitely divided into ever-smaller entities. In fact, we will discover that there are precise experiments that provide limits to the observed values for *every* domain of continuum physics. There is a limit to mass, to speed, to angular momentum, to force, to entropy and to change of charge. The consequences of these discoveries form the second step in our description of motion: quantum theory and relativity. Quantum theory is based on lower limits; relativity is based on upper limits. The third and last step of our description of motion will be formed by the unification of quantum theory and general relativity.

Every domain of physics, regardless of which one of the above steps it belongs to, describes change in terms two quantities: energy, and an extensive quantity characteristic of the domain. An observable quantity is called *extensive* if it increases with system size. Table 29 provides an overview. The intensive and extensive quantities corresponding to what in everyday language is called 'heat' are *temperature* and *entropy*.

Ref. 202

How do objects get warm?

We continue our short stroll through the field of global descriptions of motion with an overview of heat and the main concepts associated with it. For our purposes we only need to know the basic facts about heat. The main points that are taught in school are almost sufficient.

Ref. 203

Macroscopic bodies, i.e. bodies made of many atoms, have temperature. The temperature of a macroscopic body is an aspect of its state. It is observed that any two bodies in contact tend towards the same temperature: temperature is contagious. In other words, temperature describes an equilibrium situation. The existence and contagiousness of temperature is often called the *zeroth principle of thermodynamics*. Heating is the increase of temperature.

How is temperature measured? The eighteenth century produced the clearest answer: temperature is best defined and measured by the *expansion of gases*. For the simplest, so-called *ideal* gases, the product of pressure p and volume V is proportional to temperature:

$$pV \sim T \, . \qquad (89)$$

The proportionality constant is fixed by the *amount* of gas used. (More about it shortly.) The ideal gas relation allows us to determine temperature by measuring pressure and volume. This is the way (absolute) temperature has been defined and measured for about a century. To define the *unit* of temperature, one only has to fix the amount of gas used. It is customary to fix the amount of gas at 1 mol; for oxygen this is 32 g. The proportionality constant, called the *ideal gas constant* R, is defined to be $R = 8.3145\,\mathrm{J/mol\,K}$. This number has been chosen in order to yield the best approximation to the independently defined Celsius temperature scale. Fixing the ideal gas constant in this way defines 1 K, or one Kelvin, as the unit of temperature. In simple terms, a temperature increase of one Kelvin is defined as the temperature increase that makes the volume of an ideal gas increase –

Ref. 204
Page 1170

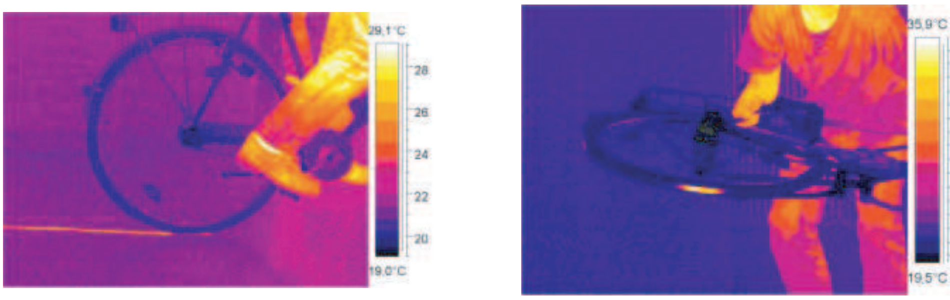


FIGURE 127 Braking generates heat on the floor and in the tire (© Klaus-Peter Möllmann and Michael Vollmer)

Challenge 489 ny keeping the pressure fixed – by a fraction of 1/273.15 or 0.3661 %.

In general, if one needs to determine the temperature of an object, one takes a mole of gas, puts it in contact with the object, waits a while, and then measures the pressure and the volume of the gas. The ideal gas relation (89) then gives the temperature. Most importantly, the ideal gas relation shows that there is a lowest temperature in nature, namely that temperature at which an ideal gas would have a vanishing volume. That would happen at $T = 0$ K, i.e. at −273.15°C. Obviously, other effects, like the volume of the atoms themselves, prevent the volume of the gas from ever reaching zero. The *third principle of thermodynamics* provides another reason why this is impossible.

Temperature

The temperature achieved by a civilization can be used as a measure of its technological achievements. One can define the Bronze Age (1.1 kK, 3500 bce) , the Iron Age (1.8 kK, 1000 bce), the Electric Age (3 kK from *c.* 1880) and the Atomic Age (several MK, from 1944) in this way. Taking into account also the quest for lower temperatures, one can define the Quantum Age (4 K, starting 1908). Ref. 205

Heating implies flow of energy. For example, friction heats up and slows down moving bodies. In the old days, the 'creation' of heat by friction was even tested experimentally. It was shown that heat could be generated from friction, just by continuous rubbing, without any limit; an example is shown in Figure 127. This 'creation' implies that heat is not a material fluid extracted from the body – which in this case would be consumed after a certain time – but something else. Indeed, today we know that heat, even though it behaves in some ways like a fluid, is due to disordered motion of particles. The conclusion of these studies is simple. Friction is the transformation of mechanical energy into *thermal energy*. Ref. 201

To heat 1 kg of water by 1 K by friction, 4.2 kJ of mechanical energy must be transformed through friction. The first to measure this quantity with precision was, in 1842, the German physician Julius Robert Mayer (1814–1878). He regarded his experiment as proof of the conservation of energy; indeed, he was the first person to state energy conservation! It is something of an embarrassment to modern physics that a medical doctor was the first to show the conservation of energy, and furthermore, that he was ridiculed by most physicists of his time. Worse, conservation of energy was accepted only when it was repeated many years later by two authorities: Hermann von Helmholtz – himself

TABLE 30 Some temperature values

Observation	Temperature
Lowest, but unattainable, temperature	0 K = −273.15°C
In the context of lasers, it sometimes makes sense to talk about negative temperature.	
Temperature a perfect vacuum would have at Earth's surface Page 893	40 zK
Sodium gas in certain laboratory experiments – coldest matter system achieved by man and possibly in the universe	0.45 nK
Temperature of neutrino background in the universe	*c.* 2 K
Temperature of photon gas background (or background radiation) in the universe	2.7 K
Liquid helium	4.2 K
Oxygen triple point	54.3584 K
Liquid nitrogen	77 K
Coldest weather ever measured (Antarctic)	185 K = −88°C
Freezing point of water at standard pressure	273.15 K = 0.00°C
Triple point of water	273.16 K = 0.01°C
Average temperature of the Earth's surface	287.2 K
Smallest uncomfortable skin temperature	316 K (10 K above normal)
Interior of human body	310.0 ± 0.5 K = 36.8 ± 0.5°C
Hottest weather measured	343.8 K = 70.7°C
Boiling point of water at standard pressure	373.13 K or 99.975°C
Liquid bronze	*c.* 1100 K
Liquid, pure iron	1810 K
Freezing point of gold	1337.33 K
Light bulb filament	2.9 kK
Earth's centre	4 kK
Sun's surface	5.8 kK
Air in lightning bolt	30 kK
Hottest star's surface (centre of NGC 2240)	250 kK
Space between Earth and Moon (no typo)	up to 1 MK
Sun's centre	20 MK
Inside the JET fusion tokamak	100 MK
Centre of hottest stars	1 GK
Maximum temperature of systems without electron–positron pair generation	ca. 6 GK
Universe when it was 1 s old	100 GK
Hagedorn temperature	1.9 TK
Heavy ion collisions – highest man-made value	up to 3.6 TK
Planck temperature – nature's upper temperature limit	10^{32} K

also a physician turned physicist – and William Thomson, who also cited similar, but later experiments by James Joule.* All of them acknowledged Mayer's priority. Publicity by William Thomson eventually led to the naming of the unit of energy after Joule.

In short, the sum of mechanical energy and thermal energy is constant. This is usually called the *first principle of thermodynamics*. Equivalently, it is impossible to produce mechanical energy without paying for it with some other form of energy. This is an important statement, because among others it means that humanity will stop living one day. Indeed, we live mostly on energy from the Sun; since the Sun is of finite size, its energy content will eventually be consumed. Can you estimate when this will happen? Challenge 490 n

There is also a second (and the mentioned third) principle of thermodynamics, which will be presented later on. The study of these topics is called *thermostatics* if the systems concerned are at equilibrium, and *thermodynamics* if they are not. In the latter case, we distinguish situations *near* equilibrium, when equilibrium concepts such as temperature can still be used, from situations *far* from equilibrium, such as self-organization, where such concepts often cannot be applied. Page 263

Does it make sense to distinguish between thermal energy and heat? It does. Many older texts use the term 'heat' to mean the same as thermal energy. However, this is confusing; in this text, 'heat' is used, in accordance with modern approaches, as the everyday term for entropy. Both thermal energy and heat flow from one body to another, and both accumulate. Both have no measurable mass.** Both the amount of thermal energy and the amount of heat inside a body increase with increasing temperature. The precise relation will be given shortly. But heat has many other interesting properties and stories to tell. Of these, two are particularly important: first, heat is due to particles; and secondly, heat is at the heart of the difference between past and future. These two stories are intertwined.

Entropy

> “– It's irreversible.
> – Like my raincoat!
> Mel Brooks, *Spaceballs*, 1987”

Ref. 202 Every domain of physics describes change in terms of two quantities: energy, and an extensive quantity characteristic of the domain. Even though heat is related to energy, the quantity physicists usually call heat is *not* an extensive quantity. Worse, what physicists call heat is not the same as what we call heat in our everyday speech. The extensive quantity corresponding to what we call 'heat' in everyday speech is called *entropy*.*** Entropy describes heat in the same way as momentum describes motion. When two objects differing in temperature are brought into contact, an entropy flow takes place between them,

* Hermann von Helmholtz (b. 1821 Potsdam, d. 1894 Berlin), important Prussian scientist. William Thomson (later William Kelvin) (1824–1907), important Irish physicist. James Prescott Joule (1818–1889), English physicist. Joule is pronounced so that it rhymes with 'cool', as his descendants like to stress. (The pronunciation of the name 'Joule' varies from family to family.)

** This might change in future, when mass measurements improve in precision, thus allowing the detection of relativistic effects. In this case, temperature increase may be detected through its related mass increase. However, such changes are noticeable only with twelve or more digits of precision in mass measurements. Page 322

*** The term 'entropy' was invented by the German physicist Rudolph Clausius (1822–1888) in 1865. He formed it from the Greek ἐν 'in' and τρόπος 'direction', to make it sound similar to 'energy'. It has always had the meaning given here.

TABLE 31 Some measured entropy values

Process/System	Entropy value
Melting of 1 kg of ice	1.21 kJ/K kg = 21.99 J/K mol
Water under standard conditions	70.1 J/K mol
Boiling of 1 kg of liquid water at 101.3 kPa	6.03 kJ/K= 110 J/K mol
Iron under standard conditions	27.2 J/K mol
Oxygen under standard conditions	161.1 J/K mol

like the flow of momentum that take place when two objects of different speeds collide. Let us define the concept of entropy more precisely and explore its properties in some more detail.

Entropy measures the degree to which energy is *mixed up* inside a system, that is, the degree to which energy is spread or shared among the components of a system. Therefore, entropy adds up when identical systems are composed into one. When two litre bottles of water at the same temperature are poured together, the entropy of the water adds up.

Like any other extensive quantity, entropy can be accumulated in a body; it can flow into or out of bodies. When water is transformed into steam, the entropy added into the water is indeed contained in the steam. In short, entropy is what is called 'heat' in everyday speech.

In contrast to several other important extensive quantities, entropy is not conserved. The sharing of energy in a system can be increased, for example by heating it. However, entropy is 'half conserved': in closed systems, entropy does not decrease; mixing cannot be undone. What is called equilibrium is simply the result of the highest possible mixing. In short, the entropy in a closed system increases until it reaches the maximum possible value.

When a piece of rock is detached from a mountain, it falls, tumbles into the valley, heating up a bit, and eventually stops. The opposite process, whereby a rock cools and tumbles upwards, is never observed. Why? The opposite motion does not contradict any Challenge 491 ny rule or pattern about motion that we have deduced so far.

Rocks never fall upwards because mountains, valleys and rocks are made of many particles. Motions of many-particle systems, especially in the domain of thermostatics, are called *processes*. Central to thermostatics is the distinction between *reversible* processes, such as the flight of a thrown stone, and *irreversible* processes, such as the aforementioned tumbling rock. Irreversible processes are all those processes in which friction and its generalizations play a role. They are those which increase the sharing or mixing of energy. They are important: if there were no friction, shirt buttons and shoelaces would Ref. 206 not stay fastened, we could not walk or run, coffee machines would not make coffee, and Page 826 maybe most importantly of all, we would have no memory.

Irreversible processes, in the sense in which the term is used in thermostatics, transform macroscopic motion into the disorganized motion of all the small microscopic components involved: they increase the sharing and mixing of energy. Irreversible processes are therefore not *strictly* irreversible – but their reversal is extremely improbable. We can say that entropy measures the 'amount of irreversibility': it measures the degree of mixing or decay that a collective motion has undergone.

Entropy is not conserved. Entropy – 'heat' – can appear out of nowhere, since energy sharing or mixing can happen by itself. For example, when two different liquids of the same temperature are mixed – such as water and sulphuric acid – the final temperature of the mix can differ. Similarly, when electrical current flows through material at room temperature, the system can heat up or cool down, depending on the material.

The *second principle of thermodynamics* states that 'entropy ain't what it used to be.' More precisely, *the entropy in a closed system tends towards its maximum*. Here, a *closed system* is a system that does not exchange energy or matter with its environment. Can you think of an example?

Challenge 492 ny

Entropy never decreases. Everyday life shows that in a closed system, the disorder increases with time, until it reaches some maximum. To reduce disorder, we need effort, i.e. work and energy. In other words, in order to reduce the disorder in a system, we need to connect the system to an energy source in some clever way. Refrigerators need electrical current precisely for this reason.

Because entropy never decreases, *white colour does not last*. Whenever disorder increases, the colour white becomes 'dirty', usually grey or brown. Perhaps for this reason white objects, such as white clothes, white houses and white underwear, are valued in our society. White objects defy decay.

Entropy allows to define the concept of *equilibrium* more precisely as the state of maximum entropy, or maximum energy sharing.

Flow of entropy

We know from daily experience that transport of an extensive quantity always involves friction. Friction implies generation of entropy. In particular, the flow of entropy itself produces additional entropy. For example, when a house is heated, entropy is produced in the wall. Heating means to keep a temperature difference ΔT between the interior and the exterior of the house. The heat flow J traversing a square metre of wall is given by

$$J = \kappa \Delta T = \kappa (T_\mathrm{i} - T_\mathrm{e}) \tag{90}$$

where κ is a constant characterizing the ability of the wall to conduct heat. While conducting heat, the wall also *produces* entropy. The entropy production σ is proportional to the difference between the interior and the exterior entropy flows. In other words, one has

$$\sigma = \frac{J}{T_\mathrm{e}} - \frac{J}{T_\mathrm{i}} = \kappa \frac{(T_\mathrm{i} - T_\mathrm{e})^2}{T_\mathrm{i} T_\mathrm{e}} . \tag{91}$$

Note that we have assumed in this calculation that everything is near equilibrium in each slice parallel to the wall, a reasonable assumption in everyday life. A typical case of a good wall has $\kappa = 1\,\mathrm{W/m^2K}$ in the temperature range between 273 K and 293 K. With this value, one gets an entropy production of

$$\sigma = 5 \cdot 10^{-3}\,\mathrm{W/m^2K} . \tag{92}$$

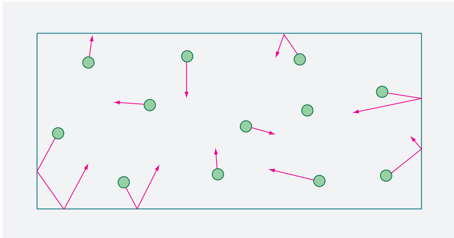


FIGURE 128 The basic idea of statistical mechanics about gases

Can you compare the amount of entropy that is produced in the flow with the amount that
Challenge 493 ny is transported? In comparison, a good goose-feather duvet has $\kappa = 1.5\,\mathrm{W/m^2K}$, which in shops is also called 15 tog.*

There are two other ways, apart from heat conduction, to transport entropy: *convection*, used for heating houses, and *radiation*, which is possible also through empty space. For example, the Earth radiates about $1.2\,\mathrm{W/m^2K}$ into space, in total thus about 0.51 PW/K. The entropy is (almost) the same that the Earth receives from the Sun. If more entropy had to be radiated away than received, the temperature of the surface of the Earth would have to increase. This is called the *greenhouse effect*. (It is also called *global warming*.) Let's hope that it remains small in the near future.

Do isolated systems exist?

In all our discussions so far, we have assumed that we can distinguish the system under investigation from its environment. But do such *isolated* or *closed* systems, i.e. systems not interacting with their environment, actually exist? Probably our own human condition was the original model for the concept: we do experience having the possibility to act independently of our environment. An isolated system may be simply defined as a system not exchanging any energy or matter with its environment. For many centuries, scientists saw no reason to question this definition.

The concept of an isolated system had to be refined somewhat with the advent of quantum mechanics. Nevertheless, the concept provides useful and precise descriptions of nature also in that domain. Only in the third part of our walk will the situation change drastically. There, the investigation of whether the universe is an isolated system will lead
Challenge 494 n to surprising results. (What do you think?)** We'll take the first steps towards the answer shortly.

Why do balloons take up space? – The end of continuity

* That unit is not as bad as the official (not a joke) BthU · h/sqft/cm/°F used in some remote provinces of our galaxy.

The insulation power of materials is usually measured by the constant $\lambda = \kappa d$ which is independent of the thickness d of the insulating layer. Values in nature range from about 2000 W/K m for diamond, which is the best conductor of all, down to between 0.1 W/K m and 0.2 W/K m for wood, between 0.015 W/K m and 0.05 W/K m for wools, cork and foams, and the small value of $5 \cdot 10^{-3}$ W/K m for krypton gas.

** A strange hint: your answer is almost surely wrong.


Daniel Bernoulli

Heat properties are material-dependent. Studying them should therefore enable us to understand something about the constituents of matter. Now, the simplest materials of all are gases.* Gases need space: an amount of gas has pressure and volume. Indeed, it did not take long to show that gases *could not* be continuous. One of the first scientists to think about gases as made up of atoms was Daniel Bernoulli.** Bernoulli reasoned that if atoms are small particles, with mass and momentum, he should be able to make quantitative predictions about the behaviour of gases, and check them with experiment. If the particles fly around in a gas, then the *pressure* of a gas in a container is produced by the steady flow of particles hitting the wall. It was then easy to conclude that if the particles are assumed to behave as tiny, hard and perfectly elastic balls, the pressure p, volume V and temperature T must be related by

Challenge 495 ny

$$pV = \frac{3k}{2} N T \qquad (93)$$

where N is the number of particles contained in the gas. (The Boltzmann constant k, one of the fundamental constants of nature, is defined below.) A gas made of particles with such textbook behaviour is called an *ideal gas*. Relation (93) has been confirmed by experiments at room and higher temperatures, for all known gases.

Bernoulli thus derived the gas relation, with a specific prediction for the proportionality constant, from the single assumption that gases are made of small massive constituents. This derivation provides a clear argument for the existence of atoms and for their behaviour as normal, though small objects. (Can you imagine how N might be determined experimentally?)

Challenge 496 ny

The ideal gas model helps us to answer questions such as the one illustrated in Figure 129. Two *identical* rubber balloons, one filled up to a larger size than the other, are connected via a pipe and a valve. The valve is opened. Which one deflates?

Challenge 497 n

The ideal gas relation states that hotter gases, at given pressure, need more volume. The relation thus explains why winds and storms exist, why hot air balloons rise, why car engines work, why the ozone layer is destroyed by certain gases, or why during the extremely hot summer of 2001 in the south of Turkey, oxygen masks were necessary to walk outside around noon.

Challenge 498 e

* By the way, the word *gas* is a modern construct. It was coined by the Brussels alchemist and physician Johan Baptista van Helmont (1579–1644), to sound similar to 'chaos'. It is one of the few words which have been invented by one person and then adopted all over the world.

** Daniel Bernoulli (b. 1700 Bâle, d. 1782 Bâle), important Swiss mathematician and physicist. His father Johann and his uncle Jakob were famous mathematicians, as were his brothers and some of his nephews. Daniel Bernoulli published many mathematical and physical results. In physics, he studied the separation of compound motion into translation and rotation. In 1738 he published the *Hydrodynamique*, in which he deduced all results from a single principle, namely the conservation of energy. The so-called *Bernoulli's principle* states that (and how) the pressure of a fluid decreases when its speed increases. He studied the tides and many complex mechanical problems, and explained the Boyle–Mariotte gas 'law'. For his publications he won the prestigious prize of the French Academy of Sciences – a forerunner of the Nobel Prize – ten times.

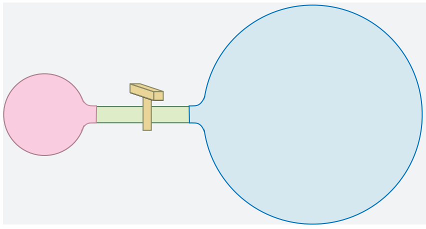


FIGURE 129 Which balloon wins?

Now you can take up the following challenge: how can you measure the weight of a car or a bicycle with a ruler only? Challenge 499 n

The picture of gases as being made of hard constituents without any long-distance interactions breaks down at very low temperatures. However, the ideal gas relation (93) can be improved to overcome these limitations by taking into account the deviations due to interactions between atoms or molecules. This approach is now standard practice and allows us to measure temperatures even at extremely low values. The effects observed below 80 K, such as the solidification of air, frictionless transport of electrical current, or frictionless flow of liquids, form a fascinating world of their own, the beautiful domain of low-temperature physics; it will be explored later on. Ref. 207 Ref. 208 Page 865, page 868

Brownian motion

It is easy to observe, under a microscope, that small particles (such as pollen) in a liquid never come to rest. They seem to follow a random zigzag movement. In 1827, the English botanist Robert Brown (1773–1858) showed with a series of experiments that this observation is independent of the type of particle and of the type of liquid. In other words, Brown had discovered a fundamental noise in nature. Around 1860, this motion was attributed to the molecules of the liquid colliding with the particles. In 1905 and 1906, Marian von Smoluchowski and, independently, Albert Einstein argued that this theory could be tested experimentally, even though at that time nobody was able to observe molecules directly. The test makes use of the specific properties of thermal noise. Ref. 209

It had already been clear for a long time that if molecules, i.e. indivisible matter particles, really existed, then heat had to be disordered motion of these constituents and temperature had to be the average energy per degree of freedom of the constituents. Bernoulli's model of Figure 128 implies that for monoatomic gases the kinetic energy T_{kin} per particle is given by Challenge 500 ny

$$T_{\text{kin}} = \frac{3}{2} kT \qquad (94)$$

where T is temperature. The so-called *Boltzmann constant* $k = 1.4 \cdot 10^{-23}$ J/K is the standard conversion factor between temperature and energy.* At a room temperature of 293 K, the kinetic energy is thus 6 zJ.

* The important Austrian physicist Ludwig Boltzmann (b. 1844 Vienna, d. 1906 Duino) is most famous for his work on thermodynamics, in which he explained all thermodynamic phenomena and observables, including entropy, as results of the behaviour of molecules. Planck named the Boltzmann constant after his investigations. He was one of the most important physicists of the late nineteenth century and stimulated

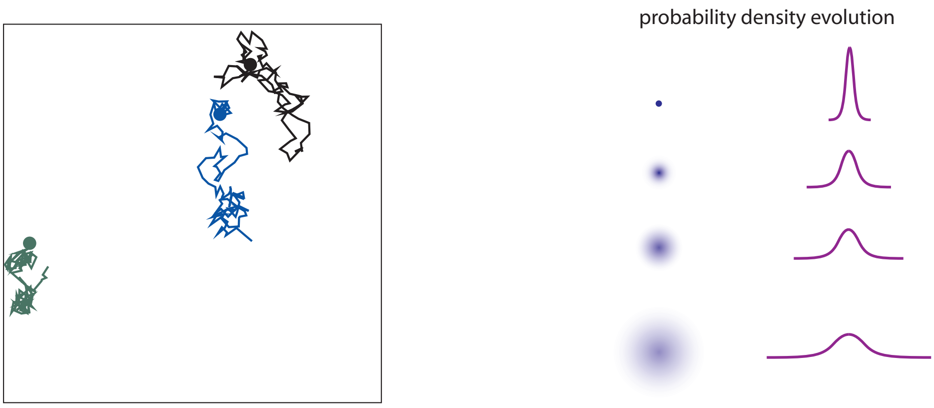


FIGURE 130 Example paths for particles in Brownian motion and their displacement distribution

Challenge 501 ny

Using relation (94) to calculate the speed of air molecules at room temperature yields values of several hundred metres per second. Why then does smoke from a candle take so long to diffuse through a room? Rudolph Clausius (1822–1888) answered this question in the mid-nineteenth century: diffusion is slowed by collisions with air molecules, in the same way as pollen particles collide with molecules in liquids.

At first sight, one could guess that the average distance the pollen particle has moved after n collisions should be zero, because the molecule velocities are random. However, this is wrong, as experiment shows.

Challenge 502 ny

An average *square* displacement, written $\langle d^2 \rangle$, is observed for the pollen particle. It cannot be predicted in which direction the particle will move, but it does move. If the distance the particle moves after one collision is l, the average square displacement after n collisions is given, as you should be able to show yourself, by

$$\langle d^2 \rangle = n l^2 . \tag{95}$$

For molecules with an average velocity v over time t this gives

$$\langle d^2 \rangle = n l^2 = v l t . \tag{96}$$

Ref. 210

In other words, the average square displacement increases proportionally with time. Of course, this is only valid if the liquid is made of separate molecules. Repeatedly measuring the position of a particle should give the distribution shown in Figure 130 for the probability that the particle is found at a given distance from the starting point. This is called the *(Gaussian) normal distribution*. In 1908, Jean Perrin* performed extensive experiments

many developments that led to quantum theory. It is said that Boltzmann committed suicide partly because of the resistance of the scientific establishment to his ideas. Nowadays, his work is standard textbook material.

* Jean Perrin (1870–1942), important French physicist, devoted most of his career to the experimental proof of the atomic hypothesis and the determination of Avogadro's number; in pursuit of this aim he perfected the use of emulsions, Brownian motion and oil films. His Nobel Prize speech (http://nobelprize.org/physics/laureates/1926/perrin-lecture.html) tells the interesting story of his research. He wrote the influential book

TABLE 32 Some typical entropy values per particle at *standard* temperature and pressure as multiples of the Boltzmann constant

Material	Entropy per particle
Monoatomic solids	0.3 k to 10 k
Diamond	0.29 k
Graphite	0.68 k
Lead	7.79 k
Monoatomic gases	15-25 k
Helium	15.2 k
Radon	21.2 k
Diatomic gases	15 k to 30 k
Polyatomic solids	10 k to 60 k
Polyatomic liquids	10 k to 80 k
Polyatomic gases	20 k to 60 k
Icosane	112 k

in order to test this prediction. He found that equation (96) corresponded completely with observations, thus convincing everybody that Brownian motion is indeed due to collisions with the molecules of the surrounding liquid, as Smoluchowski and Einstein had predicted.* Perrin received the 1926 Nobel Prize for these experiments.

Einstein also showed that the same experiment could be used to determine the number of molecules in a litre of water (or equivalently, the Boltzmann constant k). Can you work out how he did this?

Challenge 503 d

Entropy and particles

Once it had become clear that heat and temperature are due to the motion of microscopic particles, people asked what entropy *was* microscopically. The answer can be formulated in various ways. The two most extreme answers are:

— Entropy is the expected number of yes-or-no questions, multiplied by k ln 2, the answers of which would tell us everything about the system, i.e. about its microscopic state.
— Entropy measures the (logarithm of the) number W of possible microscopic states. A given macroscopic state can have many microscopic realizations. The logarithm of this number, multiplied by the Boltzmann constant k, gives the entropy.**

Les atomes and founded the Centre National de la Recherche Scientifique. He was also the first to speculate, in 1901, that an atom is similar to a small solar system.

Ref. 211 Page 267

* In a delightful piece of research, Pierre Gaspard and his team showed in 1998 that Brownian motion is also chaotic, in the strict physical sense given later on.

** When Max Planck went to Austria to search for the anonymous tomb of Boltzmann in order to get him buried in a proper grave, he inscribed the formula $S = k \ln W$ on the tombstone. (Which physicist would

In short, the higher the entropy, the more microstates are possible. Through either of these definitions, entropy measures the quantity of randomness in a system. In other words, it measures the transformability of energy: higher entropy means lower transformability. Alternatively, entropy measures the *freedom* in the choice of microstate that a system has. High entropy means high freedom of choice for the microstate. For example, when a molecule of glucose (a type of sugar) is produced by photosynthesis, about 40 bits of entropy are released. This means that after the glucose is formed, 40 additional yes-or-no questions must be answered in order to determine the full microscopic state of the system. Physicists often use a macroscopic unit; most systems of interest are large, and thus an entropy of 10^{23} bits is written as 1 J/K.*

To sum up, entropy is thus a specific measure for the characterization of disorder of
Ref. 212 thermal systems. Three points are worth making here. First of all, entropy is not *the* measure of disorder, but *one* measure of disorder. It is therefore *not* correct to use entropy as a *synonym* for the concept of disorder, as is often done in the popular literature. Entropy is only defined for systems that have a temperature, in other words, only for systems that are in or near equilibrium. (For systems far from equilibrium, no measure of disorder has been found yet; probably none is possible.) In fact, the use of the term entropy has degenerated so much that sometimes one has to call it *thermodynamic* entropy for clarity.

Secondly, entropy is related to information *only if* information is defined also as $-k \ln W$. To make this point clear, take a book with a mass of one kilogram. At room temperature, its entropy content is about 4 kJ/K. The printed information inside a book, say 500 pages of 40 lines with each containing 80 characters out of 64 possibilities, corresponds to an entropy of $4 \cdot 10^{-17}$ J/K. In short, what is usually called 'information' in everyday life is a negligible fraction of what a physicist calls information. Entropy is defined using the *physical* concept of information.

Finally, entropy is also *not* a measure for what in normal life is called the *complexity*
Ref. 213 of a situation. In fact, nobody has yet found a quantity describing this everyday notion.
Challenge 505 ny The task is surprisingly difficult. Have a try!

In summary, if you hear the term entropy used with a different meaning than $S = k \ln W$, beware. Somebody is trying to get you, probably with some ideology.

The minimum entropy of nature – the quantum of information

Before we complete our discussion of thermostatics we must point out in another way the importance of the Boltzmann constant k. We have seen that this constant appears whenever the granularity of matter plays a role; it expresses the fact that matter is made of small basic entities. The most striking way to put this statement is the following: *There is a smallest entropy in nature.* Indeed, for all systems, the entropy obeys

$$S \geqslant \frac{k}{2} . \qquad (97)$$

This result is almost 100 years old; it was stated most clearly (with a different numerical
Ref. 214 factor) by the Hungarian–German physicist Leo Szilard. The same point was made by the

finance the tomb of another, nowadays?)

Challenge 504 ny * This is only approximate. Can you find the precise value?

Ref. 215 French physicist Léon Brillouin (again with a different numerical factor). The statement can also be taken as the *definition* of the Boltzmann constant.

The existence of a smallest entropy in nature is a strong idea. It eliminates the possibility of the continuity of matter and also that of its fractality. A smallest entropy implies that matter is made of a finite number of small components. The limit to entropy expresses the fact that matter is made of particles.* The limit to entropy also shows that Galilean physics cannot be correct: Galilean physics assumes that arbitrarily small quantities do exist. The entropy limit is the first of several limits to motion that we will encounter until we finish the second part of our ascent. After we have found all limits, we can start the third and final part, leading to unification.

The existence of a smallest quantity implies a limit on the precision of measurement. Measurements cannot have infinite precision. This limitation is usually stated in the form of an indeterminacy relation. Indeed, the existence of a smallest entropy can be rephrased as an indeterminacy relation between the temperature T and the inner energy U of a system:

$$\Delta \frac{1}{T} \, \Delta U \geqslant \frac{k}{2} \, . \qquad (98)$$

Ref. 216 This relation** was given by Niels Bohr; it was discussed by Werner Heisenberg, who Page 1099 called it one of the basic indeterminacy relations of nature. The Boltzmann constant (di- Ref. 218 vided by 2) thus fixes the smallest possible entropy value in nature. For this reason, Gilles Ref. 215 Cohen-Tannoudji calls it the *quantum of information* and Herbert Zimmermann calls it the *quantum of entropy*.

The relation (98) points towards a more general pattern. For every minimum value for an observable, there is a corresponding indeterminacy relation. We will come across this several times in the rest of our adventure, most importantly in the case of the quantum Page 720 of action and Heisenberg's indeterminacy relation.

The existence of a smallest entropy has numerous consequences. First of all, it sheds light on the third principle of thermodynamics. A smallest entropy implies that absolute zero temperature is not achievable. Secondly, a smallest entropy explains why entropy values are finite instead of infinite. Thirdly, it fixes the absolute value of entropy for every system; in continuum physics, entropy, like energy, is only defined up to an additive constant. The entropy limit settles all these issues.

The existence of a minimum value for an observable implies that an indeterminacy relation appears for any two quantities whose product yields that observable. For example, entropy production rate and time are such a pair. Indeed, an indeterminacy relation connects the entropy production rate $P = \mathrm{d}S/\mathrm{d}t$ and the time t:

$$\Delta P \, \Delta t \geqslant \frac{k}{2} \, . \qquad (99)$$

From this and the previous relation (98) it is possible to deduce all of statistical physics,

* The minimum entropy implies that matter is made of tiny spheres; the minimum *action*, which we will encounter in quantum theory, implies that these spheres are actually small clouds.

Ref. 217 ** It seems that the historical value for the right hand side, given by k, has to be corrected to $k/2$.

Ref. 218, Ref. 217 i.e., the precise theory of thermostatics and thermodynamics. We will not explore this further here. (Can you show that the zeroth principle follows from the existence of a Challenge 506 ny smallest entropy?) We will limit ourselves to one of the cornerstones of thermodynamics: the second principle.

Why can't we remember the future?

> It's a poor sort of memory which only works backwards.
>
> Lewis Carroll, *Alice in Wonderland*

Page 44 When we first discussed time, we ignored the difference between past and future. But obviously, a difference exists, as we do not have the ability to remember the future. This is not a limitation of our brain alone. All the devices we have invented, such as tape recorders, photographic cameras, newspapers and books, only tell us about the past. Is there a way to build a video recorder with a 'future' button? Such a device would have to solve a Challenge 507 ny deep problem: how would it distinguish between the near and the far future? It does not take much thought to see that any way to do this would conflict with the second principle of thermodynamics. That is unfortunate, as we would need precisely the same device to Challenge 508 ny show that there is faster-than-light motion. Can you find the connection?

In summary, the future cannot be remembered because entropy in closed systems tends towards a maximum. Put even more simply, memory exists because the brain is made of many particles, and so the brain is limited to the past. However, for the most simple types of motion, when only a few particles are involved, the difference between past and future disappears. For few-particle systems, there is no difference between times gone by and times approaching. We could say that the future differs from the past only in our brain, or equivalently, only because of friction. Therefore the difference between the past and the future is not mentioned frequently in this walk, even though it is an essential part of our human experience. But the fun of the present adventure is precisely to overcome our limitations.

Is everything made of particles?

> A physicist is the atom's way of knowing about atoms.
>
> Ref. 219 George Wald

Historically, the study of statistical mechanics has been of fundamental importance for physics. It provided the first demonstration that physical objects are made of interacting particles. The story of this topic is in fact a long chain of arguments showing that all the properties we ascribe to objects, such as size, stiffness, colour, mass density, magnetism, thermal or electrical conductivity, result from the interaction of the many particles they consist of. The discovery that *all objects are made of interacting particles* has often been called the main result of modern science.

Page 240 How was this discovery made? Table 29 listed the main extensive quantities used in physics. Extensive quantities are able to flow. It turns out that all flows in nature are *composed* of elementary processes, as shown in Table 33. We have seen that the flow of mass, volume, charge, entropy and substance are composed. Later, quantum theory will show

TABLE 33 Some minimum flow values found in nature

Observation	Minimum value
Matter flow	one molecule, one atom or one particle
Volume flow	one molecule, one atom or one particle
Momentum flow	Planck's constant divided by wavelength
Angular momentum flow	Planck's constant
Chemical amount of substance	one molecule, one atom or one particle
Entropy flow	minimum entropy
Charge flow	elementary charge
Light flow	Planck's constant divided by wavelength

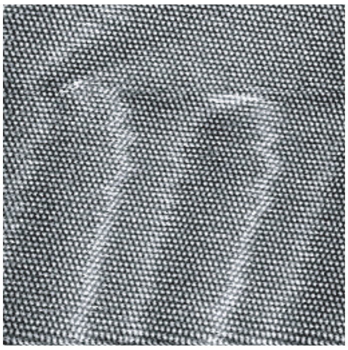


FIGURE 131 A 111 surface of a gold single crystal, every bright dot being an atom, with a surface dislocation (© CNRS)

the same for the flow of linear and angular momentum. *All flows are made of particles.*

This success of this idea has led many people to generalize it to the statement: 'Everything we observe is made of parts.' This approach has been applied with success to chemistry with molecules, materials science and geology with crystals, electricity with electrons, atoms with elementary particles, space with points, time with instants, light with photons, biology with cells, genetics with genes, neurology with neurons, mathematics with sets and relations, logic with elementary propositions, and even to linguistics with morphemes and phonemes. All these sciences have flourished on the idea that everything is made of related *parts*. The basic idea seems so self-evident that we find it difficult even to formulate an alternative. Just try!

Ref. 220

Challenge 509 ny

However, in the case of the *whole* of nature, the idea that nature is a sum of related parts is incorrect. It turns out to be a prejudice, and a prejudice so entrenched that it retarded further developments in physics in the latter decades of the twentieth century. In particular, it does *not* apply to elementary particles or to space-time. Finding the correct description for the whole of nature is the biggest challenge of our adventure, as it requires

Page 1067

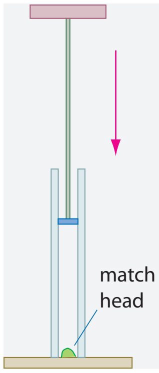


FIGURE 132 The fire pump

a complete change in thinking habits. There is a lot of fun ahead.

> “Jede Aussage über Komplexe läßt sich in eine Aussage über deren Bestandteile und in diejenigen Sätze zerlegen, welche die Komplexe vollständig beschreiben.*
> Ludwig Wittgenstein, *Tractatus*, 2.0201”

Why stones can be neither smooth nor fractal, nor made of little hard balls

The exploration of temperature yields another interesting result. Researchers first studied gases, and measured how much energy was needed to heat them by 1 K. The result is simple: all gases share only a few values, when the number of molecules N is taken into account. Monoatomic gases (in a container with constant volume) require $3Nk/2$, diatomic gases (and those with a linear molecule) $5Nk/2$, and almost all other gases $3Nk$, where $k = 1.4 \cdot 10^{-23}$ J/K is the Boltzmann constant. Page 249

The explanation of this result was soon forthcoming: each thermodynamic degree of freedom** contributes the energy $kT/2$ to the total energy, where T is the temperature. So the number of degrees of freedom in physical bodies is finite. Bodies are not continuous, nor are they fractals: if they were, their specific thermal energy would be infinite. Matter is indeed made of small basic entities.

All degrees of freedom contribute to the specific thermal energy. At least, this is what classical physics predicts. Solids, like stones, have 6 thermodynamic degrees of freedom and should show a specific thermal energy of $3Nk$. At high temperatures, this is indeed observed. But measurements of solids at room temperature yield lower values, and the lower the temperature, the lower the values become. Even gases show values lower than those just mentioned, when the temperature is sufficiently low. In other words, molecules

* Every statement about complexes can be resolved into a statement about their constituents and into the propositions that describe the complexes completely.

** A *thermodynamic degree of freedom* is, for each particle in a system, the number of dimensions in which it can move plus the number of dimensions in which it is kept in a potential. Atoms in a solid have six, particles in monoatomic gases have only three; particles in diatomic gases or rigid linear molecules have five. The number of degrees of freedom of larger molecules depends on their shape. Ref. 221

and atoms behave differently at low energies: atoms are not immutable little hard balls. The deviation of these values is one of the first hints of quantum theory.

Curiosities and fun challenges about heat

Compression of air increases its temperature. This is shown directly by the fire pump, a variation of a bicycle pump, shown in Figure 132. (For a working example, see the website http://www.tn.tudelft.nl/cdd). A match head at the bottom of an air pump made of transparent material is easily ignited by the compression of the air above it. The temperature of the air after compression is so high that the match head ignites spontaneously.

* *

If heat really is disordered motion of atoms, a big problem appears. When two atoms collide head-on, in the instant of smallest distance, neither atom has velocity. Where does the kinetic energy go? Obviously, it is transformed into potential energy. But that implies that atoms can be deformed, that they have internal structure, that they have parts, and thus that they can in principle be split. In short, if heat is disordered atomic motion, *atoms are not indivisible*! In the nineteenth century this argument was put forward in order to show that heat cannot be atomic motion, but must be some sort of fluid. But since we know that heat really is kinetic energy, atoms must indeed be divisible, even though their name means 'indivisible'. We do not need an expensive experiment to show this.

* *

Ref. 222 In 1912, Emile Borel noted that if a gram of matter on Sirius was displaced by one centimetre, it would change the gravitational field on Earth by a tiny amount. This tiny change would be sufficient to make it impossible to calculate the poth of molecules in a gas after a fraction of a second.

* *

Not only gases, but also most other materials expand when the temperature rises. As a result, the electrical wires supported by pylons hang much lower in summer than in winter. True? Challenge 510 n

* *

Ref. 223 Challenge 511 ny The following is a famous Fermi problem. Given that a human corpse cools down in four hours after death, what is the minimum number of calories needed per day in our food?

* *

The energy contained in thermal motion is not negligible. A 1 g bullet travelling at the speed of sound has a kinetic energy of only 0.01 kcal.

* *

Challenge 512 n How does a typical, 1500 m^3 hot-air balloon work?

* *

If you do not like this text, here is a proposal. You can use the paper to make a cup, as

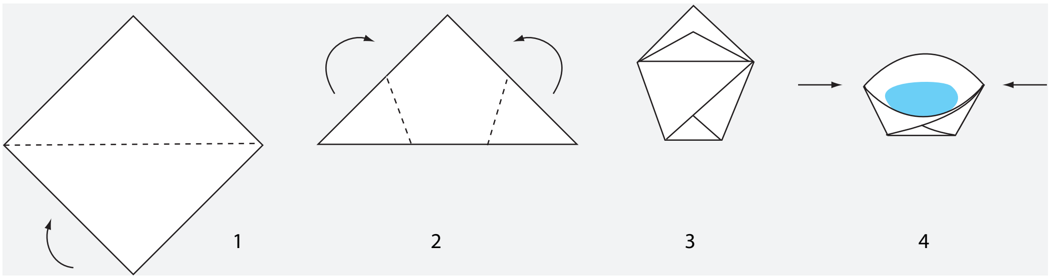


FIGURE 133 Can you boil water in this paper cup?

shown in Figure 133, and boil water in it over an open flame. However, to succeed, you have to be a little careful. Can you find out in what way?

Challenge 513 n

* *

Mixing 1 kg of water at 0 °C and 1 kg of water at 100 °C gives 2 kg of water at 50 °C. What is the result of mixing 1 kg of *ice* at 0 °C and 1 kg of water at 100 °C?

Challenge 514 ny

* *

Ref. 224

The highest recorded air temperature in which a man has survived is 127 °C. This was tested in 1775 in London, by the secretary of the Royal Society, Charles Blagden, together with a few friends, who remained in a room at that temperature for 45 minutes. Interestingly, the raw steak which he had taken in with him was cooked ('well done') when he and his friends left the room. What condition had to be strictly met in order to avoid cooking the people in the same way as the steak?

Challenge 515 n

* *

Why does water boil at 99.975 °C instead of 100 °C?

Challenge 516 n

* *

Can you fill a bottle precisely with 1 ± 10^{-30} kg of water?

Challenge 517 n

* *

One gram of fat, either butter or human fat, contains 38 kJ of chemical energy (or, in ancient units more familiar to nutritionists, 9 kcal). That is the same value as that of petrol. Why are people and butter less dangerous than petrol?

Challenge 518 n

* *

In 1992, the Dutch physicist Martin van der Mark invented a loudspeaker which works by heating air with a laser beam. He demonstrated that with the right wavelength and with a suitable modulation of the intensity, a laser beam in air can generate sound. The effect at the basis of this device, called the *photoacoustic effect*, appears in many materials. The best laser wavelength for air is in the infrared domain, on one of the few absorption lines of water vapour. In other words, a properly modulated infrared laser beam that shines through the air generates sound. Such light can be emitted from a small matchbox-sized semiconductor laser hidden in the ceiling and shining downwards. The sound is emitted

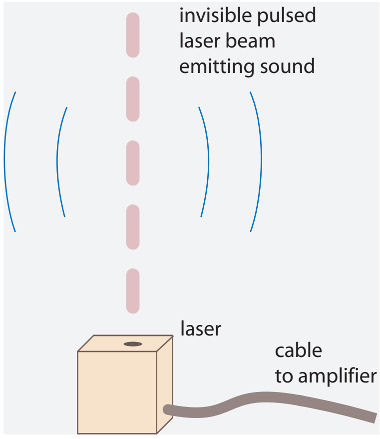


FIGURE 134 The invisible loudspeaker

in all directions perpendicular to the beam. Since infrared laser light is not visible, Martin van der Mark thus invented an invisible loudspeaker! Unfortunately, the efficiency of present versions is still low, so that the power of the speaker is not yet sufficient for practical applications. Progress in laser technology should change this, so that in the future we should be able to hear sound that is emitted from the centre of an otherwise empty room.

* *

A famous exam question: How can you measure the height of a building with a barometer, a rope and a ruler? Find at least six different ways.
Challenge 519 n

* *

What is the approximate probability that out of one million throws of a coin you get exactly 500 000 heads and as many tails?
Challenge 520 ny

You may want to use Stirling's formula $n! \approx \sqrt{2\pi n}\,(n/e)^n$ to calculate the result.*

* *

Challenge 521 n
Does it make sense to talk about the entropy of the universe?

* *

Challenge 522 ny
Can a helium balloon lift the tank which filled it?

* *

All friction processes, such as osmosis, diffusion, evaporation, or decay, are *slow*. They take a characteristic time. It turns out that any (macroscopic) process with a time-scale is irreversible. This is no real surprise: we know intuitively that undoing things always takes more time than doing them. That is again the second principle of thermodynamics.

* There are many improvements to Stirling's formula. A simple one is $n! \approx \sqrt{(2n+1/3)\pi}\,(n/e)^n$. Another is $\sqrt{2\pi n}\,(n/e)^n e^{1/(12n+1)} < n! < \sqrt{2\pi n}\,(n/e)^n e^{1/(12n)}$.

* *

Ref. 225 It turns out that *storing* information is possible with negligible entropy generation. However, *erasing* information requires entropy. This is the main reason why computers, as well as brains, require energy sources and cooling systems, even if their mechanisms would otherwise need no energy at all.

* *

Challenge 523 ny When mixing hot rum and cold water, how does the increase in entropy due to the mixing compare with the entropy increase due to the temperature difference?

* *

Challenge 524 n Why aren't there any small humans, e.g. 10 mm in size, as in many fairy tales? In fact, there are no warm-blooded animals of that size. Why not?

* *

Shining a light onto a body and repeatedly switching it on and off produces sound. This is called the *photoacoustic effect*, and is due to the thermal expansion of the material. By changing the frequency of the light, and measuring the intensity of the noise, one reveals a characteristic photoacoustic spectrum for the material. This method allows us to detect gas concentrations in air of one part in 10^9. It is used, among other methods, to study the gases emitted by plants. Plants emit methane, alcohol and acetaldehyde in small quantities; the photoacoustic effect can detect these gases and help us to understand the processes behind their emission.

* *

Challenge 525 ny What is the rough probability that all oxygen molecules in the air would move away from a given city for a few minutes, killing all inhabitants?

* *

Challenge 526 ny If you pour a litre of water into the sea, stir thoroughly through all the oceans and then take out a litre of the mixture, how many of the original atoms will you find?

* *

Challenge 527 ny How long would you go on breathing in the room you are in if it were airtight?

* *

Challenge 528 ny What happens if you put some ash onto a piece of sugar and set fire to the whole? (Warning: this is dangerous and not for kids.)

* *

Challenge 529 ny Entropy calculations are often surprising. For a system of N particles with two states each, there are $W_{\text{all}} = 2^N$ states. For its most probable configuration, with exactly half the particles in one state, and the other half in the other state, we have $W_{\text{max}} = N!/((N/2)!)^2$. Now, for a macroscopic system of particles, we might typically have $N = 10^{24}$. That gives $W_{\text{all}} \gg W_{\text{max}}$; indeed, the former is 10^{12} times larger than the latter. On the other hand, we find that $\ln W_{\text{all}}$ and $\ln W_{\text{max}}$ agree for the first 20 digits! Even though the configuration

with exactly half the particles in each state is much more rare than the general case, where
Challenge 530 ny the ratio is allowed to vary, the entropy turns out to be the same. Why?

* *

If heat is due to motion of atoms, our built-in senses of heat and cold are simply detectors
Challenge 531 ny of motion. How could they work?

By the way, the senses of smell and taste can also be seen as motion detectors, as they
Challenge 532 ny signal the presence of molecules flying around in air or in liquids. Do you agree?

* *

The Moon has an atmosphere, although an extremely thin one, consisting of sodium (Na) and potassium (K). This atmosphere has been detected up to nine Moon radii from its surface. The atmosphere of the Moon is generated at the surface by the ultraviolet radiation
Challenge 533 n from the Sun. Can you estimate the Moon's atmospheric density?

* *

Does it make sense to add a line in Table 29 for the quantity of physical action? A column?
Challenge 534 ny Why?

* *

Diffusion provides a length scale. For example, insects take in oxygen through their skin. As a result, the interiors of their bodies cannot be much more distant from the surface than about a centimetre. Can you list some other length scales in nature implied by diffu-
Challenge 535 n sion processes?

* *

Rising warm air is the reason why many insects are found in tall clouds in the evening. Many insects, especially that seek out blood in animals, are attracted to warm and humid air.

* *

Thermometers based on mercury can reach 750°C. How is this possible, given that mer-
Challenge 536 n cury boils at 357°C?

* *

Challenge 537 n What does a burning candle look like in weightless conditions?

* *

It is possible to build a power station by building a large chimney, so that air heated by the Sun flows upwards in it, driving a turbine as it does so. It is also possible to make a power station by building a long vertical tube, and letting a gas such as ammonia rise into it which is then liquefied at the top by the low temperatures in the upper atmosphere; as it falls back down a second tube as a liquid – just like rain – it drives a turbine. Why are
Challenge 538 n such schemes, which are almost completely non-polluting, not used yet?

* *

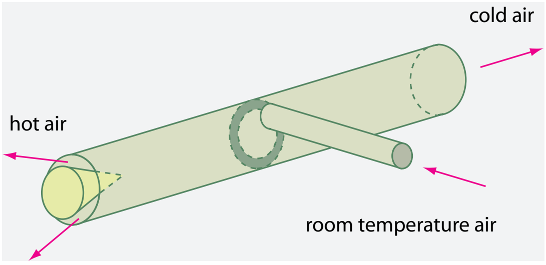


FIGURE 135 The Wirbelrohr or Ranque–Hilsch vortex tube

One of the most surprising devices ever invented is the *Wirbelrohr* or Ranque–Hilsch vortex tube. By blowing compressed air at room temperature into it at its midpoint, two flows of air are formed at its ends. One is extremely cold, easily as low as −50°C, and one extremely hot, up to 200°C. No moving parts and no heating devices are found inside.
Challenge 539 n How does it work?

* *

It is easy to cook an egg in such a way that the white is hard but the yolk remains liquid.
Challenge 540 n Can you achieve the opposite?

* *

Thermoacoustic engines, pumps and refrigerators provide many strange and fascinating applications of heat. For example, it is possible to use loud sound in closed metal chambers to move heat from a cold place to a hot one. Such devices have few moving parts and
Ref. 226 are being studied in the hope of finding practical applications in the future.

* *

Challenge 541 ny Does a closed few-particle system contradict the second principle of thermodynamics?

* *

What happens to entropy when gravitation is taken into account? We carefully left gravitation out of our discussion. In fact, many problems appear – just try to think about the issue. For example, Jakob Bekenstein has discovered that matter reaches its highest
Challenge 542 ny possible entropy when it forms a black hole. Can you confirm this?

* *

The numerical values (but not the units!) of the Boltzmann constant $k = 1.38 \cdot 10^{-23}$ J/K and the combination h/ce agree in their exponent and in their first three digits, where h is
Challenge 543 ny Planck's constant and e is the electron charge. Can you dismiss this as mere coincidence?

Figures to be added

FIGURE 136 Examples of self-organization for sand

TABLE 34 Sand patterns in the sea and on land

Pattern	Period	Amplitude	Origin
sand banks	2 to 10 km	2 to 20 m	tides
sand waves	100 to 800 m	5 m	tides
megaribbles	1 m	0.1 m	tides
ribbles	5 cm	5 mm	waves
singing sand	95 to 105 Hz	up to 105 dB	wind on sand dunes, avalanches making the dune vibrate

Self-organization and chaos

> To speak of non-linear physics is like calling zoology the study of non-elephant animals.
>
> Stanislaw Ulam

Ref. 227

In our list of global descriptions of motion, the high point is the study of self-organization. Self-organization is the appearance of order. *Order* is a term that includes *shapes*, such as the complex symmetry of snowflakes; *patterns*, such as the stripes of zebras; and *cycles*, such as the creation of sound when singing. Every example of what we call *beauty* is a combination of shapes, patterns and cycles. (Do you agree?) Self-organization can thus be called the study of the origin of beauty.

Ref. 228

Challenge 544 n

The appearance of order is a general observation across nature. Fluids in particular exhibit many phenomena where order appears and disappears. Examples include the more or less regular flickering of a burning candle, the flapping of a flag in the wind, the regular stream of bubbles emerging from small irregularities in the surface of a champagne glass, and the regular or irregular dripping of a water tap.

The appearance of order is found from the cell differentiation in an embryo inside a woman's body; the formation of colour patterns on tigers, tropical fish and butterflies; the symmetrical arrangements of flower petals; the formation of biological rhythms; and so on.

All growth processes are self-organization phenomena. Have you ever pondered the

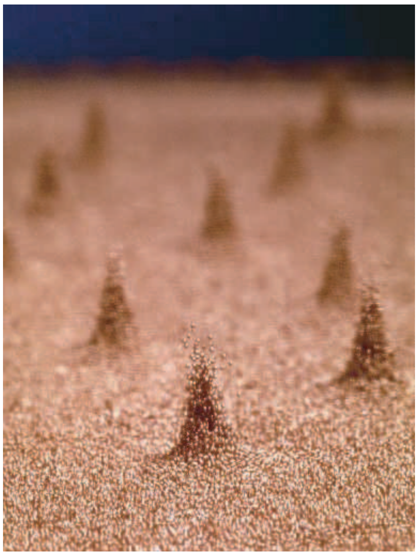


FIGURE 137 Oscillons formed by shaken bronze balls; horizontal size is about 2 cm (© Paul Umbanhowar)

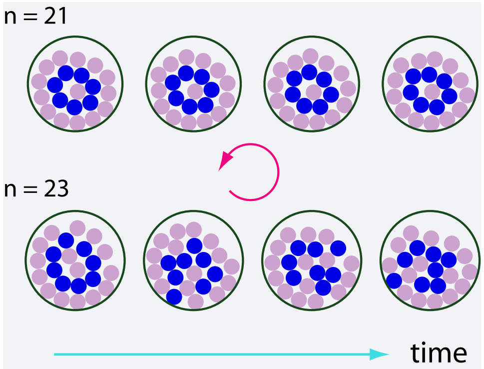


FIGURE 138 Magic numbers: 21 spheres, when swirled in a dish, behave differently from non-magic numbers, like 23, of spheres (redrawn from photographs, © Karsten Kötter)

incredible way in which teeth grow? A practically inorganic material forms shapes in the upper and the lower rows fitting exactly into each other. How this process is controlled is still a topic of research. Also the formation, before and after birth, of neural networks in the brain is another process of self-organization. Even the physical processes at the basis of thinking, involving changing electrical signals, is to be described in terms of self-organization.

Biological evolution is a special case of growth. Take the evolution of animal shapes. It turns out that snake tongues are forked because that is the most efficient shape for following chemical trails left by prey and other snakes of the same species. (Snakes smell with help of the tongue.) The fixed numbers of fingers in human hands or of petals of flowers are also consequences of self-organization.

Ref. 229
Page 715

Many problems of self-organization are mechanical problems: for example, the formation of mountain ranges when continents move, the creation of earthquakes, or the creation of regular cloud arrangements in the sky. It can be fascinating to ponder, during an otherwise boring flight, the mechanisms behind the formation of the clouds you see from the aeroplane.

Challenge 545 e

Studies into the conditions required for the appearance or disappearance of order have shown that their description requires only a few common concepts, independently of the details of the physical system. This is best seen looking at a few examples.

All the richness of self-organization reveals itself in the study of plain sand. Why do sand dunes have ripples, as does the sand floor at the bottom of the sea? We can also study how avalanches occur on steep heaps of sand and how sand behaves in hourglasses, in mixers, or in vibrating containers. The results are often surprising. For example, as recently as 1996 Paul Umbanhowar and his colleagues found that when a flat container

Ref. 230

FIGURE 139 Self-organization: a growing snow flake (quicktime © Kenneth Libbrecht)

holding tiny bronze balls (around 0.165 mm in diameter) is shaken up and down in vacuum at certain frequencies, the surface of this bronze 'sand' forms stable heaps. They are shown in Figure 137. These heaps, so-called *oscillons*, also bob up and down. Oscillons can move and interact with one another.

Oscillons in sand are simple example for a general effect in nature: *discrete* systems with nonlinear interactions can exhibit localized excitations. This fascinating topic is just Ref. 231 beginning to be researched. It might well be that it will yield results relevant to our understanding of elementary particles.

Sand shows many other pattern-forming processes. A mixture of sand and sugar, when poured onto a heap, forms regular layered structures that in cross section look like zebra stripes. Horizontally rotating cylinders with binary mixtures inside them separate the mixture out over time. Or take a container with two compartments separated by a 1 cm wall. Fill both halves with sand and rapidly shake the whole container with a machine. Over time, all the sand will spontaneously accumulate in one half of the container. As another example of self-organization in sand, people have studied the various types of Ref. 232 sand dunes that 'sing' when the wind blows over them. In fact, the behaviour of sand and dust is proving to be such a beautiful and fascinating topic that the prospect of each human returning dust does not look so grim after all.

Another simple and beautiful example of self-organization is the effect discovered in Ref. 233 1999 by Karsten Kötter and his group. They found that the behaviour of a set of spheres swirled in a dish depends on the number of spheres used. Usually, all the spheres get continuously mixed up. But for certain 'magic' numbers, such as 21, stable ring patterns emerge, for which the outside spheres remain outside and the inside ones remain inside. The rings, best seen by colouring the spheres, are shown in Figure 138.

These and many other studies of self-organizing systems have changed our understanding of nature in a number of ways. First of all, they have shown that patterns and shapes are similar to cycles: all are due to motion. Without motion, and thus without history, Ref. 234 there is no order, neither patterns nor shapes. Every pattern has a history; every pattern is a result of motion. An example is shown in Figure 139.

Secondly, patterns, shapes and cycles are due to the organized motion of large numbers of small constituents. Systems which self-organize are always composite: they are *cooperative structures*.

Thirdly, all these systems obey evolution equations which are *nonlinear* in the configuration variables. Linear systems do not self-organize. Many self-organizing systems also show *chaotic* motion.

Fourthly, the appearance and disappearance of order depends on the strength of a driving force, the so-called *order parameter*. Often, chaotic motion appears when the driving is increased beyond the value necessary for the appearance of order. An example of chaotic motion is turbulence, which appears when the order parameter, which is proportional to the speed of the fluid, is increased to high values.

Moreover, all order and all structure appears when two general types of motion compete with each other, namely a 'driving', energy-adding process, and a 'dissipating', braking mechanism. Thermodynamics plays a role in all self-organization. Self-organizing systems are always *dissipative systems*, and are always far from equilibrium. When the driving and the dissipation are of the same order of magnitude, and when the key behaviour of the system is not a linear function of the driving action, order may appear.*

All self-organizing systems at the onset of order appearance can be described by equations for the pattern amplitude A of the general form

$$\frac{\partial A(t,x)}{\partial t} = \lambda A - \mu |A|^2 A + \kappa\, \Delta A + \text{higher orders}\ . \qquad (100)$$

Here, the – possibly complex – observable A is the one that appears when order appears, such as the oscillation amplitude or the pattern amplitude. The first term λA is the driving term, in which λ is a parameter describing the strength of the driving. The next term is a typical nonlinearity in A, with μ a parameter that describes its strength, and the third term $\kappa\, \Delta A = \kappa(\partial^2 A/\partial x^2 + \partial^2 A/\partial y^2 + \partial^2 A/\partial z^2)$ is a typical dissipative (and diffusive) term.

One can distinguish two main situations. In cases where the dissipative term plays no role ($\kappa = 0$), one finds that when the driving parameter λ increases above zero, a *temporal* oscillation appears, i.e. a stable cycle with non-vanishing amplitude. In cases where the diffusive term does play a role, equation (100) describes how an amplitude for a *spatial* oscillation appears when the driving parameter λ becomes positive, as the solution $A = 0$ then becomes spatially unstable.

Challenge 546 ny

Challenge 547 ny

In both cases, the onset of order is called a *bifurcation*, because at this critical value of the driving parameter λ the situation with amplitude zero, i.e. the homogeneous (or unordered) state, becomes unstable, and the ordered state becomes stable. *In nonlinear systems, order is stable.* This is the main conceptual result of the field. Equation (100)

* To describe the 'mystery' of human life, terms like 'fire', 'river' or 'tree' are often used as analogies. These are all examples of self-organized systems: they have many degrees of freedom, have competing driving and braking forces, depend critically on their initial conditions, show chaos and irregular behaviour, and sometimes show cycles and regular behaviour. Humans and human life resemble them in all these respects; thus there is a solid basis to their use as metaphors. We could even go further and speculate that pure beauty *is* pure self-organization. The lack of beauty indeed often results from a disturbed equilibrium between external braking and external driving.

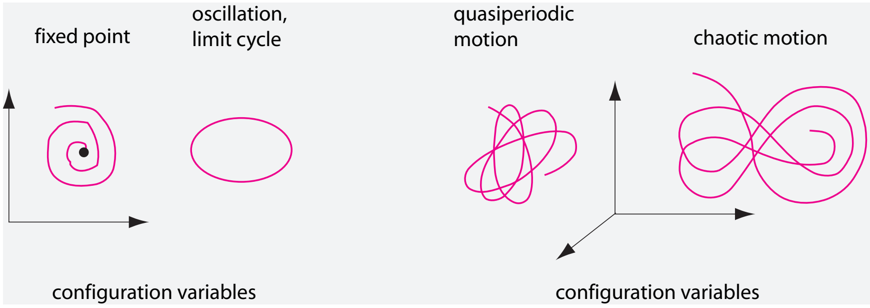


FIGURE 140 Examples of different types of motion in configuration space

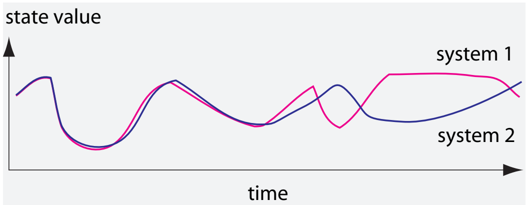


FIGURE 141 Sensitivity to initial conditions

and its numerous variations allow us to describe many phenomena, ranging from spirals, waves, hexagonal patterns, and topological defects, to some forms of turbulence. For every physical system under study, the main task is to distil the observable A and the parameters λ, μ and κ from the underlying physical processes. Ref. 235

Self-organization is a vast field which is yielding new results almost by the week. To discover new topics of study, it is often sufficient to keep one's eye open; most effects are comprehensible without advanced mathematics. Good hunting! Challenge 548 ny

Most systems that show self-organization also show another type of motion. When the driving parameter of a self-organizing system is increased to higher and higher values, order becomes more and more irregular, and in the end one usually finds chaos. For physicists, chaotic motion is the most irregular type of motion.* Chaos can be defined independently of self-organization, namely as that motion of systems for which small changes in initial conditions evolve into large changes of the motion (exponentially with time), as shown in Figure 141. More precisely, *chaos* is irregular motion characterized by a positive *Lyapounov exponent*. The weather is such a system, as are dripping water-taps, the fall of dice, and many other common systems. For example, research on the mechanisms by which the heart beat is generated has shown that the heart is not an oscillator, but a chaotic system with irregular cycles. This allows the heart to be continuously ready for demands for changes in beat rate which arise once the body needs to increase or decrease its efforts. Ref. 236

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

Incidentally, can you give a simple argument to show that the so-called *butterfly effect* does not exist? This 'effect' is often cited in newspapers: the claim is that nonlinearities imply that a small change in initial conditions can lead to large effects; thus a butterfly wing beat is alleged to be able to induce a tornado. Even though nonlinearities do indeed lead to growth of disturbances, the butterfly effect has never been observed; it does not exist.

Challenge 549 n

There is chaotic motion also in machines: chaos appears in the motion of trains on the rails, in gear mechanisms, and in fire-fighter's hoses. The precise study of the motion in a zippo cigarette lighter will probably also yield an example of chaos. The mathematical description of chaos – simple for some textbook examples, but extremely involved for others – remains an important topic of research.

Challenge 550 ny

All the steps from disorder to order, quasiperiodicity and finally to chaos, are examples of self-organization. These types of motion, illustrated in Figure 140, are observed in many fluid systems. Their study should lead, one day, to a deeper understanding of the mysteries of turbulence. Despite the fascination of this topic, we will not explore it further, because it does not lead towards the top of Motion Mountain.

Ref. 237

But self-organization is of interest also for a more general reason. It is sometimes said that our ability to formulate the patterns or rules of nature from observation does not imply the ability to predict *all* observations from these rules. According to this view, so-called 'emergent' properties exist, i.e. properties appearing in complex systems as something *new* that cannot be deduced from the properties of their parts and their interactions. (The ideological backdrop to this view is obvious; it is the last attempt to fight the determinism.) The study of self-organization has definitely settled this debate. The properties of water molecules do allow us to predict Niagara Falls.* Similarly, the diffusion of signal molecules do determine the development of a single cell into a full human being: in particular, cooperative phenomena determine the places where arms and legs are formed; they ensure the (approximate) right–left symmetry of human bodies, prevent mix-ups of connections when the cells in the retina are wired to the brain, and explain the fur patterns on zebras and leopards, to cite only a few examples. Similarly, the mechanisms at the origin of the heart beat and many other cycles have been deciphered.

Self-organization provides general principles which allow us in principle to predict the behaviour of complex systems of any kind. They are presently being applied to the most complex system in the known universe: the human brain. The details of how it learns to coordinate the motion of the body, and how it extracts information from the images in the eye, are being studied intensely. The ongoing work in this domain is fascinating. If you plan to become a scientist, consider taking this path.

Challenge 552 ny

Such studies provide the final arguments that confirm what J. Offrey de la Mettrie in 1748 stated and explored in his famous book *L'homme machine*: humans are complex machines. Indeed, the lack of understanding of complex systems in the past was due mainly to the restrictive teaching of the subject of motion, which usually concentrated – as we do in this walk – on examples of motion in *simple* systems. Even though the subject

* Already small versions of Niagara Falls, namely dripping water taps, show a large range of cooperative phenomena, including the chaotic, i.e. non-periodic, fall of water drops. This happens when the water flow has the correct value, as you can verify in your own kitchen. Several cooperative fluid phenomena have been simulated even on the molecular level.

Ref. 238
Challenge 551 ny

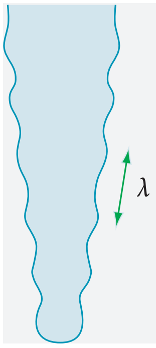


FIGURE 142 The wavy surface of icicles

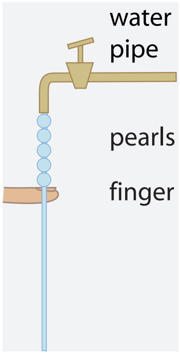


FIGURE 143 Water pearls

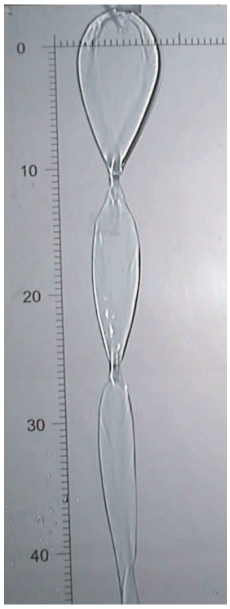


FIGURE 144 A braiding water stream (© Vakhtang Putkaradze)

of self-organization provides fascinating insights, and will do so for many years to come, we now leave it. We continue with our own adventure exploring the basics of motion.*

> Ich sage euch: man muss noch Chaos in sich haben, um einen tanzenden Stern gebären zu können. Ich sage euch: ihr habt noch Chaos in euch.
>
> Friedrich Nietzsche, *Also sprach Zarathustra.*

Curiosities and fun challenges about self-organization

All icicles have a wavy surface, with a crest-to-crest distance of about 1 cm, as shown in Figure 142. The distance is determined by the interplay between water flow and surface cooling. How?

Challenge 553 ny

* *

When wine is made to swirl in a wine glass, after the motion has calmed down, the wine flowing down the glass walls forms little arcs. Can you explain in a few words what forms them?

Challenge 554 ny

* *

How does the average distance between cars parked along a street change over time, assuming a constant rate of cars leaving and arriving?

Challenge 555 d

Ref. 239

* An important case of self-organization is *humour*.

* *

Challenge 556 d When a fine stream of water leaves a water tap, putting a finger in the stream leads to a wavy shape, as shown in Figure 143. Why?

* *

When water emerges from a oblong opening, the stream forms a braid pattern, as shown in Figure 144. This effect results from the interplay and competition between inertia and surface tension: inertia tends to widen the stream, while surface tension tends to narrow it. Predicting the distance from one narrow region to the next is still a topic of research. (Ref. 240)

If the experiment is done in free air, without a plate, one usually observes an additional effect: there is a *chiral* braiding at the narrow regions, induced by the asymmetries of the water flow. You can observe this effect in the toilet! Scientific curiosity knows no limits: are you a right-turner or a left-turner, or both? On every day? (Challenge 557 ny)

* *

Gerhard Müller has discovered a simple but beautiful way to observe self-organization in solids. His system also provides a model for a famous geological process, the formation of hexagonal columns in basalt, such as the Devil's Staircase in Ireland. Similar formations are found in many other places of the Earth. Just take some rice flour or corn starch, mix it with about half the same amount of water, put the mixture into a pan and dry it with a lamp. Hexagonal columns form. The analogy works because the drying of starch and the cooling of lava are diffusive processes governed by the same equations, because the boundary conditions are the same, and because both materials respond with a small reduction in volume. (Ref. 241) (Challenge 558 e)

* *

Water flow in pipes can be laminar (smooth) or turbulent (irregular and disordered). The transition depends on the diameter d of the pipe and the speed v of the water. The transition usually happens when the so-called *Reynolds number* – defined as $R = vd/\eta$ (η being the kinematic viscosity of the water, around $1\,\mathrm{mm^2/s}$) – becomes greater than about 2000. However, careful experiments show that with proper handling, laminar flows can be produced up to $R = 100\,000$. A linear analysis of the equations of motion of the fluid, the Navier–Stokes equations, even predicts stability of laminar flow for *all* Reynolds numbers. This riddle was solved only in the years 2003 and 2004. First, a complex mathematical analysis showed that the laminar flow is not always stable, and that the transition to turbulence in a long pipe occurs with travelling waves. Then, in 2004, careful experiments showed that these travelling waves indeed appear when water is flowing through a pipe at large Reynolds numbers. (Ref. 242)

* *

For some beautiful pictures on self-organization in fluids, see the http://serve.me.nus.edu.sg/limtt website. Among others, it shows that a circular vortex can 'suck in' a second one behind it, and that the process can then repeat.

* *

Also *dance* is an example of self-organization. Self-organization takes part in the brain.

Like for all complex movements, learning then is often a challenge. Nowadays there are beautiful books that tell how physics can help you improve your dancing skills and the grace of your movements. Ref. 243

* *

Do you want to enjoy working on your PhD? Go into a scientific toy shop, and look for a toy that moves in a complex way. There are high chances that the motion is chaotic; explore the motion and present a thesis about it. (Or simply explore the motion of a hanging wire whose upper end is externally driven.)

4. FROM THE LIMITATIONS OF PHYSICS TO THE LIMITS OF MOTION

> I only know that I know nothing.
>
> Socrates, as cited by Plato

Socrates' saying applies also to Galilean physics, despite its general success in engineering and in the description of everyday life. We will now give a short overview of the limitations of the field.

Research topics in classical dynamics

Even though the science of mechanics is now several hundred years old, research into its details is still continuing.

— We have already mentioned above the issue of the stability of the solar system. The long-term future of the planets is unknown. In general, the behaviour of few-body systems interacting through gravitation is still a research topic of mathematical physics. Ref. 244 Answering the simple question of how long a given set of bodies gravitating around each other will stay together is a formidable challenge. The history of this so-called *many-body problem* is long and involved. Interesting progress has been achieved, but the final answer still eludes us.

— Many challenges remain in the fields of self-organization, of nonlinear evolution equations, and of chaotic motion; and they motivate numerous researchers in mathematics, physics, chemistry, biology, medicine and the other sciences.

What is contact?

> Democritus declared that there is a unique sort of motion: that ensuing from collision.
>
> Simplicius, *Commentary on the Physics of Aristotle*, 42, 10 Ref. 245

Of the questions unanswered by classical physics, the details of contact and collisions are among the most pressing. Indeed, we defined mass in terms of velocity changes during collisions. Page 77 But why do objects change their motion in such instances? Why are collisions between two balls made of chewing gum different from those between two stainless-steel balls? What happens during those moments of contact?

Contact is related to material properties, which in turn influence motion in a complex way. The complexity is such that the sciences of material properties developed independently from the rest of physics for a long time; for example, the techniques of metallurgy (often called the oldest science of all) of chemistry and of cooking were related to the properties of motion only in the twentieth century, after having been independently pursued for thousands of years. Since material properties determine the essence of contact, we *need* knowledge about matter and about materials to understand the notion of mass, and thus of motion. The second part of our mountain ascent will reveal these connections.

Precision and accuracy

When we started climbing Motion Mountain, we stated that to gain height means to increase the *precision* of our description of nature. To make even this statement itself more precise, we distinguish between two terms: *precision* is the degree of reproducibility; *accuracy* is the degree of correspondence to the actual situation. Both concepts apply to measurements,* to statements and to physical concepts.

Appendix B

At present, the record number of digits ever measured for a physical quantity is 14. Why so few? Classical physics doesn't provide an answer. What is the maximum number of digits we can expect in measurements; what determines it; and how can we achieve it? These questions are still open at this point in our ascent; they will be covered in the second part of it.

Challenge 559 n

On the other hand, statements with false accuracy abound. What should we think of a car company – Ford – who claim that the drag coefficient c_w of a certain model is 0.375? Or of the official claim that the world record in fuel consumption for cars is 2315.473 km/l? Or of the statement that 70.3 % of all citizens share a certain opinion? One lesson we learn from investigations into measurement errors is that we should never provide more digits for a result than we can put our hand into fire for.

Challenge 560 n

Is it possible to draw or produce a rectangle for which the ratio of lengths is a real number, e.g. of the form 0.131520091514001315211420010914..., whose digits encode a book? (A simple method would code a space as 00, the letter 'a' as 01, 'b' as 02, 'c' as 03, etc. Even more interestingly, could the number be printed inside its own book?)

In our walk we aim for precision and accuracy, while avoiding false accuracy. Therefore, concepts have mainly to be *precise*, and descriptions have to be *accurate*. Any inaccuracy is a proof of lack of understanding. To put it bluntly, 'inaccurate' means *wrong*. Increasing the accuracy and precision of our description of nature implies leaving behind us all the mistakes we have made so far. That is our aim in the following.

Can all of nature be described in a book?

Let us have some fun with a puzzle related to our adventure. Could the perfect physics publication, one that describes all of nature, exist? If it does, it must also describe itself, its own production – including its readers and its author – and most important of all, its own contents. Is such a book possible? Using the concept of information, we can state that

* For measurements, both precision and accuracy are best described by their *standard deviation*, as explained in Appendix B, on page 1180.

such a book should contain all information contained in the universe. Is this possible? Let us check the options.

If nature requires an *infinitely* long book to be fully described, such a publication obviously cannot exist. In this case, only approximate descriptions of nature are possible and a perfect physics book is impossible.

If nature requires a *finite* amount of information for its description, there are two options. One is that the information of the universe cannot be summarized in a book; them a perfect physics book is again impossible. The other option is that the universe does contain a finite amount of information and that it can be summarized in a few small statements. This would imply that the rest of the universe would not add to the information already contained in the perfect physics book. In this case, it seems that the entropy of the book and the entropy of the universe must be similar. This seems quite unlikely.

We note that the answer to this puzzle also implies the answer to another puzzle: whether a brain can contain a full description of nature. In other words, the real question is: can we understand nature? Is our hike to the top of motion mountain possible? We usually believe this. In short, we believe something which is rather unlikely: we believe that the universe does not contain more information than what our brain could contain or even contains already.

Do we have an error in our arguments? Yes, we do. The terms 'universe' and 'information' are not used correctly in the above reasoning, as you might want to verify. We will solve this puzzle later in our adventure. Until then, do make up your own mind.

Page 1070
Challenge 561 e

Why is measurement possible?

In the description of gravity given so far, the one that everybody learns – or should learn – at school, acceleration is connected to mass and distance via $a = GM/r^2$. That's all. But this simplicity is deceiving. In order to check whether this description is correct, we have to measure lengths and times. However, it is *impossible* to measure lengths and time intervals with any clock or any ruler based on the gravitational interaction alone! Try to conceive such an apparatus and you will be inevitably be disappointed. You always need a non-gravitational method to start and stop the stopwatch. Similarly, when you measure length, e.g. of a table, you have to hold a ruler or some other device near it. The interaction necessary to line up the ruler and the table cannot be gravitational.

Challenge 562 n

A similar limitation applies even to mass measurements. Try to measure mass using gravitation alone. Any scale or balance needs other – usually mechanical, electromagnetic or optical – interactions to achieve its function. Can you confirm that the same applies to speed and to angle measurements? In summary, whatever method we use, *in order to measure velocity, length, time, and mass, interactions other than gravity are needed.* Our ability to measure shows that gravity is not all there is.

Challenge 563 n
Challenge 564 n

Is motion unlimited?

Galilean physics does not explain the ability to measure. In fact, it does not even explain the existence of standards. Why do objects have fixed lengths? Why do clocks work with regularity? Galilean physics cannot explain these observations.

Galilean physics also makes no clear statements on the universe as a whole. It seems to suggest that it is infinite. Finitude does not fit with the Galilean description of motion. Galilean physics is thus limited in its explanations because it disregards the limits of motion.

We also note that the existence of infinite speeds in nature would not allow us to define time sequences. Clocks would then be impossible. In other words, a description of nature that allows unlimited speeds is not precise. Precision requires limits. To achieve the highest possible precision, we need to discover all limits to motion. So far, we have discovered only one: there is a smallest entropy. We now turn to another, more striking one: the limit for speed. To understand this limit, we will explore the most rapid motion we know: the motion of light.

Bibliography

145 Giuseppe Fumagalli, *Chi l'ha detto?*, Hoepli, 1983. Cited on page 175.

146 The beautiful story of the south-pointing carriage is told in Appendix B of James Foster & J.D. Nightingale, *A Short Course in General Relativity*, Springer Verlag, 2nd edition, 1998. Such carriages have existed in China, as told by the great sinologist Joseph Needham, but their construction is unknown. The carriage described by Foster and Nightingale is the one reconstructed in 1947 by George Lancaster, a British engineer. Cited on page 176.

147 See for example Z. Ghahramani, Building blocks of movement, *Nature* **407**, pp. 682–683, 2000. Researchers in robot control are also interested in such topics. Cited on page 176.

148 G. Gutierrez, C. Fehr, A. Calzadilla & D. Figueroa, Fluid flow up the wall of a spinning egg, *American Journal of Physics* **66**, pp. 442–445, 1998. Cited on page 177.

149 A historical account is given in Wolfgang Yourgray & Stanley Mandelstam, *Variational Principles in Dynamics and Quantum Theory*, Dover, 1968. Cited on pages 178 and 184.

150 Max Päsler, *Prinzipe der Mechanik*, Walter de Gruyter & Co., 1968. Cited on page 182.

151 The relations between possible Lagrangians are explained by Herbert Goldstein, *Classical Mechanics*, 2nd edition, Addison-Wesley, 1980. Cited on page 183.

152 C.G. Gray, G. Karl & V.A. Novikov, From Maupertius to Schrödinger. Quantization of classical variational principles, *American Journal of Physics* **67**, pp. 959–961, 1999. Cited on page 185.

153 The Hemingway statement is quoted by Marlene Dietrich in Aaron E. Hotchner, *Papa Hemingway*, Random House, 1966, in part 1, chapter 1. Cited on page 185.

154 J.A. Moore, An innovation in physics instruction for nonscience majors, *American Journal of Physics* **46**, pp. 607–612, 1978. Cited on page 185.

155 See e.g. Alan P. Boss, Extrasolar planets, *Physics Today* **49**, pp. 32–38. September 1996. The most recent information can be found at the 'Extrasolar Planet Encyclopaedia' maintained at http://www.obspm.fr/planets by Jean Schneider at the Observatoire de Paris. Cited on page 188.

156 A good review article is by David W. Hughes, Comets and Asteroids, *Contemporary Physics* **35**, pp. 75–93, 1994. Cited on page 188.

157 G.B. West, J.H. Brown & B.J. Enquist, A general model for the origin of allometric scaling laws in biology, *Science* **276**, pp. 122–126, 4 April 1997, with a comment on page 34 of the same issue. The rules governing branching properties of blood vessels, of lymph systems and of vessel systems in plants are explained. For more about plants, see also the paper G.B. West, J.H. Brown & B.J. Enquist, A general model for the structure and allometry of plant vascular systems, *Nature* **400**, pp. 664–667, 1999. Cited on page 189.

158 J.R. Banavar, A. Martin & A. Rinaldo, Size and form in efficient transportation networks, *Nature* **399**, pp. 130–132, 1999. Cited on page 189.

159 N. Moreira, New striders - new humanoids with efficient gaits change the robotics landscape, *Science News Online* 6th of August, 2005. Cited on page 190.

160 John Mansley Robinson, *An Introduction to Early Greek Philosophy*, Houghton Muffin 1968, chapter 5. Cited on page 191.

161 See e.g. B. Bower, A child's theory of mind, *Science News* **144**, pp. 40–41. Cited on page 193.

162 The most beautiful book on this topic is the text by BRANKO GRÜNBAUM & G.C. SHEPHARD, *Tilings and Patterns*, W.H. Freeman and Company, New York, 1987. It has been translated into several languages and republished several times. Cited on page 194.

163 U. NIEDERER, The maximal kinematical invariance group of the free Schrödinger equation, *Helvetica Physica Acta* 45, pp. 802–810, 1972. See also the introduction by O. JAHN & V.V. SREEDHAR, The maximal invariance group of Newton's equations for a free point particle, http://arxiv.org/abs/math-ph/0102011. Cited on page 199.

164 The story is told in the interesting biography of Einstein by A. PAIS, *'Subtle is the Lord...' – The Science and the Life of Albert Einstein*, Oxford University Press, 1982. Cited on page 200.

165 W. ZÜRN & R. WIDMER-SCHNIDRIG, Globale Eigenschwingungen der Erde, *Physik Journal* 1, pp. 49–55, 2002. Cited on page 206.

166 N. GAUTHIER, What happens to energy and momentum when two oppositely-moving wave pulses overlap?, *American Journal of Physics* 71, pp. 787–790, 2003. Cited on page 210.

167 An informative and modern summary of present research about the ear and the details of its function is http://www.physicsweb.org/article/world/15/5/8. Cited on page 212.

168 A.L. HODGKIN & A.F. HUXLEY, A quantitative description of membrane current and its application to conduction and excitation in nerve, *Journal of Physiology* 117, pp. 500–544, 1952. This famous paper of theoretical biology earned the authors the Nobel Prize in Medicine in 1963. Cited on page 215.

169 T. FILIPPOV, *The Versatile Soliton*, Springer Verlag, 2000. See also J.S. RUSSEL, Report of the Fourteenth Meeting of the British Association for the Advancement of Science, Murray, London, 1844, pp. 311–390. Cited on pages 215 and 217.

170 N.J. ZABUSKY & M.D. KRUSKAL, Interaction of solitons in a collisionless plasma and the recurrence of initial states, *Physical Review Letters* 15, pp. 240–243, 1965. Cited on page 217.

171 O. MUSKENS, De kortste knal ter wereld, *Nederlands tijdschrift voor natuurkunde* pp. 70–73, 2005. Cited on page 217.

172 E. HELLER, Freak waves: just bad luck, or avoidable?, *Europhysics News* pp. 159–161, September/October 2005, downloadable at www.europhysicsnews.org. Cited on page 218.

173 For more about the ocean sound channel, see the novel by TOM CLANCY, *The Hunt for Red October*. See also the physics script by R.A. MULLER, Government secrets of the oceans, atmosphere, and UFOs, http://web.archive.org/web/*/http://muller.lbl.gov/teaching/Physics10/chapters/9-SecretsofUFOs.html 2001. Cited on page 220.

174 B. WILSON, R.S. BATTY & L.M. DILL, Pacific and Atlantic herring produce burst pulse sounds, *Biology Letters* 271, number S3, 7 February 2004. Cited on page 220.

175 See for example the article by G. FRITSCH, Infraschall, *Physik in unserer Zeit* 13, pp. 104–110, 1982. Cited on page 221.

176 Wavelet transformations were developed by the French mathematicians Alex Grossmann, Jean Morlet and Thierry Paul. The basic paper is A. GROSSMANN, J. MORLET & T. PAUL, Integral transforms associated to square integrable representations, *Journal of Mathematical Physics* 26, pp. 2473–2479, 1985. For a modern introduction, see STÉPHANE MALLAT, *A Wavelet Tour of Signal Processing*, Academic Press, 1999. Cited on page 221.

177 JAY INGRAM, *The Velocity of Honey*, Viking, 2o03. Cited on page 222.

178 M. BOITI, J. J.-P. LEON, L. MARTINA & F. PEMPINELLI, Scattering of localized solitons in the plane, *Physics Letters A* 132, pp. 432–439, 1988, A.S. FOKAS & P.M. SANTINI, Coherent structures in multidimensions, *Physics Review Letters* 63, pp. 1329–1333, 1989, J. HIETARINTA & R. HIROTA, Multidromion solutions to the Davey–Stewartson equation, *Physics Letters A* 145, pp. 237–244, 1990. Cited on page 222.

179 M. Ausloos, *Proceedings of the Royal Society in London A* **400**, pp. 331–350, 1985. Cited on page 223.

180 T.A. McMahob & J. Tyler Bonner, *Form und Leben – Konstruktion vom Reißbrett der Natur*, Spektrum Verlag, 1985. Cited on page 225.

181 G.W. Koch, S.C. Sillett, G.M. Jennings & S.D. Davis, The limits to tree height, *Nature* **428**, pp. 851–854, 2004. Cited on page 226.

182 A simple article explaining the tallness of trees is A. Mineyev, Trees worthy of Paul Bunyan, *Quantum* pp. 4–10, January–February 1994. (Paul Bunyan is a mythical giant lumberjack who is the hero of the early frontier pioneers in the United States.) Note that the transport of liquids in trees sets no limits on their height, since water is pumped up along tree stems (except in spring, when it is pumped up from the roots) by evaporation from the leaves. This works almost without limits because water columns, when nucleation is carefully avoided, can be put under tensile stresses of over 100 bar, corresponding to 1000 m. See also P. Nobel, *Plant Physiology*, Academic Press, 2nd Edition, 1999. Cited on page 226.

183 Such information can be taken from the excellent overview article by M.F. Ashby, On the engineering properties of materials, *Acta Metallurgica* **37**, pp. 1273–1293, 1989. The article explains the various general criteria which determine the selection of materials, and gives numerous tables to guide the selection. Cited on page 226.

184 For a photograph of a single Barium *atom* – named Astrid – see Hans Dehmelt, Experiments with an isolated subatomic particle at rest, *Reviews of Modern Physics* **62**, pp. 525–530, 1990. For an earlier photograph of a Barium *ion*, see W. Neuhauser, M. Hohenstatt, P.E. Toschek & H. Dehmelt, Localized visible Ba^+ mono-ion oscillator, *Physical Review A* **22**, pp. 1137–1140, 1980. Cited on page 228.

185 Holograms of atoms were first produced by Hans-Werner Fink & al., Atomic resolution in lens-less low-energy electron holography, *Physical Review Letters* **67**, pp. 1543–1546, 1991. Cited on page 228.

186 A single–atom laser was built in 1994 by K. An, J.J. Childs, R.R. Dasari & M.S. Feld, Microlaser: a laser with one atom in an optical resonator, *Physical Review Letters* **73**, p. 3375, 1994. Cited on page 229.

187 The photograph on page 229 is the first image that showed subatomic structures (visible as shadows on the atoms). It was published by F.J. Giessibl, S. Hembacher, H. Bielefeldt & J. Mannhart, Subatomic features on the silicon (111)-(7x7) surface observed by atomic force microscopy, *Science* **289**, pp. 422 – 425, 2000. Cited on page 229.

188 See for example C. Schiller, A.A. Koomans, T.L. van Rooy, C. Schönenberger & H.B. Elswijk, Decapitation of tungsten field emitter tips during sputter sharpening, *Surface Science Letters* **339**, pp. L925–L930, 1996. Cited on page 229.

189 U. Weierstall & J.C.H. Spence, An STM with time-of-flight analyzer for atomic species identification, MSA 2000, Philadelphia, *Microscopy and Microanalysis* **6**, Supplement 2, p. 718, 2000. Cited on page 229.

190 The present state of our understanding of turbulence is described in ... Cited on page 231.

191 K. Weltner, A comparison of explanations of aerodynamical lifting force, *American Journal of Physics* **55**, pp. 50–54, 1987, K. Weltner, Aerodynamic lifting force, *The Physics Teacher* **28**, pp. 78–82, 1990. See also the http://user.uni-frankfurt.de/~weltner/Flight/PHYSIC4.htm and the http://www.av8n.com/how/htm/airfoils.html websites. Cited on page 231.

192 Lydéric Bocquet, The physics of stone skipping, *American Journal of Physics* **17**, pp. 150–155, 2003. The present recod holder is Kurt Steiner, with 40 skips. See http://pastoneskipping.com/steiner.htm and http://www.stoneskipping.com. The site http://www.yeeha.net/nassa/guin/g2.html is by the a previous world record holder, Jerdome Coleman–McGhee. Cited on page 233.

193 S.F. Kistler & L.E. Scriven, The teapot effect: sheetforming flows with deflection, wetting, and hysteresis, *Journal of Fluid Mechanics* **263**, pp. 19–62, 1994. Cited on page 235.

194 E. Hollander, Over trechters en zo ..., *Nederlands tijdschrift voor natuurkunde* **68**, p. 303, 2002. Cited on page 235.

195 P. Krehl, S. Engemann & D. Schwenkel, The puzzle of whip cracking – uncovered by a correlation of whip-tip kinematics with shock wave emission, *Shock Waves* **8**, pp. 1–9, 1998. The authors used high-speed cameras to study the motion of the whip. A new aspect has been added by A. Goriely & T. McMillen, Shape of a cracking whip, *Physical Review Letters* **88**, p. 244301, 2002. This article focuses on the tapered shape of the whip. However, the neglection of the tuft – a piece at the end of the whip which is required to make it crack – in the latter paper shows that there is more to be discovered still. Cited on page 238.

196 S. Dorbolo, H. Caps & N. Vandewalle, Fluid instabilities in the birth and death of antibubbles, *New Journal of Physics* **5**, p. 161, 2003. Cited on page 236.

197 Z. Sheng & K. Yamafuji, Realization of a Human Riding a Unicycle by a Robot, *Proceedings of the 1995 IEEE International Conference on Robotics and Automation*, Vol. 2, pp. 1319 – 1326, 1995. Cited on page 239.

198 On human uncicycling, see Jack Wiley, *The Complete Book of Unicycling*, Lodi, 1984, and Sebastian Hoeher, *Einradfahren und die Physik*, Reinbeck, 1991. Cited on page 239.

199 W. Thomson, Lecture to the Royal Society of Edinburgh, 18 February 1867, *Proceedings of the Royal Society in Edinborough* **6**, p. 94, 1869. Cited on page 240.

200 T.T. Lim, A note on the leapfrogging between two coaxial vortex rings at low Reynolds numbers, *Physics of Fluids* **9**, pp. 239–241, 1997. Cited on page 237.

201 D. Karstädt, F. Pinno, K.P. Möllmann & M. Vollmer, Anschauliche Wärmelehre im Unterricht: ein Beitrag zur Visualisierung thermischer Vorgänge, *Praxis der Naturwissenschaften Physik* **5-48**, pp. 24–31, 1999, K.-P. Möllmann & M. Vollmer, Eine etwas andere, physikalische Sehweise - Visualisierung von Energieumwandlungen und Strahlungsphysik für die (Hochschul-)lehre, *Physikalische Bllätter* **56**, pp. 65–69, 2000, D. Karstädt, K.P. Möllmann, F. Pinno & M. Vollmer, There is more to see than eyes can detect: visualization of energy transfer processes and the laws of radiation for physics education, *The Physics Teacher* **39**, pp. 371–376, 2001, Cited on page 242.

202 F. Herrmann, Mengenartige Größen im Physikunterricht, *Physikalische Blätter* **54**, pp. 830–832, September 1998. See also his lecture notes on general introductory physics on the website http://www-tfp.physik.uni-karlsruhe.de/~didaktik. Cited on pages 241 and 244.

203 Thermostatics is difficult to learn also because it was not discovered in a systematic way. See C. Truesdell, *The Tragicomical History of Thermodynamics 1822–1854*, Springer Verlag, 1980. An excellent advanced textbook on thermostatics and thermodynamics is Linda Reichl, *A Modern Course in Statistical Physics*, Wiley, 2nd edition, 1998. Cited on page 241.

204 Gas expansion was the main method used for the definition of the official temperature scale. Only in 1990 were other methods introduced officially, such as total radiation thermometry (in the range 140 K to 373 K), noise thermometry (2 K to 4 K and 900 K to 1235 K), acoustical thermometry (around 303 K), magnetic thermometry (0.5 K to 2.6 K) and optical radiation thermometry (above 730 K). Radiation thermometry is still the central method in the range from about 3 K to about 1000 K. This is explained in detail in R.L. Rusby, R.P. Hudson, M. Durieux, J.F. Schooley, P.P.M. Steur & C.A. Swenson, The basis of the ITS-90, *Metrologia* **28**, pp. 9–18, 1991. On the water boiling point see also page 1188. Cited on page 241.

205 See for example the captivating text by Gino Segrè, *A Matter of Degrees: What Temperature Reveals About the Past and Future of Our Species, Planet and Universe*, Viking, New York, 2002. Cited on page 242.

206 B. Polster, What is the best way to lace your shoes?, *Nature* **420**, p. 476, 5 December 2002. Cited on page 245.

207 See for example the article by H. Preston-Thomas, The international temperature scale of 1990 (ITS-90), *Metrologia* **27**, pp. 3–10, 1990, and the errata H. Preston-Thomas, The international temperature scale of 1990 (ITS-90), *Metrologia* **27**, p. 107, 1990, Cited on page 249.

208 For an overview, see Christian Enss & Siegfried Hunklinger, *Low-Temperature Physics*, Springer, 2005. Cited on page 249.

209 The famous paper on Brownian motion which contributed so much to Einstein's fame is A. Einstein, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, *Annalen der Physik* **17**, pp. 549–560, 1905. In the following years, Einstein wrote a series of further papers elaborating on this topic. For example, he published his 1905 Ph.D. thesis as A. Einstein, Eine neue Bestimmung der Moleküldimensionen, *Annalen der Physik* **19**, pp. 289–306, 1906, and he corrected a small mistake in A. Einstein, Berichtigung zu meiner Arbeit: 'Eine neue Bestimmung der Moleküldimensionen', *Annalen der Physik* **34**, pp. 591–592, 1911, where, using new data, he found the value $6.6 \cdot 10^{23}$ for Avogadro's number. Cited on page 249.

210 The first tests of the prediction were performed by J. Perrin, *Comptes Rendus de l'Académie des Sciences* **147**, pp. 475–476, and pp. 530–532, 1908. He masterfully sums up the whole discussion in Jean Perrin, *Les atomes*, Librarie Félix Alcan, Paris, 1913. Cited on page 250.

211 Pierre Gaspard & al., Experimental evidence for microscopic chaos, *Nature* **394**, p. 865, 27 August 1998. Cited on page 251.

212 These points are made clearly and forcibly, as is his style, by N.G. van Kampen, Entropie, *Nederlands tijdschrift voor natuurkunde* **62**, pp. 395–396, 3 December 1996. Cited on page 252.

213 This is a disappointing result of all efforts so far, as Grégoire Nicolis always stresses in his university courses. Seth Lloyd has compiled a list of 31 proposed definitions of complexity, containing among others, fractal dimension, grammatical complexity, computational complexity, thermodynamic depth. See, for example, a short summary in *Scientific American* p. 77, June 1995. Cited on page 252.

214 Minimal entropy is discussed by L. Szilard, Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen, *Zeitschrift für Physik* **53**, pp. 840–856, 1929. This classic paper can also be found in English translation in his collected works. Cited on page 252.

215 G. Cohen-Tannoudji, *Les constantes universelles*, Pluriel, Hachette, 1998. See also L. Brillouin, *Science and Information Theory*, Academic Press, 1962. Cited on page 253.

216 See for example A.E. Shalyt-Margolin & A.Ya. Tregubovich, Generalized uncertainty relation in thermodynamics, http://arxiv.org/abs/gr-qc/0307018, or J. Uffink & J. van Lith-van Dis, Thermodynamic uncertainty relations, *Foundations of Physics* 29, p. 655, 1999. Cited on page 253.

217 H.W. Zimmermann, Particle entropies and entropy quanta IV: the ideal gas, the second law of thermodynamics, and the P-t uncertainty relation, *Zeitschrift für physikalische Chemie* 217, pp. 55–78, 2003, and H.W. Zimmermann, Particle entropies and entropy quanta V: the P-t uncertainty relation, *Zeitschrift für physikalische Chemie* 217, pp. 1097–1108, 2003. Cited on pages 253 and 254.

218 B. Lavenda, *Statistical Physics: A Probabilistic Approach*, Wiley-Interscience, New York, 1991. Cited on pages 253 and 254.

219 The quote is found in the introduction by George Wald to the text by Lawrence J. Henderson, *The Fitness of the Environment*, Macmillan, New York, 1913, reprinted 1958. Cited on page 254.

220 A fascinating introduction to chemistry is the text by John Emsley, *Molecules at an Exhibition*, Oxford University Press, 1998. Cited on page 255.

221 An excellent introduction into the physics of heat is the book by Linda Reichl, *A Modern Course in Statistical Physics*, Wiley, 2nd edition, 1998. Cited on page 256.

222 Emile Borel, *Introduction géométrique à la physique*, Gauthier-Villars, 1912. Cited on page 257.

223 See V.L. Telegdi, Enrico Fermi in America, *Physics Today* 55, pp. 38–43, June 2002. Cited on page 257.

224 K. Schmidt-Nielsen, *Desert Animals: Physiological Problems of Heat and Water*, Oxford University Press, 1964. Cited on page 258.

225 Why entropy is created when information is erased, but not when it is acquired, is explained in Charles H. Bennett & Rolf Landauer, Fundamental Limits of Computation, *Scientific American* 253:1, pp. 48–56, 1985. The conclusion: we should pay to throw the newspaper away, not to buy it. Cited on page 260.

226 See, for example, G. Swift, Thermoacoustic engines and refrigerators, *Physics Today* 48, pp. 22–28, July 1995. Cited on page 262.

227 Quoted in D. Campbell, J. Crutchfield, J. Farmer & E. Jen, Experimental mathematics: the role of computation in nonlinear science, *Communications of the Association of Computing Machinery* 28, pp. 374–384, 1985. Cited on page 263.

228 For more about the shapes of snowflakes, see the famous book by W.A. Bentlye & W.J. Humphreys, *Snow Crystals*, Dover Publications, New York, 1962. This second printing of the original from 1931 shows a large part of the result of Bentley's lifelong passion, namely several thousand photographs of snowflakes. Cited on page 263.

229 K. Schwenk, Why snakes have forked tongues, *Science* 263, pp. 1573–1577, 1994. Cited on page 264.

230 P.B. Umbanhowar, F. Melo & H.L. Swinney, Localized excitations in a vertically vibrated granular layer, *Nature* 382, pp. 793–796, 29 August 1996. Cited on page 264.

231 D.K. Campbell, S. Flach & Y.S. Kivshar, Localizing energy through nonlinearity and discreteness, *Physics Today* 57, pp. 43–49, January 2004. Cited on page 265.

232 B. Andreotti, The song of dunes as a wave-particle mode locking, *Physical Review Letters* 92, p. 238001, 2004. Cited on page 265.

233 K. Kötter, E. Goles & M. Markus, Shell structures with 'magic numbers' of spheres in a swirled disk, *Physical Review E* **60**, pp. 7182–7185, 1999. Cited on page 265.

234 A good introduction is the text by Daniel Walgraef, *Spatiotemporal Pattern Formation, With Examples in Physics, Chemistry and Materials Science*, Springer 1996. Cited on page 265.

235 For an overview, see the Ph.D. thesis by Joceline Lega, *Défauts topologiques associés à la brisure de l'invariance de translation dans le temps*, Université de Nice, 1989. Cited on page 267.

236 An idea of the fascinating mechanisms at the basis of the heart beat is given by A. Babloyantz & A. Destexhe, Is the normal heart a periodic oscillator?, *Biological Cybernetics* **58**, pp. 203–211, 1989. Cited on page 267.

237 For a short, modern overview of turbulence, see L.P. Kadanoff, A model of turbulence, *Physics Today* **48**, pp. 11–13, September 1995. Cited on page 268.

238 For a clear introduction, see T. Schmidt & M. Mahrl, A simple mathematical model of a dripping tap, *European Journal of Physics* **18**, pp. 377–383, 1997. Cited on page 268.

239 An overview of science humour can be found in the famous anthology compiled by R.L. Weber, edited by E. Mendoza, *A Random Walk in Science*, Institute of Physics, 1973. It is also available in several expanded translations. Cited on page 269.

240 K. Mertens, V. Putkaradze & P. Vorobieff, Braiding patterns on an inclined plane, *Nature* **430**, p. 165, 2004. Cited on page 270.

241 G. Müller, Starch columns: analog model for basalt columns, *Journal of Geophysical Research* **103**, pp. 15239–15253, 1998. Cited on page 270.

242 B. Hof, C.W.H. van Doorne, J. Westerweel, F.T.M. Nieuwstadt, H. Wedin, R. Kerswell, F. Waleffe, H. Faisst & B. Eckhardt, Experimental observation of nonlinear traveling waves in turbulent pipe flow, *Science* **305**, pp. 1594–1598, 2004. Cited on page 270.

243 A fascinating book on the topic is Kenneth Laws & Martha Swope, *Physics and the Art of Dance: Understanding Movement*, Oxford University Press 2002. Cited on page 271.

244 J.J. Lissauer, Chaotic motion in the solar system, *Reviews of Modern Physics* **71**, pp. 835–845, 1999. Cited on page 271.

245 See Jean-Paul Dumont, *Les écoles présocratiques*, Folio Essais, Gallimard, 1991, p. 426. Cited on page 271.