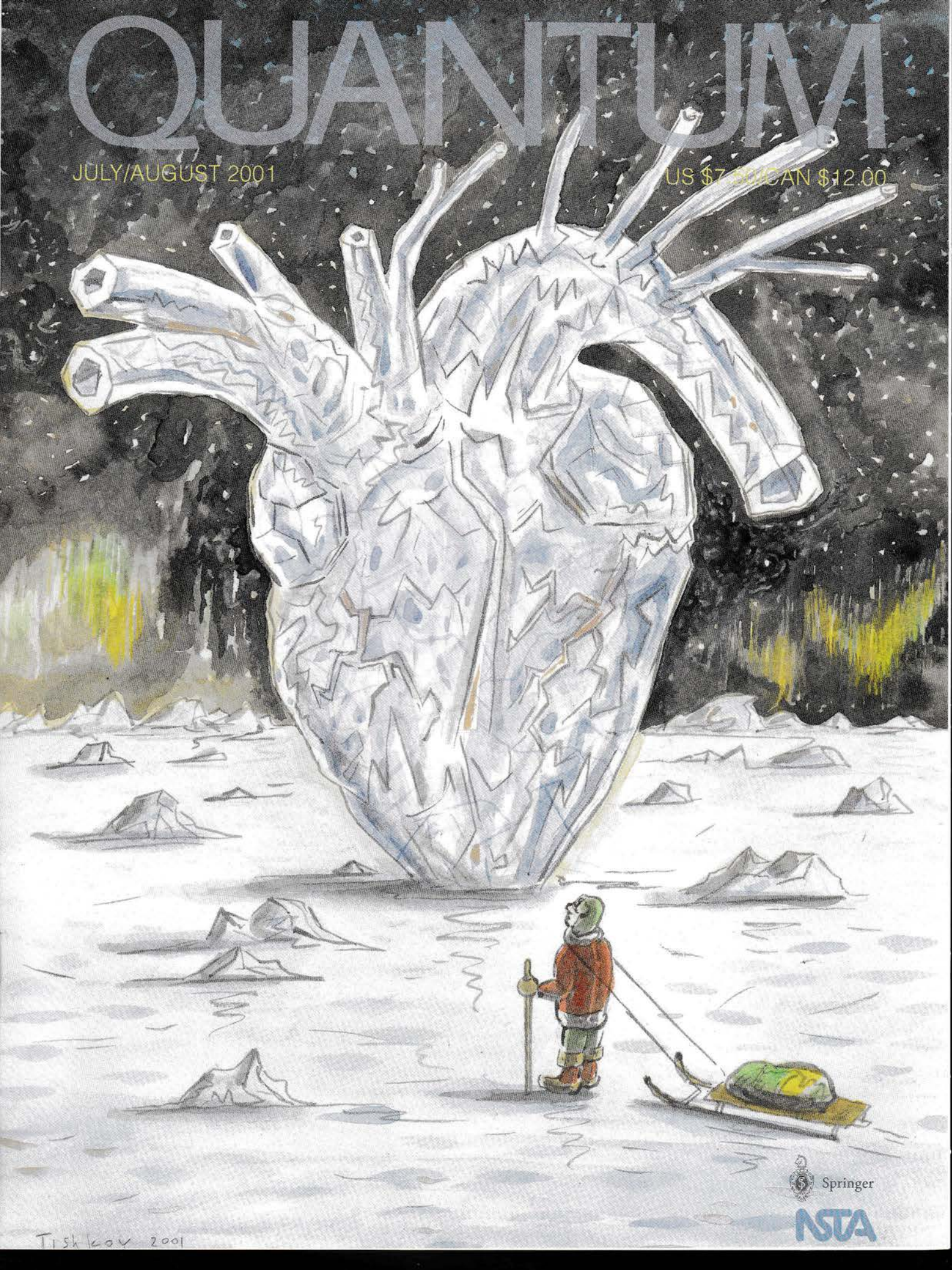
QUANTUM

JULY/AUGUST 2001

US $7.50/CAN $12.00



Tishkov 2001

Springer

NSTA

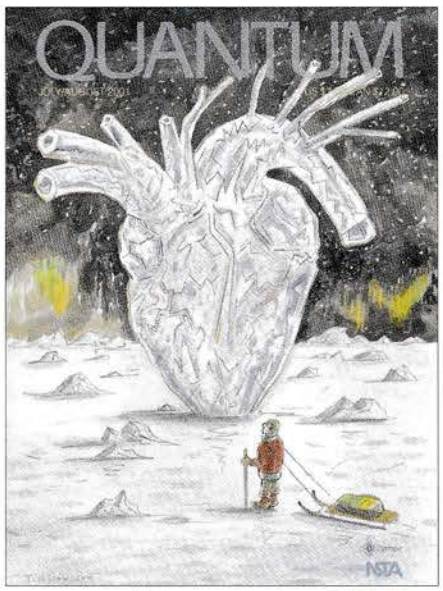
Oil on canvas, 34 7/8 × 34 7/8, Gift of Mrs. Robert W. Schuette, ©2001 Board of Trustees, National Gallery of Art, Washington, D.C.

The Magic Lantern (1764) by Charles Amédée Philippe Vanloo

PEOPLE HAVE BEEN FASCINATED BY FLICKERING images ever since shadows were first cast on cave walls by primitive peoples sitting around a fire. As time passed, technology allowed us to gain more control over the creation of these images—and some would say these images gained more control over us. The magic lantern pictured above was the 18th century equivalent of today's 52", high-definition, surround-sound television sets that dominate many family rooms. Of course, the lantern's flicker was due to the inconstant flame of a candle, while the rapid flash of the TV screen is precisely controlled by high-tech electronics. To learn how the eyes of couch potatoes have processed the images of idiot boxes both ancient and modern, turn to page 30.

QUANTUM

JULY/AUGUST 2001 VOLUME 11, NUMBER 6



Cover art by Leonid Tishkov

Why, you may wonder, is *Quantum* using a snowy scene for its July/August cover? The answer is twofold. First, the scene ties in nicely with our article on the physics of ice crystals, which can be found on page 6. And second, it marks the passage of *Quantum* magazine into the winter of its existence. With this final issue, we mark the end of *Quantum*'s 12-year journey into the most challenging areas of math and physics. Thanks for sharing the journey with us. Perhaps one day we'll meet again further down the road. (See notice below.)

NOTICE TO SUBSCRIBERS

We regret to inform our loyal readership that this is the last issue of *Quantum* published by the National Science Teachers Association. NSTA is proud of its 12-year history of producing the magazine and is grateful to its colleagues in Russia and the U.S. for their dedication and hard work. We hope to work with other groups to bring *Quantum* back as a Web-only resource, free to all—but we cannot predict the likelihood of success in developing financial support for this endeavor. Please continue to check for updates at the *Quantum* website (www.nsta.org/quantum).

FEATURES

DEPARTMENTS

Indexed in *Magazine Article Summaries, Academic Abstracts, Academic Search, Vocational Search, MasterFILE,* and *General Science Source.* Available in microform, electronic, or paper format from University Microfilms International.

NSTA *Quantum* (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 11 (6 issues) will be published in 2000–2001. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau (Moscow, Russia), as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000; telephone: (703) 243-7100; e-mail: quantum@nsta.org. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York NY 10010-7858.

Periodicals postage paid at New York, NY, and additional mailing offices. **Postmaster:** send address changes to: *Quantum*, Springer-Verlag New York, Inc., Journal Fulfillment Services Department, P. O. Box 2485, Secaucus NJ 07096-2485. Printed in U.S.A.

Subscription Information:
North America: Student rate: $18; Personal rate (nonstudent): $25. This rate is available to individual subscribers for personal use only from Springer-Verlag New York, Inc., when paid by personal check or charge. Subscriptions are entered with prepayment only. Institutional rate: $56. Single Issue Price: $7.50. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Mail order and payment to: Springer-Verlag New York, Inc., Journal Fulfillment Services Department, PO Box 2485, Secaucus, NJ 07094-2485, USA. Telephone: 1(800) SPRINGER; fax: (201) 348-4505; e-mail: custserv@springer-ny.com.

Outside North America: Personal rate: Please contact Springer-Verlag Berlin at subscriptions@springer.de. Institutional rate is US$57; airmail delivery is US$18 additional (all rates calculated in DM at the exchange rate current at the time of purchase). SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, Germany.

Advertising:
Representatives: (Washington) Paul Kuntzler (703) 243-7100; (New York) M.J. Mrvica Associates, Inc., 2 WestTaunton Avenue, Berlin, NJ 08009. Telephone (856)768-9360, fax (856)753-0064; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, Germany, telephone 49 (0) 30-827 87-0, telex 185 411.

Printed on acid-free paper.

Visit us on the internet at www.nsta.org/quantum/.

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

A publication of the National Science Teachers Association (NSTA)
& Quantum Bureau of the Russian Academy of Sciences
in conjunction with
the American Association of Physics Teachers (AAPT)
& the National Council of Teachers of Mathematics (NCTM)

The mission of the National Science Teachers Association is to promote excellence and innovation in science teaching and learning for all.

Time to move on...

You say goodbye and I say hello
Hello hello
I don't know why you say goodbye,
I say hello.
—Lennon/McCartney

THIS LAST ISSUE OF *QUANTUM* is a celebration of excellence that our magazine has brought to its readers for the past 12 years.

Our success was bringing quality physics and mathematics material that would stretch the most interested and talented students. Our failure was never reaching a circulation that could financially sustain the effort. It was a great run.

One puzzle that still confounds me is the response by scientists and mathematicians. In Russia, the most accomplished people of science and math contributed to *Kvant*. Their efforts formed the resource of many of our articles. These Russian luminaries felt honored to have articles published in a student magazine. From what I have heard, they exhibited great pride in having their work so published. This pride emerged not only from the service they were providing but also from the satisfaction of having successfully communicated complex ideas in a manner accessible to students. The American scientists and mathematicians have never viewed *Quantum* in this way. I will go so far as to say that they do not view any popularization of their ideas in this way.

Years ago, C.P. Snow talked about two cultures. For Snow, the cultures identified were people learned in science and those learned in humanities. There are another two cultures. The first is those academics who value communicating their work to students and others in the general community. In contrast are the others who shun popularization because they think it detracts from their work or, worse, they believe that communication with the public will harm their career options. Why does this second culture emerge? Why is it that successful writers are often considered "less serious" scientists because of their success in communicating their ideas to a wider audience? Are the universities and tenure criteria to blame? How can we get the wide range of talented professors (some of whom were *Quantum* readers when students) to value good communication and efforts in both formal classroom environments and informal arenas of magazines, museums, and media?

If the high quality of *Quantum* articles provides an example of successful interchange of ideas and spurs American scientists to place value on similar efforts, then we will see the emergence of new *Quantum* initiatives in the near future. We at NSTA are hopeful.

I was first introduced to *Kvant* by my friend Sergey Krotov of the former Soviet Union during our years as academic directors for our respective Physics Olympiad teams. The problems, articles, and humor in *Kvant* seemed like something we could import and massage for our United States audience. Lots of interested people stepped up to the plate. Bill Aldridge, Executive Director of NSTA, led the charge to create a magazine of "the highest quality." Bill came through on his commitment. He enlisted the help of Sheldon Glashow, a Nobel Laureate in physics; William Thurston, a Fields medalist in mathematics; and Yuri Ossipyan, vice-president of the Academy of Sciences of the USSR to launch the magazine. Edward Lozansky was right there as an international consultant, and Tim Weber took on the challenge of managing editor. NSTA, under Bill Aldridge's leadership and NSF support, committed resources to ensuring that *Quantum* met the needs of our intended audience. He also brought the AAPT and NCTM on board.

In the first issue, Bill Aldridge quotes the great Russian scientist and poet Michail Lomonosov as he viewed the Northern Lights: "Nature, where are your laws? The dawn appears from the dark northern climes! Does not the sun there set up its throne? Are not the ice-bound seas emitting fire? Behold, a cold

flame has covered us! Behold, the day has trod the earth at night." The 12 years of *Quantum* have attempted to answer some of these questions while illuminating the minds of so many who will one day provide us with another glimpse into the wonder of the universe.

Bill Thurston, in that same first issue, reflected on the beginning of his illustrious career as a mathematician. "As a child, I often hated arithmetic and mathematics in school. Pages of exercises were tedious and dull.... I stared out the window and let my mind wander. Sometimes I tried to puzzle something out.... Might the square root of 2 eventually be periodic if you write it out in base 12 instead of base 10? How many ways are there to fold a map into sixteenths, in quarters each way?" This spirit of inquiry pervaded the many issues of *Quantum* and stimulated many readers to invent their own questions and have their minds wander.

It is now time to say goodbye to *Quantum* and time to move on to the next challenge. The high school students who first subscribed to *Quantum* are now 30 years old. Some have received their doctoral degrees, and I read their scientific papers and biographies in magazines and newspapers. Others have chosen different paths in law, commerce, history, philosophy, and education. Everyone was enriched by *Quantum*.

It is strange that the last issue should come on my watch, as I serve as president of NSTA. The contest problem that my colleague and friend Larry Kirkpatrick and I wrote each issue was both challenging and rewarding. We were always pleased to see the creativity that Tomas Bunk added to our work with his illustration. And his illustrations were an enjoyable contrast to the work of Russian artists that has graced so many pages of our magazine.

NSTA loved *Quantum*, as did its readers. NSTA will continue to search for ways to engage the students, teachers, and others in the enjoyment of math and science. NSTA will be looking to say hello, once again. ◼

—Arthur Eisenkraft
NSTA President 2000–2001

HAPPENINGS

CyberTeaser winners

MANY OF YOU WERE ABLE TO get all of your ducks in a row to correctly answer this month's cyberteaser. Some of you ran a-fowl during your calculations, but the majority of of our contestants were able to set their sights on the correct answer.

Congratulations to those of you who were able to determine the limitations of Baron Munchhausen's hyperbolic hunting claims. Here are the names of the ten contestants who drew a bead on the correct answer the quickest:

John Beam (Bellaire, Texas)

Jerold Lewandowski (Troy, New York)

Theo Koupelis (Wausau, Wisconsin)

Shvachko Valya (Fremont, California)

Anastasia Nikitina (Princeton, New Jersey)

Margarita Satraki (Athens, Greece)

Wade Roach (Anchorage, Alaska)

Lewis Mitchell (Mt. Keira, Australia)

David Yu (San Francisco, California)

Dale A. Boyd (Chesterfield, Missouri)

Congratulations to our happy hunters! Each of you will receive a copy of this issue of *Quantum* and the classic *Quantum* button. In addition, one of you will receive some reading material for those cold mornings in the duck blind: a copy of *Quantum Quandaries*. Hopefully, it will warm your brain if not your extremities.

For those of you who weren't fortunate enough to win a copy of this perplexing publication, *Quantum Quandaries* (stock # PB123X, $8.95) can be ordered from the NSTA Science Store by calling 800-877-5300, or by visiting the online store at *www.nsta.org/store*.

Although this will be the last issue of *Quantum*, we invite you to continue to visit our website at *www.nsta.org/quantum*. There you will find our archive of past cyber teasers, along with a listing of past winners. Thank you for playing along with us all these years. We hope your future is filled with challenging puzzles to ponder. ◼

Just for the fun of it!

B325

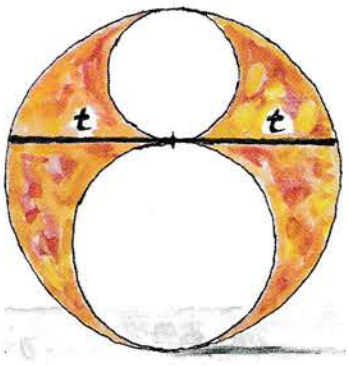
Calculating acquaintances. A teacher noticed that before class every student shook hands with 6 girls and 8 boys. The number of handshakes between boys and girls was 5 less than the number of all other handshakes. How many students were in the class? (K. Kokhas)

B326

Library Science 101. Seven volumes of an encyclopedia are arranged in the following order: 1, 5, 6, 2, 4, 3, 7. Arrange them in ascending order by volume number using the following operation. You may move three adjacent volumes to the beginning, to the end, or between any other two volumes without changing the order of the volumes in this triple. (A. Savin)

B327

A checkered career. A knight moving according to the rules of chess traversed an entire miniature chessboard consisting of 6 × 6 squares and returned to its original position after having visited each square exactly once. For some of the squares, the figure shows the number of the move when the square was visited. Restore the entire path the knight took. (A. Savin)



B328

Play this chord. The length of the chord tangent to the inscribed circles shown in the figure is $2t$. Find the area of the shaded part of the circle. (From the book *Mathematical Discovery* by George Pólya)

B329

The T of tea. Tea was poured from the same teapot into a cup with sugar in it and a cup without sugar in it. In which cup will the tea be cooler?

ANSWERS, HINTS & SOLUTIONS ON PAGE 55

Art by Pavel Chernusky

The many faces of ice

The physics of frozen water

by A. Zaretsky

ICE HAS ALWAYS BEEN CONsidered a symbol of clarity, majesty, and beauty—a beauty, however, that is also austere, carrying with it a coldness, a touch of evil, even death. It wasn't for nothing that Dante imagined that ice must be found at the center of the Earth, at the last step leading to Hell, and that Satan's treasure is hidden in a chest of ice.

It goes without saying that an artist is entitled to view natural phenomena and the laws of the universe through an emotional lens. A scientist, on the other hand, must be more impartial, both in choosing the object of study and in analyzing the mass of (sometimes contradictory) data.

What is so interesting about ice from the physical point of view?

The chemical formula for ice is H_2O. When it is cooled, water freezes—or, strictly speaking, it crystallizes. At present, thirteen different structural forms of ice are known. Those who are superstitious may have a problem with this number, but it puts ice in a special position: no other substance with such a simple chemical composition has so many phase transitions. Let this be our initial motivation in examining this curious object.

Figure 1 shows almost the entire phase diagram of H_2O. However, there are still many gaps in the picture. Suffice it to say that ice-X and ice-XI were discovered only a few years ago. The ice that everybody knows is ice-I—or, strictly speaking, ice-I_h. At normal pressure and 0°C its density is 0.917×10^3 kg · m^{-3}—it is less dense than water (that is, water's density decreases during the process of crystallization).

This paradoxical property of ice (recall that crystallization usually results in an increase in density) is of vital importance for life on Earth. The glacial armor that forms on the surface of water produces such efficient thermal insulation that lakes and reservoirs do not freeze to their full depth. It's better not to think what would happen to marine creatures if ice were just a bit more dense than water!

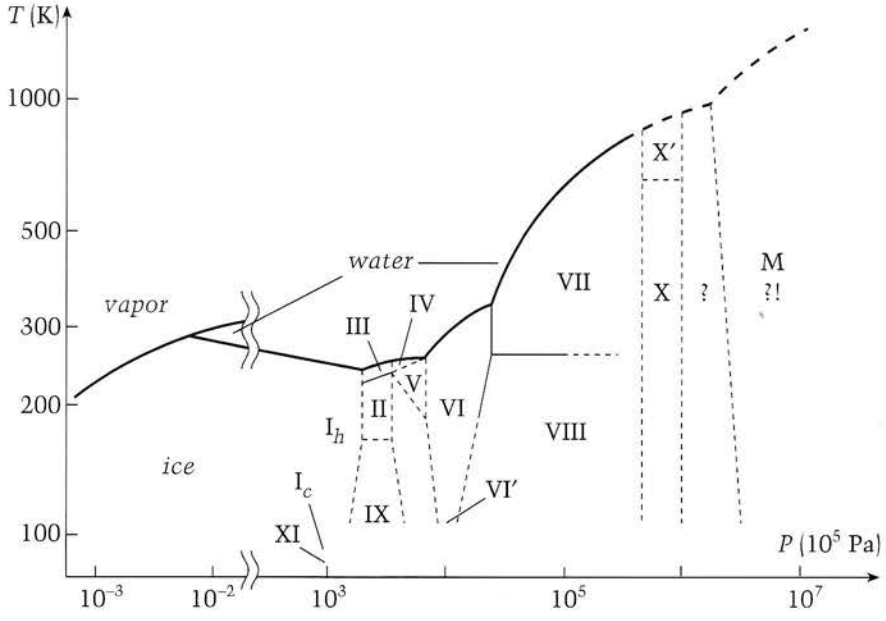
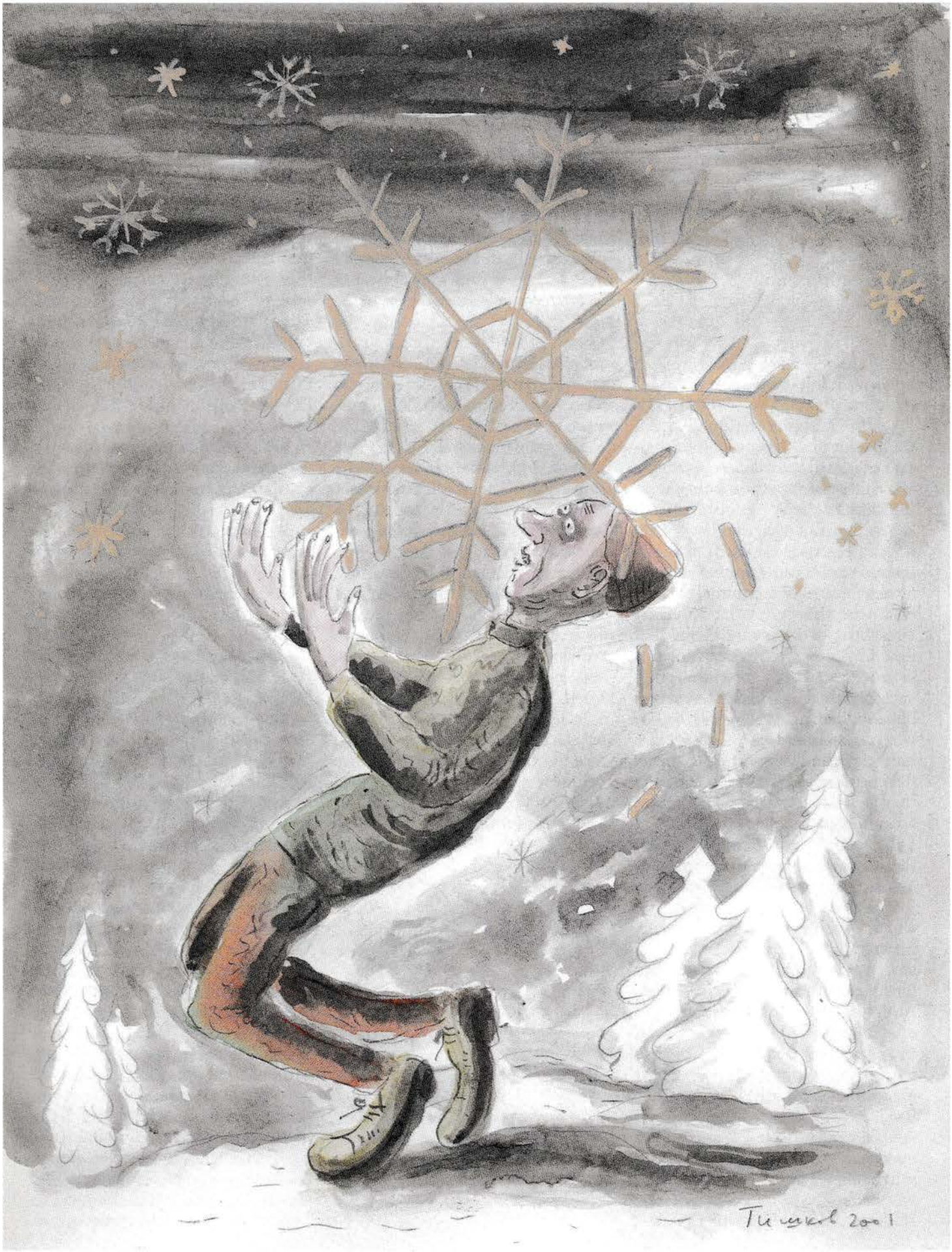


Figure 1

Art by Leonid Tishkov



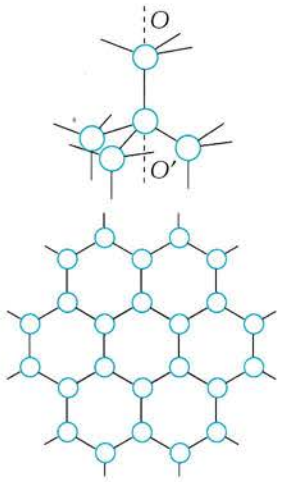


Figure 2

Now let's consider the crystal structure of ice-I_h. As early as 1917 the first X-ray diffraction analysis was performed. It wasn't too intricate a task to study the arrangement of oxygen atoms in the ice crystal. In 1922 the English physicist William Henry Bragg showed that every oxygen atom must be located approximately at the center of mass of its nearest neighbors. Soon it became clear that the oxygen lattice has a hexagonal structure and looks like figure 2. Look closely—turning the lattice 60° about the *OO′*-axis doesn't change it. This is why snowflakes (which are ice crystals, by the way) are symmetrical. (Hans Christian Andersen was mistaken: snowflakes are not ten-pointed stars—if they are star-shaped at all, they have six-points.)

It took much longer to determine the arrangement of the hydrogen atoms. The reason is that X-rays are scattered mainly by electrons. The electrons in ice are almost always concentrated near oxygen atoms, so it was very difficult to localize the hydrogen atoms by X-ray diffraction. At first it was assumed that the hydrogen atoms are located between the nearest oxygen atoms. This hypothetical structure is very symmetrical. Let's think about whether it's actually true.

First of all, it's clear that if all the hydrogen atoms are located in the middle between oxygen atoms, this would be a typical ionic crystal. However, the dielectric permeability of such substances is generally less than 10; in the case of ice, this parameter can as high as 100—that is, an order of magnitude greater. Not good for our hypothesis.

In addition, the spectra of ice, water, and water vapor (the latter is virtually a collection of individual H_2O molecules) are very similar in the infrared region. But these spectra reflect the molecular structure of a substance. Therefore, a molecule of water is "preserved" in a crystal of ice. Such crystals are called *molecular*.

Now let's do some simple arithmetic. The X-ray data tell us that the distance between oxygen atoms in ice is 27.6 nm (1 nm = 10^{-9} m). Thus, in the symmetrical model, the hydrogen atoms must be located at a distance of 13.8 nm from the oxygen atoms. But in an "isolated" water molecule the O–H distance is 9.6 nm, which is at odds with the symmetrical model.

Clearly we need to find a structure in which a certain degree of "independence" is guaranteed for the H_2O molecules. Several such structures were proposed by the English physicists John Desmond Bernal and Richard Gildart Fowler in 1933. The lattices were very complicated, and perhaps for this reason at the end of their paper the authors proposed a very unusual model for ice. According to their hypothesis, in the premelting zone the ice is crystalline only with respect to the location of whole water molecules, but the orientation of these molecules can be arbitrary to a certain degree.

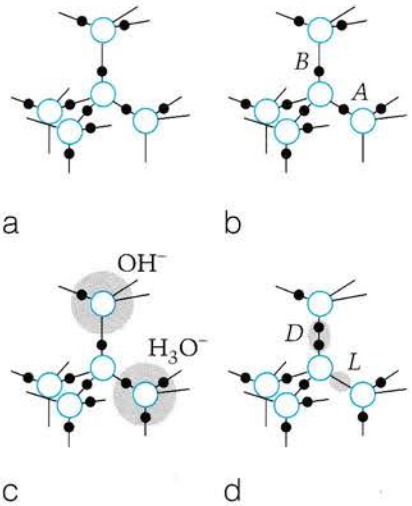


Figure 3

This important and interesting hypothesis was further developed by the American physicist and chemist Linus Pauling. He proposed that ice-I_h is crystalline only with respect to the oxygen atoms (which means that only these atoms are arranged in a certain order, forming in the total structure an independent crystal lattice—that is, a "sublattice"). The hydrogen atoms are not ordered, but their coordinates are not quite as arbitrary: they are subjected to certain rules known as "Bernal–Fowler–Pauling" (BFP) rules. Here they are:

1. The protons are arranged on the line connecting the oxygen atoms at a distance of 0.95 Å from an atom of oxygen.

2. There are two and only two protons located near every oxygen atom.

3. One and only one proton is located between neighboring oxygen atoms.

Figure 3a shows schematically a piece of ice lattice that satisfies the BFP rules.

Thus, in the classical sense of the terms, ice-I_h is neither a crystal (the hydrogen sublattice is disordered) nor an amorphous solid body (the oxygen sublattice is ordered). Again we see the need for physical research to clarify matters!

But wait—are we saying that ice is never a "true crystal"? To answer this question, we need to specify what kind of ice we mean. Ice-II, ice-VIII, ice-IX, and ice-XI are "true crystals." In these types of ice, both the hydrogen and the oxygen atoms are ordered and located at quite definite places.

Ice-X is even more interesting. Although the structure of this type is a point of vigorous debate, we can be reasonably sure that in ice-X the hydrogen atoms are located right between neighboring oxygen atoms. But look at the phase diagram. To obtain this kind of ice, we need to compress water to a pressure of

about 50 GPa ($50 \cdot 10^9$ Pa)—that is, we need to apply a 5-ton load per 1 mm^2 of its surface!

What will happen to the ice if the pressure is increased further? Unfortunately, at present it's impossible to answer this question experimentally. But scientists suppose that at ten times the pressure the ice will become... a metal. Metallic ice! It sounds unbelievable. We'll have to wait and see.

Let's return to our "ordinary" ice—that is, ice-I_h. As a general rule, solid bodies tend to become more ordered at lower temperatures. Thus, sooner or later (that is, at a sufficiently low temperature) the protons will occupy definite "crystalline positions." Theoretical estimates show that the ordering of protons becomes energetically favorable at temperatures of 60–70 K. Therefore, to convert ice-I_h to a "true crystal," we need to decrease the temperature. However, cooling results in an increase in the characteristic time of proton redistribution (the lower the temperature, the slower the motion of the protons in the oxygen sublattice). In chemically pure ice at 120 K, the protons are redistributed in 10 s, while at 100 K they undergo the same rearrangement in an hour; at 90–95 K it takes a day. At the boiling point of nitrogen (78 K), we would have to wait a year for the ordered hydrogen sublattice to be arranged, while at 70–73 K you would spend your whole life waiting for it to happen! Let's not allow the temperature to decrease to 40–45 K: at this temperature the proton redistribution time equals... the age of the Universe itself ($5 \cdot 10^{17}$ s).

The BFP rules explain many properties of ice. However, let's imagine a crystal that is in strict compliance with these rules. Look carefully at the ice lattice (figure 3a)—can the protons move in it? The answer is no. If proton *A* changes its position as shown in figure 3b, the second principle of the BFP rules will be violated. The motion of proton *B* along the path shown in figure 3b would violate the third principle.

Thus the BFP rules prohibit the motion of the protons in the ice lattice. However, ice is not an ideal dielectric—it has a measurable electric conductivity, and its conductance is protonic. So protons *can* move in ice! Does this mean that the BFP rules are wrong after all? That would be too simple an answer. The correct one is this: they are valid in almost all cases, but it is the exceptions to the rules that explain the wonderful electrical properties of ice.

At the end of the 1950s the Swiss physicist Jacquard proposed a brilliant way of describing these electrical properties in terms of the motion of particular defects that arise when the BFP rules are violated. It's a very unusual mechanism of electric conductance, so let's examine it in some detail.

To begin with, we "spoil" the ice lattice and break the second BFP rule: a proton will be taken from one oxygen atom and added to another (figure 3c). In this way we obtain *ionic defects* (they are denoted as OH^- and H_3O^+). Now we disturb the regular lattice by breaking rule 3 as shown in figure 3d. This results in the appearance of orientation defects (called *L*- and *D*-defects). The defects are always present in the real structure of ice but in very small numbers: there are only about 10^5 orientation defects and a pair of ionic defects among $3 \cdot 10^{11}$ molecules of chemically pure ice at 10°C.

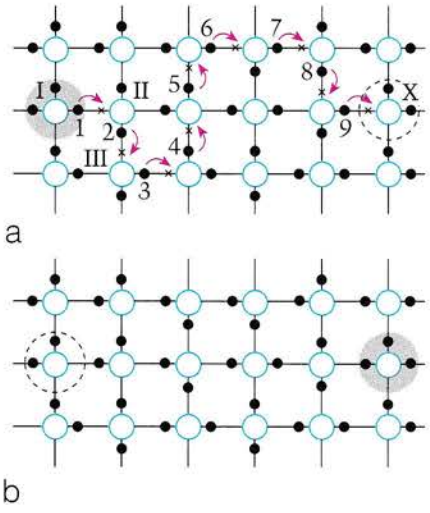


Figure 4

Figure 4a shows how the protons can move. Consider the ionic defect H_3O^+. Proton 1 moves to the place marked by the cross. Previously the H_3O^+ ion was near the oxygen atom I, and now it is at oxygen atom II. Then proton 2 moves to the place also marked with a cross—now the H_3O^+ ion is located near atom III, and so on. We see that different protons move along different segments of the same trajectory, as if they were passing the baton in a relay race. The result looks as if an individual H_3O^+ ion traveled the whole way "by itself." Figure 4b shows the ice lattice after displacement of the H_3O^+ ion from position I to position X.

For clarity, we considered a hypothetical flat, square ice lattice. Every oxygen atom in such a lattice is surrounded by four equivalent oxygen neighbors, just as in a real three-dimensional lattice. Therefore, all the derived inferences are also valid for the real ice lattice. Try to understand the motion of OH^- ions and *L*- and *D*-defects on your own.

It turned out that in some other substances the mechanism of conductivity is just the same. Nowadays even biochemists pay close attention to studies of the electric conductance mechanism in ice, since proton transfer in a number of biological objects is very similar to proton motion in ice.

Although Jacquard's theory explained many phenomena, the amount of new experimental data grows with every passing year, and they demand further elaboration of currently accepted concepts. New and sometimes paradoxical hypotheses are proposed. According to one such bold hypothesis, the "true" carrier of the electric charge in ice are electrons that have "saddled" the proton defects. Whether this is true or not, time will tell.

The volumetric properties of ice are unusual and rather exotic, but the features of its surface are even more remarkable. Wake someone up in the middle of the night and ask at what temperature ice melts. "At 0°C, of course." However, change

the wording slightly: "At what temperature does the surface of ice begin to melt?" Even specialists in the physics of ice won't be able to give you a straight answer.

From the viewpoint of fundamental science, studies of the surface of ice are very interesting. But their practical importance is even more impressive. Indeed, the movement of ice-breaking ships, the de-icing of ships and planes, construction in polar regions, the motion of glaciers: in all these processes it's important to know what takes place on the surface of ice—that is, at the boundary between ice and metal, plastic, various kinds of soil, and so on. It's necessary to know the mechanism of ice adhesion to various building materials, the value and character of friction, and similar characteristics. And the sporting goods industry is forever on the lookout for materials with a minimal coefficient of friction!

By the way, why do many materials slide on ice so easily? Perhaps many readers are ready to answer this seemingly "simple" question. Everybody knows that swift sliding takes place when there is a layer of water between the ice and the object. When the object presses hard on the ice, the ice melts, since melting under pressure occurs at lower temperatures.

Is it really so simple? Let's see. The ice phase diagram says that the melting point of ice decreases by one degree per pressure increment of $\Delta P = 10^7$ Pa. Let the mass of an ice skater be $M = 60$ kg and the area of a skate $S = 3 \cdot 200$ mm$^2 = 6 \cdot 10^{-4}$ m^2. The increase in pressure will be $\Delta P = Mg/S = 10^6$ Pa. Thus the melting point of the ice under the skate will be decreased by only 0.1°C! Can we enjoy skating at –10°C? It seems that we should look for some other reasons for efficient sliding on the ice surface.

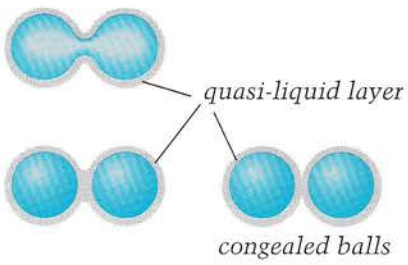


Figure 5

It's thought that at temperatures below the melting point, a thin quasi-liquid layer is formed at the surface of crystalline ice (the prefix "quasi" means "almost" or "sort of"). It's doubtful that this film is just an ordinary liquid. However, in many respects it's similar to water. In some experiments the quasi-liquid layer was detected even at –30°C—that's 30 degrees below the standard melting point! Does the quasi-liquid layer underlie the small value of ice's coefficient of friction? This hypothesis explains many features of the surface of ice, but unfortunately the true nature of these phenomena is far more complicated and may be understood more fully in the future.

By the way, the scientific debate between adherents of the "quasi-liquid layer" and "pressure melting" has gone on for more than a century. Such renowned personalities as Michael Faraday and the Thomson brothers (one of them known worldwide as Lord Kelvin) devoted some of their scientific efforts to studying ice.

No doubt many of our readers have had the pleasure of throwing a snowball. Did you ever wonder why it's so easy to make a snowball? Snow isn't dough or modeling clay, after all—it's small crystals of ice. Try making an iron "snowball" out of metal filings. Ice is much better suited to the task! Can you figure out why? Michael Faraday did, more than a hundred years ago. In 1850 he noticed the following fact: the pieces of ice congeal if they contact each other at a temperature near 0°C. Why? To explain this phenomenon, Faraday suggested the existence of a certain (quasi-liquid) layer on the surface of ice.

Look at figure 5 and you'll see immediately why the existence of a quasi-liquid layer causes two pieces of ice to adhere. The concept of a quasi-liquid layer was also used in attempting to explain the movement of glaciers.

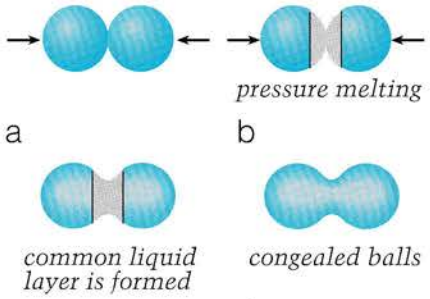


Figure 6

The authority of Faraday was very great, but it didn't stop James Thomson (brother of Lord Kelvin) from suggesting an alternative explanation. He drew attention to the part of the phase diagram that was known at the time and proved theoretically that ice must melt when it is compressed. Therefore, if two pieces of ice are pressed together, they must melt together (figure 6).

Faraday conducted a series of experiments that undercut Thomson's theory, but the controversy was still not settled. The fine theories of Faraday and Thomson are surely based on sound data, but the real mechanism of ice adhesion lies somewhere in between.

In this article we considered only some of the physical properties of ice. These properties affect many atmospheric processes and geophysical phenomena on Earth. Recently it was established that rapid crystallization of ice is accompanied by glowing. Perhaps the northern lights are related to some properties of ice?

Ice is also found widely in outer space. Mars, Jupiter, and Saturn contain huge amounts of ice, and many asteroids and even some natural satellites are made entirely of ice.

Many of the problems discussed here still await a solution. How many types and aspects of ice are still hidden from science? There is no shortage of problems and hypotheses about the many "faces" of ice-I, ice-II, ice-III... It's quite possible that some of our readers will succeed in mapping the unexplored regions of ice physics. Perhaps you'll discover yet another "face" of ice. ■

Challenges

Physics

P326

Cat and two mice. A cat is sunning himself on the roof of a barn, at the very edge. Two nasty mice shot a pebble at him with a slingshot. They missed—after describing a parabola, the pebble recoiled elastically from the inclined roof near the cat's paws. The recoil occurred at t_1 = 1.2 s. After a period t_2 = 1.0 s the pebble hit the paw of the mouse who shot it (figure 1). What was the distance s between the cat and the mice? (D. Aleksandrov and V. Slobodyanin)

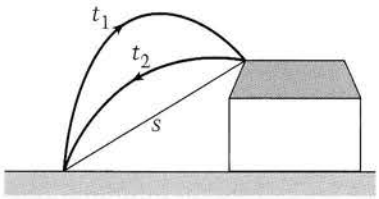


Figure 1

P327

Sinking barge. In the middle of the bottom of a rectangular barge (length a = 80 m, width b = 10 m, and height c = 5 m), a hole with diameter d = 1 cm was punched. Determine the time needed to sink the barge if water is not pumped out of it. The top of the barge is open, the barge is unloaded, and the initial height of the sides above the water's surface was h = 3.75 m.

(S. Varlamov)

P328

Two electric hemispheres. Find the force of interaction between two electrically isolated hemispheres of radii R and r carrying charges Q and q, respectively, uniformly distributed over their surfaces (figure 2). The centers and the maximum cross-sectional planes of the hemispheres coincide. (G. Grigoryan)

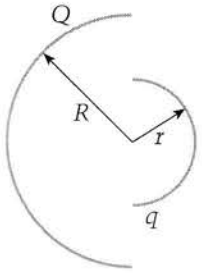


Figure 2

P329

Large capacitor. Estimate the steady-state charge accumulated on the capacitor with capacitance 1000C shown in figure 3. (O. Shvedov)

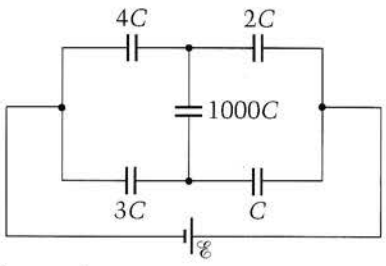


Figure 3

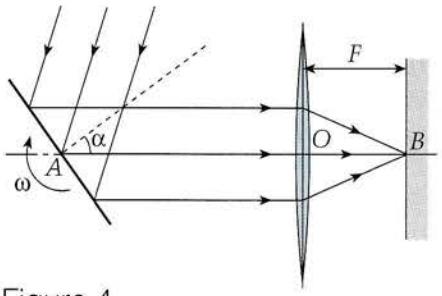


Figure 4

P330

Lens and dancing mirror. A plane mirror rotates with an angular speed ω about an axis that is perpendicular to the optic axis AB of a converging lens with a focal length F (figure 4). A parallel beam of rays hits the mirror, where it is reflected and focused on a screen situated at the focal plane of the lens. Find the speed of the light spot on the screen at the moment it passes the focal plane of the lens. (E. Palchikov)

Math

M325

How elongated is it? We'll say that the "elongation" of a rectangle is the ratio of its longer side to its shorter side. Let a rectangle B be inscribed in a rectangle A so that the vertices of B lie on the sides of A. Prove that the elongation of B is not less than that of A. (N. Vasilyev)

M326

Integers and altitudes. The sides of a triangle are integers x, y, and z. It is given that one of its altitudes is equal to the sum of two others. Prove that $x^2 + y^2 + z^2$ is the square of an integer. (D. Fomin)

M327

Can't be less. Prove that $x \cdot 2^y + y \cdot 2^{-x} \geq x + y$ for any positive numbers x and y.

(N. Vasilyev, V. Prasolov, and V. Senderov)

CONTINUED ON PAGE 15

Stretching exercise

Give your mind a workout

by Donald Barry

THERE ARE TWO MAIN REASONS why textbooks should include challenging problems. The first is that students gain a valuable and authentic sense of accomplishment whenever they conquer something truly difficult. The second is that the more difficult problems have a greater variety of solutions, providing students with the opportunity to experience their own creativity in mathematics. Unfortunately, since present textbooks don't stretch our students, we need to create challenging problem sets.

Here's one such problem along with a solution, eight annotated proofs, and some additional problems. I gave it to a class of strong ninth-grade geometry students, following work with right triangle trigonometry, the Sine and Cosine Laws, inscribed angles, and the Power of a Point Theorem.

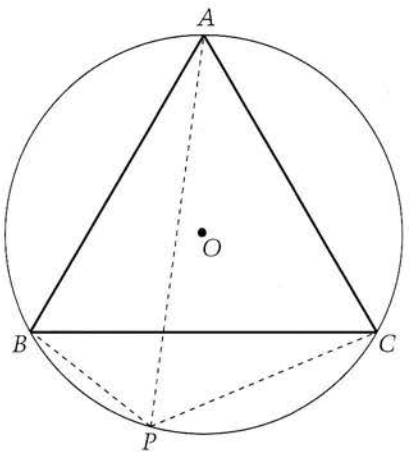


Figure 1

Problem: *Equilateral triangle ABC is inscribed in circle O(see figure 1). Point P lies on* $\overset{\frown}{BC}$ *and PA = 3. Determine the value of PB + PC.*

The students first reactions was that the problem was impossible because neither P's location nor the size of the circle or triangle were given. But when you give students challenging problems, suspicions of old tricks arise, and so it wasn't long before my students guessed that the answer had to be invariant. They slid P until it coincided with B, making $PB = 0$. This made $PC = PA = 3$, and if the answer was invariant then $PB + PC = 3$ as well. But they felt uneasy. They weren't certain that $PB + PC$ always equaled PA, and some of them realized that their special case solution involved changing the size of the figure. We had a double period, and within 25 minutes, Soojin Park and Jen Wong came up with the following fairly complicated proof. Luckily, I'd done the problem the same way the night before so that I could point them in the right direction when they became confused.

Proof 1: A clever student solution.

$$\Delta BDP \sim \Delta ADC \rightarrow \frac{BP}{AC} = \frac{BD}{AD}$$

$$\rightarrow BP = \frac{AC \cdot BD}{AD}.$$

$$\Delta PDC \sim \Delta BDA \rightarrow \frac{PC}{BA} = \frac{DC}{DA}$$

$$\rightarrow PC = \frac{BA \cdot DC}{DA}.$$

Since $AC = BA$ and $AD = DA$ we have:

$$BP + PC = \frac{AC \cdot BD + AC \cdot DC}{AD}$$

$$= \frac{AC(BD + DC)}{AD} = \frac{AC \cdot BC}{AD} = \frac{AC^2}{AD}.$$

This is interesting, suggesting the presence of a geometric mean. Aha, consider:

$$\Delta PAC \sim \Delta CAD \rightarrow \frac{PA}{AC} = \frac{AC}{AD}$$

$$\rightarrow PA = \frac{AC^2}{AD}.$$

Hence, $PB + PC = PA$ (figure 2).

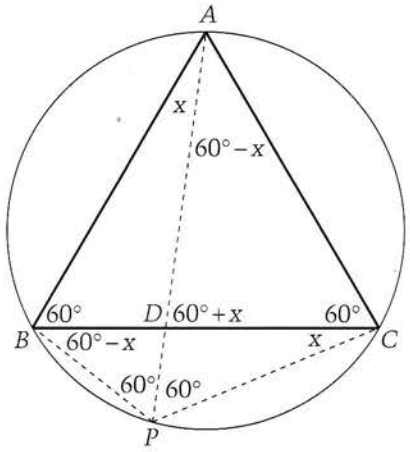


Figure 2

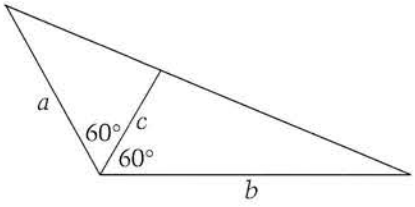


Figure 3

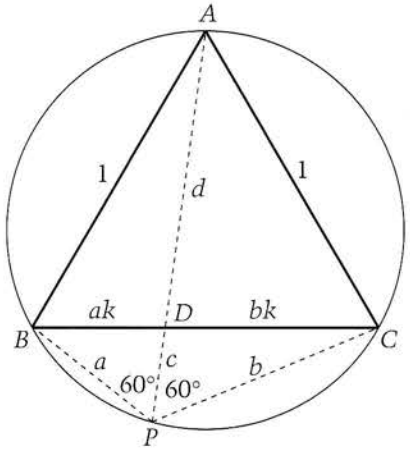


Figure 4

Proof 2: The kitchen sink. For my part, I looked at the problem and felt that surely there ought to be a proof involving the Triangle Angle Bisector Theorem, the Power of a Point Theorem, and perhaps the invariance involving the bisector of a 120° angle. It took a while but I finally put them all together in the following proof, which reveals facets of the problem that no other method illuminates.

First, given figure 3, it is true that

$$\frac{1}{a}+\frac{1}{b}=\frac{1}{c}.$$

This gives

$$a+b=\frac{ab}{c}.$$

Second, referring to Figure 4, we know that

$$\frac{PB}{PC}=\frac{BD}{DC}$$

by the Triangle Angle Bisector Theorem. Thus, let $BD = ak$ and $DC = bk$.

Third, let $PD = c$ and $DA = d$. Then, since the products of the segments of concurrent chords are equal (the Power of a Point Theorem), we have $(ak)(bk) = cd$, giving $(abk)k = cd$.

Fourth, without loss of generality, let the sides of equilateral triangle ABC be 1. Since $\triangle ADC \sim \triangle BDP$,

$$\frac{DC}{DP}=\frac{AC}{BP}\rightarrow\frac{bk}{c}=\frac{1}{a}\rightarrow abk=c.$$

From $(abk)k = cd$ we now have $k = d$, giving $abd = c$ or

$$\frac{ab}{c}=\frac{1}{d}.$$

Drawing upon the fact that

$$a+b=\frac{ab}{c},$$

we now have

$$a+b=\frac{1}{d}.$$

That's interesting, a theorem in its own right.

Fifth and last, since $\triangle ADC \sim \triangle ACP$, we have

$$\frac{AC}{AD}=\frac{AP}{AC}\rightarrow\frac{1}{d}=\frac{c+d}{1}.$$

Thus, both $c + d$ and $a + b$ equal $1/d$, and so they are equal. If the side of ABC equals m, then

$$c+d=a+b=\frac{m^2}{d},$$

not as pretty a result but still nifty.

Working independently, Ben Bloom, a ninth grader, developed a proof much like mine. He didn't know that $1/a + 1/b = 1/c$ and didn't use it, but he did discover that if the side of the triangle is 1, then both PA and $PB + PC$ equal the reciprocal of AD. His proof is shorter than mine, and I'm grateful for having learned something.

Proof 3: Create a new equilateral triangle. One of my colleagues, Bill Scott, was so excited by the proof he

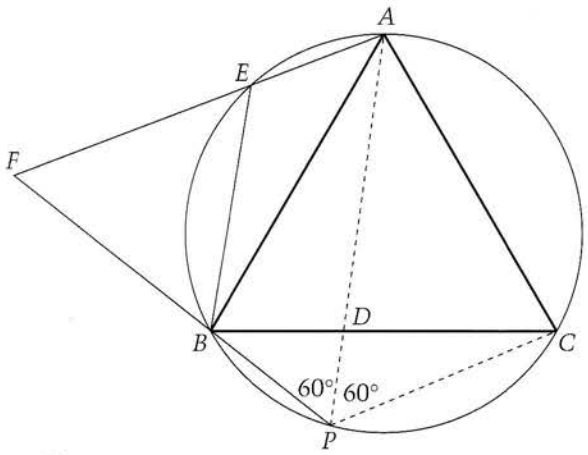


Figure 5

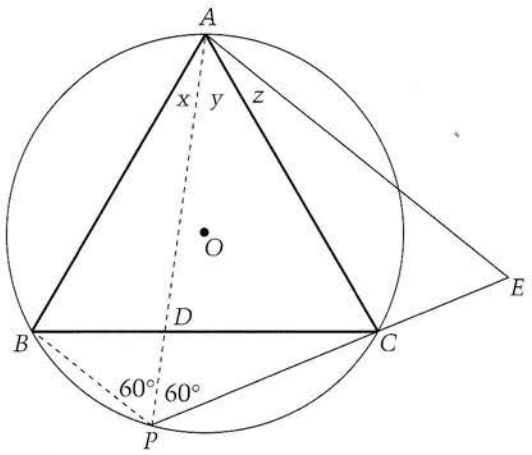


Figure 6

discovered that he burst into my Analytic Geometry class to share the proof right then and there in front of all my students. I loved the interruption; I think it is important for students to see our passion and excitement for mathematics, and Bill more than filled that bill. I also like his proof because it is very visually satisfying.

In figure 5 mark off chord $\overline{AE}$ equal to chord $\overline{BP}$, making $\overset{\frown}{AE} \cong \overset{\frown}{BP}$. Since $\overset{\frown}{AB} \cong \overset{\frown}{BC}$ we have $\overset{\frown}{BE} \cong \overset{\frown}{PC}$ and so $BE = PC$. Extend both $\overline{AE}$ and $\overline{BP}$ until they meet at F. Since $EAPB$ is an isosceles trapezoid and $\angle BPA = 60°$, then $\angle PAF = 60°$ and triangle FPA is an equilateral triangle with $FP = PA$. But $\overline{BE}$ is parallel to $\overline{AP}$, making triangle FBE an equilateral triangle. Thus, $FB = BE$, and since we showed that $BE = PC$ we know that $FB + BP = PC + BP$. Since $FB + BP = PA$ then $PB + PC = PA$.

Proof 4: A second equilateral triangle construction. Extend $\overline{PC}$ to E (figure 6) so that $PA = PE$. Since $\angle APC = 60°$ then $\triangle PAE$ is equilateral. If $\angle BAP = x$, $\angle PAC = y$, and $\angle CAE = z$, then $x + y = 60°$ and $y + z = 60°$, making $x = z \rightarrow \angle BAP = \angle CAE$. Since $\angle BPA = \angle E = 60°$ and $AB = AC$, then $\triangle PBA \sim \triangle ECA$ by AAS, giving $PB = CE$. Thus, $PB + PC = CE + PC = PE$. Since $PE = PA$, then $PB + PC = PA$.

Proof 5: Law of Cosines. I suggested that the students look for a trigonometric solution. As an aside I should say that I think our current trigonometry textbooks miss all

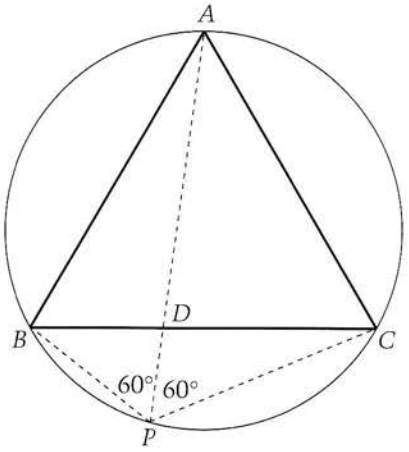


Figure 7

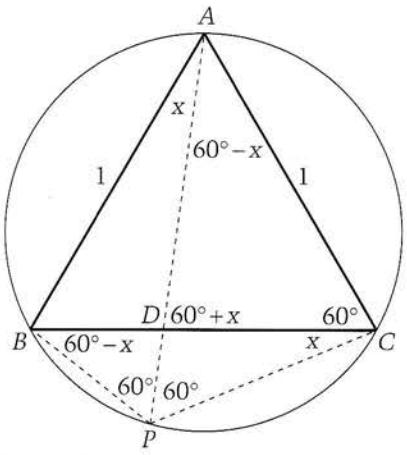


Figure 8

sorts of opportunities to tie trigonometry to the previous work students do in geometry. However, no one in the class found a trigonometric solution. They ran into roadblocks when they tried to find PC using the Law of Cosines rather than finding setting AB and AC equal, as in the following solution (figure 7):

$$AB^2 = PB^2 + PA^2 - 2(PB)(PA)\cos 60°$$

gives $AB^2 = PB^2 + PA^2 - (PB)(PA)$. Likewise: $AC^2 = PC^2 + PA^2 - (PC)(PA)$. Since $AB = AC$

$$PB^2 + PA^2 - (PB)(PA) = PC^2 + PA^2 - (PC)(PA),$$

$$PB^2 - PC^2 = (PB)(PA) - (PC)(PA),$$

$$(PB + PC)(PB - PC) = PA(PB - PC).$$

Therefore, $PB + PC = PA$.

Proof 6: Law of Sines and Summation formula. No one tried to use either Law of Sines or the summation formula for sines, but such an approach was, in fact, quite accessible as the following proof shows:

Without loss of generality let the side of the triangle be 1(figure 8).

Using $\triangle APC$:

$$\frac{1}{\sin 60°} = \frac{PC}{\sin(60° - x)}$$

$$\rightarrow \sin 60°\cos x - \cos 60°\sin x = (\sin 60°)PC.$$

Thus,

$$\frac{\sqrt{3}}{2}\cos x - \frac{1}{2}\sin x = \frac{\sqrt{3}}{2}PC \rightarrow$$

$$\sqrt{3}\cos x - \sin x = \sqrt{3}(PC).$$

Using $\triangle ABP$:

$$\frac{1}{\sin 60°} = \frac{PB}{\sin x} \rightarrow 2\sin x = \sqrt{3}(PB).$$

Using $\triangle APC$:

$$\frac{1}{\sin 60°} = \frac{PA}{\sin(60° + x)}$$

$$\rightarrow \sin 60°\cos x + \cos 60°\sin x = (\sin 60°)PA.$$

Thus,

$$\frac{\sqrt{3}}{2}\cos x + \frac{1}{2}\sin x = \frac{\sqrt{3}}{2}PA$$

$$\rightarrow \sqrt{3}\cos x + \sin x = \left(\sqrt{3}\right)PA.$$

Clearly,

$$\sqrt{3}(PB) + \sqrt{3}(PC) = \sqrt{3}\cos x + \sin x = \sqrt{3}(PA),$$

so $PA = PB + PC$.

Proof 7: A third equilateral triangle construction. No one discovered the following proof, although we were beginning to nudge around it. I love this one because it extends our understanding of auxiliary lines; they can be more than parallel or perpendicular lines.

Construct an equilateral triangle (figure 9): draw CE so that $\angle PCE = 60°$. If we let $\angle DCP = x$, then $\angle DCE = 60° - x$, but since $\angle ACD = 60°$, that makes $\angle ACE = x$, so $\angle DCP = \angle ACE$. Also, $\angle BPC = 120° = \angle AEC$, and $AC = BC$, so by AAS, $\triangle ACE \sim \triangle BCP$ and $AE = BP$. Since $\triangle ECP$ is equilateral, $PC = PE$, so $PB + PC = AE + PE = PA$.

Proof 8: Ptolemy's theorem. Our discussion of the equilateral prob-

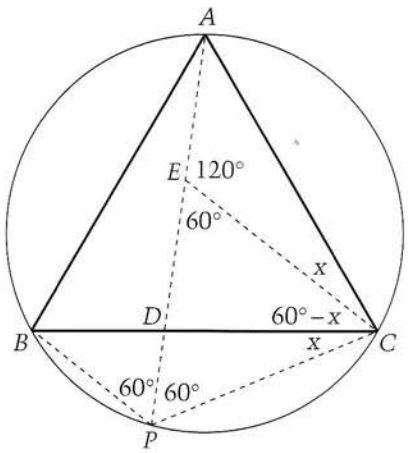


Figure 9

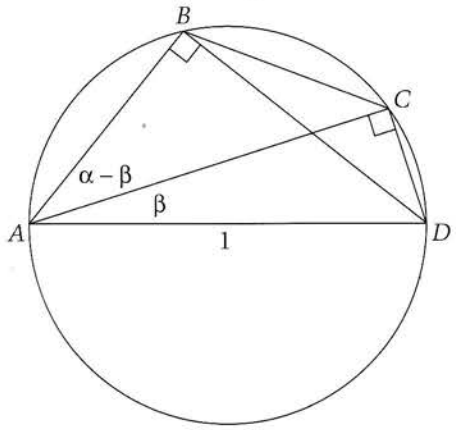


Figure 10

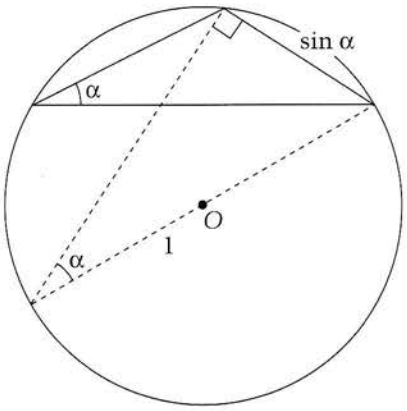


Figure 11

lem paved the way nicely for a discussion of Ptolemy's theorem, and, additionally, proof 7 primed my students for proof 8.

Ptolemy's theorem states that if $ABCD$ is inscribed in a circle, then the sum of the product of the opposite sides equals the product of the diagonals; i.e.

$$AB \cdot DC + AD \cdot BC = AC \cdot DB.$$

We apply it in this problem using quadrilateral $ACPB$:

$$AC \cdot PB + AB \cdot PC = AP \cdot BC.$$

Since $AC = AB = BC$, we divide to obtain $PB + PC = PA$. This is the simplest proof, and suddenly it became clear that the problem we'd started with was just a special case of Ptolemy's theorem.

Ptolemy's theorem and the construction of trigonometric tables. Once we had Ptolemy's theorem, I could return to a previous discussion of the origins of trigonometric tables and demonstrate how this result could have been used to develop the difference formula for sines.

In Figure 10, let AD, a segment of unit length, be a diameter of the given circle. In such a circle, it is not difficult to see that a chord subtended by an angle of measure α has length $\sin \alpha$.[1] Thus $BC = \sin(\alpha - \beta)$, $CD = \sin \beta$, $BD = \sin \alpha$, and (from right triangles ABD, ACD) $AB = \cos \alpha$, $AC = \cos \beta$. Substituting into $AB \cdot CD + BC \cdot AD = BC \cdot AC$, we find that $\cos \alpha \sin \beta + 1 \cdot \sin(\alpha - \beta) = \sin \alpha \cos \beta$, and we can solve for $\sin(\alpha - \beta)$.

With this result, we can now construct tables of the values of trigonometric functions. For example, using special triangles, one can find exact values for 15° and 18° angles, and the formula then yields an exact value for a 3° angle. The half angle formula then gets exact trigonometric values for the 1.5° and .75° angles, and one is well on the way to constructing accurate trigonometric tables.

Invariance. It should be clear that the original equilateral triangle problem was not a dead-end—it led us on a merry search for proofs and then back to the origins of trigonometric tables. In general, challenging problems do just that—they are fertile and productive, they reveal connections, suggest questions, pique interest, and demonstrate the creative potential of mathematics. I have found that geometry problems that involve an invariant relationship are a particularly good source of such materials. Initially, they catch students' attention because they are provocative—they appear to be unsolvable. Yet by using either a special case or an extreme case, students can often obtain a numerical answer, but they don't trust their work, and they are keen to solve the general problem to know for sure that they are correct. Furthermore, once the general problem has been solved, they've often discovered a significant result.

I'd like to close this article with three problems that involve invariants. The first leads to the Parallelogram Law, the second raises all sorts of questions, and the third may be familiar.

Problem 1. If $ABCD$ is a parallelogram with $AB = 7$ and $BC = 24$, determine the value of $AC^2 + BD^2$.

Problem 2. If $ABCD$ is a rectangle and point P is somewhere in space such that $PA = 9$, $PB = 7$, and $PC = 2$, find PD. Is there an analogous result for trapezoids, for just isosceles trapezoids, or for hexagons?

Problem 3. In $\triangle ABC$, $\angle ABC = 120°$, and D lies on $\overline{AC}$ so that $\overline{BD}$ bisects ABC.

a) If $DB = 2$, find the value of

$$\frac{1}{AB} + \frac{1}{BC}.$$

b) If $DB = 12$, find all pairs of integral values for AB and BC. ■

CONTINUED FROM PAGE 11

M328

Numerous nines. The sequence $\{a_n\}$ is defined by the following rules: $a_0 = 9$, $a_{k+1} = 3a_k^4 + 4a_k^3$ for any $k > 0$. Prove that a_{10} contains more than 1,000 instances of the digit 9 in decimal notation.

(N. Vyalyi)

M329

Perfect chromatic balance. Every face of a convex polyhedron is a polygon with an even number of sides. Is it always true that its edges can be colored in two different colors so that each face has an equal number of edges of every color?

(S. Tokarev)

ANSWERS, HINTS & SOLUTIONS ON PAGE 52

[1] Indeed, if one side of the angle is a diameter, this result follows from the trigonometry of the right triangle. If not, then we can construct an inscribed angle subtending the same arc for which one side is a diameter, and which has the same measure (figure 11).

Many happy returns

The tricky business of returning from space

by Albert Stasenko

"HURRAH!" CRIED THE INhabitants of Havr, crowded on every Havrian embankment.... A black body splashed into the bay. In the middle of it were three floundering men.

"We had no food for fifty-seven days!" mumbled Mr. Lund, who was skinny as a starving artist, and he explained what happened.

—Anton Chekhov,
"The Flying Isles"

Strange as it may seem, returning to Earth is by no means a simpler task than space travel in a rocket. You'll know the answer to this question: How much energy is needed to lift a load (you, for example) h meters? Yes: mgh. Accordingly, this potential energy will be transformed into kinetic energy and finally into heat during and after a fall from this same height h. How can one use this energy if the altitude h equals, say, the height of a 10-floor building? Our aim is to make a soft landing with almost zero speed. This means that the kinetic energy that is continually gained from the loss of potential energy must be dissipated in one way or another—for example, to overcome the frictional forces of a rope hanging from the tenth floor.

In the same way, a spacecraft that must make a soft landing on the Moon or a planet without atmosphere (Mercury or Mars) must decrease its speed, and its corresponding kinetic energy, by firing its engines to oppose the spacecraft's motion.

However, this reasoning is correct only if we heed the advice so often given in physics textbooks: "neglect air resistance." In reality, no experienced spacecraft designer neglects it: many hovering devices—parachutes, gliders, and airplanes with their engines turned off, which for better or worse fly like gliders—land without using an ounce of fuel. Their speed is decreased by aerodynamic resistance—that is, by the force acting on the vehicle due to the air molecules. Can one use the force of the Earth's atmosphere to make a landing from outer space "free of charge"? It's not as trivial a problem as it may seem at first glance. Indeed, recall the tragic destiny of most meteorites, which cannot land safely! Only miserable remnants manage to reach the Earth's surface. We, on the other hand, want to land in one piece, without giving up any of our mass or expending any energy.

Can we do it?

To answer this question, let's come up with a plan and break the problem down into manageable pieces. We'll recall how air resistance—the aerodynamic force—depends on atmospheric density and the object's speed; and how an object heats up and cools down as it moves in the atmosphere. Finally we'll take the last step and draw our conclusion.

Atmospheric density

First of all we'll consider the atmosphere on Earth. In this limited space we can't hope to cover every sort of atmospheric phenomenon—winds, thunderstorms, typhoons, clouds, and so on. There are entire books devoted to these topics. However, we can do what physicists do whenever things get too complicated (unfortunately, only complicated processes are left, because the simple ones were understood long before modern physics was born): we'll construct a mathematical model of the atmosphere. A good model should describe the most salient features of the phenomenon, or aspects that we'll use later in our investigations. What is most characteristic about the atmosphere? Its

Art by Vasily Vlasov


B
A
D
F
C
E

density, for one thing—after all, we'll be examining rapid motions in the atmosphere, and intuition clearly tells us that the force of air resistance should depend on the medium's density. Indeed, it's much harder to move our hand quickly in water than in air.

We know that the density of the atmosphere must decrease with altitude. The dependence of density on the height above sea level is approximately described by the Boltzmann barometric formula:

$$\rho \cong \rho_0 e^{-h/h^*}. \qquad (1)$$

This can be used for altitudes up to $h \cong 80$–100 km. In this formula $\rho_0 = 1.3$ kg/m^3 is the density of air at sea level. The value $h^* = 7.16$ km in the denominator of the exponent is the density scaling coefficient. Clearly the atmospheric density at height h^* is less than that at sea level by a factor of $e = 2.72$.

Aerodynamic force

A moving object is affected by the surrounding medium—in this case, air. You can feel the resistive force of air by putting your hand out the window of a moving car. Turn your hand about the horizontal axis—the drag force presses it either upward (when the wind hits your palm) or downward (when the wind hits the back of your hand). In the first case the angle of attack (between the plane of your palm and the velocity **v** of the incident wind) is considered positive; in the second case it's negative (figure 1).

What does this force depend on? We'll use dimensional analysis to help us find the answer. First let's write down the characteristics of the processes that are supposed to determine the value of this force.

Certainly it must depend on speed v—you know that quite different forces affect your outstretched hand when you're moving up an escalator or rushing along in a car. Now, if you move your hand with the same speed in air and in water, the resistive force is greater in water. Thus the force of resistance depends also on the density ρ of the surrounding medium. Have we missed some important factor? It seems so. Try waving your hand in the air with the same speed with and without a fan—it doesn't take long to realize it's harder with the fan. Therefore, the aerodynamic force also depends on the characteristic size L of the moving object.

Now let's write down these variables and their dimensions:

$$[v] = \text{m/s}, [F] = \text{kg/m}^3, [L] = \text{m}.$$

The next step is to compose a combination of these variables that has the dimensions of force—that is, the newton: $[F] = N = \text{kg} \cdot \text{m/s}^2$. Clearly the unit of mass [kg] is present only in the density ρ, and the unit of time [s] is present only in the velocity v (here it is to the first power, but we need it to the second power). Therefore, the force will surely be proportional to the product ρv^2. However, the dimensions for this combination is $\text{kg/s}^2 \cdot \text{m}$, but we need [m] in the denominator. So we need to multiply our formula by the square of the length (which has the dimensions of area). Thus the only combination of the parameters v, ρ, L that results in the dimensions of force is $F \sim \rho v^2 L^2$. Of course, dimensional analysis can't provide us with a dimensionless factor of proportionality.

"Wait a minute!" the thoughtful reader will exclaim. "What about the viscosity of air—the resistive force certainly depends on that!" Well, this is true. Not only that, in many cases viscosity provides the largest contribution to the resistive force: for example, a pellet sinking in honey is affected by the Stokes force, which is proportional not to area but to the linear radius of the sphere and also to its velocity—but this time to the first power only. The flow of honey in this case is so slow, it's called "creeping." Such motion in a thick viscous medium has little to do with the flight of large aircraft at supersonic speeds.

A similar phenomenon—that is, the radical influence of an object's size on the character of its motion (the scaling effect)—can be observed in other cases. Consider, for example, a steel needle, which can float on the surface of water. Can an iron crowbar float in the same way? Obviously not. In the case of a needle the lifting force is the surface tension, which is proportional to the first power of an object's size. Another force, gravity, which is proportional to the third power of an object's size, plays the key role in the experiment with the crowbar, and this force cannot be counterbalanced either by surface tension or by buoyancy.

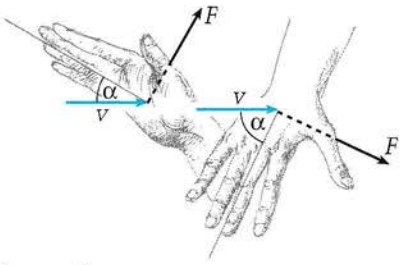


Figure 1

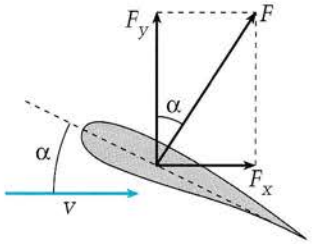


Figure 2

Let's break down the net aerodynamic force **F** affecting a streamlined body into two components: F_y, which is perpendicular to the velocity vector v, and F_x, which is directed along this vector (figure 2). The first component (F_y) is lift, and the second one is drag. Clearly they have the same dimensions (newtons, N), so $F_y \sim \rho v^2 L^2$ and $F_x \sim \rho v^2 L^2$. We can guess that some dimensionless coefficients must be inserted into these formulas, which depend at least on the angle of attack. Indeed, put a wing or your palm perpendicular to the velocity v—you'll get drag only and no lift. We'll call the respective dimensionless values the lift coefficient (C_y) and the drag coefficient (C_x). Now we have

$$F_y = C_y \rho v^2 L^2, \; F_x = C_x \rho v^2 L^2. \qquad (2)$$

As we mentioned, these aerodynamic coefficients, C_x and C_y, de-

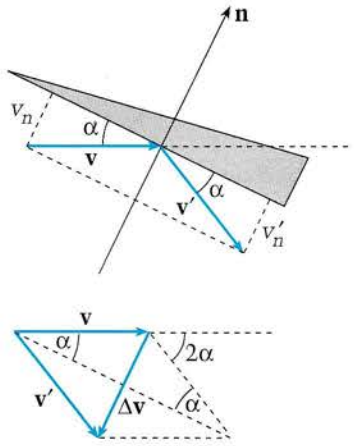


Figure 3

pend on the angle of attack. How can we know what these dependencies are? In some cases they can be calculated. For example, calculations can be made for the case of an aircraft flying at a very high altitude where the air is thin, so the air molecules don't collide with one another but collide only with the aircraft. This approach greatly simplifies the calculations.

Now let's consider motion in the reference frame fixed to an object that flies in the atmosphere with a velocity **u**. In this frame the object is at rest and the air molecules fly into it. We assume that u is much larger than the thermal velocity of the air molecules; therefore, we can neglect the stochastic thermal motion of the molecules and assume that all of them move toward the body with velocity $\mathbf{v} = -\mathbf{u}$.

As a first step we consider the collision of a single molecule with the hard surface of an object inclined to the velocity vector by an angle of attack α. Assuming the collision to be absolutely elastic, we find that the speed does not change after the impact, so the vector **v** turns through the angle 2α (figure 3). It's not difficult to find the resulting change in the velocity of the molecule (see figure 3):

$$\Delta v = 2v_n = 2v \sin \alpha.$$

The corresponding change in the momentum of a molecule with mass m is

$$\Delta p_m = 2mv \sin \alpha.$$

According to momentum conservation and Newton's third law, the same momentum will be imparted to the body. How many molecules are incident on a unit area per unit time? In a unit of time the area S (figure 4) will be struck by all the molecules that are located at a distance v from it—that is, by molecules located in a volume $Sv \sin \alpha$. If n is the concentration of the molecules, $N = nSv \sin \alpha$ molecules will strike the area S per unit time.

Thus the net momentum imparted to the object per unit time is directed perpendicular to the surface and is equal to

$$\Delta p_m \cdot N = 2mv \sin \alpha \cdot nSv \sin \alpha = 2\rho v^2 S \sin^2 \alpha,$$

since $nm = \rho$. Recall that the rate of change of momentum is equivalent to the force acting on the object. Therefore, we have obtained the formula for the aerodynamic force:

$$F = 2\rho v^2 S \sin^2 \alpha.$$

The next step is to break this force down into two components (lift and drag forces):

$$F_y = F \cos \alpha = 2\rho v^2 S \sin^2 \alpha \cos \alpha,$$

$$F_x = F \sin \alpha = 2\rho v^2 S \sin^3 \alpha. \qquad (3)$$

The dependencies of F_y and F_x on the angle of attack α are shown in figure 5. We can see that lift is maximum at some angle of attack α_0. Its

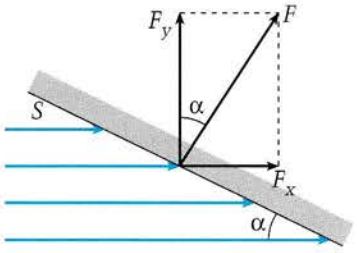


Figure 4

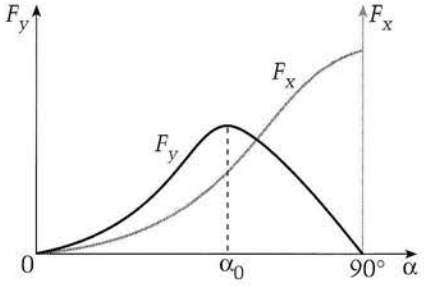


Figure 5

value can be found by plotting the function $F_y(\alpha)$ point by point, using a calculator that provides the values of trigonometric functions. Those who know how to find the extremum of a function can determine the maximum by setting the first derivative equal to zero. Both methods yield

$$\tan \alpha_0 = \sqrt{2}.$$

Since

$$\cos \alpha \frac{1}{\sqrt{1+\tan^2 \alpha}}$$

and

$$\sin \alpha = \frac{\tan \alpha}{\sqrt{1+\tan^2 \alpha}},$$

we obtain the maximum lift by inserting the value $\tan \alpha_0 = \sqrt{2}$ into these formulas:

$$F_{y\max} = \frac{4}{3\sqrt{3}} \rho v^2 S = 0.77 \rho v^2 S.$$

In cases where the aerodynamic coefficients cannot be calculated, their dependence on the angle of attack must be studied experimentally. Hundreds of wind tunnels have been constructed all over the world for this purpose. Even in these cases dimensional analysis has saved huge amounts of money and effort, because we know beforehand that the force is proportional to the medium's density, the area of the object's surface, and the square of the speed.

Shock wave

Until now we've considered a flying object in the form of a plate of area S making an angle α with the direction of the flow of molecules. The velocities of the molecules were assumed to be equal in magnitude and direction, and the object's speed was considered far greater than the speed of thermal motion of the molecules. We weren't interested in what was going on behind the streamlined body: the molecules never got there, so this space could be filled by, say, a wedge with the same angle α at its vertex. Now let this wedge sink into denser and denser atmospheric layers and thus

lose speed, which is still greater than the mean thermal speed of the molecules and consequently greater than the speed of sound. Indeed, such an organized process as sound cannot travel faster than the molecules themselves, which move at thermal speed.

In order to make the braking more pronounced, let's turn the plane of the wedge to form a larger angle with the direction of flow (remember, the drag force F_x is proportional to $\sin^3 \alpha$). After this angle is increased, the molecules will rebound from the front of the wedge at an angle approximating a right angle, so a compressed layer of air (shock wave) will be formed at the front of the wedge. When the plane is positioned perpendicular to the flow, the molecules striking the wedge far from its edge won't be able to get away from it—they'll be packed in like sardines! Arriving molecules, which "know" nothing about this obstacle (because the speed of the flow is greater than the speed of sound), will plunge into this overcrowded layer. A compressed layer will be formed at plane S, separated from the undisturbed gas by a *normal shock wave* (figure 6). This shock wave is characterized by an abrupt change in the speed of the flow—from supersonic to subsonic.

What's going on with the temperature? Let's estimate it. Consider an object moving very quickly with a speed v (more precisely, with a speed $u = v$) 20 times the speed of sound c. In other words, the ratio of the speed of the flow v to the speed of sound c is $M = 20$ (this ratio is called the Mach number). Each molecule in the undisturbed incident flow has a kinetic energy that is approximately equal to

$$\frac{mv^2}{2} = \frac{mc^2}{2}M^2.$$

Inside the compressed layer the molecules are almost motionless with respect to the streamlined object; therefore, their kinetic energy is transformed almost entirely into heat. To estimate the temperature T_1 in the shock wave, we use the following approximate equation (for the case of diatomic air molecules):

$$\frac{mc^2}{2}M^2 \approx \frac{5}{2}kT_1,$$

from which we get

$$T_1 \approx \frac{mc^2M^2}{5k} = \frac{mN_Ac^2M^2}{5R} = \frac{\mu_{air}c^2M^2}{5R}.$$

Taking the molar mass of the air to be $\mu_{air} = 29 \cdot 10^{-3}$ kg/mole, and $c \cong 300$ m/s, we have

$$T_1 \approx \frac{29 \cdot 10^{-3} \cdot (300)^2 \cdot (20)^2}{5 \cdot 8.31}\text{K} \approx 25{,}000\text{ K}.$$

This is an amazing result: the temperature in a shock wave traveling with a speed of 20M is higher than the temperature at the Sun's surface! At this temperature the molecules will disintegrate, dissociating into individual atoms. What else does this mean? Let's compare the kinetic energy of a molecule with its ionization energy. Reference books give this value per molecule of a given sort (for example, the ionization energy of nitrogen, the main component of the Earth's atmosphere, is $\phi = 2.5 \cdot 10^{-18}$ J). Now let's calculate the kinetic energy of an incident nitrogen molecule:

$$E = \frac{m_{N_2}c^2}{2}M^2 = \frac{1}{2} \cdot 4.7 \cdot 10^{-26} \cdot (300)^2 \cdot (20)^2 \text{ J} \approx 0.85 \cdot 10^{-18} \text{ J}.$$

We see that E and ϕ are of the same order of magnitude, so we can expect to find free electrons and ions in the shock wave.

Clearly dissociation and atomic ionization will consume some of the primary kinetic energy of the incident molecules, so the temperature in the shock wave will be somewhat less than that provided by our rough estimate. Nevertheless, it will be of the order of a thousand degrees.

Although we considered the formation of a shock wave at a plane, which we gradually oriented at a right angle to the flow, our reasoning is valid for any moving object that has a portion of its surface oriented perpendicular to the flow. In aerodynamics such objects are called "blunt." However, the existence of such a plane isn't enough: to produce a shock wave similar to what was just described, it's necessary to have a rather large radius of curvature of the object's surface near the front—it must be greater than the thickness of the shock wave. In this case the shock wave will be a line stretching some distance across the flow. An example is shown in figure 7, where a sphere is placed in a supersonic flow. We can see that near the sphere's leading point (the braking point), the shock wave is almost flat (this region is enclosed by the dashed rectangle). Such objects might be offended by the name given them: "obtuse." Note that from the aerodynamic point of view, any human being (even a very clever one) is just an obtuse object.

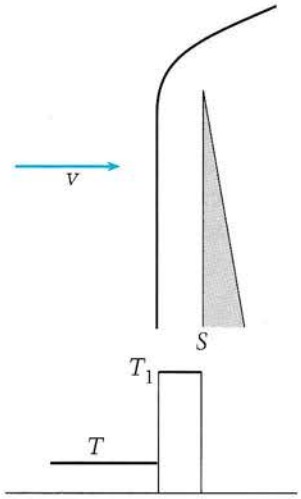


Figure 6

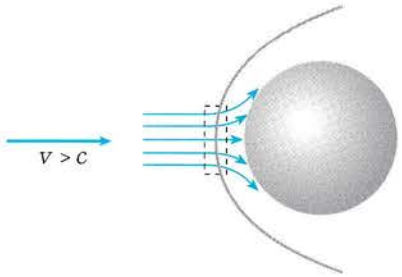


Figure 7

In short, a compressed and strongly heated gas layer (shock wave) is formed in front of any blunt or obtuse object in a supersonic flow.

Conclusion: the reentry corridor

At the beginning of this article we discussed the problem of a soft landing on a planet without an atmosphere or with a very thin one. To accomplish this, almost the same amount of energy must be expended as is needed for taking off from such a planet. However, if the atmosphere is rather dense, this energy (and precious fuel) can be saved by using the drag force F_x for braking and the lift force F_y for supporting the vehicle during reentry. But what does it mean to have a sufficiently thick atmosphere? Is atmospheric density the only parameter to be taken into consideration? What about the planet's size, its mass (these parameters determine the acceleration due to gravity at the planet's surface). What are the effects of the atmosphere's thickness and composition? And how do the characteristics of the reentry vehicle affect the process of a soft landing?

We'll consider these questions one by one.

Let a vehicle of mass M and wing area S move with speed v in the Earth's atmosphere at an altitude h, where the atmospheric density is ρ. The lift is $F_y = C_y\rho v^2 S$. Flying by the planet, the vehicle moves along a curved trajectory. Simplifying, we'll assume that this trajectory is close to being a circle of radius $R + h$ (R is the Earth's radius). Then the centripetal acceleration of the vehicle will be $v^2/(R + h)$. The value of this acceleration is determined by the force of gravity and the lift force F_y:

$$Mg - F_y = \frac{Mv^2}{R+h}.$$

Taking into account that the thickness of the atmosphere is small compared to the Earth's radius, we rewrite the last equation in the form of

$$C_y\rho v^2 S \approx M\left(g - \frac{v^2}{R}\right).$$

Accordingly, at very high altitudes, where the density of the atmosphere and lift are almost zero, the right-hand side of the equation is zero, so

$$v_1 = \sqrt{Rg} = 7.8 \text{ km/s},$$

the orbital velocity.

What will happen during the descent of the reentry vehicle? The velocity will decrease due to the drag force F_x. As the altitude decreases, the density of the atmosphere increases very rapidly (see formula (1)). Therefore, notwithstanding the decrease in speed, the lift force F_y increases. This force can be used to slow the "fall" of the vehicle, which actually becomes a glider. In its flight the force of gravity is counterbalanced by the lift force:

$$Mg = C_y\rho v^2 S.$$

The wings of a space vehicle can be used at altitudes where the atmospheric density is larger than

$$\rho_{\min} \approx \frac{Mg/S}{C_y v^2}$$

(the numerator in this fraction—that is, the force of gravity per unit area of the wing—is called the *net wing loading*, so we wrote the fraction in this particular way). Formula (1) gives the limiting altitude above which a vehicle cannot be supported by wings at a given speed v:

$$\frac{Mg/S}{C_y v^2} \approx \rho_0 e^{-\frac{h_{\max}}{h^*}}$$

$$\Rightarrow \frac{h_{\max}}{h^*} \approx \ln\frac{\rho_0 C_y v^2}{Mg/S},$$

from which we get

$$h_{\max}(v) \approx h^* \ln\frac{\rho_0 C_y v^2}{Mg/S}.$$

In other words, if we expect the vehicle not to "drop" at a particular altitude h, its speed must be greater than

$$v_{\min}(h) \approx \sqrt{\frac{Mg/S}{C_y\rho_0} e^{\frac{h}{h^*}}}.$$

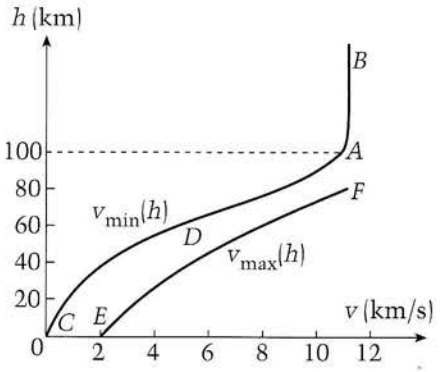


Figure 8

Figure 8 shows the plot $BADC$ corresponding to the reentry of a space vehicle into the Earth's atmosphere with the escape speed v_e = 11.2 km/s. The formula we just wrote describes the portion CD corresponding to flight in the lower atmosphere.

Thus we obtained the curve for the altitude above which a vehicle cannot be supported by the atmosphere: at any given speed (less than the orbital speed), the lift force will be smaller than needed if we ascend higher than this limiting altitude. Therefore, the region of "forbidden" heights is shown as slashes above the curve.

Perhaps we should fly faster at high altitudes? Careful! At high speeds a shock wave will "attach" itself to the blunt edges of the vehicle, and the air behind it (not even the air, but a mixture of its molecular fragments) will be heated so much that the reentry vehicle might be burnt to a crisp like a meteorite. This is a problem that complicates reentry into the atmosphere: the heat barrier. How can this heat be removed? Every possible way must be used—the thermal conductivity of the vehicle itself, which lets the heat flow from "stem to stern"; partial melting of the hull (which modifies its shape, making the vehicle more "obtuse"); thermal radiation of "white-hot" incandescent parts; and so on.

We won't discuss the heat barrier problem in detail, but in order to estimate the allowable speed of

CONTINUED ON PAGE 27

Elementary functions

Definitions from two perspectives

by A. Veselov and S. Gindikin

IN THIS ARTICLE WE DISCUSS exponential, logarithmic, and trigonometric functions. These functions belong to the class of *elementary* functions. This class also includes the linear function $y = ax + b$, the power function $y = x^n$, and various combinations of these functions—sum, difference, product, quotient, and composition. These functions are called "elementary" because they were the first functions that were studied in mathematics, and for a long time the arsenal of mathematicians didn't extend beyond them.

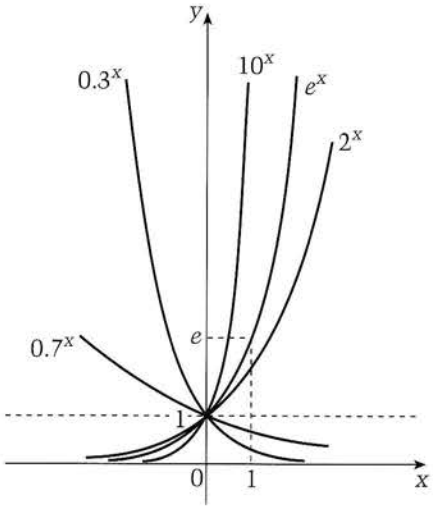


Figure 1. *Graphs of several exponential functions.*

We'll consider two basic approaches to a definition of the exponential, logarithmic, and trigonometric functions: axiomatic and kinematic. We'll see that both approaches lead to the same functions, and we'll discuss how they relate to the "textbook" definitions.

Axiomatic definition of the exponential function

The exponential function $f(x) = a^x$ is a function that has the following properties:

1. f is defined for all x;
2. $f(1) = a > 0$ $(a \neq 1)$;
3. for all x, y the relation

$$f(x + y) = f(x)f(y);$$

holds;

4. $f(x)$ is a monotonic function.

This method of defining an object by listing its properties is called *axiomatic*. It's important to understand that the axiomatic definition of the exponential function does not, in essence, differ from the textbook definition. Property 3 is called the *functional equation* for the exponential function. Functional equations relate the functions' values at different points, and the functions that satisfy functional equations are called their *solutions*. In our case, exponential functions are the solutions to the functional equation (property 3), and there are no other solutions under the additional constraints listed above.

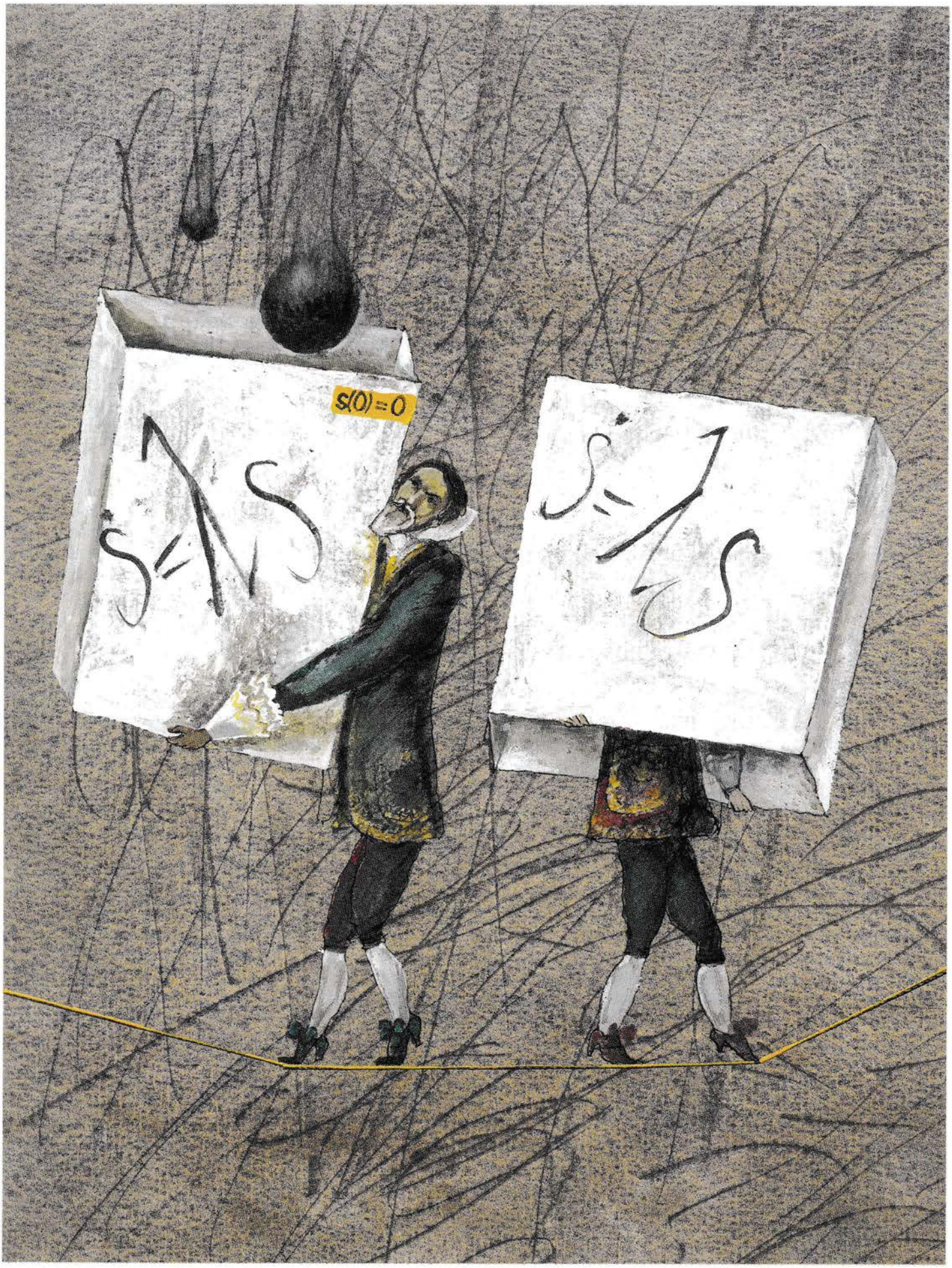
Problem 1. Prove that any solution to equation (property 3) that is defined for all x is either identically zero or everywhere positive. This implies that an exponential function that is defined everywhere exists only for nonnegative values of a.

The axiomatic approach to the definition of the exponential function that is used nowadays did not appear right away. It's remarkable that the exponential function first appeared as the solution to a differential equation. This story is worth recalling.

Kinematic definition of the exponential function

At the very beginning of the 17th century, Galileo (1564–1642) wanted to establish a law describing the free fall of objects. He hoped to establish this law in a purely speculative way, guided only by considerations of simplicity. Thus he assumed that the speed of the free fall is proportional to the distance traveled:

Art by Pavel Chernusky


S(0) = 0

$$\dot{s} = \lambda s. \qquad (1)$$

From physical considerations, $l > 0$ (the speed increases) and $s(0) = 0$ (the body is at rest at the initial moment). Later, Galileo was surprised to discover that the differential equation (1) cannot have a solution with $s(0) = 0$ that was different from $s \equiv 0$ (that is, movement with the properties that Galileo first ascribed to free fall is impossible).

Problem 2. Prove Galileo's assertion.

Hint. Let $\lambda > 0$. Then the speed $\dot{s}$ increases as s increases. If $s(t_0) = c > 0$ at some point t_0, then $\dot{s}(t_0) = \lambda s$, $s(t) < \lambda s$ for $t < t_0$; thus $t_0 > 1/\lambda$, no matter what the value of c may be. From this fact, it's easy to see that $s \equiv 0$.

This observation induced Galileo to change his proposed law to $\dot{s} = gt$, which led him to discover the law of free fall. He lost interest in the motion described by equation (1).

Here we should recall another remarkable mathematician: John Napier, Baron Murchiston (1550–1617), who studied motion according to law (1) almost at the same time as Galileo, but for quite different reasons. While Galileo wanted to obtain a mathematical description of an actual mechanical phenomenon, Napier was heading in just the opposite direction: he sought a mechanical interpretation of purely mathematical procedures. Even before Napier, some mathematicians had tried to use the relationship between geometric and arithmetic progressions to simplify calculations (to replace multiplication with addition). To increase accuracy, ever more "dense" progressions were needed, and Napier's brilliant conjecture consisted in the fact that one needs to go all the way to the end in this process and replace discrete progressions with continuous magnitudes—that is, with functions. Considering an imaginary motion according to law (1) for this purpose and interpreting time as a continuous analogue of an arithmetic progression, Napier satisfied himself that in this case the distance traveled was an analogue of a geometric progression. That is, if the time t increases by k units, the distance traveled is multiplied by q^k for some q. If we make the time units smaller, the geometric progression transformation still corresponds to the changes in time.

The main difference between Napier's approach and Galileo's consists in the fact that Napier lifts the restriction $s(0) = 0$. Let λ be fixed. Then Galileo's assertion implies an important property of motion (1). For a fixed λ, the motion according to law (1) is completely determined by the value of s at any fixed moment t_0. Indeed, if $s_1(t)$, $s_2(t)$ are two functions with $s_1(t_0) = s_2(t_0)$, then letting $s(t) = s_1(t) - s_2(t)$, we find that $s(t_0) = 0$, and that $s(t)$ satisfies equation (1) as well. Then, by Galileo's result, $s(t)$ is identically 0, so $s_1(t) = s_2(t)$ for all values of t. Now, if s is a solution to (1), then $s_1(t + t_0)$ is also a solution. So all the solutions to (1) differ only by a constant.

Let $S_\lambda(t)$ be a solution to (1) such that $S_\lambda(0) = 1$. Let's set

$$s_1(t) = S_1(t + t_0),$$

$$s_2(t) = S_1(t)S_1(t_0).$$

Both these functions are solutions to (1), and they take one and the same value $S_1(t_0)$ at $t = 0$; therefore they must be identical. Thus we obtain

$$S_1(t + t_0) = S_1(t)S_1(t_0). \qquad (2)$$

The function S_λ is differentiable, so it is continuous. By Galileo's result, it never equals zero, and so never changes sign. Since $S_\lambda(0) = 1 > 0$, this sign is always positive. Equation (1) implies that S'_λ (the derivative) has the same sign as λ, so that S_λ increases for $\lambda > 0$ and decreases for $\lambda < 0$. The results of this paragraph show that the solution to equation (1) with $s(0) = 1$ satisfies the axiomatic definition of an exponential function.

We should discuss one more question. We've studied the solutions to (1) under the assumption they exist. This assumption follows from a general property of differential equations, which we're not going to discuss here. If we make this assumption, we obtain a proof of the existence of the exponential function that is different from the proof discussed above. On the other hand, the existence and differentiability of the exponential function implies the solvability of equation (1).

Napier's number *e*

We now have two approaches to the definition of the exponential function. With the axiomatic definition, it would be quite natural to distinguish various exponential functions by their base $f(1) = a$. However, if equation (1) is used for the definition, it would be more natural to use the coefficient λ to distinguish them. What is the relationship between these two constants? First of all, with the kinematic approach, it would be quite natural to single out the exponential function with $\lambda = 1$. It's conventional to denote its base by e, and the function itself is sometimes denoted by $\exp t$ (after the beginning of the word exponent). Thus, for the motion governed by the law $\dot{s} = s$, for which $s(0) = 1$, we have $s(t) = e^t$ ($= \exp t$).

Let us examine the number e. This is the distance from the origin at which the point will be at time $t = 1$. Since the speed of the motion is greater than 1 and the motion starts at the distance 1 from the origin, we find that at $t = 1$ the distance $e > 1 + 1 = 2$. Let's now show that $e < 3$ (that is, the point will not have time to get to the point on the x-axis with coordinate 3). To prove this fact, let's split the distance from 1 to 3 into eight equal parts (the length of each part being 1/4). The point passes the first part with a speed that is not greater than its speed at the right-hand endpoint of this part, where it is equal to 5/4 because of equation (1); therefore, the time it takes the point to pass this part of the path is not less than 1/5. Similarly, it takes the point not less than 1/6 to pass the second part of the entire distance, not less than 1/7 to pass the third part, and so on. The total time needed to travel the distance from point 1 to point 3 is not less than

$1/5+1/6+1/7+1/8+1/9+1/10+1/11+1/12$.

This sum is greater than 1; thus e is nearer than 3.

Now let's split the time interval [0, 1] into n equal parts. We can see that the speed of the point is greater than 1 during the first time interval $\Delta t = 1/n$, so the point travels a distance greater than $1/n$ and ends up at a distance greater than $1 + 1/n$ from the origin. Its speed at this moment is greater than $1 + 1/n$ because of equation (1). During the next time interval $\Delta t = 1/n$, the point travels a distance greater than $(1 + 1/n)/n$, ends up farther than $(1 + 1/n)^2$ from the origin, and has a speed greater than $(1 + 1/n)^2$, and so on. So, at the moment $t = 1$, the point is farther than $(1 + 1/n)^n$ from the origin.

It seems quite plausible from this reasoning that $a_n = (1 + 1/n)^n$ increases as n increases (prove that this is indeed so) and tends to e. In a textbook on algebra and mathematical analysis it's proved that

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n.$$

This formula makes it possible to evaluate e with any accuracy desired. It turns out that

$$e = 2.718281828459045...$$

Similarly, it can be proved that

$$e^t = \lim_{n\to\infty}\left(1+\frac{t}{n}\right)^n. \qquad (3)$$

The reader is invited to think about the kinematic interpretation of this equation.

Problem 3. Give a kinematic proof of the inequality

$$e < (1 + 1/n)^{n+1}.$$

Hint. Split the interval $[1, (1 + 1/n)^{n+1}]$ into $(n + 1)$ segments of the form $[(1 + 1/n)^{k-1}, (1 + 1/n)^k]$. When the point is passing the kth segment of length $1/(n + 1)(1 + 1/n)^k$ according to law (1), its speed does not exceed $(1 + 1/n)^k$ and, accordingly, the time taken is greater than $1/n + 1$, so that the total time is greater than 1.

Let's return to the question with which we started our discussion of the number e. Let $S_1(t) = a^t$. What is the relationship between the constants λ and a? Note that if $s(t)$ is a solution to equation (1) for some λ, then $s(ct)$ is also a solution to (1) with $\lambda' = c\lambda$ and with the same value at zero. So $S_{\lambda c}(t) = S(ct)$, from which we get

$$S_\lambda(t) = a^t, \text{ where } a = e^\lambda.$$

Thus the relationship between a and λ can be very simply written in terms of e.

Let's sum up our consideration of equation (1). We have proven that the general solution to this equation has the form

$$s(t) = s_0 \exp(\lambda t),\ s_0 = s(0).$$

Equation (1) is called the *equation of exponential growth* (or *decay*, if $\lambda < 0$). It describes many processes encountered in nature: radioactive decay, change of temperature with time and of pressure with height, many laws of biological and social evolution, and many others. The most salient feature of the exponential function, which manifests itself in the phenomena it describes, is its rapid growth (for $\lambda > 0$). In particular, exponential growth (for $\lambda > 0$) exceeds any polynomial growth, and exponential decay (for $\lambda < 0$) exceeds any polynomial decay. That is to say,

$$\lim_{t\to+\infty}\frac{t^k}{e^{\lambda k}} = 0 \quad (\lambda > 0),$$

$$\lim_{t\to+\infty} e^{\lambda k} t^k = 0 \quad (\lambda < 0).$$

The logarithmic function $\log_a x$, $a > 0$, $a \neq 1$

This function is defined as the inverse of the exponential function a^x, $a \neq 1$. All properties of the logarithmic function follow immediately from this definition. It is defined for all $x > 0$, its range of values is the whole set of real numbers, $\log_a x$ increases for $a > 1$ and decreases for $a < 1$, and $\log_a 1 = 0$. But the main thing is that the functional equation

$$\log_a xy = \log_a x + \log_a y, \qquad (4)$$

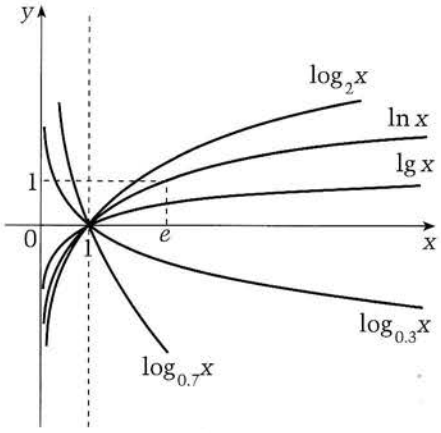


Figure 2. *Graphs of several logarithmic functions.*

(which is called *the fundamental property of the logarithm* in some textbooks) holds. This property immediately follows from the functional equation for the exponential function (1).

Problem 4. Prove directly that the inverse function to the function $y = a^x$ in fact has all these properties.

The logarithmic function can also be described axiomatically: $\log_a x$ is a monotonic function that is equal to 1 at $x = a$, satisfies (4), and is defined for all $x > 0$. Note that when we pass to the logarithmic function, we essentially use the strict monotonicity of a^x for $a \neq 1$ (this guarantees that the inverse function exists); for this reason, $a = 1$ is excluded.

The natural logarithm $\ln x$ is defined as the inverse function of e^x—that is, as the time taken by the point that moves according to the law $\dot{s} = s$ to get from 1 to x.

It must be said that it was the logarithmic function, not the exponential one, that was Napier's aim. However, the model that he constructed was somewhat complicated. He considered equation (1) with $\lambda = 10^{-7}$ and took its solution with $s(0) = 10^7$—that is, $s(t) = 10^7\exp(10^{-7}t)$. The corresponding inverse function—Napier's logarithm—had the form $t = \text{Nap log} = 10^7\ln 10 + 10^7\ln x$, and for this function the equation $t(x \cdot y) = t(x) + t(y)$, and $t(1) = 0$, does not hold. By the end of his twenty years of preparing logarithmic tables, Napier was undoubtedly not satisfied with the sys-

tem he had adopted. He already had the idea of base-10 logarithms, but he was exhausted and couldn't perform the necessary calculations.

The corresponding tables were prepared by the London mathematician Henry Briggs, who had an opportunity to discuss the idea of common logarithms with Napier during the last years of Napier's life. The common logarithms are still sometimes called Briggsian (the book *Arithmetic of Logarithms* by Briggs, printed in 1624, was devoted to common logarithms). Natural logarithms first appeared in 1619 in a book by the little-known teacher of mathematics Speidel (a short, anonymous table appeared the year before in the appendix to the second edition of Napier's book).

Using the rule of differentiation of inverse functions, we derive from the kinematic definition of logarithms that

$$(\ln x)' = 1/x.$$

Furthermore, the aforementioned relationship between λ and the base of the solution $S_\lambda(t)$ means that $(a^t)' = \ln a \cdot a^t$, from which we get

$$(\log_a x)' = \frac{1}{x \ln a}.$$

Problem 4. Find the maximum of numbers of the form $\sqrt[n]{n}$ (n is a natural number).

Hint. Test the function

$$f(x) = x^{\frac{1}{x}} = \exp\left(\frac{\ln x}{x}\right)$$

or, what is the same, the function $g(x) = \ln x/x$, for extrema.

The function $\ln x$ is convenient for describing different characteristics of processes described by equation (1). For example, if the time τ during which a magnitude doubles (or halves) is taken as the measure of growth (decay), we obtain $\tau = \ln 2/\lambda$ for $\lambda > 0$ and $\tau = -\ln 2/\lambda$ for $\lambda < 0$, since $e^{\lambda\tau} = 2$. In the case of radioactive disintegration, τ is called the half-life ($\lambda < 0$).

Problem 5. An account in a savings bank grows by 3% a year. How many years will it take for the account to double?

Finally, we note that since the exponential function grows very fast, the logarithmic function, on the contrary, grows very slowly. It grows more slowly than any positive power of x:

$$\lim_{x\to+\infty} \frac{\ln x}{x^\alpha} = 0 \qquad \alpha < 0.$$

Trigonometric functions

Both methods (axiomatic and kinematic) that we discussed in connection with exponential function are also possible for defining trigonometric functions. However, neither of them was historically the first. Originally, it was natural to define trigonometric functions using ratios of the sides of a triangle or in a circle. True enough, this method gives functions that take numeric values while their argument takes non-numerical, angular values. To pass from the functions of a numeric argument, one must first learn how to measure angles.

Although trigonometry has been developing since the second century B.C. and acquired great importance in connection with astronomical calculations, our modern definition of trigonometric functions dates only to the 18th century (mainly to Euler's work). Before his time, various tables were used—for instance, tables of chords corresponding to angles. It should be remembered that for most mathematicians, functions were limited to algebraic functions only and it did not occur to them that trigonometry has any connection with analysis.

Let's look at a kinematic definition of trigonometric functions. Consider one of the simplest laws: acceleration (or force) is proportional to distance:

$$\ddot{x} + \omega^2 x = 0. \qquad (5)$$

This equation is called the *equation of harmonic oscillations*: it describes the motion of a ball on a spring, the oscillation of a pendulum about its equilibrium point, and so on. First let's restrict our attention to the case $\omega = 1$:

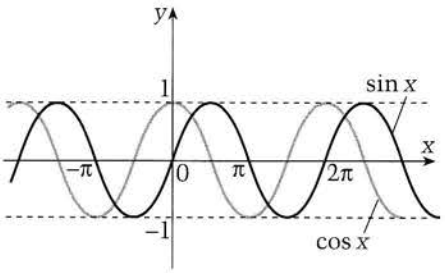


Figure 3. *Graphs of several trigonometric functions.*

$$\ddot{x} + x = 0. \qquad (6)$$

We know that $\sin t$ and $\cos t$ are solutions to this equation, but let's try to obtain an independent definition of these functions. First, we must obtain an analogue of the Galileo's theorem for (1)—the uniqueness theorem. We must prove that the unique solution to (6) with $x(0) = \dot{x}(0) = 0$ is $x \equiv 0$. Notice that we have to specify two conditions at $t = 0$. This is because (6) involves the second derivative. The proof follows from the conservation of the quantity $E = 1/2(\dot{x}^2 + x^2)$.

Problem 6. Show that equation (6) implies that $\dot{E} = 0$, and so is constant.

Denote by $s(t)$ the solution to (6) subject to the initial conditions $s(0) = 0$, $\dot{s}(0) = 1$. Let's analyze the properties of this solution (which, in fact, is identical with $\sin t$, but we want to infer all its properties from (6)).

(1) Let $c(t) = \dot{s}(t)$; then $c(t)$ is a solution to (6) subject to the initial conditions $c(0) = 1$, $\dot{c}(0) = 0$ (this fact can be verified by direct differentiation: $\ddot{c} = \dddot{s} = -\dot{s} = -c$, $\dot{c} = -s$).

(2) $s(t) = -s(-t)$, $c(t) = c(-t)$ (it suffices to verify that $x(t) = -s(-t)$ is a solution to (6) with the same initial conditions at $t = 0$ as for $s(t)$; the same for $c(t)$).

(3) The following summation theorems hold:

$$s(t + t_0) = s(t)c(t_0) + s(t_0)c(t), \qquad (7)$$

$$c(t + t_0) = c(t)c(t_0) - s(t)s(t_0). \qquad (8)$$

It's sufficient to verify that the left-and right-hand sides are solutions to (6) for some fixed t_0 with identical initial conditions at $t = 0$.

A lot of trigonometric formulas can be derived from the summation

theorems (in fact, almost all of them—we'll discuss this below). For example, setting $t_0 = -t$ in (8), we obtain

$$c^2(t) + s^2(t) = 1. \qquad (9)$$

It follows from (9) that $|c(t)| \le 1$, $|s(t)| \le 1$.

(4) It can be proved that $c(t)$ vanishes at some positive value of the argument.

If this were not so, $c(t)$ would be positive for $t > 0$; this would mean that the function $s(t)$ grows monotonically and, consequently, takes positive values for $t > 0$. Let's fix some $t = t_0 > 0$. Then $s(t_0) = s_0 > 0$. The equation $\dot{c}(t) = -s(t)$ means that the rate of decrease of $c(t)$ is greater than s_0 beginning at the moment t_0 and, consequently, $c(t)$ will go to zero not later than time $c(t_0)/s(t_0)$, which contradicts the assumption that $c(t)$ is positive.

Let τ be the smallest positive number such that $c(\tau) = 0$. (The number τ is really $\pi/2$, but we must prove this fact from what we already know.)

It follows from (9) that $s(\tau) = \pm 1$. Since $s(\tau)$ increases on the interval $(0, \tau)$, and since $\dot{s}(t) = c(t) > 0$, we have $s(\tau) = 1$. From the summation theorems (7) and (8), we see that

$$s(t + \tau) = c(t),$$
$$c(t + \tau) = -s(t).$$

Therefore,

$$s(t + 2\tau) = c(t + \tau) = -s(t),$$
$$c(t + 2\tau) = -s(t + \tau) = -c(t),$$
$$s(t + 4\tau) = -s(t + 2\tau) = s(t),$$
$$c(t + 4\tau) = -c(t + 2\tau) = c(t).$$

Thus the solutions $c(t)$ and $s(t)$ are periodic with the period $T = 4\tau$. It can be proved that there is no shorter period.

Problem 7. Prove this statement.

Thus we have shown that the theory of trigonometric functions can be built exclusively on the analysis of equation (6). As for equation (5), its general solution has the form

$$x(t) = a \cos \omega t + b \sin \omega t,\ a = x(0),$$

$$b = \pi \frac{\dot{x}(0)}{\omega}.$$

Functional equations for trigonometric functions

The fact that trigonometric functions can be described by means of functional equations is fundamental. Roughly speaking, this means that if trigonometric functions are defined geometrically, it suffices geometrically to prove the summation theorems only. All the other innumerable trigonometric formulas (except for the formulas involving π) can be formally derived from the summation theorems without using geometric reasoning.

The exact assertion (compare it with the similar axiomatic description of the exponential function) is the subject of the following problem.

Problem 8. Let $s(t)$ and $c(t)$ be functions such that (1) they are defined for all t; (2) $s(0) = 0$, $c(0) = 1$; (3) they satisfy the functional equations (7) and (8); (4) they are continuous. Then $c(t) = \cos \omega t$, $s(t) = \sin \omega t$ for some real number ω.

Hint. Use the fact that every continuous function is completely determined by its values at the points of the form $t = mt_0/2^k$ (m, k are integers, t_0 is an arbitrary fixed real number distinct from 0). ■

CONTINUED FROM PAGE 21

flight in the atmosphere, we'll just consider the "worst-case scenario," where all the incoming heat is removed from the vehicle by radiation only.

The Stefan–Boltzmann law says that the power taken away by radiation from one square meter of a radiating surface is proportional to the fourth power of the temperature: $q = \sigma T^4$ J/(m$^2 \cdot$ s). The proportionality factor—the Stefan–Boltzmann constant σ—can be found in a physics reference book: $\sigma = 5.67 \cdot 10^{-8}$ W/(m$^2 \cdot$ s $\cdot$ K^4). This is a very steep dependence: given a threefold increase in the object's temperature, the amount of irradiated heat increases by a factor of $3^4 = 81$—almost a hundredfold! According to this law, a body must be "white-hot" to be efficiently cooled by radiation.

Now let's estimate the maximum speed of atmospheric flight in which the reentry vehicle does not melt. The air mass striking a unit area of the vehicle's surface per unit time is ρv. This air carries an energy of $\rho v(v^2/2)$, which is almost entirely converted into heat. We've proposed that all this heat will be removed by thermal radiation, which carries away energy σT^4 from a unit area per unit time. The surface temperature of the vehicle must not be greater than the melting point T_m:

$$\frac{v^2}{2}\rho v \le \sigma T_m^4.$$

Thus the vehicle's velocity cannot exceed the maximum value corresponding to any given altitude:

$$v \le v_{max}(h) = \left[\frac{2\sigma T_m^4}{\rho(h)}\right]^{1/3}.$$

The plot $v_{max}(h)$ is given in figure 8—this is the *EF* curve.

Therefore, if we want to use the "supporting" and "braking" properties of the atmosphere—that is, if we want to land on a planet in a winged vehicle and use no fuel to brake in the air—then, in the altitude–velocity coordinate axes, the trajectory cannot pass higher than the curve *CDAB* (the wings will not support the vehicle) and lower than the curve *EF* (vehicle will burn up). Both these curves are plotted for landing on Earth. We can see that increasing the speed reduces the distance between the boundary curves to a minimum—that is, the reentry corridor becomes narrowest at high speeds. There's no room for error here! Fortunately, the boundary curves don't meet or cross at any altitude. It's as if Nature wanted to make sure that intelligent creatures, having left Earth to explore, could return home in winged vehicles. ■

Physics without

SUMMER IS A CAREFREE time for students. Textbooks are put away; homework is a hazy memory. How nice to just splash in a pool, or wander in the woods, or lie on the grass and look at the sky, not thinking about anything in particular. But if you take a good look around, you'll find that many things that seem ordinary and even boring are really pretty interesting.

You're standing at the edge of a pond. See how the water strider runs across the glassy top of the water? Don't be in a hurry—stop and watch how the insect calmly takes step after step without piercing the surface. Surface "tension"? Not for this little creature!

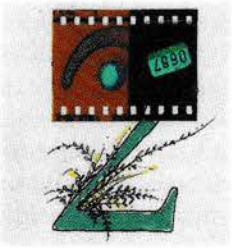
On a summer evening, you can find glow-worms in the thick grass or on tree stumps in the woods. These living "lamps" produce chemical substances that "burn" (that is, are oxidized) in the air. The energy released is radiated in the form of visible light. So a glow-warm "creates" a source of light within its own body.

Or take the cuttlefish: it squirts liquid from somewhere inside its body and moves according to one of Newton's laws (the one about action and reaction). Another clever creature!

The inorganic world is also full of interesting things. Here are some experiments you can do on your own or with your friends. Don't be upset if you can't explain the results. Sooner or later, a new school year will start, and many things seen in the bright light of summer will be illuminated further with explanations.

Art by Yuri Vaschenko

Not many people who live in temperate climes have ever seen a mirage. But if you want to see one, you don't need to go to Saudi Arabia. All you need is a little patience and a bit of preparation. First, find a smooth wall illuminated by the Sun. Wait for a clear, calm day. Stand with a friend close to the wall, at opposite ends of it. Look along the wall and you'll see two friends instead of one! The wall turned into a mirror and reflected your friend's image. Again—you'll need to be patient for this to work. Good experiments don't always work the first time, after all.



If you watch the Moon regularly, sooner or later you'll probably see a colored ring (or a couple of them, if you're lucky) around it. They're particularly distinct when a light, semitransparent cloud comes between you and the Moon. The light from the Moon is dispersed and scattered by random heterogeneities in the cloud, which produces a diffraction pattern in the form of colored iridescent rings.

If you're burning with curiosity to see these iridescent patterns and there's no Moon in the sky, you can create them artificially. All you need are pieces of photographic film spoiled by exposure to light (like the bits from the beginning and end of the roll) and a sewing needle.

Scratch a line on the emulsion side with the needle. You've created a so-called diffraction slit. Bring this slit close to your eye and look through it at some bright source of light (for example, a bright lamp). You'll see bright, multicolored bands off to both sides. Turn the film, and these bands will also turn so as to remain parallel to the slit. Now scratch a small circle on the film. In this case, the diffraction pattern will be composed of concentric multicolored rings. Play with other lines and see the resulting diffraction patterns.

A rainbow can also be seen using a water drop. Set the drop on a rod or a blade of grass. Turn your back to the Sun and carefully raise the drop. When the Sun's rays form an 42° angle with the line between your eye and the drop, the transparent drop suddenly gives off a bright, pure color. What color? Any color you want! Move the drop carefully along an arc and you can see all the colors of the rainbow.

Once I was gathering mushrooms in a forest and witnessed an amusing scene. My son was looking attentively at the dewdrops on the grass and on the pine needles. His movements were rather strange: he slowly sat down, while looking off at seemingly nothing. Then another mushroom

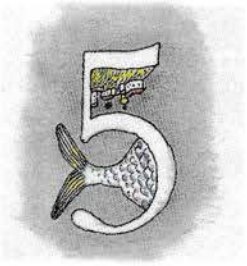
ut fancy tools

lover came into the clearing—a very respectable-looking gentleman. He watched my son's slow-motion antics for quite some time, clearly at a loss as to what he was up to. Finally he broke down and asked the boy to explain his strange behavior. Then I watched with amusement as both naturalists periodically moved carefully, then froze in awkward positions, near an ordinary fir tree.

A small drop of water can be turned into a magnifying glass. Take a sheet of dense paper and pierce it with a thick needle. Then place a droplet onto the hole. Your magnifying glass is ready!

Move your improvised magnifying glass very close to a newspaper. You'll see the magnified image of a letter. The smaller the drop, the higher the magnification. After you "fine-tune" your magnifier by adjusting the drop size, you're ready to examine any small object in more detail.

By the way, Antony van Leeuwenhoek (1632–1723) constructed the first microscope using the same design, except that he used a "drop" of glass.



Speaking of small droplets, perhaps you have a *huge* drop of water in your house—a round fishbowl (filled with water, of course). It's a wonderful optical instrument—you can see some very unusual things with it.

First of all, the apparent size of the fish inside depends strongly on their location in the fishbowl. How strongly? And why?

Second, the fishbowl inverts the image of distant objects. Look out at the street through the fishbowl. Note the direction a car's image travels when the car itself is moving, say, from left to right.

Third, when the fish swim to the side walls (relative to our viewpoint), they suddenly lose their heads—all we can see are their tails! After a moment, the whole fish disappears from view. Where is it hiding? Recall the law of total internal reflection and explain this phenomenon. Rack your brain (even though you're on vacation) and calculate the region of fishbowl where the fish can "hide."

Fourth, look at the street lamps in the evening—once directly through the window, then through the fishbowl. Has the brightness of the lamps changed?

In novels and short stories (and even in the pages of our magazine) you may have encountered descriptions of the so-called "green flash" that appears at sunset. (See the list of *Quantum* articles below.) As a rule, the green flash is observed at sea, where the air is very clean. When the Sun drops down on the sharply delineated horizon, an emerald green flash may appear for the briefest moment. This is due to the dispersion of light—that is, the angular separation of its component colors (the colors of the rainbow) as the light passes through a thick layer of air that is nonuniform in density. In principle, all colors should be observed, but against the background of the Sun, which is giving off a yellowish light, the human eye perceives the green color more intensely.

In the summer you can try to catch this rare phenomenon. If you don't live near the sea (or aren't vacationing there), choose a place with a sharply etched horizon. You can use the edge of a roof or the slope of a nearby hill.

If the summer passes and you haven't seen the Sun in its unnatural green disguise, don't despair. You can keep trying all year round. Actually, winter is a good time to see a green Sun. Look at the rising or setting Sun through a frosted window. If you find a suitable place on the glass, the ice crystals will do their thing—they'll break the sunlight down into all the colors of the rainbow. You can see a green Sun, or a blue Sun, or a red Sun... ■

—A. Dozorov

Quantum for naturalists—young or experienced:

D. Tarasov and L. Tarasov, "The Play of Light," May/June 1996, pp. 10–13.

L. Tarasov, "The Green Flash," January/February 1997, pp. 38–39.

A. Eisenkraft and L. D. Kirkpatrick, "Color Creation," May/June 1997, pp. 36–39.

V. Novoseltzev, "Visionary Science," May/June 1998, pp. 21–25.

An arresting sight

by V. Uteshev

IMAGINE A PIECE OF WOOD being rotated on a lathe. Let it be illuminated not by sunlight but by rapidly repeating flashes of light. If its rotational frequency f satisfies the condition $f = n\nu$ ($n = 1, 2, 3$ ...), where ν is the frequency of the flashes, we'll always see only one side of the piece. This "experiment" is actually dangerous, because we might mistakenly think the piece isn't rotating and try to grab it. This optical illusion occurs because the piece makes n complete turns during the dark period between flashes, so we always see the same side of the piece.

This example demonstrates the stroboscopic effect. The word "stroboscopy" comes from the Greek words *strobos* = rotation and *skopeo* = to look. It refers to the optical illusion of arrested rotation. If the frequency f differs slightly from $n\nu$, we see the object turning slowly with each successive flash. It is easy to understand that the object will turn in the direction of the actual rotation if $f > n\nu$ and in the opposite direction if $f < n\nu$.

The stroboscopic effect is widely used in science and technology for precise measurements of rotational frequencies or periods of periodic processes. Note, however, that the greater n, the poorer the sharpness and brightness of the "frozen" object. Usually, $n \leq 5$ is chosen in practice.

Cinematography is a wonder of the 20th century, and it owes its existence to the stroboscopic effect.

"The moment that you had pronounced him one,
Presto! his face changed, and he was another,
And surely he was some mother's cousin's brother."
—George Gordon, Lord Byron (1788–1824)

Photographic film is pulled in a precise, jerky manner, frame by frame, in front of the lens of the projector at a rate of 24 frames per second. The images of successive frames are almost identical, so we see continuous motion on the screen due to the inertial lag of our visual system.

Now let's take a closer look at how the human visual system operates.

A few words on physiology

Beams of light from an object pass through the optical system of the eye, consisting of a convex cornea, a crystalline lens, and a semiliquid, transparent vitreous humor. The beams are focused on the retina, which is the biological "screen" of this natural "camera." The retina has a complex structure consisting of several layers of nerve cells.

Figure 1 is a schematic of a cross section of the retina. The light is incident from the right. In the outer layer (1), which is in direct contact with the vascular envelope, there are cells containing black pigment. Next come the basic visual perception elements (2), called rods and cones because of their shapes. Layers 3–5 consist of nerve fibers connected to the rods and cones. Behind these layers are the so-called granular layers 6 and 7, which are also related to the nerve cells and their axons. Layer 8 is formed by ganglion cells, each of which is connected to

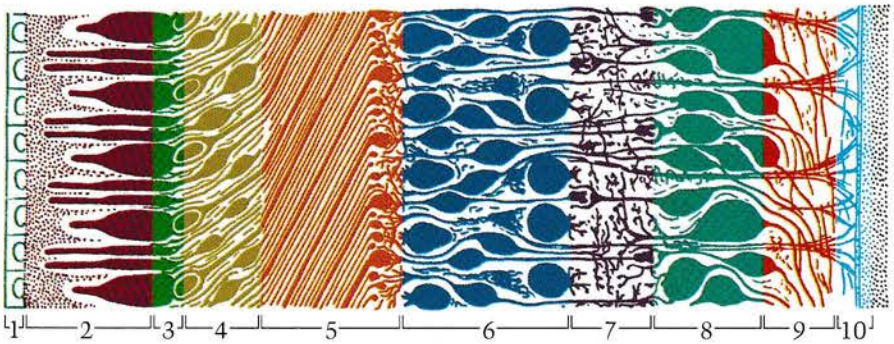
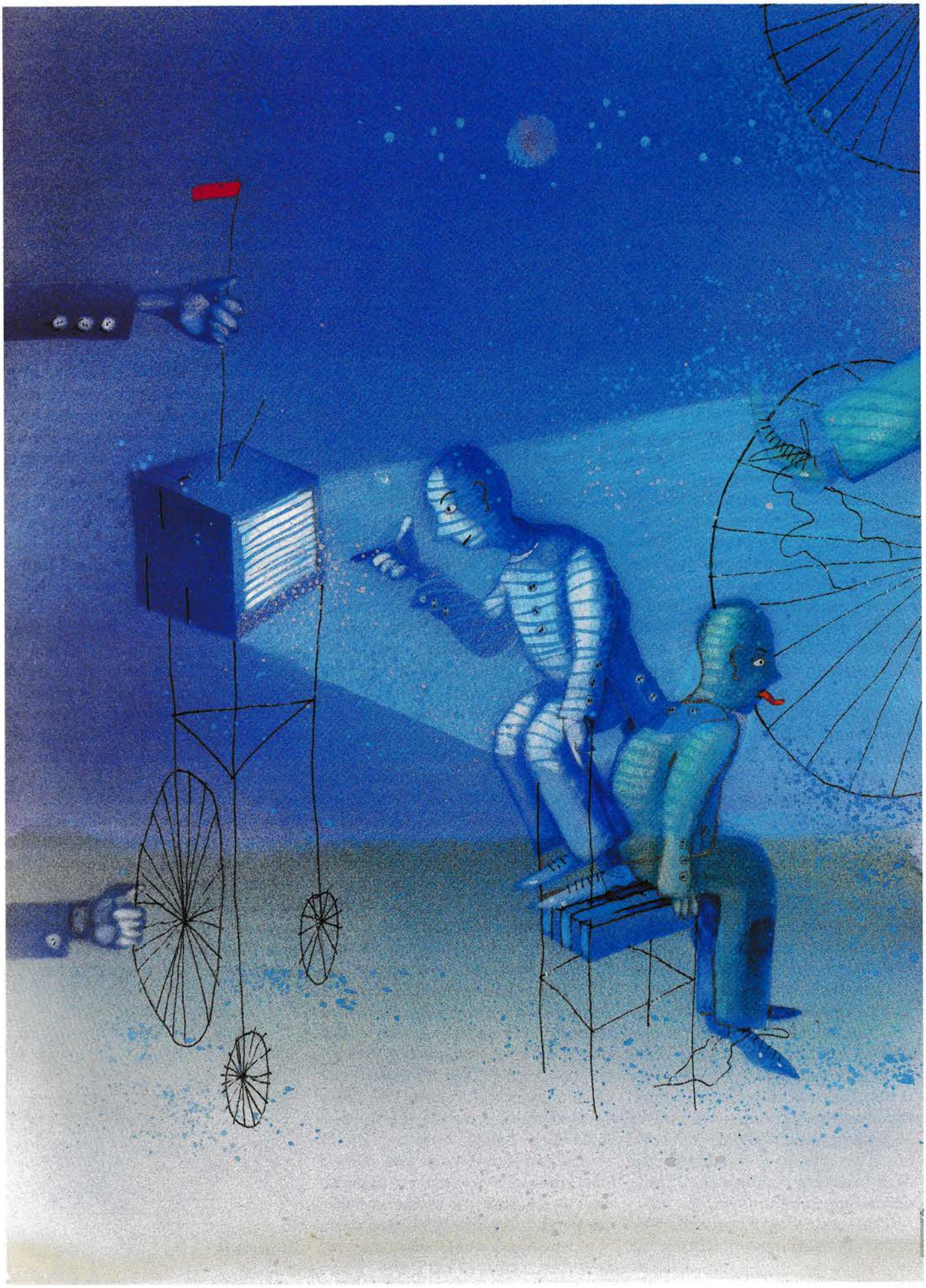


Figure 1

Art by Ekaterina Silina



a nerve fiber in layer 9. Layer 10 is the internal limiting membrane. Every nerve fiber terminates with a cone or a group of rods. (Figure 1 and the description above are taken from a book by the famous Russian ophthalmologist Sergey Ivanovich Vavilov (1891–1951), *The Eye and the Sun*.)

The human eye contains about 7 million cones and 130 million rods. Biophysicists and physiologists discovered that rods are more sensitive to light than cones, although cones can differentiate colors. In what follows we'll use the term "rods" to denote both types of light-sensitive elements of the eye.

When light strikes the rods, it splits the light-sensitive protein rhodopsin into retinene and opsin. As a result, the rods become excited and generate nerve impulses that go to special vision analyzers in the brain. A chain of specific chemical reactions converts retinene and opsin back into rhodopsin.

The time interval (about 0.2 s) needed to restore rhodopsin molecules is called the "dead time." During this time the rod is characterized by the so-called refractory (non-excitable) state, in which the rod doesn't respond to light. The dead time depends primarily on the intensity of the incident light. All rods work independently and have different dead times (although the difference is not great). As a result, the composite retina has no dead (blind) time at all.

Visual inertia, which is widely used (particularly in the movies), arises due to the fact that the visual system can "remember" an image for only a short time—about 0.2 s. During this period it remembers the previous image and cannot process the next one. If a shorter time passes between successive signals than the time necessary to retain an image, the successive frames fuse and visual perception becomes continuous. This is why we don't notice the dark intervals between the frames on the screen in a movie.

The sensitivity of our visual system is amazing. It can reliably detect (without special contrivances) 10 photons and notice a flash in darkness with a duration of only 0.000001 s.

We should note that informational "stitching" of the image frames underlies the optical illusion that converts the jerking frames into the continuous motion of an object. It's important that this illusion (or a sophisticated analysis of video information, one might say) is created by the brain, while the visual system plays no role in these extremely complicated processes.

In addition to the phenomena described, the eye has many other wonderful properties. For instance, it can produce the stroboscopic effect without any physical devices. Let's see how.

The amazing "eye–voice" effect

This interesting phenomenon was first reported in 1967. It was stated that a rotating object can be "arrested" by a voice. Moreover, unlike the usual stroboscopy, which requires a flashing light, physiological eye–voice stroboscopy can be demonstrated in ordinary sunlight.

Here's how it works. Imagine that we observe alternating white and black bands moving downward. The images of these bands will also move on our retina. Now we need to produce an oscillation of our eye in the plane of the bands. The amplitude and frequency of vibration must be chosen to keep the image motionless on the retina during most of the oscillation period. If we can do that, the illusion of arrested bands will occur.

Vibration with the necessary parameters can be produced by the human voice. Oscillations of the vocal cords are transmitted to the eye through the skull bones, so by changing the volume and frequency of your voice, you can arrest the motion of the bands.

Instead of your voice you can use the vibrations of your tongue as you roll an "r," but the result won't be as good. It's also possible to use a loudspeaker connected to a sound generator, a motor with a regulated angular velocity, or even a vibrating massager. By pressing your head against any vibrating device, you can obtain the stroboscopic effect.

This "physiological" stroboscopy can arrest the rotating propellers of an airplane, the spokes of a bicycle wheel, and other vibrating or rotating mechanisms. Some people can even produce bands on a television screen or change the quality of the image by humming. However, such feats require patience and skill, so don't be surprised if you can't do it.

Conventional stroboscopy redux

Let's consider another interesting manifestation of stroboscopy. When you watch TV, try to move your forefinger up and down in the space between the screen and your eyes as shown in figure 2. You'll see several "motionless" fingers. When your hand moves downward, these fingers will be wider and more luminous than when your hand is lifted up. Why?

The image on the TV appears due to the persistent scanning motion of the electron beam from left to right and from top to bottom on the surface of the luminescent screen. When the beam hits a region of the screen, it produces a brief flash at that spot. Sweeping across the screen, the electron beam leaves illuminating lines behind it. After a moment these lines fade, so one might say that light "frames" are moving on the screen (as with a movie). When watching TV, we don't see the flashes produced by the scanning beam. The frequency of these flashes is too high for our vi-

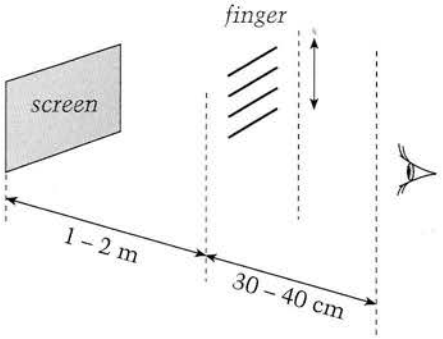


Figure 2

sual system (50 Hz), so the dark intervals appearing on the screen between frames simply elude our perception.

Now let's make our finger move in front of the screen for about 0.2 s. Given a frame refresh rate of 50 frames per second, this means that 10 frames move down the screen during this time. If you move your finger downward, its motion is slower relative to the downward movement of the scanning electron beam compared to the motion of your finger moving upward. Therefore, your finger is illuminated for a longer time, and the shadow produced by your finger exists for a longer time. So when your finger moves downward, its shadow produces a wider band on your retina, and your fingers seem to be wider and brighter.

By changing the speed of your finger, you can regulate the number of frames illuminating it and thus the number of separated shadows produced by the finger on your retina. It's possible to choose a finger speed for which the different shadows will be resolved by your eyes (that is, they will be viewed separately from one another). Since 10 frames move down the screen during the chosen period of 0.2 s, we can see 10 motionless fingers in this case.

How to make a stroboscope

To make your own stroboscope, buy an electric motor and a battery. You'll also need a piece of cardboard or thin metal, wires, a variable resistor (a rheostat rated at several ohms), and a film projector. Cut out a cardboard disk with a diameter of 30 cm and make 6–8 equally spaced slits not far from the edge of the disk. Attach the center of the disk to the shaft of the motor and set the film projector in such a way as to make

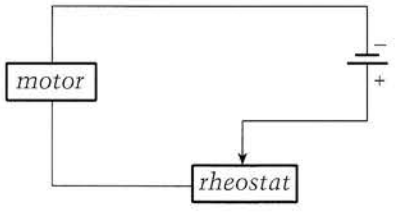


Figure 3

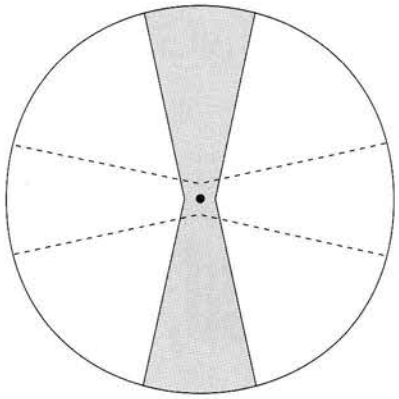


Figure 4

the slits pass in front of the objective. Now assemble the electrical circuit shown in figure 3, and your stroboscope is ready for use.

By rotating the knob of the rheostat, you can regulate the rotational velocity of the disk and thus the flicker frequency. At some rotational frequency you'll be able to arrest a revolving (or arbitrarily moving) object observed through the openings of the disk. You can observe a fan, water dripping from a faucet, waves on water, a swinging pendulum, and so on. If you know the rotational frequency of the disk, you can determine the frequency of the object's periodic oscillations. To do this, you'll need to choose the minimal rotational frequency at which the object seems to stop moving. Thus you can determine one frequency by means of another.

You can do a lot of interesting experiments with a stroboscope. For example, you can demonstrate that an ordinary incandescent bulb "twinkles" (its luminance periodically changes ever so slightly). Not only that, you can show that the blinking frequency is double that of the alternating current in the wires. To detect the flashing, you should draw one or two black bands on the stroboscope's disk (figure 4) and illuminate it with a table lamp. After starting the motor at some rotational frequency, you'll see arrested or slowly moving dark bands. They'll be rather faint, so you should look carefully. If you repeat this experiment with sunlight, you won't see any dark bands.

Stroboscopy and life

The stroboscopic effect has many practical uses, mostly in fields where it's necessary to measure high-frequency rotations or oscillations. It's also used to create the illusion of motion in neon signs, to measure rotational speeds in electronic music devices, to emphasize the fluid motion of performers on stage, and, as we said above, to create virtual life on TV and in the movies.

However, even more original and extravagant projects are also based on the stroboscopic effect. For example, imagine a long set of pictures drawn on the walls of a subway tunnel. When the train moves at a certain speed (and, possibly, if the wall is illuminated by a flashing light), these pictures will "come to life," so the passengers can look out the window and be treated to an animated cartoon or advertisement! ◙

Quantum on vision and optical illusions:

B. Fabrikant, "Through a Glass Brightly," September/October 1990, pp. 34–38.

M. Berkinblit and E. Glagoleva, "Mathematics in Living Organisms," November/December 1992, pp. 34–39.

V. M. Bolotovsky, "What's That You See?" March/April 1993, pp. 5–8.

P. Bliokh, "What Little Stars Do and the Big Old Planets Don't," March/April 1994, pp. 22–27.

V. Surdin, "Optics for a Stargazer," September/October 1994, pp. 18–21.

D. Tarasov and L. Tarasov, "The Play of Light," May/June 1996, pp. 10–13.

A. Eisenkraft and L. D. Kirkpatrick, "Color Creation," May/June 1997, pp. 36–39.

A. Mitrofanov, "Can You See the Magnetic Field?" July/August 1997, pp. 18–22.

V. Novoseltzev, "Visionary Science," May/June 1998, pp. 21–25.

V. Surdin, "The Eye and the Sky," January/February 2000, pp. 16–20.

Volta, Oersted, and Faraday

by A. Vasilyev

"IN A WORD, VOLTA INvented the electric battery, and Oersted short-circuited it with a wire and saw that a compass needle deflects. Faraday, on the other hand, forced current to flow in a wire by moving a magnet and discovered electromagnetic induction."—*from a conversation with Moscow University students before a physics exam*

"If you rub a resin disc with cat fur and put an iron coin on it, the electric charge that has collected on the coin can be used to charge a Leyden jar." Thus the Italian physicist Alessandro Volta (1745–1827) reported, in letters to the leading scientists of the time, his discovery of the electrophorus. This electrophorus was the first device that made it possible to accumulate electric charge (if only in small amounts) and use it, say, to obtain a spark or cure a paralyzed finger.

But Volta's greatest achievement was his invention of the first electric battery in 1799, which became widely known as the "voltaic pile." This battery consisted of several dozen alternating copper and zinc discs with skin or cardboard separators between them. To activate the voltaic pile, the separators were soaked with alkaline or salt solution, and all the elements were compressed. The voltaic pile needed no periodic recharging and, in Volta's words, it "caused shaking" whenever a person touched it. Further improvements in the voltaic pile led to a cup-shaped version—the forerunner of the modern storage battery. Volta immersed silver and zinc plates in the cups and connected them in series with metal wires to build up the electric effect.

From the viewpoint of modern science, the induction of an electric spark by shorting the edge plates of a voltaic pile results from chemical reactions in the battery that generate a potential difference across its terminals. Imagine a zinc plate immersed in a solution of sulfuric acid (H_2SO_4). The process of zinc dissolving is complex: it is doublycharged positive ions (Zn^{2+}) that enter the solution, not neutral atoms. As a result, solution in the immediate vicinity of the plate becomes positively charged, the zinc plate becomes negatively charged, and the metal acquires an electrochemical potential relative to the electrolyte solution. The sign and value of this potential depend not only on the nature of the acid and the metal but also on the concentration of ions in the electrolyte. If plates of different metals are immersed in the solution, the potential difference generated between them will be equal to the difference in their electrochemical potentials.

For example, the electrochemical potential of zinc immersed in sulfuric acid, in which each liter contains one mole of dissolved metal ions, is equal to –0.5 V. By contrast, under the same conditions the electrochemical potential of a copper plate will be +0.6 V. The point is that the copper plate accepts positive ions and acquires a positive charge, thereby giving a negative charge to the adjacent solution. The voltage difference between the zinc and copper plates (the emf of this metal pair) is 1.1 V. (Note that the great Italian physicist is immortalized in the unit of measurement used: the volt.)

As early as 1800, Volta's discovery made it possible to break down water and ammonia. It also led to new technologies such as silver, copper, and zinc plating. In short, it opened the new age of electrical science: electrodynamics.

In 1820, the permanent Secretary to the Danish Royal Society, Hans Christian Oersted (1777–1851), delivered a lecture on physics and during a demonstration found that the magnetic needle of a compass was deflected when the terminals of a battery were connected by a wire. He had a strong battery that made the wire red-hot. At first, Oersted thought that a high temperature was needed for the electric current to produce the magnetic effect, but soon it became clear that the needle "felt" even small currents. The author named this phenomenon "an electric conflict," according to the popular philosopher Friedrich Schelling (1775–1854), who believed that everything in this world occurs

Art by Vadim Ivanyuk


W
S

due to a collision of polar-opposite entities. True, the wire in Oersted's experiment connected the opposite poles of the battery (negative and positive), but it was far more important that the "conflict" between electricity and magnetism occurred not only in the metal wire but in the surrounding space as well.

The effect of the electric current on the magnetized needle was very unusual. Indeed, all the forces known at that time resulted in either attraction or repulsion. By contrast, the magnetic needle was neither attracted to nor repulsed from the current-carrying wire. It simply turned so that it is perpendicular to the wire. Noting this feature, Oersted wrote that "according to the facts described, the electric conflict produces a vortex around the wire. Otherwise it cannot be understood why the same fragment of wire placed under the magnetic pole drives it eastward, and when lifted above the pole, it is driven westward." Here Oersted encapsulates the idea that electric current is encircled by magnetic lines of force.

The discovery of the Danish physicist ignited an unprecedented interest in the scientific community and, in particular, among French scientists. Soon after it was published, Jean Baptist Biot (1774–1862) and Felix Savart (1791–1841) found the mathematical expression for the force exerted by the electric current on the magnetic pole. Dominique Francois Arago (1786–1853) discovered that iron filings are magnetized by a current-carrying wire, and André Marie Ampère (1775–1836) obtained the expression for the force between electric currents and revealed the close "genetic" relationship (not conflict!) between electrical and magnetic processes.

The next great discovery involved the rotation of a magnet around a current-carrying conductor, and the honor belongs not to a Frenchman but to the Englishman Michael Faraday (1791–1867). The experiments conducted by Faraday in 1831 to demonstrate electromagnetic rotation were brilliant. To perform such an experiment (in other words, to construct the first electric motor in history), he needed to come up with an arrangement of the magnet and the current in which the current would act on only one pole of the magnetic. To accomplish this, Faraday directed the current through cups filled with mercury into which wires were lowered. In one cup the wire was set along the axis of the cup, and a magnet floated in the mercury, half-submerged (with one pole above the surface). In this setup the current affected only the upper pole of the magnet and forced it to circle around the wire. In the other cup the magnet was fixed along the cup's axis, and it was the current-carrying wire that circulated about the magnet.

Having demonstrated that electromagnetic rotation was possible in principle, Faraday set himself the task of "converting magnetism into electricity." Many physicists worked on this problem and tried to produce an electric spark or another manifestation of the electric force by winding wire on a magnetized piece of iron. All these experiments were unsuccessful, because a permanent magnet failed to generate electricity.

However, in the interest of historical justice, we should note that while European physicists made one fruitless attempt after another, the American scientist Joseph Henry (1797–1878) observed that current is induced in a coil when a magnet moves in it. As Henry prepared to publish the results of his experiments, Faraday published his paper on the electromagnetic induction he had discovered.

Here are some excerpts from Faraday's journal, written in 1831:

> I took an iron ring and wound two coils on it, each of copper wire insulated with cotton cloth.
>
> Then I charged a battery made of 10 pairs of plates with an area of 4 square inches. The ends of one coil were closed with a copper wire passing just above a magnetic needle. When the other coil was connected to the battery, an immediate effect on the needle was observed: it started to oscillate and eventually returned to its original position. The same thing happened when the second coil was disconnected from the battery.

In the following passage, Faraday describes the effect of induction produced by permanent magnets:

> If a magnet is inserted into a spiral with a single continuous motion, the needle turns to one side, then it suddenly stops, and finally it starts to move in the opposite direction.

Elsewhere Faraday writes:

> The results that I have already obtained in my experiments with magnets led to the idea that the current transmitted by a battery through one conductor actually induces a similar current in another conductor, but this latter current lasts only for a brief moment, and by its nature, it resembles rather the electric wave generated by the discharge of an ordinary Leyden jar than the current produced by a galvanic battery....
>
> When the wires are brought close to one another, the induced current has the direction opposite the direction of the inducing current. When the wires move away from one another, the induced current has the same direction as the inducing current. If the wires remain motionless, there is no induction current.

These excerpts from Faraday's journal describe one of the greatest discoveries in human history, one that had an enormous impact on science and technology. However, historical justice again demands that we recall some other wonderful discoveries made by this scientist.

For example, Faraday proved that all the types of electricity known in his time were identical: animal (torpedo ray and electric eel), magnetic, galvanic, thermal (thermoelectricity), and frictional (triboelectricity). In his attempt to discover the nature of electricity, he conducted experiments on the transmission of electric current through solutions of salts, acids, and alkalis. As a result, he found the laws of electrolysis. Faraday discovered the effect of dielectrics on electrostatic interaction and introduced the concept of di-

CONTINUED ON PAGE 40

The fellowship of the rings

by S. Shabanov and V. Shubin

WE'LL SHOW YOU HOW TO make air and water vortex rings in the lab. We'll consider their properties and their interactions with obstacles and with one another.

When we were still in grade school, we noticed some interesting features in the behavior of vortex rings. We were able to explain almost all of them on a qualitative level.

Formation of vortex rings

We made air vortices in the laboratory with a Tait's vortex machine (figure 1), which is a horizontal cylinder closed on one end by an elastic membrane (made of leather, for example) and having an opening through a circular diaphragm on the other end.

Inside the cylinder are two vessels: one contains hydrochloric acid (HCl), and the other ammonium hydroxide (NH_4OH). These two chemical agents produce a thick fog or smoke of ammonium chloride (NH_4Cl) particles.

Striking the membrane imparts a velocity to the adjacent layer of smoke. When set in motion, this layer compresses the neighboring layer, which in turn transfers the compression to the next layer, and so on. When the sound wave reaches the diaphragm, the smoke will escape from the hole, disturb the previously calm air, and curl itself into a ring due to viscous friction.

What factors affect the formation of vortex rings? Might it be that the most important role in this process is played by the rim of the orifice? To test this hypothesis, we replaced the usual diaphragm in the Tait's machine by a sieve with a number of holes. If our hypothesis is correct, we should obtain many small rings. However, experiments show that the sieve produces only a single large vortex ring (figure 2).

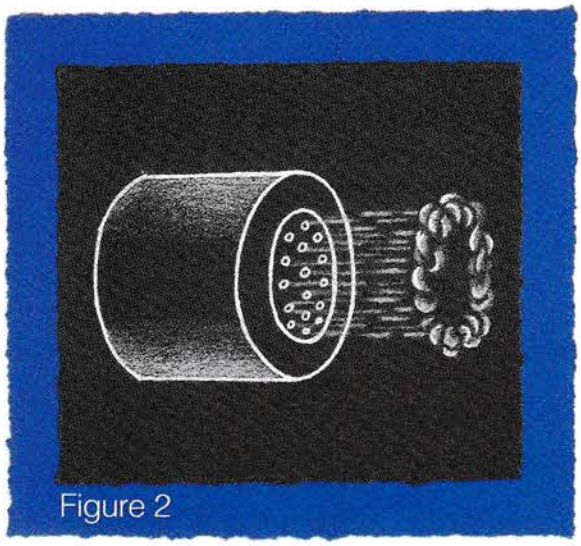


Figure 2

It is very important that the smoke escapes from the hole in discrete bursts of a continuous stream. If the membrane is replaced by a moving piston, there will be no rings, and only continuous smoke will curl from the hole.

Vortices in water can be produced by a common pipette filled with ink. Take some ink up into a pipette and squeeze a drop from a height of 2–3 cm into an aquarium containing static water (that is, without convective or other flows produced, say, by an illuminating lamp, aerating device, etc.). The resulting ink rings are clearly seen in the transparent water (figure 3).

We can modify the experiment and squeeze ink directly into the water (figure 4). In this case the rings will have somewhat larger diameters.

The formation of vortex rings in water and air is similar: the ink in water plays the same role as the smoke in air. In both cases a key role is played by forces of viscous friction. However, the similarity of

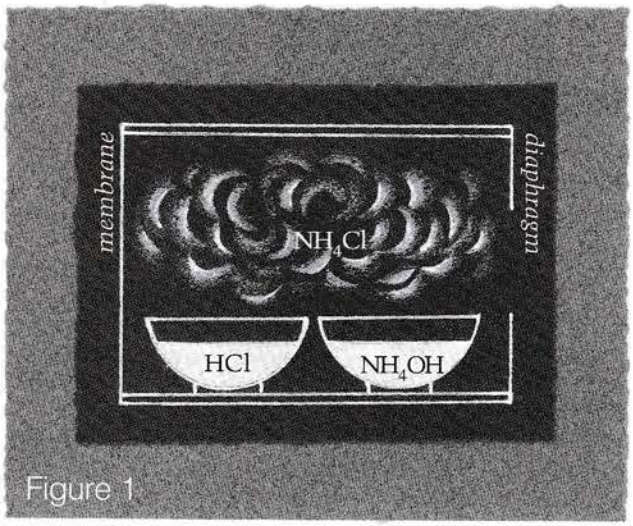


Figure 1

Diagrams by M. Konstantinova

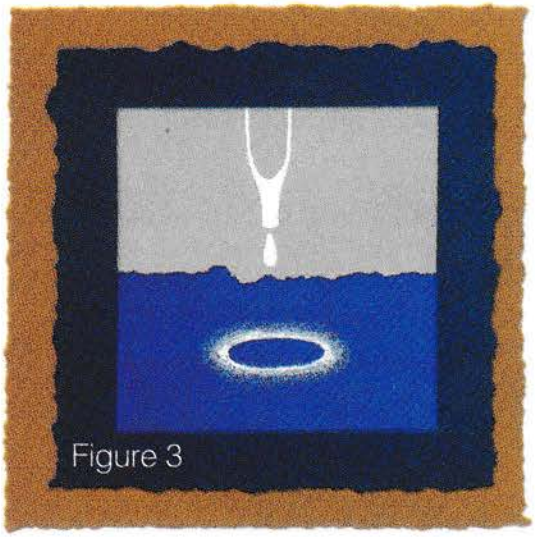


Figure 3

these cases is not perfect: the two types of rings look similar only for a short time after formation. After that, the behavior of the rings in water and air begins to differ.

Motion of the medium about the vortex rings

What is going on in the medium where a vortex ring is produced? Let's study this problem with some experiments.

Put a lighted candle 2–3 cm away from the diaphragm of the Tait's machine. Aim the smoke ring so that it passes near the flame. We see that the flame will either die or fluctuate vigorously. This indicates that more than the visible part of the ring is set in motion. The air layers near the visible ring are also moving.

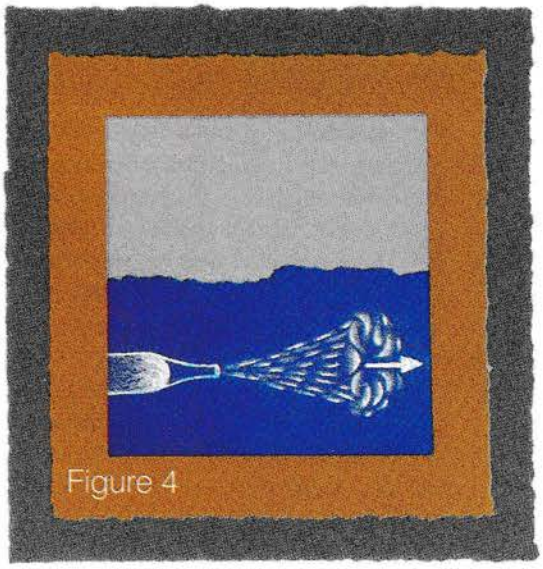


Figure 4

How do they move? Take two small rags soaked in hydrochloric acid (one rag) and ammonium hydroxide (another rag). Suspend the rags at a distance of 10–15 cm apart. Immediately the air between the rags will be filled with smoke (fumes of ammonium chloride). Now direct a smoke ring from the apparatus into this cloud of smoke. After passing the smoke cloud the ring will grow in size, and the cloud will be set into rotational motion. Therefore, the air near the vortex ring must be rotating (figure 5).

A similar experiment can be carried out with water. While slowly stirring the water in a glass, drop some ink into it. Stop stirring, and after a while "threads" of ink will form in the water. At this time, squeeze an ink ring into the water. When the ring passes near the threads, the ring will spin the threads around itself.

Vortex rings in water

Let's consider the evolution of vortex rings in water. As we saw above, if an ink drop falls from a height of 2–3 cm into an aquarium, it generates an ink vortex ring. What will happen to this ring; that is, how will it evolve?

Paradoxically, after a while the vortex ring will decay into several smaller rings, and these rings will also decay into smaller rings, etc. As a result, a beautiful "castle" will appear in the water (Figure 6).

We saw that the split-up of the first ring was preceded by the formation of bulbs on the ring, which later gave birth to the second generation of rings. How can this be explained? The outer parts of the ink vortex ring are constantly mixing with water. That means that the ring as a whole moves in a heterogeneous medium, so some parts of it get ahead of others. Since the ink is heavier than water, it spills into these advanced parts of the ring. At this moment the surface tension comes into play and produces the round bulbs. Shortly thereafter these bulbs become isolated drops of ink, which produce a new generation of vortex rings. The process repeats itself several times, producing a number of vortex generations. Alas, we could not find any regularity in this hydromechanical fission: in 10 experiments the number of rings in the fourth generation was too variable to draw quantitative conclusions. Perhaps our readers will have better luck in finding the mean number of aqueous vortex rings in each generation.

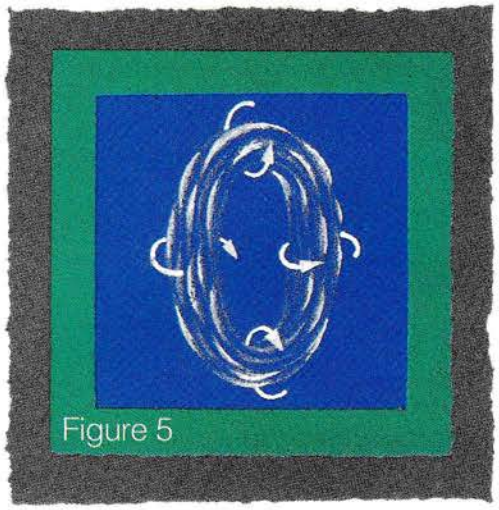


Figure 5

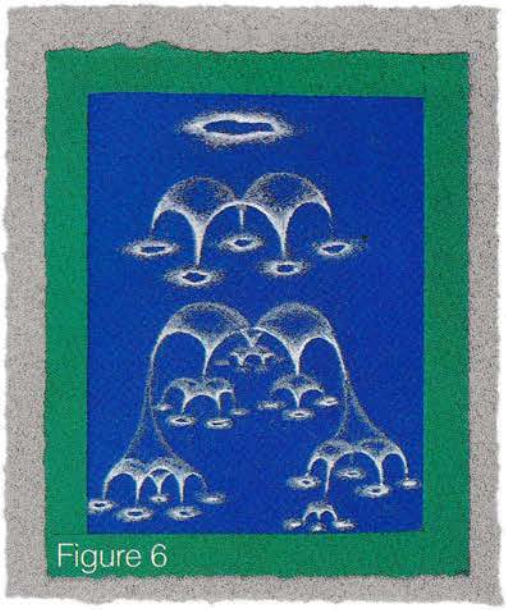


Figure 6

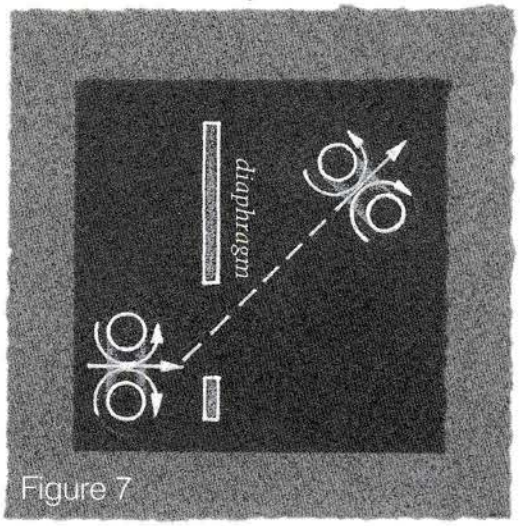


Figure 7

In other experiments we found that the existence of a vortex ring is possible only within some minimum "vital" volume. For this we set tubes of various diameters along the trajectory of aqueous vortex rings. When the diameter of the tube was slightly larger than the size of the ring, it decayed within the tube, and then a smaller ring was generated from its remnants. When the tube's diameter was about four times larger than that of the ring, the ring would pass unscathed through the tube. The vortex was practically free of any external influences.

Scattering of the smoke rings

We carried out several experiments on the interaction of smoke vortex rings with a plane and with diaphragms of various diameters. We refer to the processes occurring here as the *scattering* of vortex rings from obstacles.

Imagine a ring impinging on a diagram whose diameter is smaller than that of the ring. We consider two cases: 1) a central collision, which occurs when the ring's translation velocity is normal to the diaphragm and the ring's center travels through the center of diaphragm, and 2) an off-center collision, when the ring's center doesn't pass through the center of the diaphragm.

In the first case, the impinging ring is scattered, but another (smaller) ring is generated on the other side of the diaphragm. The mechanism of its generation is the same as in the Tait's machine: the air moving around the primary ring flows into the orifice and carries with it the smoke of the scattered vortex.

If the diameter of the diaphragm is equal to or somewhat larger than that of the ring, a central collision occurs in a similar way.

The off-center collision is far more interesting: in this case the newly generated ring travels at some angle to the initial trajectory (figure 7). Try to figure out why this happens.

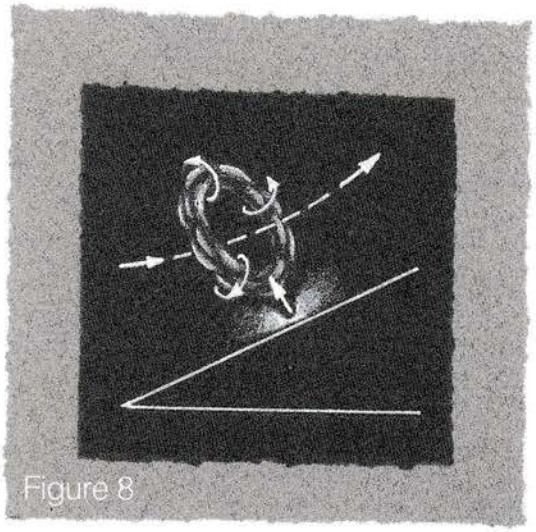


Figure 8

Now let's consider the interaction of a vortex ring with a plane. The experiments showed that if the obstructing plane is perpendicular to the velocity of the ring, the ring swells uniformly while maintaining its shape. We explain this phenomenon as follows: the flow of air inside the ring produces a high-pressure region that causes the entire ring to expand uniformly.

When the obstructing plane is inclined at an angle to its initial position, the incident vortex is repelled by it (figure 8). This repulsion can also be explained by the formation of a high-pressure region between the vortex ring and the plane.

Interaction of the vortex rings

Surely, the most interesting experiments were those in which two vortex rings interacted with each other. We conducted such experiments both in air and water.

Let an ink drop fall from a height of 1–2 cm and another from a slightly greater height of 2–3 cm. These drops produce two vortices in the water, which move with different speeds ($v_2 > v_1$). When the two rings are at the same depth, they start to interact with each other.

Three types of vortex interaction are possible. In the first case, the delayed ring passes the first ring without "touching" it (Figure 9a). Even so, the two rings interact. The flows of water in the two rings repel each other. In addition, there is a spillover of ink from the first ring to the second ring: the latter has a more in-

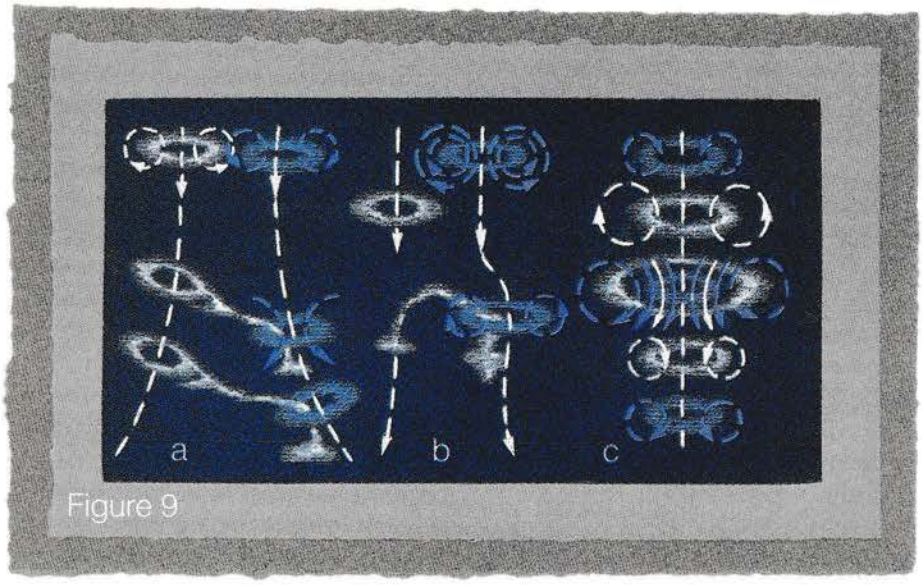


Figure 9

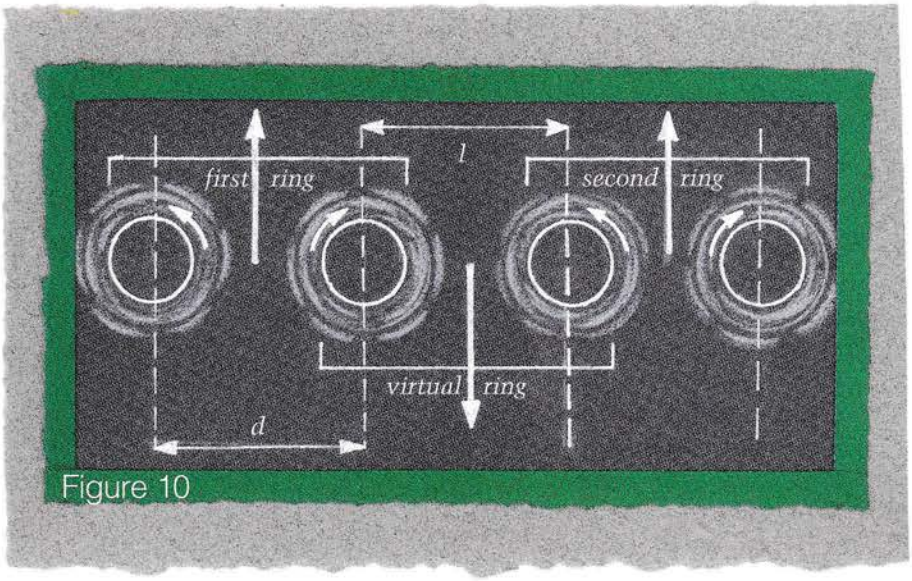


Figure 10

tense flow that entrains some ink from the first ring. Sometimes a portion of this ink passes through the second ring and produces a new ringlet. Then the rings begin to break up and we could no longer observe anything interesting.

In the second case, the second ring overtakes and touches the first ring (figure 9b). As a result, the more intense flows in the second ring destroy the first ring. As a rule, new small vortices are formed from the fragments of the first ring.

Finally, in the third case the rings collide centrally (figure 9c). The second ring passes through the first ring and shrinks in size, while the first ring expands. As in the previous cases, these changes result from the interaction of the flows of water around the two rings. As usual, the process culminates in the breakup of the rings.

The interaction of smoke rings was studied with the help of a modified Tait's machine that had two holes instead of one. It turned out that the outcome of an experiment was heavily dependent on the force and duration of the blow delivered to the membrane. In our setup the membrane was struck by a heavy pendulum.

When the distance l between the holes was smaller than the diameter of a hole ($l < d$), the two air streams intermixed and produced a single vortex ring. As a rule, no ring could be generated if $d < l < 1.5d$. In all other cases two rings were generated. When l was larger than $4d$, the rings did not "feel" each other. At $1.5d < l < 4d$ the rings initially drew together, and then (in some cases) they moved apart before disintegrating.

The attraction of the rings can be explained by the formation of a sort of "virtual ring" between the real ones, moving in the opposite direction (figure 10). Consequently, the planes of the real rings turn toward each other, and these rings draw together.

We could not explain what happens to the rings at the end of their short and turbulent life. ◙

Quantum on vortices and turbulence:

S. Kuzmin, "Spinning in a Jet Stream," September/October 1994, pp. 49–52.

L. Leonovich, "Fluids and Gases on the Move," January/February 1996, pp. 28–29.

J. Raskin, "Foiled by the Coanda Effect," January/February 1994, pp. 5–11.

A. Stasenko, "Airplanes in Ozone," May/June 1995, pp. 20–25.

A. Stasenko, "Whirlwinds Over the Runway," July/August 1997, pp. 36–39.

CONTINUED FROM PAGE 36

electric permeability. Later he experimentally proved the law of conservation of electric charge and came very near to discovering the law of energy conversion and conservation. In 1845 he discovered diamagnetism, which is the quality that causes a substance to be pulled out of a magnetic field, and in 1847 he observed paramagnetism—the property of substance that causes it to be drawn into a magnetic field. In addition, Faraday discovered the rotation of the plane of polarization in a magnetic field, which was the first evidence of the electromagnetic nature of light and laid the groundwork for magneto-optics, an entirely new branch of modern physics.

Finally, it was Michael Faraday who advanced the extremely fruitful concept of physical fields. According to Albert Einstein, the idea of a field was the most original of Faraday's achievements and the most important discovery since the time of Newton. All of Faraday's predecessors considered space a passive witness to the processes going on between the bodies and charges. By contrast, in Faraday's world space is an active participant in the physical game. "One must have had a powerful gift of scientific foresight," Einstein wrote, "to perceive that in describing electrical phenomena it is not the charges and particles that are essentially responsible for phenomena, but rather the space between the charges and particles." ◙

Quantum on the history of electromagnetism:

S. Filonovich, "The Modest Experimentalist, Henry Cavendish," January/February 1991, pp. 41–44.

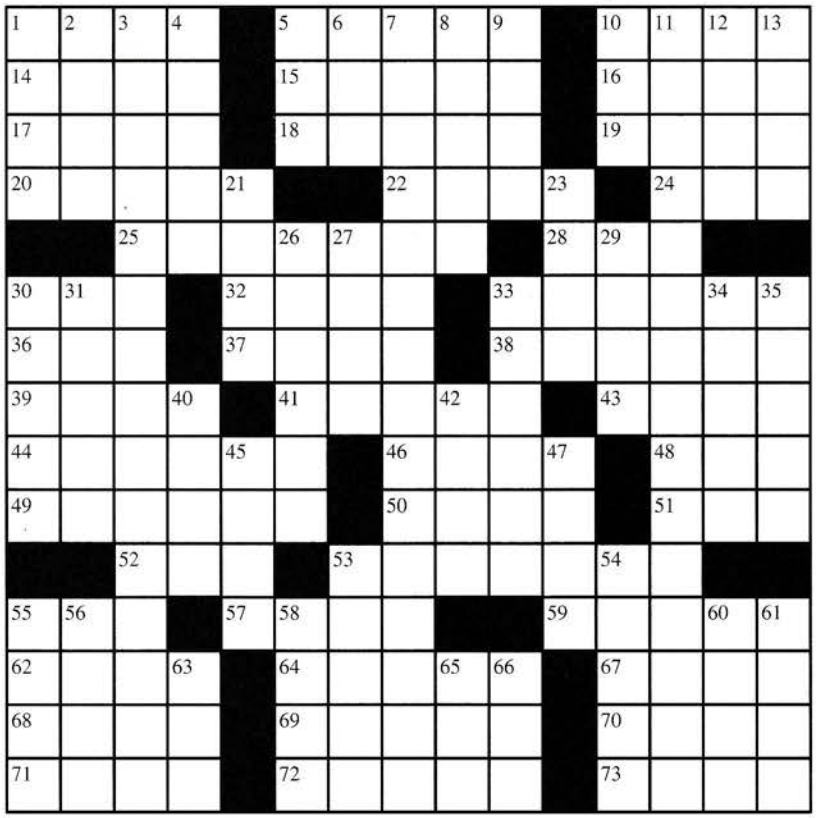
A. Byalko, "Backtracking to Faraday's Law," January/February 1994, pp. 20–23.

P. Bliokh, "The Advent of Radio," November/December 1996, pp. 4–9.

A. Leonovich, "Of Combs and Coulombs," January/February 1997, pp. 28–29.

Criss × cross science

by David R. Martin



ACROSS

1 Alphabet run
5 Out-of-date
10 Far and ___
14 Jai ___
15 Existed
16 Mature
17 64,203 (in base 16)
18 Real or virtual follower
19 Famous lion
20 Small children
22 Point of minimum disturbance
24 Egg layer
25 Meas. sys. based on m.k.s.
28 2814 (in base 16)
30 Trig. function
32 Female horse
33 Bores
36 Characterized by: suff.
37 Greek portico
38 Tooth material
39 Geophysicist Harry ___ (1859–1944)
41 Tear open
43 Aspen or Sequoia
44 Crab constellation
46 Some breads
48 Type of parity
49 White feldspar
50 Poet's listen
51 Tee's predecessor
52 10^{18}: pref.
53 Electron or proton
55 It's mostly Nitrogen
57 A purely quantum mechanical attribute
59 Buddhist scripture
62 Town near L. Albert
64 Solutions with low pH
67 "___ a woman" (Ray Charles' boast)
68 Asteroid
69 Throw
70 Durable cotton cloth
71 10^{-24} cm^2
72 Unit of magnetic flux density
73 ___ Domini

DOWN

1 Handle
2 "Now ___ me down to sleep."
3 1988 Physics Nobelist
4 Nisei's opposite
5 One lb./in.
6 Point
7 Electrolytic dissociation theorist
8 Lily bulbs
9 60,909 (in base 16)
10 100 m^2
11 1901 Physics Nobelist
12 Orbital point
13 Give birth to lambs
21 Adds
23 Work for
26 Physical world
27 Element in steel
29 Decree
30 About
31 Actor Ryan ___
33 Degauss
34 City near Manchester
35 Winter vehicles
40 609
42 Hebrew month
45 Zeta followers
47 Glides over snow
53 Microfilm
54 Seance board
55 43,947 (in base 16)
56 Very small amount
58 Alliance
60 Horse color
61 Periodic table abbr.
63 Columnist ___ Landers
65 650
66 Jamaican music

SOLUTION ON PAGE 56

SOLUTION TO THE MAY/JUNE PUZZLE

S	A	L	T	■	■	E	R	G	S	■	O	A	S	T
E	P	E	E	■	O	B	O	E	S	■	S	C	I	A
L	U	D	W	I	G	B	O	L	T	Z	M	A	N	N
A	S	S	■	N	E	A	T	■	■	E	I	C	H	■
■	■	■	D	E	E	D	■	M	A	R	U	■	■	■
L	A	S	E	R	S	■	■	A	T	O	M	I	C	■
A	D	U	L	T	■	G	R	I	T	S	■	M	E	V
M	D	L	I	■	O	R	A	L	S	■	B	A	L	I
B	E	L	■	A	D	A	M	S	■	D	I	G	I	T
■	R	A	D	I	A	N	■	■	L	I	N	E	A	R
■	■	■	E	M	Y	D	■	B	O	O	S	■	■	■
■	S	A	G	E	■	■	P	O	O	D	■	P	H	D
A	L	F	R	E	D	W	H	I	T	E	H	E	A	D
N	O	D	E	■	C	O	O	L	S	■	E	T	N	A
A	B	B	E	■	L	E	N	S	■	■	N	A	S	A

Wave interference

by L. Bakanina

WAVE INTERFERENCE CONSTITUTES A large and important area of physics. Research in this area has played a significant role in the development of optics, because the discovery of light interference was a strong argument in favor of its wavelike nature.

When two waves meet, the resulting oscillation obeys the superposition principle: the resulting oscillation is the sum of the oscillations caused by all of the individual waves. However, the interference pattern appears only in cases where the individual waves are coherent—that is, all of them have the same frequency and constant phase differences. What does this pattern look like?

Imagine a sinusoidal wave propagating in some direction in the plane. At a point located a distance d from the origin the wave produces oscillations described by the following formula:

$$a = A\cos\omega\left(t - \frac{d}{v}\right) = A\cos\left(\omega t - \frac{2\pi}{T}\frac{d}{v}\right) = A\cos\left(\omega t - \frac{2\pi}{\lambda}d\right),$$

where a is the value of the oscillating physical parameter at time t, A is the amplitude, ω is the cyclic frequency, $T = 2\pi/\omega$ is the oscillation period, and v is the wave speed (for electromagnetic waves in a vacuum v is equal to the speed of light c).

The sum of two coherent waves with the same amplitude is

$$a = a_1 + a_2 = A\cos\left(\omega t - \frac{2\pi}{\lambda}d_1\right) + A\cos\left(\omega t - \frac{2\pi}{\lambda}d_2\right)$$

$$= 2A\cos\left[\frac{2\pi}{\lambda}\frac{(d_2 - d_1)}{2}\right]\cos\left(\omega t - \frac{2\pi}{\lambda}\frac{(d_1 + d_2)}{2}\right).$$

and the amplitude of the resulting oscillation is

$$A_{\text{sum}} = 2A\left|\cos\frac{2\pi}{\lambda}\frac{(d_2 - d_1)}{2}\right| = 2A\left|\cos\frac{\Delta\phi}{2}\right|,$$

where $\Delta\phi = 2\pi(d_2 - d_1)/\lambda$ is the phase difference of the two waves. Depending on the phase difference $\Delta\phi$ (and thus the path difference $d_2 - d_1$) of the waves, the resulting amplitude can vary from $A_{\text{sum max}} = 2A$ (when the phases coincide) to $A_{\text{sum min}} = 0$ (when the phases differ by π).

The human eye (or a photodetector) perceives not the amplitude of the oscillation but its intensity I, which is the energy incident on a unit area per unit time. Both the energy and the intensity are proportional to the square of the amplitude, so

$$I_{\text{sum}} \sim A_{\text{sum}}^2 \sim 4A^2\cos^2\frac{\Delta\phi}{2}.$$

Thus there are points (nodes) where the total intensity is greater than the sum of the intensities of both superimposed waves ($I_{\text{sum max}} \sim 4A^2$) and points (antinodes) where the waves nullify each other ($I_{\text{sum min}} = 0$). This spatial redistribution of energy is a characteristic feature of wave interference.

The phenomenon of interference can easily be demonstrated with waves in water or with radio waves. In contrast, in optics it's not easy to produce an interference pattern, because ordinary light sources (and not, say, lasers) emit light composed of waves with rapidly and randomly varying phases. This type of light and their sources are called incoherent. However, if we break the light emitted by such a source into two pencils of light and then superpose them, we can observe a rather clear interference pattern.

Note that in order to describe the interference pattern correctly, we need to know the phase difference with an accuracy far better than π; otherwise the maxima and minima cannot be discriminated (or resolved, as the physicists say). Correspondingly, the error in measuring the path difference of the two waves must be much less than the wavelength λ (for light waves it must be far less than 0.1 mm).

To familiarize ourselves with the phenomenon of wave interference, let's consider some problems given

Art by Tomas Bunk


THERE IS A CRACK IN EVERYTHING – THAT'S HOW THE LIGHT GETS IN · (L. COHEN)
TOM BUNK

on the physics exams at the Moscow Institute of Physics and Technology.

Problem 1. *A plane electromagnetic wave with frequency* ν *emitted by a funnel-shaped antenna falls perpendicularly on a flat reflecting screen. Find the amplitude of the reflected wave if an electric field-intensity meter shows maximum and minimum field amplitudes* A_1 *and* A_2*, respectively, as it moves between the antenna and the screen. In addition, determine the distance between two adjacent maxima of the field.*

Solution. Superposition of the incident and reflected waves occurs in the space between the antenna and the screen. At the points of maximum intensity the phases of these two waves coincide, so the resulting oscillation has the following amplitude:

$$A_1 = A_{\text{inc}} + A_{\text{ref}}.$$

At the points of minimum intensity the incident and reflected waves have opposite phases, so the total amplitude is

$$A_2 = A_{\text{inc}} - A_{\text{ref}}.$$

These equations yield

$$A_{\text{ref}} = (A_1 - A_2)/2.$$

Let's describe the oscillations at a point with coordinate x induced by the incident wave as

$$a_{\text{inc}} = A_{\text{inc}}\cos\omega\left(t - \frac{x}{c}\right) = A_{\text{inc}}\cos\left(\omega t - \frac{2\pi}{\lambda}x\right).$$

If the distance between the antenna and the screen is l, the reflected wave travels the distance $2l - x$ to this point. Therefore, the oscillations induced by this wave are described by

$$a_{\text{ref}} = A_{\text{ref}}\cos\omega\left(t - \frac{2l - x}{c}\right) = A_{\text{ref}}\cos\left(\omega t + \frac{2\pi}{\lambda}x - \frac{4\pi l}{\lambda}\right).$$

(Reality, as usual, is more complicated: depending on the properties of the screen, reflection may preserve the phase of the incident wave or change it by π; however, this won't affect our reasoning below.)

Now denote the coordinates of two neighboring maxima by x_1 and x_2. The phase differences of the oscillations induced by the incident and reflected waves at these points are

$$(\Delta\phi)_1 = \left(\omega t - \frac{2\pi}{\lambda}x_1\right) - \left(\omega t + \frac{2\pi}{\lambda}x_1 - \frac{4\pi}{\lambda}l\right) = -\frac{4\pi}{\lambda}x_1 + \frac{4\pi}{\lambda}l$$

and

$$(\Delta\phi)_2 = -\frac{4\pi}{\lambda}x_2 + \frac{4\pi}{\lambda}l,$$

respectively. The phases of the resulting oscillations at two adjacent maxima must differ by 2π. Therefore,

$$\Delta\phi = (\Delta\phi)_1 - (\Delta\phi)_2 = 2\pi,$$

or

$$\frac{4\pi}{\lambda}(x_2 - x_1) = 2\pi,$$

from which we get

$$x_2 - x_1 = \frac{\lambda}{2} = \frac{c}{2\nu}.$$

Problem 2. *A point source of light S is located at the focal point of a lens. A symmetric prism with interior angle* $\alpha = 0.01$ *rad and width* $D = 6$ cm *is situated behind the lens (figure 1). At what distance L from the prism can the largest number of interference bands be observed? How many bands can be seen on the screen? What is the width of the bands? The refractive index of the prism glass is* $n = 1.5$*, and the wavelength of the light is* $\lambda = 0.5$ *μm.*

Solution. Since the light source is located at the focal point of the lens, a parallel beam of light lands on the prism along its optic axis (figure 2). The prism splits this beam into two beams traveling at an angle β with the optic axis, which we can find using Snell's law $\sin\alpha/\sin\beta = n$. Since the angles are small,

$$\beta \approx \sin\beta = \frac{\sin\alpha}{n} \approx \frac{\alpha}{n}.$$

It's clear that the largest width of the interference pattern will be at the place corresponding to the great-

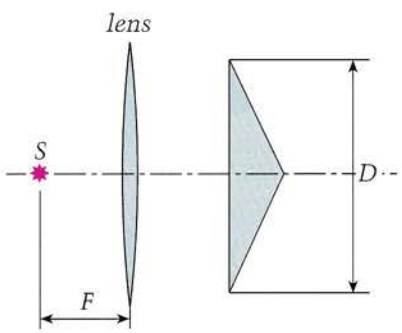


Figure 1

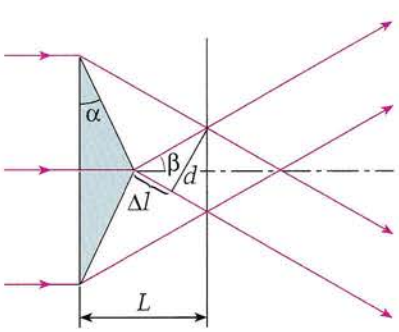


Figure 2

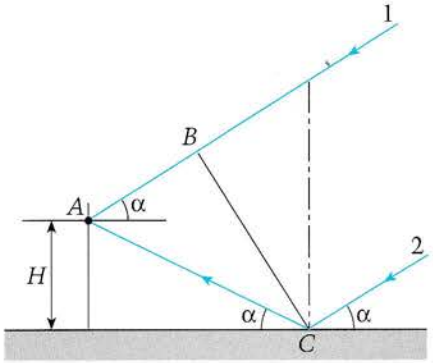


Figure 3

est cross-sectional area of the interfering beams. Figure 2 shows that this distance is

$$L \approx \frac{D}{4\tan\beta} \approx \frac{D}{4\beta} \approx \frac{Dn}{4\alpha} \approx 112.5 \text{ cm}.$$

The width of an entire interference pattern at this distance is $b = D/2 = 3$ cm.

The width of each crossing beam is $d = (D/2)\cos\beta \cong D/2$. The angle between the wave fronts of the beams is 2β, so the maximum path difference of the beams is

$$\Delta l = d \tan 2\beta \cong d \cdot 2\beta.$$

The path difference corresponding to two neighboring maxima (or minima) equals the wavelength λ. Thus the number of interference bands on the screen is

$$N = \frac{\Delta l}{\lambda} \approx \frac{2d\beta}{\lambda} \approx \frac{D\alpha}{n\lambda} \approx 800,$$

and the width of a single band is

$$b_0 = \frac{b}{N} \approx \frac{D/2}{D\alpha/(n\lambda)} \approx \frac{n\lambda}{2\alpha} \approx 37.5 \text{ μm}.$$

Problem 3. *A radio receiver that tracks the appearance of a satellite beyond the horizon is located on the shore of a lake at a height $H = 3$ m above the water's surface. As the satellite rises above the horizon, periodic changes in signal intensity are observed. Find the frequency of the radio wave emitted by the satellite if the intensity maxima are observed at angular elevations of the satellite above the horizon of $\alpha_1 = 3°$ and $\alpha_2 = 6°$. The surface of a lake can be considered an ideally reflecting mirror.*

Solution. The receiver detects both the ray traveling directly from the satellite and the ray reflected by the lake. In figure 3 these are rays 1 and 2, respectively. By drawing the wave front BC, which is perpendicular to both rays, we get the path difference between the incident and reflected rays:

$$\Delta l = |AC| - |AB| = \frac{H}{\sin\alpha} - \frac{H}{\sin\alpha}\cos 2\alpha.$$

Since all the angles in this problem are small, we have $\sin\alpha \cong \alpha$ and $\cos\alpha \cong 1 - \alpha^2/2$, so

$$\Delta l = \frac{H(1-\cos 2\alpha)}{\sin\alpha} \approx 2H\alpha.$$

The maximum intensity is observed when the path difference of the interfering rays is equal to an integer number of wavelengths:

$$2H\alpha_1 = k\lambda \text{ and } 2H\alpha_2 = (k+1)\lambda.$$

This equation yields the wavelength and frequency of the satellite's radio signal:

$$\lambda = 2H(\alpha_2 - \alpha_1)$$

and

$$\nu = \frac{c}{\lambda} = \frac{c}{2H(\alpha_2 - \alpha_1)} \approx 10^9 \text{ Hz}.$$

Exercises

1. A funnel antenna emits a plane electromagnetic wave with frequency $\nu = 9.4 \cdot 10^9$ Hz in the direction perpendicular to a reflecting screen. An electric field intensity meter establishes the result of interference of the incident and reflected waves as they move between the antenna and the screen. Determine the minimum thickness of a dielectric plate attached flat against the screen such that the meter shows a minimum electric field intensity if, before the plate is attached, it was located at a spot with maximum intensity. The refractive index of the plate is $n = 1.4$; ignore absorption by the plate and reflection from it.

2. A shortwave transmitter operates at a frequency $\nu = 30$ MHz. A receiver is located at a distance $L = 2{,}000$ km from it. Radiowaves reach the receiver after being reflected by ionospheric layers at altitudes of $h_1 = 100$ km and $h_2 = 300$ km. Find the equation describing how the intensity of the signal changes as the receiver is moved along the line connecting it to the transmitter. The displacement of the receiver is small compared to L.

3. Radiowaves from a star situated in the plane of the equator are received by two antennas located on the equator and separated by a distance $L = 200$ m. Signals from the antenna are sent along cables of equal length to a receiver. Find the equation describing the change in the amplitude of the voltage in the receiver's input circuit as a consequence of the Earth's rotation. Reception occurs at a wavelength $\lambda = 1$ m. During the time of observation, the star remains very close to the zenith. ■

Quantum on light diffraction and interference:

P. V. Bliokh, "Make Yourself Useful, Diana," March/April 1992, pp. 34–39.

V. A. Fabrikant, "Vavilov's Paradox," July/August 1992, pp. 49–50.

A. Eisenkraft and L. D. Kirkpatrick, "Rising Star," September/October 1994, pp. 44–47 and March/April 1995, pp. 37–38.

N. M. Rostovtsev, "Behind the Mirror," January/February 1996, pp. 37–38.

A. Eisenkraft and L. D. Kirkpatrick, "Do You Promise Not to Tell?" January/February 1997, pp. 30–32 and July/August 1997, pp. 32–32.

A. Eisenkraft and L. D. Kirkpatrick, "Color Creation," November/December 1997, pp. 32–33.

D. Panenko, "Diffraction in Laser Light," March/April 1999, pp. 33–35.

V. Surdin and M. Kartashev, "Light in a Dark Room," July/August 1999, pp. 40–44.

A. Stasenko, "Physical Optics and Two Camels," September/October 1999, pp. 44–47.

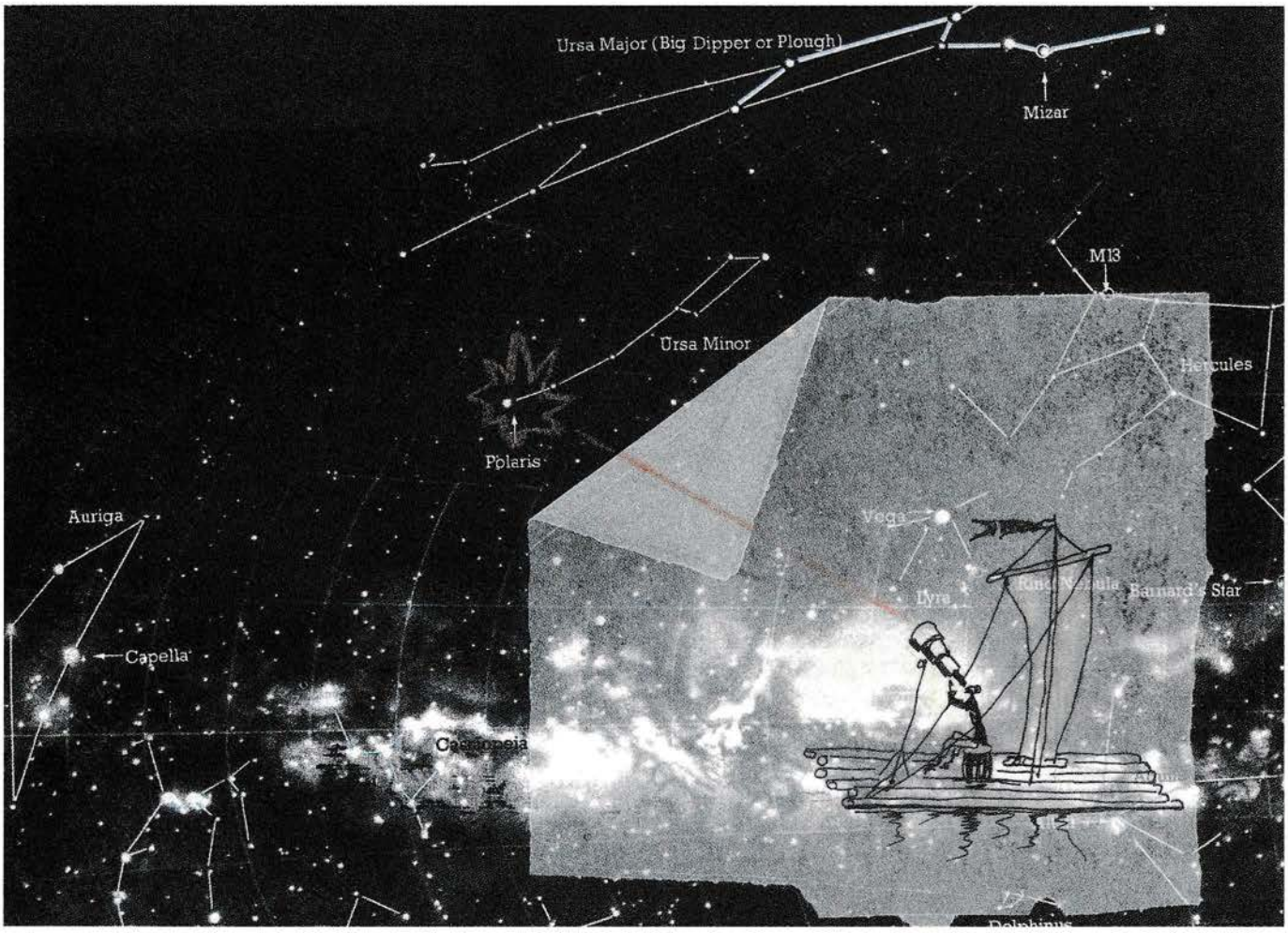
A. Vasilyev, "Ernst Abbe and 'Carl Zeiss,'" July/August 2000, pp. 46–49.

Giving astronomy its due

by A. Mikhailov

THE SIGNIFICANCE OF ASTRONOMY and its role in human history are often underestimated. Everyone allows that astronomy helped develop rules for measuring time and orienting ourselves on the Earth's surface, and that it discovered much of interest in the firmament. But the consensus seems to be that this isn't as important as what other sciences—say, physics and chemistry—have given us. Only in our age of space exploration has the value of astronomy become more obvious. And yet one might argue that if it hadn't been for astronomy, our history would have been quite different indeed.

Imagine if the entire sky were constantly blocked by thick clouds, so that we couldn't see the Sun, Moon, planets, and stars. Day would still alternate with night, and we would notice that it was brighter on some nights than others, and that this phenomenon was somehow related to the tides. We



Art by Yuri Vaschenko

would dream up all sorts of clever theories, most of them far from reality.

Thanks to astronomy, more than two thousand years ago we learned that the Earth is a sphere. This realization came from measuring changes in the midday height of the Sun and the culmination height of the stars as an observer moved along a meridian. This idea was also corroborated by the round shadow of the Earth on the Moon during lunar eclipses. The first known successful attempt to calculate the circumference of the Earth was made by Eratosthenes of Cyrene in the third century B.C. If we couldn't see the heavenly bodies, this discovery might not have been made. We might still think the Earth is flat and hemmed in by the ocean on all sides.

Some might object that the convexity of the sea and land, which can be seen when objects recede on open water or on a broad, featureless plain, is proof that the Earth is round. However, other observations may lead to the opposite conclusion—that the Earth is concave. Indeed, if we climb up to a high spot, we don't perceive the horizon as dropping off at the "edges," and the lowlands around us look like a depression in the Earth's surface. We get the impression that the Earth is like an overturned bowl—the uneven bottom is the dry land, while the sides, bent downward, are covered by the ocean.

If people hadn't known about the sphericity of the Earth, history would be quite different. America wouldn't have been discovered, in all likelihood, because seafaring would have been stunted. In ancient times, people sailed near the shore—why plunge deep into the sea and tempt fate? The sea route from Europe to India and its treasures hugged the shores of Africa and rounded its southern tip. The idea of reaching the lands of East Asia by traveling west would seem absurd if sailors were unaware that the surface of the Earth was "closed."

Could there have been some other way of discovering the sphericity of the Earth without astronomical observations? In principle, yes, there were such methods—but they were nearly impossible to implement. One such method is based on measuring how the horizon drops when observed from a high vantage point; another involves measuring the spherical excess in the angles of large triangles on dry land.

The way the horizon drops off can be depicted geometrically as in figure 1. Imagine that we're situated at point A at a height h above the surface of a sphere of radius R. Draw the tangent line to the sphere from point A to point B, which lies on the horizon. The angle BAE, which is equal to the angle α formed by the tangent line and an imaginary plane with point B on its "horizon," is the dip of the horizon at height h. Figure 1 shows that $|AB| = d = R \tan \alpha$. According to the Pythagorean theorem,

$$(R + h)^2 = R^2 + d^2 = R^2 + R^2 \tan^2 \alpha,$$

from which we get (neglecting the small factor h^2)

$$\tan \alpha = \sqrt{2h/R}.$$

Since angle α is small, $\tan \alpha \cong \alpha$, so

$$\alpha = \sqrt{2h/R} \text{ (rad)}.$$

Clearly if α is constant for a given height h at any spot on Earth, this means R = const—that is, the Earth is a sphere. Therefore, by measuring α we can discover the sphericity of the Earth. However...

If we plug the value R = 6,370 km into the last equation and convert radians to minutes of arc (1 rad = 3437.7′), we get the following result for the Earth:

$$\alpha \approx 60'.91\sqrt{h},$$

where h is measured in kilometers.

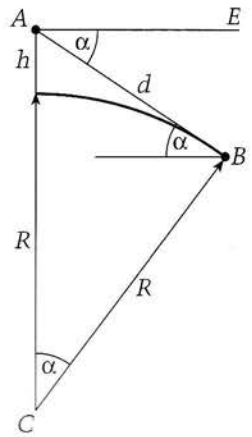


Figure 1

Assume that we perform our geodesic measurements at a height h = 1 km above sea level. In this case, our formula for the dip of the horizon yields almost 61′, which is measurable even with a small theodolite. However, in reality it's not easy to measure the dip of the horizon. This is because the rays of light are refracted when they pass through atmospheric layers of different densities, so the line of sight AB will not be straight. Usually it's curved, with its concavity directed downward. Therefore, in reality the theodolite will measure not the dip of the horizon but some angle that may differ from the correct value α by 20 percent or more. This way of measuring the Earth's curvature is very unreliable, and it cannot prove the sphericity of the Earth. Moreover, it can be implemented only on islands with high mountains, because only in such places is the horizon an uninterrupted line and the altitude can be measured reliably.

As noted previously, another method of supposedly detecting the Earth's sphericity is based on measurements of the spherical excess in the angles of a large-scale triangle. The sum of all the angles of a triangle whose sides are the arcs of the great circles of the sphere is larger than 180 degrees. Stereometry says that the difference (spherical excess) is

$$\varepsilon = 206265 \cdot S/R^2 \text{ angular seconds},$$

where S is the area of the triangle and R is the radius of the sphere. Assume that we could measure the angles of a huge equilateral triangle on the Earth with sides 100 km long. The area of this triangle is 4,330 km^2, and the spherical excess is 22″. This value can be "captured" by a good theodolite, but such instruments were created for astronomical purposes; for "rough" terrestrial measurements, such precision wasn't

CONTINUED ON PAGE 49

The Three Chords theorem

by Shikong Le and Lioukan Chen

THE GREEK MATHEMATICIAN and astronomer Ptolemy (ca. 90–168) found a theorem in geometry that bears his name. This theorem consists of two parts.

Ptolemy's theorem, Part I: If quadrilateral $ABCD$ is inscribed in a circle (figure 1), then

$$AB \cdot CD + AD \cdot BC = AC \cdot BD. \quad (1)$$

Ptolemy's theorem, Part II: If quadrilateral $ABCD$ cannot be inscribed in a circle, then

$$AB \cdot CD + AD \cdot BC > AC \cdot BD. \quad (2)$$

Ptolemy's theorem, discovered 1,800 years ago, is still considered a great result and a treasure of ancient mathematics. The Swiss mathematician Leonhard Euler (1707–1783) came up with a similar theorem: if points A, B, C, D lie on a straight line, then

$$AB \cdot CD + AD \cdot BC = AC \cdot BD.$$

Comparing Euler's and Ptolemy's theorems, we notice an interesting fact. If a straight line is viewed as a circle whose radius is infinitely long, then four points on the straight line can be seen as a circle. Euler's theorem would thus be a special case of the first part of Ptolemy's theorem.

There are many instances where the two theorems are compatible. Let's look at another.

Example. Suppose $ABCD$ is a convex quadrilateral inscribed by a circle. Then

$$\frac{AC}{BD} = \frac{BA \cdot AD + BC \cdot CD}{AB \cdot BC + AD \cdot DC}. \quad (3)$$

Proof. Suppose R is the circumradius of quadrilateral $ABCD$. We first show that

$$AB \cdot BC + AD \cdot DC = 2R \cdot BE + 2R \cdot DF \quad (4)$$

In figure 2, $BE \perp AC$ and $DF \perp AC$, we draw a diameter CM and line BM. Then we have $\angle CBM = 90°$, $\angle M = \angle BAC$. Therefore,

$$BC = CM \sin \angle M = 2R \sin \angle BAC.$$

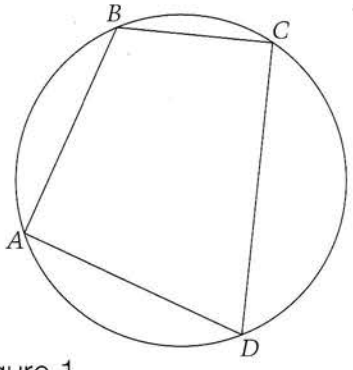


Figure 1

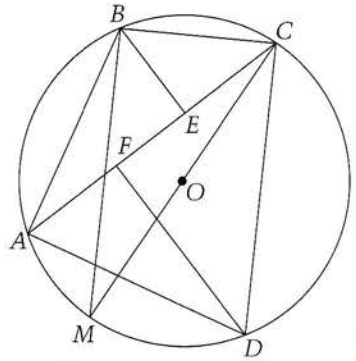


Figure 2

Similarly, by drawing line DM, we can show that

$$DC = 2R \sin \angle DAC.$$

From right triangles BEA, DAF, we have

$$BE = AB \sin \angle BAC,$$
$$DE = AD \sin \angle DAC.$$

Using these four equations, we have

$$AB \cdot BC + AD \cdot DC$$
$$= AB \cdot 2R \sin \angle BAC + AD \cdot 2R \sin \angle DAC$$
$$= 2R \cdot BE + 2R \cdot DF.$$

Now we have:

$$AC(AB \cdot BC + AD \cdot DC) = AC(2R \cdot BE + 2R \cdot DF) = 2R(AC \cdot BE + AC \cdot DF) = 4R(S_{\triangle ABC} + S_{\triangle ACD}) = 4R \cdot S_{ABCD}.$$

In fact,

$$BD(BA \cdot AD + BC \cdot CD) = 4R \cdot S_{ABCD}.$$

Thus $AC(AB \cdot BC + AD \cdot DC) = BD(BA \cdot AD + BC \cdot CD)$, and expression (3) follows.

Now we can play with result (3) a bit. Let $R \to \infty$ (thus $\angle B \to 180°$, $\angle C \to 180°$, $\angle A \to 0°$, $\angle D \to 0°$). Then the convex quadrilateral $ABCD$ turns into four points on a line. But expression (3) remains true. Thus we derive a new theorem from the example above.

Theorem: Let points A, B, C, D lie on a straight line in sequence. Then expression (3) is true.

Recently, Minghuai Hou, a teacher in Liaoning province, found a new

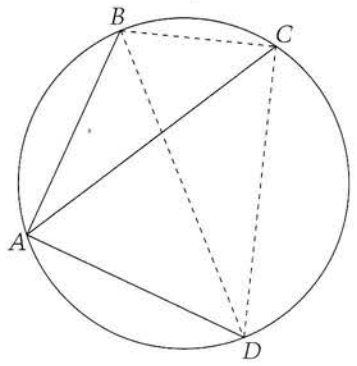


Figure 3

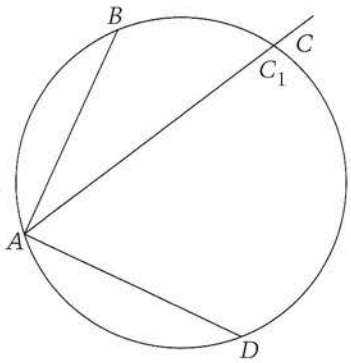


Figure 4

theorem, which we will call the Three Chords theorem.

Let point A lie on a circle, and let AB, AC, and AD be three chords of that circle (figure 3). Then

$$AB \sin \angle CAD + AD \sin \angle BAC = AC \sin \angle BAD. \quad (5)$$

It was reported that this theorem won a gold medal from an international organization. After studying it, we have concluded that this Three Chords theorem is equivalent to the first part of Ptolemy's theorem. Indeed, if we connect BC and DC, we get a quadrilateral $ABCD$ inscribed in a circle. Therefore, we can apply Ptolemy's theorem to see that expression (1) is true. Using the law of sines in $\triangle ABC$, we have

$$\frac{BC}{\sin \angle BAC} = \frac{CA}{\sin \angle CBA} = \frac{AB}{\sin \angle ACB} = 2R.$$

where R stands for the radius of the circle in figure 3. Thus

$$BC = 2R \sin \angle BAC. \quad (6)$$

Similarly, in $\triangle ACD$ and $\triangle ABD$, we get

$$CD = 2R \sin \angle CAD \quad (7)$$

and

$$BD = 2R \sin \angle BAD, \quad (8)$$

respectively. Substituting these expressions into (1), we get

$$AB \cdot 2R \sin \angle CAD + AD \cdot 2R \sin \angle BAC = AC \cdot 2R \sin \angle BAD. \quad (9)$$

Equation (5) follows if we divide by $2R$.

So the Three Chords theorem can be derived from the first part of Ptolemy's theorem. Conversely, the first part of Ptolemy's theorem can be derived from the Three Chords theorem. Indeed, we can easily derive equation (9) from equation (5). With the aid of equations (6), (7), and (8), equation (1) follows. Therefore, the Three Chords theorem is equivalent to the first part of Ptolemy's theorem.

Using Ptolemy's theorem, we can prove the converse of the Three Chords theorem.

Theorem. Let line segments AB, AC, AD, and the angles they form satisfy equation (5). Then points A, B, C, D are on the same circle—that is, line segments AB, AC, and AD are chords of the same circle (figure 4).

Proof. Draw a circle through points A, B, D. It intersects ray AC at point C_1. We shall prove that point C_1 coincides with point C.

From Ptolemy's theorem we have shown that

$$AB \sin \angle C_1AD + AD \sin \angle C_1AB = AC_1 \sin \angle DAB$$

—that is,

$$AB \sin \angle CAD + AD \sin \angle CAB = AC_1 \sin \angle DAB. \quad (10)$$

Comparing equation (10) with equation (5), we have $AC_1 \sin \angle BAD = AC \sin \angle BAD$. Obviously $\sin \angle BAC > 0$, $\sin \angle CAD > 0$, so equation (5) shows $\sin \angle BAD > 0$. Therefore, $AC_1 = AC$—that is, point C_1 coincides with point C.

As we've noted, the Three Chords theorem is an equivalent form of the first part of Ptolemy's theorem, but the Three Chords theorem is much briefer and easier to apply. For this we are grateful to Mr. Hou. ■

CONTINUED FROM PAGE 47

necessary. So the second method of proving the Earth is round could not have been realized.

Our analysis shows that the sphericity of the Earth could not have been discovered without astronomical observations. Without astronomy the history of the world would be different, and science and technology would suffer great losses. The biggest loss would be the law of universal gravitation discovered by Isaac Newton at the end of the 17th century on the basis of astronomical observations by Tycho Brahe and Johannes Kepler. This fundamental law underlies all the exact sciences, and many fields of science could not have developed without it: mechanics, the theory of magnetism and electricity, aviation—to say nothing of space flight. The theory of relativity would surely be out of the question.

The speed of light was first measured by the Danish astronomer Ole Rømer (1644–1710) in 1676 and resulted from his observations of Jupiter's moons. It's likely that the speed of light could have been measured experimentally without astronomical observations, but it would have happened much later, and this would have slowed the development of optics. Even chemistry was enriched by astronomy: recall that helium was discovered first on the Sun and only later found here on Earth. Astronomy also posed a number of mathematical problems whose solutions advanced this field of endeavor.

Perhaps someone will say, "Radio astronomy isn't bothered by your hypothetical cloud cover. It can even tell us about celestial bodies that we can't see." But radio astronomy has been around for only seventy years or so, and it appeared and developed thanks to its "granddad"—optical astronomy. ■

Card party

by Don Piele

IT'S TIME ONCE AGAIN TO HOST THE ANNUAL card party, where six couples will gather together to enjoy a night of cards. From painful past experiences, I have learned to avoid putting couples together at the same table. It's no fun watching friends leave the party not speaking to each other. To avoid this I will distribute the twelve guests around three tables so that each table has two women and two men but no couples. I wonder how many ways I can do this?

One way to model this problem is to number the tables from 1 to 3; if a women is placed at table *i* write out a slip for her partner with her table number on it. Thus, the men will all end up with slips numbered {1, 1, 2, 2, 3, 3}. If a man with slip 1 sits at table 1, he will be sitting with his partner. Any complete permutation of the slip numbers will result in a seating arrangement where no man sits with his own partner. A complete permutation (sometimes called a derangement) is any permutation where all numbers are not in their original positions. For example, {2, 2, 3, 3, 1, 1} is a complete permutation of {1, 1, 2, 2, 3, 3}, but {2, 2, 1, 3, 3, 1} is not, because one 3 has not moved.

First approach

Let's begin our investigation by constructing a function that will build all complete permutations. First, we need a function to generate the list of table numbers {1, 1, 2, 2, ..., *n*, *n*} for any *n*, in case we want to add more than three tables.

```
slipNumbers[n_]:=Flatten[Table[i,{i,n},{2}]]
slipNumbers[3]
{1,1,2,2,3,3}
```

Next, let's construct a predicate that detects if one list is a complete permutation of another by seeing if the difference of the lists contains any zeros. **FreeQ** tests to see if the difference is free of zeros.

```
completePermutationQ[L1_,L2_]:=FreeQ[L1-L2,0]
```

Let's do a test.

```
completePermutationQ[{1,1,2,2,3,3},{2,2,3,3,1,1}]
True
```

To generate all permutations we will use *Mathematica*'s built-in **Permutations** function. Let's see how many permutations there are for {1, 1, 2, 2, 3, 3}.

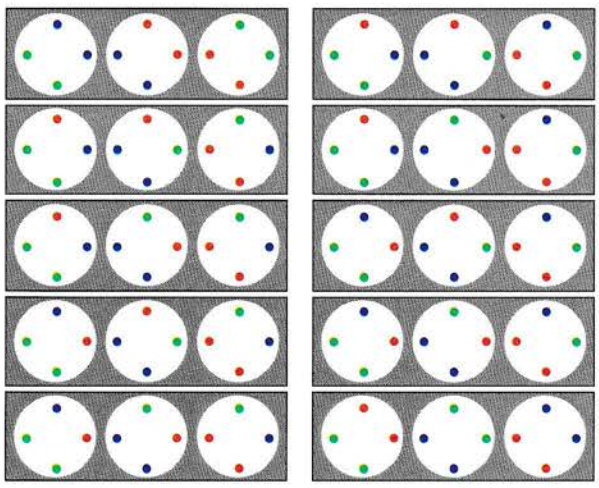
```
allPermutations=Permutations[{1,1,2,2,3,3}];
%//Shallow
{{1,1,2,2,3,3},{1,1,2,3,2,3},{1,1,2,3,3,2},{1,1,3,2,2,3},
{1,1,3,2,3,2},{1,1,3,3,2,2},{1,2,1,2,3,3},{1,2,1,3,2,3},
{1,2,1,3,3,2},{1,2,2,1,3,3}, <<80>>}
```

Only a few of the 90 total permutations are shown. Of course, it is well known that the way to compute the number of permutations of six objects if three of the objects have duplicates is `6!/(2!2!2!)`.

Next we select from all permutations those that are complete permutations of {1, 1, 2, 2, 3, 3}.

```
Select[allPermutations,
completePermutationQ[#,{1,1,2,2,3,3}]&]
{{2,2,3,3,1,1},{2,3,1,3,1,2},{2,3,1,3,2,1},{2,3,3,1,1,2},
{2,3,3,1,2,1},{3,2,1,3,1,2},{3,2,1,3,2,1},{3,2,3,1,1,2},
{3,2,3,1,2,1},{3,3,1,1,2,2}}
```

The answer to my original question of how many ways I can seat the fellows with different partners than their mates is 10. The graphic below shows the 10 arrangements with the women sitting in seats South and West and the men in seats North and East. Notice there are never more than two of a given color at any table.



Let's put our ideas together into a function that will derive the arrangements for n tables.

```
completePermutations[n_]:=
Module[{start=slipNumbers[n]},
  Select[Permutations[start],
completePermutationQ[#,start]&]]
```

The complete permutations for four tables is much higher at 297.

```
completePermutations[4]//Shallow
```

```
{{2,2,1,1,4,4,3,3},{2,2,1,3,4,4,1,3},{2,2,1,3,4,4,3,1},
{2,2,1,4,1,4,3,3},{2,2,1,4,4,1,3,3},{2,2,3,1,4,4,1,3},
{2,2,3,1,4,4,3,1},{2,2,3,3,4,4,1,1},{2,2,3,4,1,4,1,3},
{2,2,3,4,1,4,3,1},<<287>>}
```

The problem with this method of counting the complete permutations for n tables is that all permutations of $2n$ objects need to be constructed. This gets out of hand fast, and the highest number we can reach in a reasonable amount of time is $n = 5$.

```
Table[{i,Length[completePermutations[i]]},
{i,2,5}]
```

$$\begin{pmatrix} 2 & 1 \\ 3 & 10 \\ 4 & 297 \\ 5 & 13756 \end{pmatrix}$$

Second approach

One way to count the number of arrangements without completely enumerating them is to use generating functions. Suppose the card party has three tables: A, B, and C. The first empty seat at the first table can be filled by a person whose mate is at table B or C, which we represent by the factor $(b + c)$. The second empty seat at table A has the same restriction, so we repeat the term $(b + c)$. Each empty seat at table B can be filled by a person whose mate is at table A or C, and so we represent each seat at table B by a factor $(a + c)$. After continuing for table C and multiplying the factors, we get the following expression:

```
Expand[(b + c)^2 (a + c)^2 (a + b)^2]
```

$$b^2a^4 + c^2a^4 + 2bca^4 + 2b^3a^3 + 2c^3a^3 + 6bc^2a^3 + 6b^2ca^3 + b^4a^2 + c^4a^2 + 6bc^3a^2 + 10b^2c^2a^2 + 6b^3ca^2 + 2bc^4a + 6b^2c^3a + 6b^3c^2a + 2b^4ca + b^2c^4 + 2b^3c^3 + b^4c^2.$$

The coefficient of $a^2b^2c^2$ in the expression above is the number of ways to seat six people so that no person is sitting at the same table as his or her partner.

```
Coefficient[%, a^2 b^2 c^2]
```

10

If we move up to four tables and denote the last table by D, then we can pick a man from table B, C, or D for each of the two seats at table A. Continuing the same argument as used above for three tables, we have the corresponding expression

```
Expand[(b + c + d)^2 (a + c + d)^2 (a + b +
d)^2 (a + b + c)^2];
```

The number of seating arrangements for four tables is the coefficient of $a^2b^2c^2d^2$.

```
Coefficient[%, a^2 b^2 c^2 d^2]
```

297

For five tables, add another letter E, and expand the expression to get the coefficient of $a^2b^2c^2d^2e^2$.

```
Expand[(b + c + d + e)^2 (a + c + d + e)^2 (a +
b + d + e)^2 (a + b + c + e)^2 (a + b + c + d)^2];

Coefficient[%, a^2 b^2 c^2 d^2 e^2]
```

13756

Even though *Mathematica* is able to expand these terms faster than it can generate complete permutations, it will quickly run up against an exponential explosion in the number of computations required. For five tables the computer must perform over one million computations. At seven tables the number goes up to over 78 billion computations. We have just hit the wall!

We can automate the expression expansion approach to create a new **completePermutationsII**.

```
completePermutationsII[n_] :=
Module[{L = Array[A, {n}], K, P}, K =
Table[Drop[L, {i}], {i, n}];
P = Times@@(Plus@@#&/@K)^2;
Coefficient[Expand[P], Times@@L, 2]]
```

Now we can extend the number of seating arrangements to seven tables—more than I care to worry about.

```
Join[{{"tables","arrangements"}},
Table[{n,completePermutationsII[n]},{n,2,7}]]
```

tables	arrangements
2	1
3	10
4	297
5	13756
6	925705
7	85394646

Final thoughts

This is the last in a series of 10 Informatics and 18 Cowculations columns that have appeared in *Quantum* magazine over the past five years. I have made all the *Mathematica* notebooks available on my homepage at **www.uwp.edu/academic/mathematics/faculty/piele.html**.

You can keep abreast of developments in the USA Computing Olympiad through the web page at **www.usaco.org**. ◙

ANSWERS, HINTS & SOLUTIONS

Physics

P326

Let $\mathbf{u}_0$ be the pebble's initial velocity and $\mathbf{u}_p$ be the pebble's velocity at the moment it hit the mouse's paw. The crucial point here is the choice of the reference system. Let's direct the X-axis along the slope of the roof and the Y-axis perpendicular to the roof through the cat's paw (figure 1). Energy conservation says that

$$u_0 = u_p.$$

The projection of the pebble's velocity on the x-axis immediately before the impact equals the projection of its velocity just after the impact, so

$$u_{0x} = u_{px}.$$

Therefore,

$$u_{0y} = u_{py}.$$

Thus the x- and y-projections of the pebble obey the following equations:

$$u_{0x}t + \frac{g_x t^2}{2} = s_x,$$

$$u_{0y}t + \frac{g_y t^2}{2} = s_y,$$

where g_x and g_y are the projections of the vector $\mathbf{g}$ onto the respective axes. According to Viete's theorem, the equations can be transformed into

$$s_x = -\frac{g_x t_1 t_2}{2},$$

$$s_y = -\frac{g_y t_1 t_2}{2},$$

from which the distance s can be obtained:

$$s = \sqrt{s_x^2 + s_y^2} = \frac{g t_1 t_2}{2}.$$

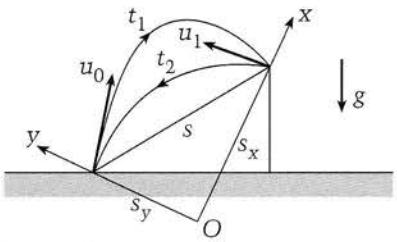


Figure 1

P327

Since the buoyant force acting on the barge equals the weight of the displaced water, the increase in the force exerted by the external water on the barge will be proportional to the amount of water that has entered the barge through the hole.

First we'll show that the difference in the water level inside and outside the barge will not change as the barge sinks. Denote the mass of the barge by m. With no water inside, the barge will float if

$$mg = \rho g a b (c - h),$$

where ρ is the density of water. Let the barge sink in such a way that the height of its sides above the water surface becomes h_1, while the depth of the water inside the barge becomes l. Now the equilibrium equation looks like this:

$$mg + \rho g a b l = \rho g a b (c - h_1).$$

From these equations we get

$$c - h_1 - l = c - h,$$

which means that the difference in the water levels inside and outside the barge at any point in time (as long as the barge is afloat) is constant and equal to the difference between the barge's height and the height of the side of a nonleaking barge. Therefore, water will flow into the barge with a constant speed, which can be calculated according to Bernoulli's equation.

In calculating the potential energy, we'll say that the surface of the water outside the barge is the zero level. Then we can write Bernoulli's equation for a tube of water flowing from the surface to the hole in the bottom of the barge:

$$P_0 - \rho g(c - h_1 - l) + \frac{\rho v^2}{2} = P_0,$$

where P_0 is the atmospheric pressure and v is the speed of the water as it enters the hole. In writing Bernoulli's equation we've taken into account the fact that the surface of the water inside the barge is much greater than the cross-sectional area of the hole at the bottom. This means that the speed at which the water rises inside the barge is far less than the entry speed of the water v, and so it can be neglected. In this way, we get the water's entry speed:

$$v = \sqrt{2g(c - h_1 - l)} = \sqrt{2g(c - h)}.$$

The barge will sink with drop below the level of the external water. At that moment, the height of the internal water is h, and its volume is

$$V = abh = Sv\Delta t,$$

where $S = \pi d^2/4$ is the cross-sectional area of the hole, and Δt is the time we seek. Finally, we get

$$\Delta t = \frac{4abh}{\pi d^2 \sqrt{2g(c - h)}} \approx 7.6 \cdot 10^6 \text{ s}$$

$$\approx 2111 \text{ h} \approx 88 \text{ days} \approx 3 \text{ months}.$$

P328

Figure 2 shows the directions of the forces F acting on the hemispheres if their charges have the same sign. If we add a hemisphere of

radius R carrying an electric charge Q to the system as shown in figure 3, the force affecting the hemisphere of radius r will be zero, since there is no electric field inside a charged sphere. Therefore, the right and the left halves of the composite sphere

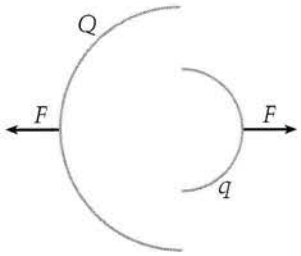


Figure 2

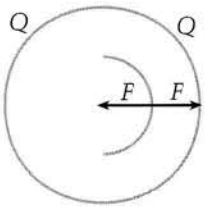


Figure 3

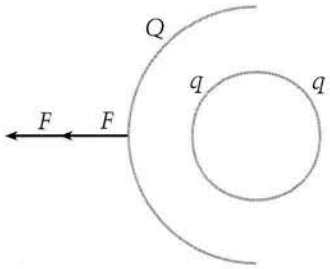


Figure 4

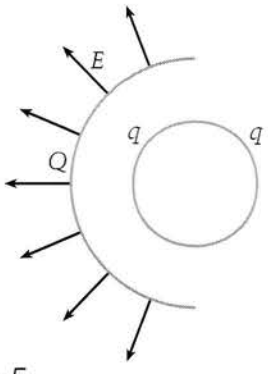


Figure 5

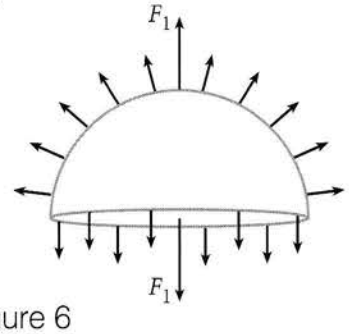


Figure 6

affect the small hemisphere with forces F that are equal in magnitude and opposite in direction.

Now let's add a hemisphere of radius r and charge q to the initial hemispheres (figure 4). From our reasoning above it's clear that two hemispheres carrying charge q will affect the large hemisphere with charge Q with forces F that are equal in magnitude and oriented in the same direction, so the force between the sphere of radius r carrying charge $2q$ and the concentric hemisphere of radius R with charge Q will be $2F$.

However, this force can easily be calculated from the electric field at the surface of the large hemisphere (figure 5):

$$E = k\frac{2q}{R^2},$$

which yields the pressure P on the large hemisphere:

$$P = \sigma E,$$

where $\sigma = Q/2\pi R^2$ is the surface density of the electric charge. The resulting force equals the pressure times the area of the plane subtending the hemisphere (figure 6):

$$F_1 = \pi R^2 P.$$

Taking into consideration that $F_1 = 2F$, we finally get

$$F = k\frac{qQ}{2R^2}.$$

P329

Charge conservation tells us that the charges on all of the capacitors are approximately equal (to an order of magnitude). However, the voltage drop across the $1000C$ capacitor is negligibly small (compared to the other capacitors in the circuit), because its capacitance is 1000 times that of its neighbors. Therefore, as a first approximation, we may consider the large $1000C$ capacitor to be short-circuited. The equivalent circuit is shown in figure 7.

Denote the voltages across the C and $2C$ capacitors by V_1, and that across the $3C$ and $4C$ capacitors by V_2. Charge conservation says that the total charge on the connected plates is zero, so

$$3CV_2 + 3CV_2 = CV_1 + 2CV_1,$$

from which we get

$$V_2 = {}^3/_7 V_1.$$

Taking into consideration that

$$\mathcal{E} = V_1 + V_2,$$

we get

$$V_1 = 0.7\mathcal{E},\ V_2 = 0.3\mathcal{E}.$$

Now we can return to the original scheme and evaluate the charges accumulated on the capacitors. The charge on the $3C$ capacitor is approximately equal to

$$q_1 = 3CV_2 = 0.9\mathcal{E}C,$$

while that on capacitor C is

$$q_2 = CV_1 = 0.7\mathcal{E}C.$$

Therefore, the charge stored in capacitor $1000C$ is

$$q \equiv q_1 - q_2 = 0.2\mathcal{E}C.$$

It's curious that our estimate is close to the precise value

$$q = \frac{\mathcal{E}C}{5 + 12/1000},$$

which can be obtained by routine calculations. Surprisingly, our estimate differs from the correct value by only 0.2%.

P330

We assume that the initial angle between the optic axis of the lens and the perpendicular to the mirror is α, while the light is focused at point B, where the optic axis intersects the screen (figure 8). At this moment, the reflected rays travel parallel to the principal optical axis. After a small time interval Δt, the mirror will turn through a small angle $\Delta\alpha = \omega\Delta t$. The reflected beam

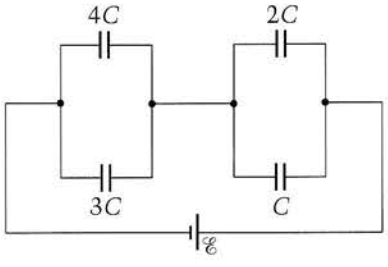


Figure 7

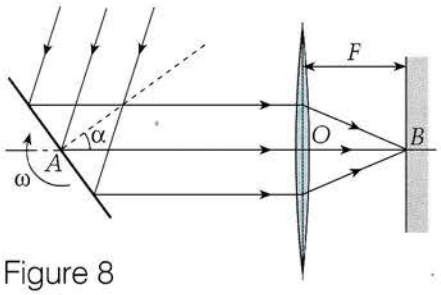


Figure 8

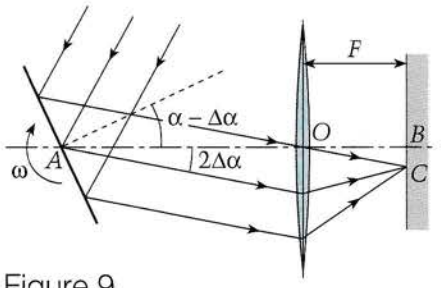


Figure 9

will "turn" through the angle $2\Delta\alpha$ relative to the axis (figure 9), so the light spot on the screen will shift to point C. Let's find the displacement $|BC|$ of the spot during time Δt:

$$|BC| = |OB| \tan 2\Delta\alpha = F \tan 2\Delta\alpha \cong 2F\Delta\alpha.$$

As usual, we consider $\Delta\alpha$ to be small, so $\tan \Delta\alpha \cong \Delta\alpha$. Now it's child's play to find the instantaneous speed of the light spot at point B:

$$v = \frac{|BC|}{\Delta t} = \frac{2F\Delta\alpha}{\Delta t} = \frac{2F\omega\Delta t}{\Delta t} = 2F\omega.$$

Math

M325

Figure 10 shows a smaller rectangle inscribed in a larger, as the problem describes. Consider α, the angle formed by the two longer sides of the rectangles. Since it is the smaller angle in right triangle EBC, $\alpha \le \pi/4$. Now let $CD = 1$ and $BC = k$, so that the elongation of $ABCD$ is k. Then $EC = k \sin\alpha$ and $BE = k \cos\alpha$. Since $\angle DCF$ is also α, $CF = \cos\alpha$ and $DF = HB = \sin\alpha$. Then the elongation of $EFGH$ is

$$\frac{HE}{EF} = \frac{k\cos\alpha + \sin\alpha}{\cos\alpha + k\sin\alpha},$$

and we want to compare this with k. That is, we need to know whether $k\cos\alpha + \sin\alpha \le k^2 \sin\alpha + k\cos\alpha$, or if $k^2 \sin\alpha \ge \sin\alpha$. Since $k \ge 1$, this is certainly true, so the elongation of $ABCD$ is not less than that of $EFGH$.

M326

Let z be the smallest side of triangle ABC, and let S be its area. Since

$$S = \frac{1}{2}h_z \cdot z = \frac{1}{2}h_x \cdot x = \frac{1}{2}h_y \cdot y,$$

then

$$\frac{2S}{z} = \frac{2S}{x} + \frac{2S}{y}$$

—that is, $xy - xz - yz = 0$. This product reminds us of the cross products that occur when we square a trinomial. Indeed,

$$(x + y - z)^2 = x^2 + y^2 + z^2 + 2(xy - xz - yz) = x^2 + y^2 + z^2,$$

Since $x + y - z$ is an integer, this proves the assertion of the problem.

M327

Solution 1. Consider the function $f(x) = x^a b + x^{-b} a$, where $a > 0$, $b > 0$, and $x \ge 1$. Then the derivative

$$f'(x) = ab(x^{a-1} - x^{-b-1}).$$

Now for $x > 1$, $x^{a-1} > x > 1$, and $x^{-b-1} = 1/x^{b+1} < 1$, so $f'(x) > 0$, so $f(x)$ is an increasing function. Therefore,

$$f(2) \ge f(1).$$

Thus we have the inequality

$$b2^a + a2^{-b} \ge a + b.$$

It remains to replace b with x and a with y in this inequality.

Solution 2 (for experts in calculus). Since the function $f(t) = 2^t$ is convex downward (its derivative is increasing), any chord that connects two points of its plot lies above the plot (see figure 11). Taking two positive numbers x and y, the points $(-x, 0)$ and $(y, 0)$ are located as in figure 11, and the chord AB is divided by K in the same ratio as CD is divided by O. Using a well-known formula from analytic geometry, this means that the inequality

$$1 = f(0) < OK = \frac{f(-x)y + f(y)x}{x + y}$$

holds.

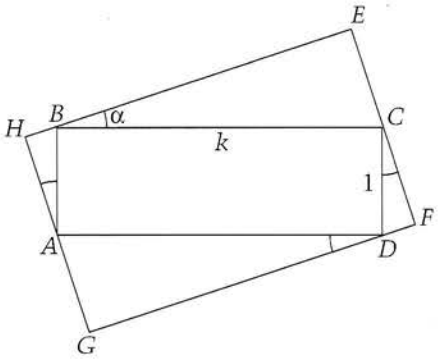


Figure 10

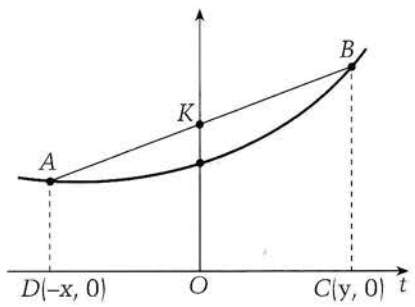


Figure 11

M328

We prove that if a_k ends in t instances of the digit 9—that is, has the form $a_k = a \cdot 10^t - 1$, where t is an integer—then a_{k+1} has at least $2t$ instances of the digit 9 at the end. Indeed,

$$a_{k+1} = 3(a \cdot 10^t - 1)^4 + 4(a \cdot 10^t - 1)^3.$$

Removing the parentheses and collecting similar terms, we obtain

$$3(a \cdot 10^t - 1)^4 + 4(a \cdot 10^t - 1)^3 = 3a^4 \cdot 10^{4t} - 8a^3 \cdot 10^{3t} + 6a^2 \cdot 10^{2t} - 1 = 10^{2t}(3a^4 \cdot 10^{2t} - 8a^3 \cdot 10^t + 6a^2) - 1 = P \cdot 10^{2t} - 1,$$

for a certain integer P.

This means that the last $2t$ digits of a_{k+1} are nines. Since $a_1 = 9$, then a_{10} has at least $2^{10} > 1{,}000$ nines at the end.

M329

The answer is no.

Let M be a polyhedron with edges colored as described in the statement of the problem. We'll prove that M has an even number of edges. Indeed, if x_i edges have color i (for $i = 1, 2$), then the total number of edges of this color is $2x_i$ (since every

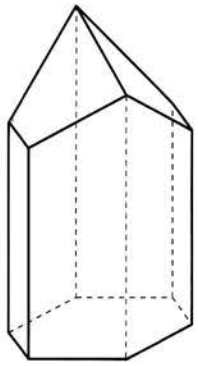


Figure 12

edge serves as a side for two faces). Since every face has the same number of edges of every color, then $2x_1 = 2x_2$: that is, $x_1 = x_2$, and the total number of edges, $x_1 + x_2$, is even.

The answer to the question posed in the problem is no. Indeed, we'll give an example of a polyhedron such that its every face has an even number of edges and the total number of edges is odd. Figure 12 depicts a 9-hedron with 19 edges. It is bounded by one hexagonal face and eight quadrilateral faces. Such a polyhedron can be obtained by cutting off two parts of a hexagonal prism—namely, the parts that lie above two planes passing through one of the longer diagonals of the upper base and crossing three lateral faces each.

Brainteasers

B325

Let B be the number of boys in the class, and let G be the number of girls. Since every girl shook hands with 8 boys, the number of handshakes between boys and girls was $8G$. However, this number can also be written as $6B$, since every boy shook hands with 6 girls.

How many handshakes were exchanged by boys only? Each boy shook hands with 8 other boys, and there are B boys. This would be $8B$ handshakes, except that we have counted each one twice (once for each participant). Thus the number of handshakes between boys is $4B$. Similarly, the number of handshakes between girls is $6G/2 = 3G$. From the statement of the problem we obtain the equation $8G + 5 = 4B + 3G$. Solving the system consisting of this equation and the equation $8G = 6B$, we find that $G = 15$ and $B = 20$.

B326

The operations can be performed in the following order: 15**62**437—16**24**537—12**45**637—1234567.

B327

See figure 13.

17	24	3	32	11	26
2	31	18	25	4	33
23	16	1	10	27	12
30	9	28	19	34	5
15	22	7	36	13	20
8	29	14	21	6	35

Figure 13

B328

Designating the radii of the inscribed circles by x and y, we see that the area to be determined can be written as

$$S = \pi(x + y)^2 - \pi x^2 - \pi y^2 = 2\pi xy.$$

Now let t be the altitude in the right triangle shown in figure 14.

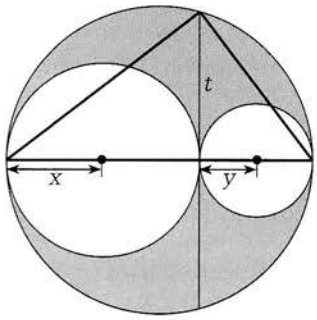


Figure 14

Then a well-known theorem of geometry tells us that $t^2 = 4xy$. Thus $S = \pi t^2/2$.

B329

The tea will be cooler in the cup with sugar in it.

Physics Contest

The fundamental particles

In the March/April issue of *Quantum* we asked our readers to use the idea of quarks as fundamental particles to build some of the hadrons we find in nature. We learned that baryons consist of three quarks and that mesons consist of a quark and an antiquark.

Readers were asked to use the down, up, and strange quarks to build the first group of particles. The properties of these three quarks are given in table 1. The only two properties of a hadron that are needed to do this are its charge (in units of the electronic charge) and its strangeness.

The neutron, as its name implies, is neutral. Because it is a baryon, it must be composed of three quarks, and because it does not have any strangeness, it must be composed of only down and up quarks. Since the down quark has a charge of $-1/3$ and the up quark has a charge of $+2/3$, the neutron is composed of two down quarks and an up quark (n = ddu).

The negative pion has a charge of -1 and a strangeness of zero. Once again it must be composed of only down and up quarks. However, the pion is a meson and consists of a quark–antiquark pair. Because the antiquark has the opposite charge of a quark, we need a down quark and an up antiquark to get a charge of -1 $\left(\pi^- = d\bar{u}\right)$.

The neutral kaon is a meson with a strangeness of +1. The only way we can get this strangeness is by using a strange antiquark, which converts the strangeness of -1 to $+1$. This gives us a charge of $+1/3$. Therefore

Name	*Symbol*	*Charge*	*Strange*
Down	d	$-1/3$	0
Up	u	$+2/3$	0
Strange	s	$-1/3$	-1

Table 1

we need a down quark to give us an overall charge of zero $\left(K^{o} = d\bar{s}\right)$.

The lambda baryon only comes with a zero charge. Its strangeness of –1 requires that we use a strange quark, giving us a charge of –1/3. The other two quarks must be down and up quarks. The only combination of two of these that gives us a charge of +1/3 is one of each. Therefore $\Lambda^0 = dus$.

The antineutron is an antibaryon and must be composed of three antiquarks. We simply use the antiquark version of each quark in the neutron to build the antineutron. Therefore $\bar{n} = \bar{d}\bar{d}\bar{u}$.

The cascade minus has a strangeness of –2 and a charge of –1. This requires two strange quarks for a charge of –2/3. We need a down quark to give us the additional charge of –1/3. Thus, we have $\Xi^- = dss$.

Up to this point we have not had to worry about the Pauli exclusion principle. However, to complete the picture, we need to add a new quantum number. Color comes in three varieties: red, green, and blue. Let's use the subscripts r, g, and b, respectively, to represent these three values, and the subscripts c, m, and y to represent the complimentary colors cyan, magenta, and yellow, respectively. In this scheme all hadrons must be white if we imagine the colors to combine as lights; that is, additively.

Four of the particles are repeats. All we need to do is to make sure that we include all combinations of the color quantum number.

$$n = d_r d_g u_b + d_g d_b u_r + d_b d_r u_g,$$

$$\bar{n} = \bar{d}_m \bar{d}_c \bar{u}_y + \bar{d}_c \bar{d}_y \bar{u}_m + \bar{d}_y \bar{d}_m \bar{u}_c,$$

$$\pi^- = d_r \bar{u}_m + d_g \bar{u}_c + d_b \bar{u}_y,$$

$$\Xi^- = d_r s_g s_b + d_g s_b s_r + d_b s_r s_g.$$

The delta plus plus is a baryon without strangeness. Therefore, it is composed of only down and up quarks. The only combination that gives us a charge of +2 is three up quarks. Thus, we have

$$\Delta^{++} = u_r u_g u_b.$$

Finally, the antilambda is the antiparticle of the lambda. All we need to do is to change each quark to the antiquark and each color to the complementary color.

$$\overline{\Lambda^0} = \bar{d}_m \bar{u}_c \bar{s}_y + \bar{d}_c \bar{u}_y \bar{s}_m + \bar{d}_y \bar{u}_m \bar{s}_c.$$

—Larry D. Kirkpatrick and Arthur Eisenkraft

Criss × cross science

H	I	J	K	■	P	A	S	S	E	■	A	W	A	Y
A	L	A	I	■	L	I	V	E	D	■	R	I	P	E
F	A	C	B	■	I	M	A	G	E	■	E	L	S	A
T	Y	K	E	S	■	■	N	O	D	E	■	H	E	N
■	■	S	I	U	N	I	T	S	■	A	F	E	■	■
C	O	T	■	M	A	R	E	■	D	R	I	L	L	S
I	N	E	■	S	T	O	A	■	E	N	A	M	E	L
R	E	I	D	■	U	N	R	I	P	■	T	R	E	E
C	A	N	C	E	R	■	R	Y	E	S	■	O	D	D
A	L	B	I	T	E	■	H	A	R	K	■	E	S	S
■	■	E	X	A	■	F	E	R	M	I	O	N	■	■
A	I	R	■	S	P	I	N	■	■	S	U	T	R	A
B	O	G	A	■	A	C	I	D	S	■	I	G	O	T
A	T	E	N	■	C	H	U	C	K	■	J	E	A	N
B	A	R	N	■	T	E	S	L	A	■	A	N	N	O

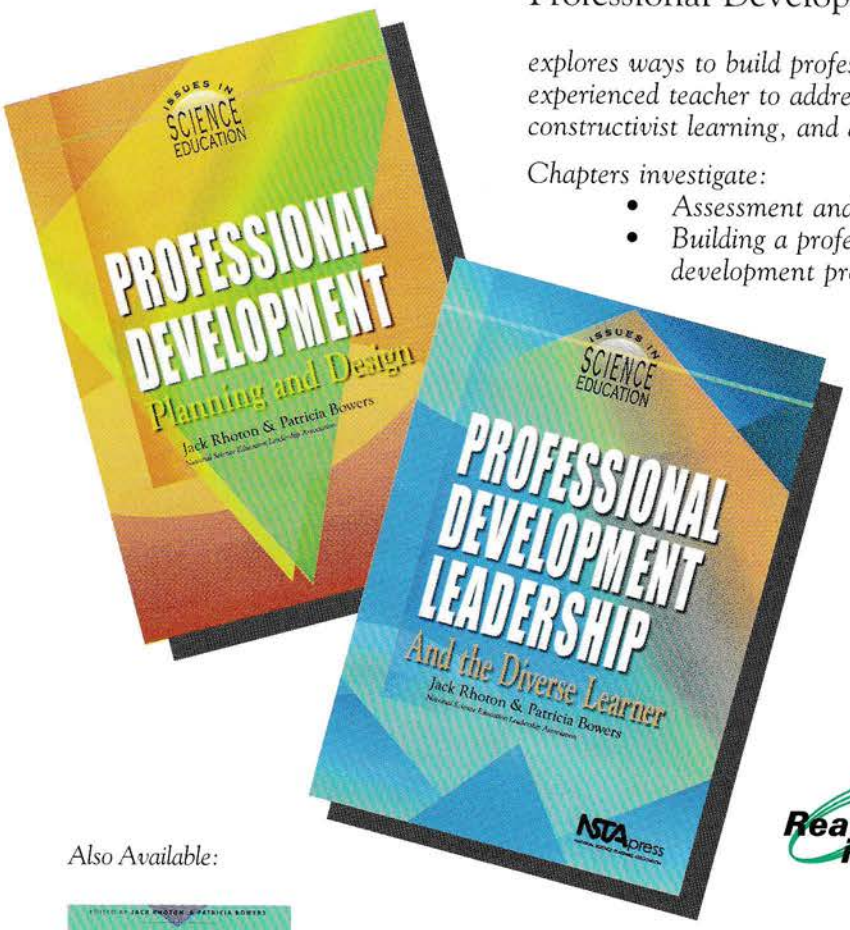
INDEX

Volume 11 (2000–2001)

ISSUES IN SCIENCE EDUCATION

Brought to you by ...

The National Science Teachers Association and
The National Science Education Leadership Association



Professional Development Planning and Design

explores ways to build professional development for the new and experienced teacher to address national standards, reform efforts, constructivist learning, and assessment strategies.

Chapters investigate:

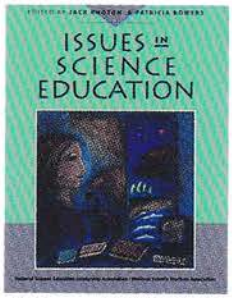
- *Assessment and evaluation*
- *Building a professional development program*
- *Curriculum development*
- *Education reform*
- *Online learning*

208 pp., © 2001 NSTA
#PB127X2, ISBN 0-87355-185-0
Members: $22.46
Non-Members: $24.95

To order, call
1-800-277-5300
or visit
www.nsta.org/store/

Read it! Read these entire books online before you order—for free!—http://www.nsta.org/store/

Also Available:



176 pp., © 1996 NSTA
#PB127X
Members: $23.36
Non-Members: $25.95

Professional Development Leadership and the Diverse Learner

discusses specific ways that professional development methods—classes, workshops, collaborative work—can give teachers the skills, resources, and knowledge necessary to help learners from different backgrounds and with different learning styles achieve scientific literacy.

Chapters investigate:

- *Community collaboration*
- *Equity/diversity*
- *Leadership development*
- *Motivation*
- *Teaching techniques*

176 pp., © 2001 NSTA
#PB127X3, ISBN 0-87355-186-9
Members: $22.46
Non-Members: $24.95

NSTApress™
NATIONAL SCIENCE TEACHERS ASSOCIATION