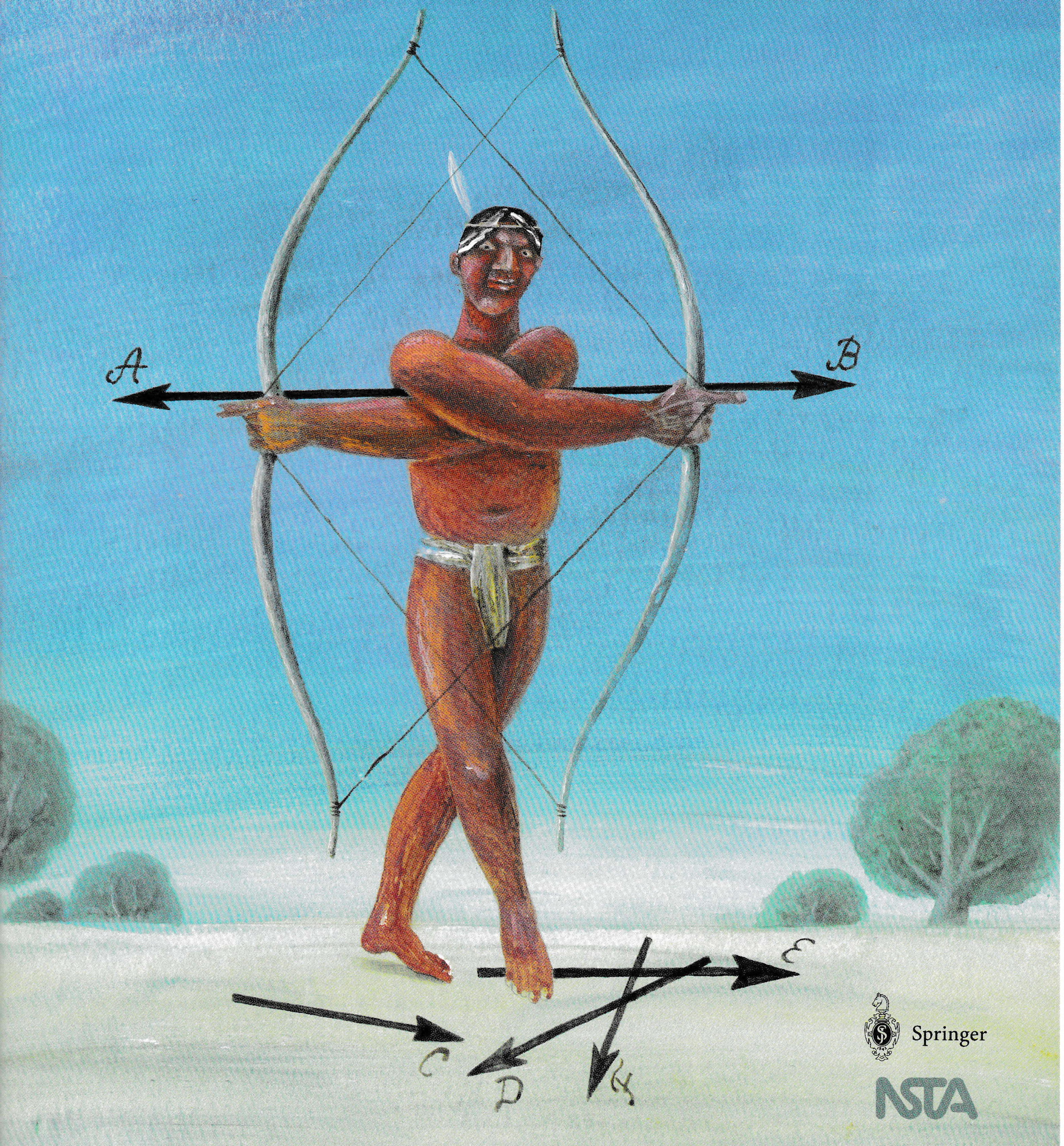
QUANTUM

MARCH/APRIL 2000

US $7.50/CAN $10.50



Springer

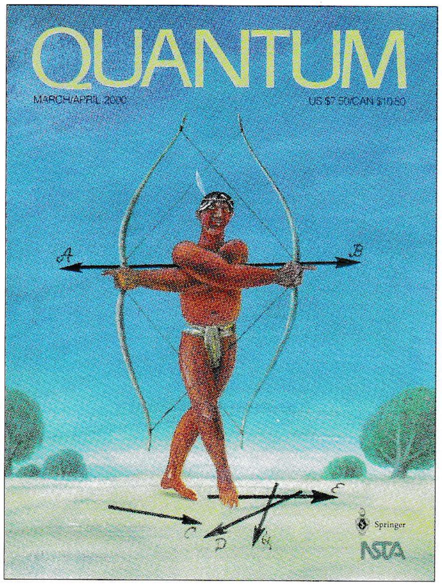
NSTA

Trumpeters of Napoleon's Imperial Guard (1812/1814) by Théodore Gericault

THOUGH SEEMINGLY DRESSED MORE FOR showing than blowing, these mounted imperial guardsmen were state-of-the-art communications equipment during the Napoleonic Wars. The trumpeter's call ordered charges, issued retreats, and rallied the troops—assuming, of course, that they could be heard. Find out how atmospheric quiet zones may have decided the fate of Europe at the battle of Waterloo by turning to page 48. You'll also be given some sound advice on the causes of other auditory phenomena as well as an explanation of some of the environmental conditions that can affect our view of the world.

QUANTUM

MARCH/APRIL 2000 VOLUME 10, NUMBER 4



Cover art by Jose Garcia

When too many people lend a hand, their efforts can often be counterproductive. Unfortunately, this is not the case with the underhanded bowman pictured on our cover. It looks as though he could use a little help. For a more typical example of "too many cooks spoiling the soup," turn to page 18 and find out how one menagerie managed to halt the wheels of progress.

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.

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 10 (6 issues) will be published in 1999–2000. *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; electronic 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: $49. 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) Jay Feinman (212) 460-1682; 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.

Just for the fun of it!

B286

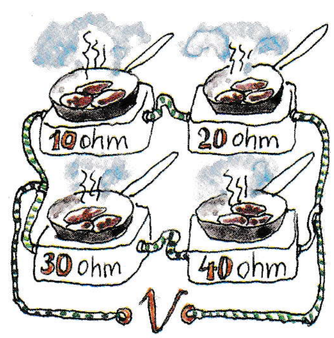
Can you place it? Let us successively write all the natural numbers beginning with 1. Which digit is at the 500,000th place in this sequence?

B287

Also ran. At a certain sporting event, 100 students took part in track-and-field events, 50 students participated in swimming, and 48 in sharpshooting. It turned out that the number of students who took part in only one event was twice as high as that who took part in just two events and three times as high as the number of students who participated in exactly three events. What is the total number of students that participated in the events?

B288

Cook-off calculations. Four hot plates are used simultaneously to fry cutlets. Connections of the hot plates and their resistances are as shown. On which hot plate will the cutlets be ready first?

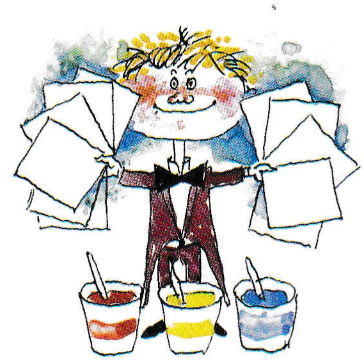


B289

Smooth ascent. Two circular towers are the same height but have different diameters. A spiral staircase runs around each of the towers from top to bottom. The slopes of the staircases are identical and constant. Which staircase is longer?

B290

Color bind. Arrange 11 nonoverlapping equal squares on the plane in pairs such that, for any coloring of these squares in three colors, there exist two identically painted squares that touch each other along a side.



Art by Pavel Chernusky

ANSWERS, HINTS & SOLUTIONS ON PAGE 61

Mathematics: 1900–1950

An overview of the first half of the 20th century

by V. Tikhomirov

THE END OF THE 19TH CENTURY and the beginning of the 20th century were marked by unprecedented development in science and technology. In 1895, Roentgen discovered x-rays. Popov and Marconi invented radio. In 1896, Antoine-Henri Becquerel discovered natural radioactivity of uranium salts. In 1900, Planck developed the theory of heat emission based on the quantum hypothesis. At the same time, the term *gene* was introduced, and the splendid field of genetics was born. In 1903, the first plane built by the Wright brothers performed its 59-second flight. In 1905, Einstein developed the special theory of relativity (similar ideas were developed simultaneously by Poincaré), gave new impetus to the quantum theory (explained the theory of the photoelectric effect), developed the fundamentals of the theory of Brownian motion, and published the formula $E = mc^2$.

At that time, many people thought that progress would bring about global prosperity and the reign of reason.

Alas, these expectations didn't come true. Many tragic events happened in the 20th century: wars, genocide, environmental degradation, bloodcurdling crimes, and so on. At the present time, humankind faces the most serious problems and, if we are unable to unite and listen to reason, we may perish. In the previous century, it was impossible to imagine such a situation: the Earth seemed infinitely rich, and nobody felt a threat to the very existence of life.

Science (and, in particular, mathematics) played an important role in all the changes that took place in this century. What changes occurred in mathematics during this century? In this article, we have a look at the history of mathematics in the first half of the 20th century.

The accomplishments of mathematics in the 20th century perhaps exceed its achievements during the previous two and a half thousand years. How should we assess the accomplishments of this science? First, we discuss the advantages that mathematics can bring to humankind.

Goals of mathematics

In the previous century, there was an argument between two prominent scientists: the French Jean Fourier and the German Karl Jacobi. Fourier took the position that the goal of mathematics is to help study the laws of nature. Jacobi argued that the goal is to glorify the human mind. Thus, Jacobi suggested that mathematics has a certain "inner" sense which is as difficult to explain as the sense of poetry or art.

In addition to the study of nature and the "glorification of the mind," practical applications (in engineering, technology, economics, and biology) have stimulated mathematical studies in modern times. Mathematics has also contributed to the philosophical understanding of the world.

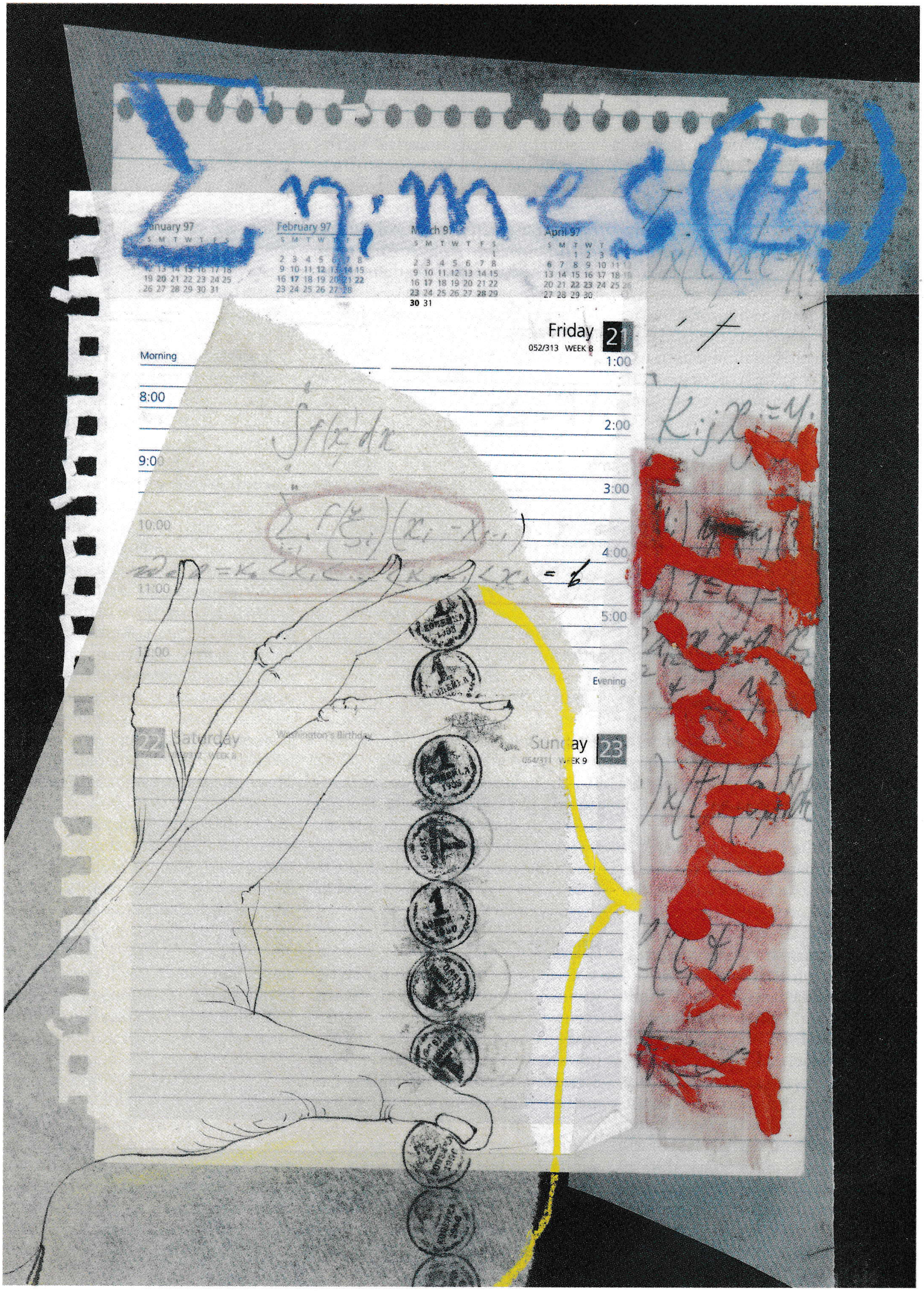
Below, we outline the historical development of mathematics in the 20th century. The world was changing fast, and so was mathematics.

Mathematical schools

Up to the beginning of the 20th century, mathematics developed mainly within several countries.

In the 19th century, two mathematical schools competed: the French and the German. F. Klein gave an impressive, though biased, account of this competition in his very interesting book *Development of Mathematics in the 19th Century*. At the beginning of this cen-

Art by Yuri Vaschenko



January 97
February 97
April 97
Friday 21
052/313 WEEK 8
Morning
8:00
9:00
10:00
11:00
1:00
2:00
3:00
4:00
5:00
Evening
23

tury, Gauss reigned in mathematics, and at the end of the century, it was Poincaré.

At the turn of the century, other mathematical schools appeared, such as the Italian, Hungarian, Austrian, Swedish, and others. In the middle of the century, the Russian school (chiefly, in Petersburg) appeared. In the second decade of the 20th century it was supplemented by the Moscow school, which became the most prominent mathematical school in the world in the 1930s. At the turn of the century, the first prominent mathematicians appeared in America; after World War I, the Polish school appeared.

Such was the situation at the beginning of the century. Nowadays, this situation is changing. Mathematics is becoming a truly international science. Hilbert's vision of the world uniting as a single mathematical community is being realized. Lines of investigation have become more diverse, and priorities have changed.

New developments in mathematics at the beginning and end of the century

We can get an idea of which branches of mathematics were most important at the beginning of the 20th century by looking at the list of sections at the Second Paris Congress of 1900. That congress had a great impact on the history of mathematics because Hilbert formulated his famous problems there. Four main sections were represented at the congress: arithmetic and algebra, calculus, geometry, and mechanics and mathematical physics. There were two additional sections: history and bibliography, and teaching and methodology.

We can assess the changes that occurred in mathematics during the 20th century by looking at the list of sections of modern mathematical congresses: mathematical logic and foundations of mathematics, algebra, number theory, geometry, topology, algebraic geometry, complex analysis, Lie groups and the theory of representations, real and functional analysis, probability theory and mathematical statistics, partial differential equations, ordinary differential equations, mathematical physics, numerical methods and the theory of computing, discrete mathematics and combinatorics, mathematical aspects of information science, applications of mathematics to nonphysical sciences, the history of mathematics, and teaching.

Many of these branches of mathematics evolved only in the 20th century. In addition, the priorities have changed: before World War II, analysis and its branches (equations of mathematical physics, probability theory, and complex-variable theory) were the main issues of mathematics; after the war, the interests of many mathematicians moved to topology, multidimensional complex analysis, algebraic geometry, Lie groups, and the theory of representations. The most resounding successes and prestigious awards came to mathematicians working in these fields.

However, the change of priorities occurred after World War II, which is beyond the scope of this article. Which new branches of mathematics evolved at the beginning of the century? First of all, these are *functional analysis, topology,* and *the theory of functions.* We begin our review of the accomplishments of mathematics in the first half of the 20th century with the discussion of these branches.

Functional analysis

The advent of functional analysis was one of the most important events in the development of mathematics before World War I. This new branch of mathematics combined many concepts of classical analysis, linear algebra, and geometry.

As early as the late 19th century, similarities were found between the theory of linear algebraic equations in a finite number of variables and their infinite-dimensional analogs—linear integral equations. The pivotal advance was achieved by Fredholm in 1900. He replaced the integral equation

$$x(t) - \lambda \int_a^b K(t,\tau)x(\tau)d\tau = y(t), \qquad (1)$$

where $y(\cdot)$ is a given function and $x(\cdot)$ is the function to be found, by the following system of linear equations:

$$x_i - \lambda h \sum_{j=0}^{n} k_{ij}x_j = y_i. \qquad (2)$$

For this purpose, he replaced the integral by the integral sums

$$t_i = a + ih,\ x_i = x(t_i),\ y_i = y(t_i),$$
$$k_{ij} = K(t_i, \tau_j),\ 1 \le i, j \le n.$$

Methods for solving systems of linear equations were developed in the 18th century. One acquires an initial knowledge of these methods at high school, and the complete theory is studied in the first year at university. Using these methods and passing to the limit, Fredholm found the solvability conditions and algorithms for solving equation (1). These results stimulated the development of a theory combining algebraic and geometric methods applied to objects in an infinite-dimensional space. Thus, *linear functional analysis* appeared.

Another important part of functional analysis was the *theory of quadratic forms,* which was first developed by Hilbert in 1904–1906. Any quadratic form

$$Q(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$$

can be brought to the diagonal form $\lambda_1 y_1^2 + \lambda_2 y_2^2$ by a rotation of the axes. Hilbert proved an analog of this theorem for the quadratic form

$$Q(x(\cdot)) = \int_a^b\int_a^b K(t,\tau)x(t)x(\tau)dt\,d\tau,$$

$$K(t, \tau) = K(\tau, t),$$

where the argument is a square-integrable function $x(\cdot)$ rather than a vector $x = (x_1, x_2)$. A function is said to be square-integrable if

$$\int_a^b x^2(t)dt < \infty.$$

The set of such functions was called *Hilbert space* and is usually denoted by L_2. The theory of quadratic forms in Hilbert spaces served as a mathematical foundation of quantum mechanics.

Development of topology

The word *topology* is nowadays used to refer to two different branches of mathematics. From the beginning, a special adjective was used with the word *topology* to distinguish those two branches. One branch, first developed by Poincaré, was called *combinatorial topology;* and the other, pioneered by Cantor, was called *general* or *set-theoretic topology.*

General topology is close to set theory and lies at the foundations of mathematics according to the plan of this science developed by the successors of Cantor—Hilbert, H. Weyl, and others. This axiomatic theory has as its aim the study of concepts like limits, convergence, continuity, and so on. In the 20th century, the foundations of general topology were laid by the German mathematician Hausdorff, the Polish Kuratowski, the prominent Russian mathematician P. Alexandrov of the Moscow school, and others.

Combinatorial topology is a branch of geometry. It studies the properties of geometric figures that remain unchanged under one-to-one continuous mappings. Cantor constructed a one-to-one mapping of the segment onto the square; however, a one-to-one continuous mapping of the segment onto the square does not exist. This fact is proved in the theory of dimension. Several prominent mathematicians took part in the development of this branch of topology: Poincaré (who posed the problem and outlined a method for its solution), the Dutch mathematician Brouwer, the Frenchman Lebesgue, the Austrian Menger, and a representative of the Moscow school of mathematics, Uryson, who tragically died at the age of 26.

Presently, the adjective *combinatorial* is usually omitted when referring to geometric topology. Unless otherwise noted, the term *topology* refers to Poincaré's work.

The destiny of the two topologies turned out to be different. General topology serves mainly to glorify the human mind, but it is not directly involved in the study of the laws of nature and applied research.

For a long time, geometric topology was also considered as an abstract science; however, quite recently it turned out that it could help us understand the structure of the universe. In addition, topological methods are used in virtually all branches of mathematics: analysis, the theory of differential equations, and so on. Topology is now one of the central branches of mathematics.

The theory of functions

At the beginning of the century, Lebesgue completed the construction of the theory of measure and integration. In the 19th century, following Cauchy and Riemann, the integral

$$\int_a^b f(x)dx$$

was defined as the limit of Riemann sums. That is, the following expressions were taken as approximate values of the integral:

$$\sum_{yi=1}^{n} f(\xi_i)(x_i - x_{i-1}),$$

where $a = x_0 < x_1 < \dots x_{n-1} < x_n = b$ is a partitioning of the interval of integration and x_i is a point in the interval $[x_{i-1}, x_i]$.

Lebesgue tried a different approach. He partitioned the *y*-axis, rather than the *x*-axis, by points ... $y_{i-1} < y_i < \dots$, arguing that for discontinuous functions it was impossible to choose a point x_i that would adequately represent the function on the interval $[x_{i-1}, x_i]$. However, the sets E_i on the *x*-axis for which $y_{i-1} \le f(x) < y_i$ can be rather weird for certain complicated functions. Therefore, to develop the theory of integration, the theory of measure had to be developed first. In other words, one had to learn how such weird sets can be measured. This was done by Borel and Lebesgue.

Lebesgue defined the measure of a set E (say, on the interval $[0, 1]$) as follows. He referred to the lower bound of the sums of the lengths of intervals that cover E as the *upper measure of E.* The upper measure is defined for any set. The set E is said to be *Lebesgue measurable* if the sum of the upper measure of E and of its complement (with respect to the interval $[0, 1]$) is 1. In this case, the upper measure of E is called the *Lebesgue measure* of the set E, and is denoted by $\text{mes}(E)$.

Lebesgue replaced Riemannian sums used to define the integral by sums of the form

$$\sum_i \eta_i \text{mes}(E_i),$$

where η_i is a point on the interval $[y_{i-1}, y_i]$. He eloquently described the advantages of his method by comparing how two clerks count money. An inexperienced clerk counts coins in the order in which they come to him. An experienced and methodical clerk proceeds as follows: I have $\text{mes}(E_1)$ coins of 1-franc, whose sum is $1 \times \text{mes}(E_1)$. I have $\text{mes}(E_2)$ 2-franc coins, whose sum is $2 \times \text{mes}(E_2)$. I have $\text{mes}(E_5)$ 5-franc coins, whose sum is $5 \times \text{mes}(E_5)$, and so on. As a result, I have $1 \times \text{mes}(E_1) + 2 \times \text{mes}(E_2) + 5 \times \text{mes}(E_5) + \dots$ francs. Certainly, both clerks obtain the same result. However, in the case of an infinite number of indivisibles, the difference between these methods is fundamental. The new measure theory gave rise to a new approach in the theory of functions—the *metric theory of functions.* Set theory also underwent a transformation. The new theory was launched by three French mathematicians—Borel, Baire, and Lebesgue. It became known as *descriptive set theory,* which studies the structure of various weird sets.

In the 1920s, the leading role in the theory of functions passed to the Russian mathematical school represented by Nikolai N. Lusin and his disciples P. Aleksandrov, N. Bari, A. Kolmogorov, D. Menshov, M. Suslin, A. Khinchin, and others. These scientists founded the Moscow school of mathematics. Having taken their first steps in the theory of functions, each of Lusin's disciples proceeded to study various other fields of mathematics. Kolmogorov and Khinchin worked in probability theory, Aleksandrov and Uryson in topology, Lyusternik and Shnirelman in nonlinear analysis, Novikov in mathematical logic, and Lavrent'ev contributed to complex analysis and mechanics. Only Men'shov and Bari continued studying the theory of functions. In the 1930s, no other mathematical school in the world had such a group of brilliant mathematicians.

In the next section, we discuss the role of mathematics in studying the laws of nature.

Mathematics and physics

At the end of the 19th century, it seemed that physics was a completed field of knowledge. According to legend, a young man[1] who wanted to become a physicist once asked a renowned physicist for advice. The master answered that he saw no prospects in physics—only two problems remained unsolved: to explain the Michelson–Morley experiment and to elucidate the laws of radiation. Soon, these problems would be solved and nothing would remain to be done in physics.

Several years later, the first problem gave rise to the *special theory of relativity,* and the second to *quantum mechanics,* which overturned all our notions about the structure of the world.

The special theory of relativity was developed in 1904–1906 by Lorentz, Einstein, and Poincaré. The structure of the physical world described by this theory was quite strange. It contradicted physical intuitions evolved during the previous three centuries.

The mathematical foundations of the special theory of relativity were formulated by the prominent German mathematician Herman Minkowski, who established a connection between this theory and Lobachevsky's geometry.

We illustrate this idea as follows. Consider two airplanes that fly towards each other, one with speed v with respect to the Earth, and the other with speed v' (also with respect to the Earth). According to Newtonian mechanics, the speed of the second airplane with respect to the first one is $v + v'$; however, the special theory of relativity gives a different formula:

$$\frac{v+v'}{1+\dfrac{vv'}{c^2}},$$

where c is the speed of light. If the airplanes move in the same plane rather than along the same line, the formula for adding the velocities is related to a transformation of Lobachevsky's plane. In short, this can be formulated as follows: *the space of velocities in the special theory of relativity is realized by Lobachevsky's plane, where the formula for adding velocities is defined in terms of the motion of this plane.*

It turned out that time and space cannot be considered as separate entities—our world is four-dimensional. As a result, the multidimensional geometry was given a physical meaning.

This event was very important for mathematicians, because theories that many people considered extremely abstract and as having no connection with reality proved to be useful in describing fundamental characteristics of the universe.

Ten years later, Einstein developed the general theory of relativity, which demolished all the established notions of a "flat" world. The geometry of our world proved to be "curved" and related to gravitation. The readings of measuring devices will differ, depending on the trajectory along which they travel from one point to another. This phenomenon is closely related to one of the most important concepts of geometry—the concept of connectivity, which defines parallel displacement on curved surfaces. This notion was studied by geometers of the Italian school at the beginning of the 20th century (Levi-Civita and others).

These events gave impetus to the intensive development of geometry in the 1920s and 1930s and of topology at the present time.

In the 1920s, humankind faced another shock—the advent of quantum mechanics. One of the most stable principles of science was overturned—that of the predictability of the future on the basis of the past. It turned out that the microcosm is unpredictable in principle. Suppose that an electron passes through an opening and hits a screen. It turns out that one can predict only the *probability* of finding the electron at a certain position on the screen. This seemed unbelievable even for such a great scientist and one of the founders of quantum mechanics as Einstein—he said repeatedly that he didn't believe God played dice with the universe.

We now try to explain what kind of mathematics underlies this phenomenon. In classical mechanics, the motion of a particle is characterized by its coordinate x and its momentum p. It is assumed that they can be measured simultaneously, and that the future motion of the particle is uniquely determined by a differential equation.

In quantum mechanics, the location of a particle is defined by a (complex) wave function $X(x)$ in the Hilbert space L_2 on the line. This function satisfies the condition

$$\int_{-\infty}^{+\infty} |X(x)|^2\, dx = 1$$

and determines the probability $\mathcal{P}_X([a, b])$ of finding the particle (at a certain instant of time) in the interval $[a, b]$ according to the formula

$$\mathcal{P}_X([a,b]) = \int_a^b |X(x)|^2\, dx.$$

[1] Max Planck.

The momentum is characterized by another function $P(p)$. This function is also defined on the line and satisfies the condition

$$\int_{-\infty}^{+\infty} |P(p)|^2 \, dp = 1.$$

The probability $\mathcal{P}_p([\alpha, \beta])$ that the particle momentum lies within the range $\alpha \le p \le \beta$ is given by the equation

$$\mathcal{P}_p([\alpha,\beta]) = \int_{\alpha}^{\beta} |P(p)|^2 \, dp.$$

The motion of the particle is determined by a partial differential equation called the *Schrödinger equation*.

One of the most important points in quantum mechanics is that the wave function and the momentum function are related by the *Fourier transform*

$$P(p) = (2\pi\hbar)^{-1/2} \int_{-\infty}^{+\infty} X(x) e^{-i\frac{p}{\hbar}x} \, dx, \quad (3)$$

where $\hbar$ is Planck's constant. The most probable values of the particle coordinate and momentum (their mean values) are given by the formulas

$$\xi = \int_{-\infty}^{+\infty} x|X(x)|^2 \, dx, \quad \eta = \int_{-\infty}^{+\infty} p|P(p)|^2 \, dp.$$

If these average values are zero, then the dispersions of the coordinate and momentum are given by the formulas

$$D_X^2 = \int_{-\infty}^{+\infty} x^2 |X(x)|^2 \, dx,$$

$$D_P^2 = \int_{-\infty}^{+\infty} p^2 |P(p)|^2 \, dp.$$

Equation (3) implies the inequality

$$D_X^2 D_P^2 \ge \frac{\hbar^2}{4},$$

which is called the *Heisenberg uncertainty principle.* It reflects the fact that it is impossible to measure precisely the position and momentum of the particle simultaneously. This overturned hopes for determinism and complete knowledge of the microcosm.

It happened that the mathematical foundations of quantum mechanics were developed by Hilbert and his disciples shortly before the advent of quantum mechanics itself. In particular, the equivalence of two approaches to the description of the microcosm suggested by Heisenberg and Schrödinger was quickly established, owing to the fact that one of the founders of the new science, Max Born, had attended Hilbert's lectures on functional analysis and the theory of infinite-dimensional quadratic forms.

Here is another story. When the English botanist Brown discovered the chaotic movement of small particles in a liquid, neither mathematicians nor physicists paid much attention to his discovery. The theory of Brownian motion was first developed by Einstein (in the same year, 1905, that he created the foundation for theory of relativity and quantum mechanics) and the Polish physicist M. Smoluchowski. It was Norbert Wiener who first suggested the mathematical theory of Brownian motion. It turned out that the trajectories of Brownian particles are continuous functions that have no derivatives.

The first example of a continuous function that has no derivatives at any point was constructed by Weierstrass in 1872. Mathematicians looked skeptically at this discovery: many of them thought that this monster had no relation to reality. One of the most renowned mathematicians of the 19th century, Charles Hermite, said that he was aghast at these monsters, continuous functions without derivatives. Again, the conventional idea that everything in the world is "smooth" was overturned. It turned out that the world is populated by "monsters."

The complete mathematical theory of Brownian motion was developed by A.N. Kolmogorov. This theory is one of the most prominent achievements of mathematics in the first half of the 20th century.

Development of abstract branches of mathematics

The desire to glorify the human mind independently of any practical goal stimulated the efforts of many mathematicians and sometimes led them into "jungles" that had almost no relation to reality.

In the first half of the 20th century, the concept of an axiomatic construction of the entire body of mathematics arose. Citing Kolmogorov, according to this concept, pure set theory—this spiritual legacy of Cantor—lies at the foundation of mathematics. This theory left deep marks on the history of mathematics. It was believed that Cantor found a "paradise" for mathematicians. When absurdities in set theory were found and many scientists cast doubt on its foundation, Hilbert said: "Nobody can expel us from Cantor's paradise."

The development of axiomatics was connected with the critical analysis of the foundations of mathematics.

The extraordinary development of algebra in the 1920s led to the algebraization of all of mathematics. A considerable contribution to this process was made by Emmy Noether and her student B.L. van der Waerden. Elementary geometry (Hilbert) and probability theory (Kolmogorov) were axiomatized as well. We have mentioned earlier the fields of general topology and measure theory. Many other axiomatic theories began to be developed as well.

In the late 1930s, a group of French mathematicians decided to present mathematics on an axiomatic basis. This group wrote under the collective pseudonym of Nicolas Bourbaki, a French general. Set theory was at the basis of the whole construction. Then, the first story was constructed: ordered structures, algebra, general topology, and measure theory; next, the second story had to be constructed, where algebraic and geometric structures were to be combined with topological and ordered structures, and so on. This endeavor remained unfinished. The very idea seems Utopian, since it is not possible to imitate the development of science. However, Bourbaki's efforts were not in vain. In particular, they created a language that still serves for communication among mathematicians.

Mathematics and the military-industrial complex

Mathematics played an important role in many events of the century, some of which almost led to worldwide disaster.

In particular, many mathematicians "on opposite sides of the trenches" took part in various programs aimed at the development of modern weapons.

X-rays and radioactivity gradually led scientists to the idea of using atomic energy. Initially, physicists got along without mathematicians. However, the development of the A-bomb and especially the H-bomb required the development of complex mathematical models and massive computations. Many prominent mathematicians took part in the creation of the A-bomb. As a result, principles of computational mathematics were revised and powerful computers were developed. It seems that the time has not yet come (at least in Russia) for an assessment of the contribution of mathematicians to the creation of nuclear arms. However, there is no doubt that this contribution was considerable.

The Wright brothers built their airplane without using mathematics; however, further development of aviation stimulated the development of aerodynamics and the theory of flight. Among the classics of this science are N. Zhukovsky and his disciples Chaplygin, Golubev, and others. They applied the theory of complex functions (and developed it at the same time) to flight theory. In the 1940s, supersonic aerodynamics emerged.

The invention of radio gave rise to the development of a new field of mathematics—the theory of nonlinear oscillations. Among the creators of this theory are prominent Russian scientists: Mandelstam and his students and collaborators Papaleksi, Andronov, and others.

Problems of controlling shellfire and bombing stimulated the development of many branches of probability theory (von Neumann, Wiener, and Kolmogorov).

Problems of encrypting classified messages and efficient transmission through communication lines gave rise to a new branch of mathematics: information theory (K. Shannon) and coding theory.

Problems of automatic control in industry and cosmic navigation stimulated the development of optimal control theory (Pontryagin and Bellman). The same can be said about many other branches of pure and applied mathematics.

Much of what was done in secret laboratories later became general knowledge. The confrontation between two social systems during the cold war resulted in unprecedented development of technical facilities and the present information explosion caused by total computerization. The first computers emerged at the end of the period that we are considering in this article, and mathematicians played a key role in their development. In particular, von Neumann made the major contribution to the development of principles of computer design and of programming.

Of course, mathematics is also essential for engineering, economics, biology, and other fields of human activity. Among prominent mechanicians and engineers who made a major contribution in mathematics are Bubnov, Galerkin, Krylov, Timoshenko, G. Taylor, and von Kármán. This list can certainly be enlarged.

Mathematics and philosophy

The 20th century was a time of greatness—a time when our understanding of the world around us was changed forever. Mathematics played an important role in this process.

At the beginning of the century, it seemed that science was close to understanding the structure of the world. Rational people were sure that the laws of nature could be discovered, that the universe existed and would exist for ever, that it was unbounded both in time and in space, that the Earth evolved in a natural way and so did life, and that natural evolution resulted in everything we have before our eyes. Doubt was cast on all these facts in our century.

The general theory of relativity stimulated the development of cosmology. This led to the Big Bang theory, that posits the existence of an initial point in the life of the universe. According to modern estimates, the universe has existed for no longer than 10^{14} years. The space of the "filled" universe proved to be bounded, though expanding. The theory of a "contracting universe," which predicts that the universe will ultimately collapse into a point, remains unproved. The prominent Russian scientist Friedmann made a major contribution to the development of cosmological theories. Quite recently, fantastic theories about the multiplicity of domains in the universe that differ in the direction of time have been developed (Sakharov).

Significant drawbacks were found in the majority of theories dealing with the origin of the solar system. The origin of the Earth and of life, their evolution, and the origin of humankind seem even more enigmatic.

Doubt was also cast on many major philosophical concepts. The basic postulate of the post-Newtonian scientific philosophy stated that the world was governed by differential equations; in other words, it is completely predictable. Only a small domain of this ordered world was governed by Chaos: it seemed that only in gambling could something be unpredictable. Pascal, Fermat, J. Bernoulli, and Laplace were the first to describe the laws of chance.

However, the domain occupied by chaos has been steadily expanding. Probability theory—the science of chance—has been steadily developing. In the 20th century it took on an orderly shape. Half a century ago, it seemed that the Kingdom of Chaos and the Kingdom of Order were comparable in size. Only in our time has the situation changed.

In contrast to the standpoint of Newton and Laplace, many scientists now believe that *everything is Chaos*, and there is good reason to believe this.

Serious doubt was cast on the idea of the unlimited capabilities of humanity. We have already mentioned information theory—the new branch of mathematics that emerged in the 1940s. Norbert Wiener included information theory in a more general scientific discipline, which he called *cybernetics*. The development of this science is related to many philosophical ideas, especially to the notion of consciousness. It seemed to most people that only humans could think. However, in the 1940s, Turing and Wiener put forward the idea of modeling human consciousness. Quite recently, the idea of a computer defeating the world chess champion seemed crazy; nevertheless, it happened! The discussion of the possibility of creating artificial beings capable of thinking pertains to a new philosophy that has emerged in our time.

At the beginning of the century, many mathematicians (and particularly Hilbert) believed that "every problem can be solved." It seemed possible to design an algorithm that would enable a machine to read the formal description of an axiomatic theory and prove any theorem of this theory. It turned out that this plan could be implemented (in principle) for elementary geometry, although the machine would have to work extremely long to prove any significant part of the well-known theorems of geometry. However, this plan is actually impossible to implement for most other theories (in particular, for arithmetic). This great theorem was proved by Gödel in 1931.

Now we should say a few words about problems that were stated or solved in our century, and about the role that these problems have played in the development of science. This topic deserves a separate article. Here, we give only a short review.

Problems

The study of the laws of nature, the development of abstract mathematics, the achievements of applied mathematics, and speculation about the philosophical foundations of the world resulted in the emergence of new branches of mathematics and new fundamental concepts, the achievement of outstanding results, and the development of new theories and efficient methods.

In the introductory part of his lecture at the International Mathematical Congress in Paris in 1900, devoted to the formulation of important mathematical problems, Hilbert said: "It is impossible to deny the important role of certain problems for mathematics as a whole and for particular investigations."

The preceding centuries have left for the 20th century several great problems. The oldest of them is Fermat's last theorem, which states that, for every integer $n > 2$, the equation $x^n + y^n = z^n$ has no solutions in positive integers. This problem was posed in the 17th century. Two famous problems in number theory—Goldbach's conjecture and Euler's conjecture—have come down to us from the 18th century. Goldbach, in his letter to Euler in 1742, claimed that *every odd natural number greater than 6 is equal to the sum of three primes.* Euler noted that to prove this, it is sufficient to prove that *every even natural number greater than 2 is equal to the sum of two primes.*

Of the problems posed in the 19th century, the most famous ones are Riemann's problem about the zeros of the zeta function, and the continuum hypothesis of Cantor.

In the 20th century, the most famous set of problems is the list of Hilbert's problems, which we have already mentioned. The first place in this list is occupied by the continuum problem: *does an uncountable set exist which can be mapped in a single-valued way onto a unit interval, but such that the unit interval cannot be mapped onto this set in a single-valued way?* In other words, *does a set exist with a cardinality greater than that of a countable set but less than that of a unit interval?*

Fermat's last theorem was proved at the very end of the 20th century. Goldbach's conjecture was "almost" proved by I.M. Vinogradov, who proved in 1937 that any *sufficiently large* odd number can be represented as a sum of three primes. The Euler and Riemann problems remain open.

Let us explain how several of Hilbert's problems were solved. To a large extent, Hilbert turned out to be a good forecaster, but in several cases intuition failed him. As a rule, this was related to the optimistic view of the world that was characteristic of the previous century.

In putting forward the continuum problem, Hilbert proceeded from the assumption that it can be decided one way or the other. However, it turned out that the continuum hypothesis could neither be proved nor disproved within the framework of conventional axiomatic set theory and mathematical logic. The fact that it cannot be disproved was proved by Gödel in 1936. That it cannot be proved was shown by Cohen in 1963.

Hilbert's confidence in the unlimited possibilities of the human mind was expressed by his aphorism: "We want to know, we will know." This confidence made him sure that every mathematical problem must have a solution. Thus, he posed the following problem as his tenth: Given a polynomial P in n variables, with integer coefficients, find an algorithm to determine whether or not the equation $P = 0$ has solutions in integers. The solution to this problem also turned out to be negative as well, as proved by Matiyasevich in 1970.

CONTINUED ON PAGE 17

A star is born

Gravity backs a stellar production

by V. Surdin

WE BEGIN OUR STORY IN Great Britain at the beginning of the 20th century. Several years after a brilliant graduation from Cambridge University, James Hopwood Jeans (1877–1946) carried out a series of outstanding studies in various fields of theoretical physics: he published a monograph on the kinetic theory of gases and papers on molecular physics and the theory of radiation. Under the influence of George Howard Darwin, Professor of Physics and Astronomy at Cambridge University (the son of the famous biologist Sir Charles Robert Darwin), he completed a number of studies in theoretical astrophysics. These included a fundamental study, *"Stability of a spherical nebula,"* published in 1902 in the Proceedings of the Royal Society (London).

This work described the behavior of gaseous condensations under the action of intrinsic gravitational forces. It became the cornerstone of the modern theory of gravitational instability, which explains the origin of virtually all structural elements of the Universe, from galaxies and their clusters to the stars, planets, and their satellites. The characteristic sizes and masses of gravitationally unstable gaseous condensations are now called Jeansean and are labeled by the index J. For example, M_J is the Jeans mass and R_J is the Jeans radius.

Certainly, George Howard Darwin might be proud that he could focus the attention of the young Jeans on astronomy: his student made an amazing contribution to the development of cosmogony and stellar dynamics. However, could Darwin, who in 1899–1900 was President of the Royal Astronomical Society, foresee that Sir James would fill this post a quarter of a century later and establish the annual Darwin Lectures in honor of his tutor?

Curiously, the connection between these astronomers was even more profound. George Darwin was famous for his studies of tides and the shape of rotating fluid bodies. Jeans continued the studies of the cosmogonic role of tidal phenomena and even developed the tidal theory of the origin of the Solar System, which was very popular in the first half of the 20th century. This theory considers the birth of a planetary system to be an extremely rare event caused by a close fly-by of some star near a sun, which tore a part of the condensed protoplanetary matter from the sun's shell. Although in relation to the Solar System this hypothesis is of only historical interest, the mechanism of tidal interaction undoubtedly plays an important role in the world of galaxies, stellar clusters, and possibly during the formation of stars, that is, at the stage of so-called protostar evolution.

However, let's return to the main work of Jeans: the theory of gravitational instability. After Jeans laid the foundation at the start of the 20th century, astronomers needed another 70 years to discover the component of the interstellar medium that is directly linked to star formation. When Jeans's theory is applied to this component, it correctly predicts the parameters of newborn stars. This period between theoretical predictions and experimental findings is strikingly long for a modern, rapidly developing science. Let's consider the fundamentals of Jeans' theory.

This theory originates in the work of the great physicist Sir Isaac Newton (1643–1727). Five years after Newton published his law of gravitation, his young friend, clergyman Richard Bentley, who was then Master of Trinity College, Cambridge, asked him in a letter about the reasons for star formation. He was interested in whether the recently discovered gravitational force could be the cause of the origin of stars. (This question allows us to consider Bentley as the joint author of the idea of gravitational instability). We find Sir Isaac Newton in his first letter to Dr. R. Bentley (Dec. 10, 1692) writing as follows:

"It seems to me, that if the matter of our Sun and planets, and all the matter of the universe, were evenly scattered throughout all the heavens, and every particle had an innate gravity towards all the rest, and the whole space throughout which this matter was scattered, was but finite, the matter on the outside of this space would by its gravity tend towards all the matter on the inside, and by consequence fall down into the middle of the whole space, and there compose one great spherical mass. But if the matter were evenly disposed throughout an infinite space, it could never convene into one mass; but some of it would convene into one mass and some into another, so as to make an infinite number of great masses, scattered great distances from one to another throughout all infinite space. And thus might the Sun and fixed stars be formed, supposing the matter were of a lucid nature."

As we see, Newton elaborated the idea of gravitational condensation of the primordial matter. He considered this matter to be absolutely inert and cold, with no resistance to the compressing gravitation. Therefore, according to Newton, any region of higher density must be progressively compressed, becoming more and more dense as a result of gravity.

Toward the end of the 19th century, physicists clearly understood that any matter, including rarefied gas, is elastic: it is this property that underlies such phenomena as the existence of propagating sound waves. Therefore, Jeans concluded

Photos courtesy of NASA

Columns of cool interstellar hydrogen gas and dust are giant incubators for new stars. They are part of the Eagle Nebula, a nearby star-forming region 7,000 light-years away in the constellation Serpens.

that the gravitational compression starts only when the force of gravity overcomes the pressure of the gas.

To find conditions that make this possible, we shall obtain some simple physical estimates. Let's consider what processes are induced by a small stochastic compression of some gaseous volume with characteristic size λ and density ρ. On the one hand, the force of gravity tends to prolong this compression. Were the pressure of the gas entirely absent, all of the matter of the condensation would fall to its center during the period of free fall,

$$t_f \sim (G\rho)^{-1/2},$$

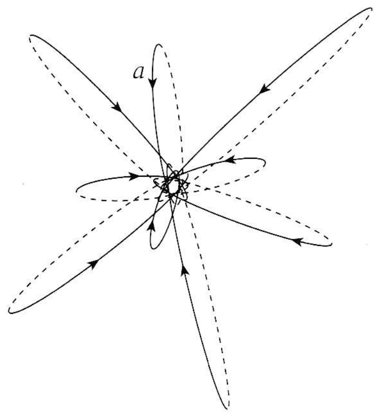
which is independent of the size of the perturbation.

This formula can be explained with the help of Kepler's third law. The fall of any particle to the cloud's center occurs under the influence of the attractive force generated by the mass $M \sim \rho a^3$, where a is the initial distance from the particle to the center. The fall time is equal to half of the period of orbiting along a very narrow ellipse, one focus of which coincides with the cloud's center. According to Kepler's third law, the period of motion along an ellipse with major axis a is equal to the period of circular revolution at radius $a/2$ around the mass M. (We recall that it is Kepler's ellipse as long as *all* the particles fall to the center of gravity and the mass of the cloud is conserved at every r.) The period of such circular motion can be easily calculated using Newton's second law.

On the other hand, were gravitation absent, the pressure of the gas would force the cloud to expand during the so-called "dynamical time," estimated as

$$t_d \sim \frac{\lambda}{v_s},$$

where v_s is the speed of sound in the gas, which is the same order of magnitude as the speed of molecular motion in this gas:



Free compression of a cloud under the action of its own gravitation can be considered as motion of all its constituent particles along strongly elongated elliptic orbits obeying Kepler's laws.

$$v_s \sim \sqrt{\frac{RT}{\mu}}.$$

(Here R is the gas constant, and μ is the molar mass.)

Evidently, if $t_d \ll t_f$, the molecular pressure will have plenty of time to redistribute the matter in such a way as to prevent its further gravitational compression. In contrast, when $t_f \ll t_d$, the gravitational compression occurs more rapidly than the molecular expansion. Let's find the ratio of t_d and t_f:

$$\frac{t_d}{t_f} \sim \frac{\lambda}{(G\rho)^{-1/2} v_s} = \frac{\lambda}{(RT/G\rho\mu)^{1/2}}.$$

We see that small-scale perturbations ($\lambda \ll (RT/G\rho\mu)^{1/2}$) are stable with respect to stochastic compression, but large-scale perturbations ($\lambda \gg (RT/G\rho\mu)^{1/2}$) are unstable: having appeared, they cannot be damped by expansion of the gas.

Thus, our reasoning yields the following estimate of the critical size λ_J:

$$\lambda_J \sim \left(\frac{RT}{G\rho\mu}\right)^{1/2},$$

from which we obtain an estimate of the critical mass M_J:

$$M_J \sim \rho\lambda_J^3 \sim \left(\frac{RT}{G\mu}\right)^{3/2} \rho^{-1/2}.$$

Apart from numerical coefficients, the last two formulas are the famous Jeans laws.

To gain a better feeling for the meaning of these results, we deduce them again from the equilibrium condition for a gaseous cloud under the action of molecular pressure and gravitation, which according to the definition of λ_J and M_J takes place precisely when $\lambda = \lambda_J$ and $M = M_J$. The characteristic value of the free-fall acceleration in the cloud is (order of magnitude)

$$g \sim G\frac{M}{\lambda^2} \sim G\rho\lambda.$$

The pressure of the gas must counterbalance the force of gravity; this pressure must vary from zero at the outer boundary of the cloud to the value

$$P \sim \rho g\lambda \sim G\rho^2\lambda^2$$

in the depth of the cloud (we recall the formula for hydrostatic pressure, $P = \rho gh$). Substituting here the value of P from the gas equation of state

$$P = \frac{\rho RT}{\mu},$$

we obtain the same estimates for λ_J and $M_J \sim \rho\lambda_J^3$ as above. Using more sophisticated mathematics, one can obtain more accurate values of λ_J and M_J (we shall not prove this here):

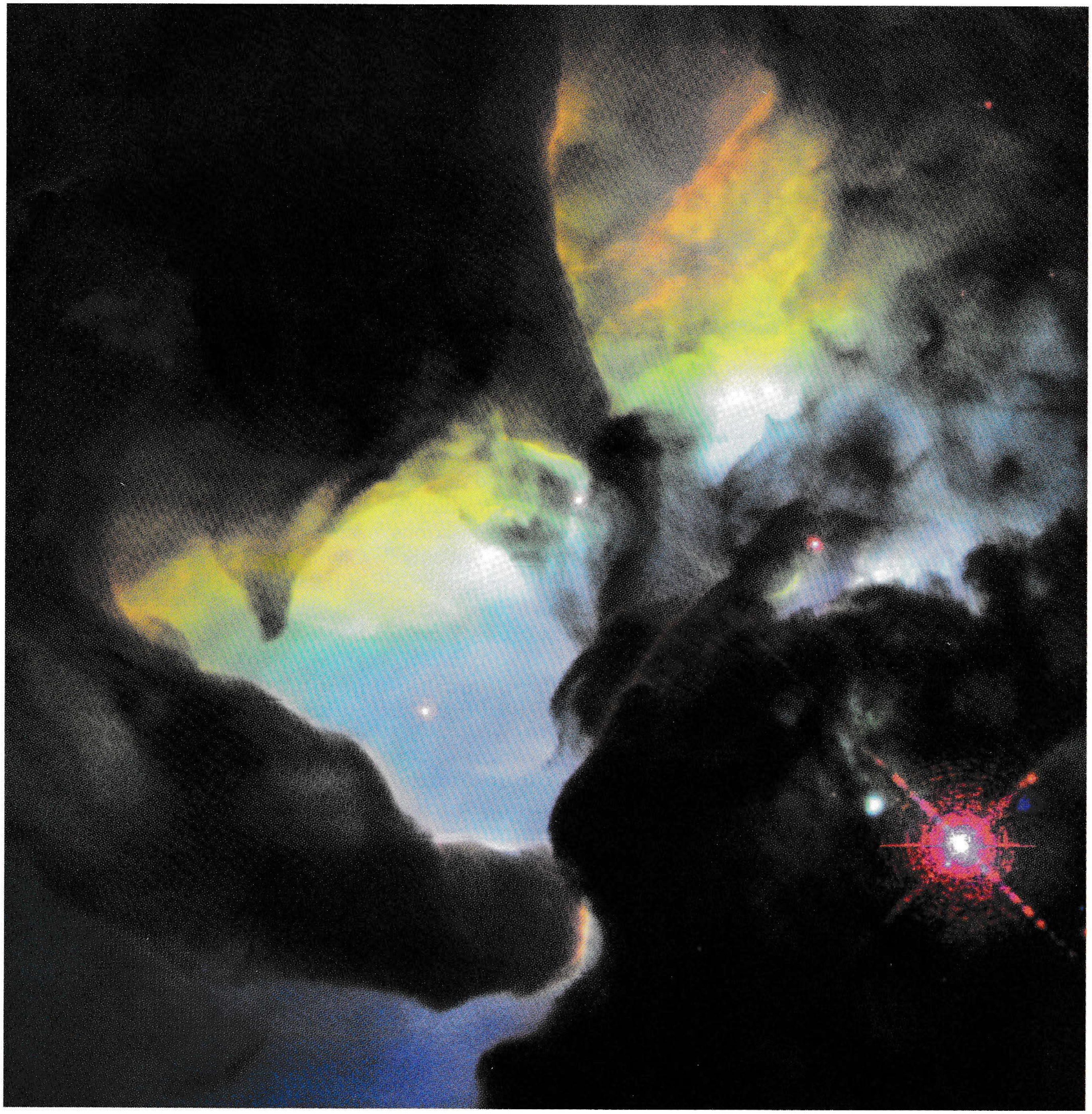
$$\lambda_J = \left(\frac{2RT}{G\rho\mu}\right)^{1/2}, \quad M_J = \left(\frac{RT}{2G\mu}\right)^{3/2} \rho^{-1/2}.$$

These simple laws are cornerstones of the theory of gravitational instability. This says that if density perturbations of various sizes arise in a gaseous medium for any reasons, the largest of them with mass greater than M_J will be irreversibly compressed into dense matter. However, large-scale density fluctuations are rare events, and usually only small-scale density fluctuations occur in the gaseous cloud. Therefore, it is most probable that gravitational compression will occur with perturbations that have mass $M \cong M_J$.

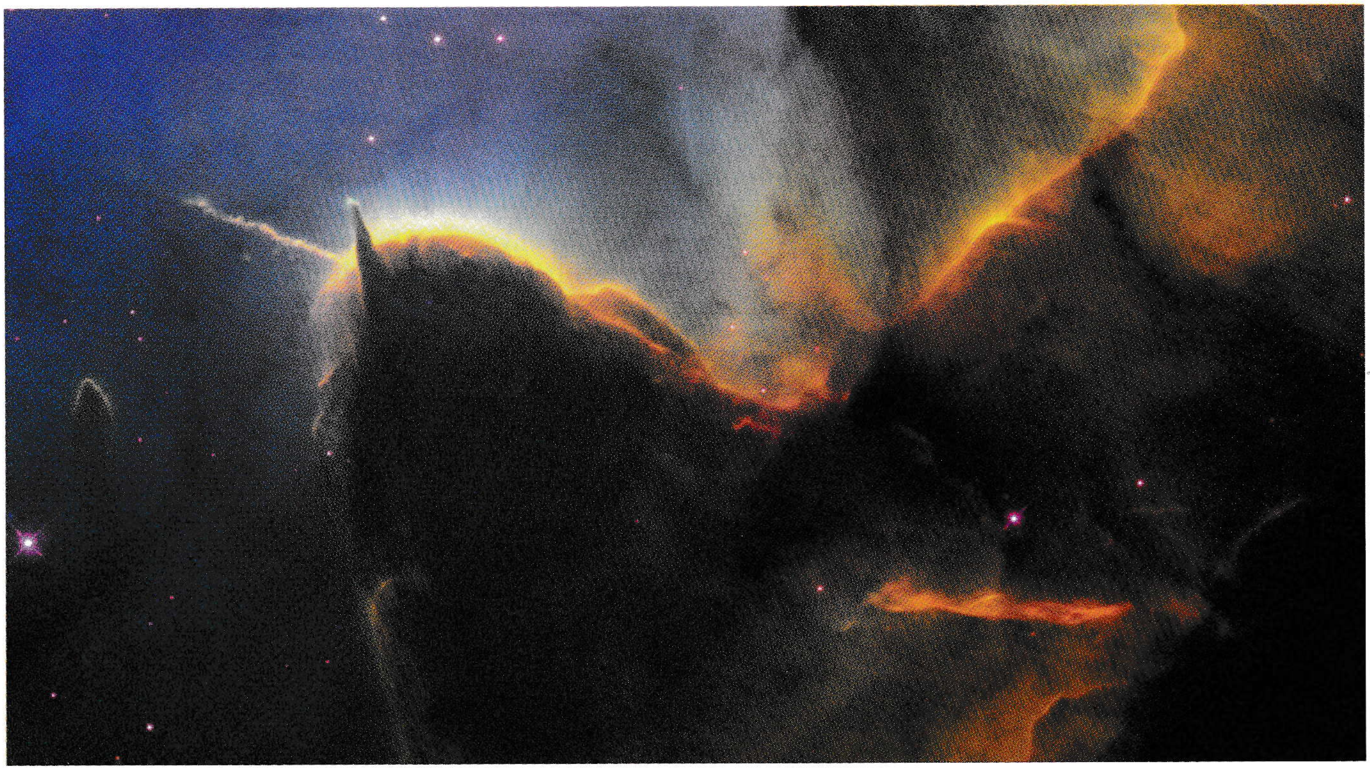
Two hundred and forty years after Newton wrote his letter to Bentley, Jeans wrote in his book *"The Stars in their Courses"* (1931): "Assume that at the beginning of time all the space was filled with gas... One can prove that this gas couldn't remain equally distributed in space, but immediately began to condense into balls. We can calculate how much gas is needed to form every ball."

Unfortunately, Jeans slightly overestimated his possibilities: he couldn't provide a sufficiently correct proof of the validity of his formulas, because in his time almost nothing was known about the physical properties and composition of the interstellar gas from which the stars are formed. However, now we can perform this work using modern experimental data.

Astronomers have shown that the chemical composition of stars and interstellar gas is very stable: every 1,000 hydrogen atoms are accompanied by 100 atoms of helium and 2–3 atoms of other (heavier) elements. In the dense cold clouds



Interstellar "twisters" in the heart of the Lagoon Nebula. The large difference in temperature between the hot surface and cold interior of the clouds, combined with the pressure of starlight, may produce strong horizontal shear to twist the clouds into their tornado-like appearance. The Lagoon Nebula and nebulae in other galaxies are sites where new stars are being born from dusty molecular clouds.



A stellar jet (the thin, wispy object pointing to the upper left) in the Trifid Nebula protrudes from the head of a dense cloud and extends three-quarters of a light-year into the nebula. The jet's source is a very young stellar object that lies buried within the cloud. Jets such as this are the exhaust gases of the star formation process. Radiation from the massive star at the center of the nebula is making the gas in the jet glow, just as it causes the rest of the nebula to glow.

where the stars are born, hydrogen exists in the molecular form H_2 with molar mass $\mu(H_2) = 2$ g/mol. With due allowance for other chemical elements, the mean molar mass of the interstellar gas is $\mu_{mean} = 2.3$ g/mol.

In astronomy, the density of interstellar gas $\rho = (\mu_{mean}/N_A)n$ is expressed in terms of the concentration of molecular hydrogen, $n(H_2)$. Taking into account the composition of interstellar gas, we get $n = 1.1n(H_2)$. Substituting ρ and n into Jeans' formulas, we obtain their modern form:

$$\lambda_J = 3.2 \cdot 10^4\text{AU}\left(\frac{T}{10\text{K}}\frac{10^4\ \text{cm}^{-3}}{n(H_2)}\right)^{1/2},$$

$$M_J = 0.3M_\odot\left(\frac{T}{10\text{K}}\right)^{3/2}\left(\frac{10^4\ \text{cm}^{-3}}{n(H_2)}\right)^{1/2}.$$

Here $M_\odot = 2 \cdot 10^{30}$ kg is the mass of the Sun, and AU (the astronomical unit) is the average distance between the Earth and the Sun (1 AU = $1.5 \cdot 10^{13}$ cm).

Recently, it was revealed that the typical regions of star formation are the small-scale condensations in the interstellar molecular clouds, where the temperature is T = 5–20 K and the hydrogen concentration is $n(H_2) \cong 10^4$–10^6 cm^{-3}. Correspondingly, $M_J = (0.02–2)M_\odot$ in these regions. Astronomical observations have shown that the masses of most stars are confined to this range. Thus, Jeans' theory works!

Another touchstone for this theory is the size of the "ancestor" cloud of the stars. For our Sun this size is estimated as 10^4 AU: this is the value of the radius of the dense interior of the Oort Cloud, which contains most of the mass of the cometary nuclei. How is this value related to the predictions of Jeans' theory?

The radius of the gaseous fragment at the beginning of its gravitational instability is naturally taken to be $R_J = \lambda_J/4$. Since the mass of the fragment must be equal to that of the Sun, and assuming T $\cong$ 20 K, we get $n(H_2) = 10^4$ cm^{-3} and $R_J = 10^4$ AU. This value is the same as the initial radius of the Oort Cloud. Need we look for other proofs of the validity of Jeans' theory?

Of course, the simple Jeans formulas do not take into account many physical processes occurring in the interstellar medium. These formulas are correct for an ideal homogeneous gas at rest, which never really exists in Nature. The real interstellar matter is in constant motion, often with speeds greater than the speed of sound. Moreover, this motion occurs under the strong influence of the magnetic field, the gravitational attraction of the neighboring stars in the Galaxy, and the radiation pressure from the brightest of them. No wonder that after Jeans many physicists developed and delineated the gravitation instability theory. Among them were E.M. Lifshits, S. Chandrasekhar, Y. B. Zeldovich, and J. Silk.

This theory is now highly developed: it takes into consideration the expansion and rotation of the gaseous medium and its interaction with magnetic fields and with external sources of gravitation. However, analysis of these additional physical phenomena did not change the fundamental theoretical conclusions: in the most modern studies the values of the Jeansean mass and radius are used to estimate the effects produced by gravitational instability. One of the reasons for such scientific longevity is the limited capabilities of modern astronomical devices: only on rare occasions is it possible to obtain some other parameters of the protostellar medium in addition to its density and temperature.

Sir James himself was filled with enthusiasm by the results of his studies, and, first of all, by their simple and clear nature. He wrote: "It is clear why all stars have similar mass; all of them were formed by the same process. Perhaps they look like the goods produced by the same machine-tool."

At present, we know that the masses of various stars may differ by as much as a thousand times. There are even greater variations in the parameters of the interstellar medium. Therefore, we can consider Jeans' enthusiasm as premature. However, this was clear to Jeans as well, who realized that only the first hurdle was left behind on the long road *per aspera ad astra*. Foreseeing future problems in the theory describing the formation of stars and galaxies, he wrote a note of warning: "At the present state of our knowledge, any attempt to dictate the final solution of the basic cosmogonic problems would be nothing but pure dogmatism."

Almost one century passed after the plight of Sir James. This century yielded plenty of data on star formation. It also demonstrated that Jeans' theory of gravitational instability stood the most difficult trial—the test of time. Isn't this wonderful in the century of quantum and relativistic physics, which made so many classical theories obsolete?

Quantum on cosmology and astrophysics:

Y. Zeldovich, "A Universe of Questions," January/February 1992, pp. 6–1.

Y. Solovyov, "The Universe Discovered," May/June 1992, pp. 12-18.

W. A. Hiscock, "The Inevitability of Black Holes," March/April 1993, pp. 26–29.

A. Sakharov, "Does the Elementary Length Exist?," May/June 1997, pp. 14–20.

S. Silich, "Interstellar Bubbles," November/December 1997, pp. 14–19.

I. D. Novikov, "The Thermodynamic Universe," March/April 1998, pp. 10–14.

A. Kuzin, "The Anthropic Principle," January/February 1999, pp. 4–9.

CONTINUED FROM PAGE 11

Hilbert was so sure that functions of three variables have a more complex structure than functions of two variables that he proposed the following hypothesis as his 13th problem: There exists a function of three variables that cannot be represented as the superposition of continuous functions of two variables. This hypothesis was radically disproved by Arnold and Kolmogorov in 1957. It turned out that any function of n variables can be represented as a superposition of the simplest function of two variables [the addition function $(x, y) \to x + y$] and continuous functions of one variable.

The emergence of topology was followed by great achievements. Here are several examples. A circle divides the plane into two parts, which means that it is impossible to connect a point outside this circle with a point inside it by a continuous curve without intersecting the circle. The 19th-century French mathematician Jordan proved that any homeomorphic (that is, continuous one-to-one) image of a circle also divides the plane into two parts. The Dutch mathematician Brouwer generalized this result for the homeomorphic image of a multidimensional sphere. In his proof, he developed Poincaré's ideas. In particular, he proved a remarkable theorem called the *Brouwer fixed-point theorem.* In the simplest case, this theorem is as follows: *Any continuous mapping of the plane disc onto itself has a fixed point.* Further development of topology led to generalizations of these results obtained by the Americans Alexander and Lefshetz, the German Horf, the Russians Aleksandrov, Kolmogorov, and Pontryagin, and others.

Several elegant topological problems were posed by Poincaré. An example is the problem of three closed geodesics. Suppose we take a smooth stone and try to fit a rubber band over it. If the rubber band doesn't slip off, we have found a closed geodesic. Poincaré conjectured that, for any smooth oval-like body, there exist three closed geodesics, and this number cannot be increased. In particular, for an ellipsoid with three different axes, this number is exactly three. This problem was solved by the Soviet mathematicians Lyusternik and Shnirelman.

* * *

We have given an account of only some of the events of our century in which mathematics played an important role; we have also discussed some topics concerning our "internal world." We hope that you have gained some feeling for the enormous amount that was done in this rather short period of time. We hope to publish in our journal articles devoted to recent discoveries so that the reader can be proud of our time. With contemporaries such as Einstein, Kolmogorov, and Sakharov, it is certainly a dynamic time to be working as a scientist. I hope you too can one day share in the excitement.

Geometry of sliding vectors

The Swan is heaven bound,
The Crayfish pulls backward, the Pike moves upstream.

by Y. Solovyov and A. Sosinskiy

THE PHRASE IN THE EPIgraph is taken from the well-known fable of the famous Russian writer I. A. Krylov. In this fable, the Swan, the Crayfish, and the Pike harnessed themselves to a cart, but could not move it because everyone pulled it in a different direction. Would you be able to determine the motion of the cart if you are given the forces and directions in which the Swan, the Crayfish, and the Pike are pulling?

This problem comes up quite often: certain forces are applied to a rigid massive body at certain points. How will this body move? Strange as it may seem, the usual vector approach studied at school in physics and mathematics classes is quite inadequate for solving such a problem.

If we must take into account not only the mass, but also the size of a body, and if we deal with a real body rather than a material point, it is not always clear how the forces applied to the body at different places should be added, and what we are allowed to do with the corresponding vectors. Indeed, what are such "real" vectors? This article presents an answer to this question in the form of a small mathematical theory—the *theory of sliding vectors.*

What kinds of vectors exist?

A vector in the plane or in space is usually depicted by a directed segment $\overrightarrow{AB}$. It is defined by two points: its *origin* (or *point of application*) *A* and its *terminus B.* If the segment *AB* is extended in both directions, we obtain a line, which is called the *line of action* of the vector *AB.*

When are two vectors considered the same? The answer to this question defines how we will think of vectors, and it depends on the kind of physical objects they represent. We can think of at least three types of vectors.

1. It may happen that two *geometrically equal* vectors depict one and the same physical or mechanical magnitude. [Remember that in geometry, two vectors (directed segments) are said to be *congruent* if their lines of action are parallel, their lengths are equal, and the order of their points defines the same direction on their lines of action; in other words, if these vectors are congruent under parallel translation.] Such vectors, which have neither a definite line of action nor a definite point of application, are said to be *free.* For example, the vectors of magnetic induction of a constant magnetic field or the velocity of an inertial coordinate system with respect to another inertial coordinate system are free vectors. We may consider them as being applied at any point. Mathematicians, as well as physicists, study free vectors. The vector that defines a translation is the simplest example. When such a vector is applied to a point, its terminus indicates the image of this point under the translation.

2. On the other hand, we sometimes encounter physical quantities represented by vectors that cannot be separated from their point of application. Such vectors are called *localized.* The vector of the instantaneous velocity of a moving point is one such example. It cannot be separated from the moving point (of course, we assume that the other points in space are not moving with the same velocity).

3. Finally, it may happen that two geometrically congruent vectors represent equal physical quantities only if they have the same line of action.

Art by Jose Garcia


B
D
A
C

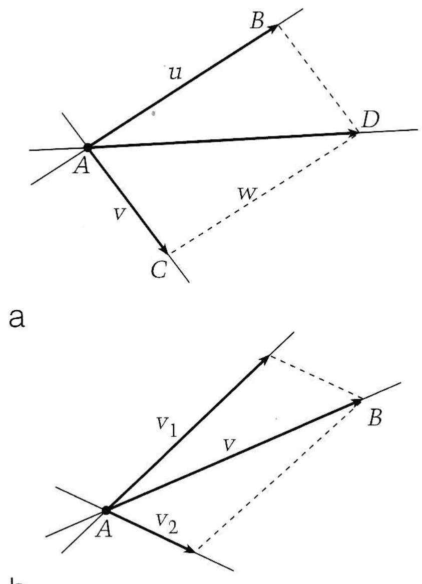


Figure 1

An example is given by vectors that represent forces acting on a rigid body. Such vectors, which cannot be separated from their line of action, are called *sliding* vectors. These vectors model forces that actually act on rigid bodies having a definite size and shape rather than on abstract and infinitesimally small points. The rigidity of the body being acted on is important to our model. Such a body neither expands nor contracts, and passes the force along its line of action without loss of magnitude. Thus, the particular point of application of the force along its line of action is irrelevant, and in our model, we allow it to "slide" along its line of action.

In this article, we deal with sliding vectors. In what follows, the word *vector*, unless otherwise noted, should be interpreted as a sliding vector; instances of free or localized vectors will be specifically indicated.

We will denote (sliding) vectors by bold letters, for example $\mathbf{v}$, or by two letters with a bar, for example $\overline{AB}$. Free vectors will be denoted by letters with an arrow above them, for example, $\vec{v}$ or $\overrightarrow{AB}$.

Systems of sliding vectors

Thus, a *(sliding) vector* $\mathbf{v} = \overline{AB}$ is defined by the line $l = AB$—its line of action—and the directed segment AB on this line. Two vectors $\mathbf{v} = \overline{AB}$ and $\mathbf{u} = \overline{CD}$ are considered identical if they define the same free vector ($\overrightarrow{AB} = \overrightarrow{CD}$) and have the same line of action (the lines AB and CD coincide).

In what follows, we will consider finite *systems* of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, rather than individual vectors. Such systems correspond to systems of forces that can be applied to a rigid body. The order in which the vectors are enumerated is irrelevant. However, one and the same vector may occur several times in the system, and it must be taken into account as many times as it occurs.

Our aim is to learn how to transform such systems of vectors and to reduce, if possible, an arbitrary system to a simpler one. For this purpose we need some elementary operations on systems of vectors and a notion of equivalence.

Elementary operations and equivalence of systems of vectors

If the lines of action of two vectors $\mathbf{v}$ and $\mathbf{u}$ intersect, these vectors can be added in a natural way: if $\mathbf{u} = \overline{AB}$ and $\mathbf{v} = \overline{CD}$ (where A is the common point of their lines of action), then the vector $\mathbf{w} = \mathbf{u} + \mathbf{v}$ is defined as the vector $\mathbf{w} = \overline{AD}$, where the point D is obtained by adding the free vectors $\overrightarrow{AB} + \overrightarrow{AC} = \overrightarrow{AD}$ by the *parallelogram law* (figure 1a). Such an addition of "intersecting" vectors is the first elementary operation. It reduces a system of two vectors $(\mathbf{u}, \mathbf{v})$ to a system consisting of the single vector $(\mathbf{w})$.

This operation has an inverse: for any vector $\mathbf{v} = \overline{AB}$, one may choose two arbitrary lines passing through an arbitrary point on the line of action (say, through point A) and *resolve* the vector $\overrightarrow{AB}$ by the parallelogram law (figure 1b). Thus, we obtain a system of two vectors $(\mathbf{v}_1, \mathbf{v}_2)$ from the single vector $(\mathbf{v})$.

Note that only vectors with intersecting or coinciding lines of action can be added. The addition of vectors that have the same line of action is especially simple. In particular (see figure 2), two opposite vectors $\mathbf{v} = \overline{AB}$ and $-\mathbf{v} = \overline{BA}$ give a null vector upon addition: $0 = \overline{AA}$.

The second elementary operation consists of *eliminating a null vector*: that is, in passing from the system $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{v}, -\mathbf{v})$ to the system $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$. In particular, for $k = 0$, the simple pair $(\mathbf{v}, -\mathbf{v})$ is reduced to the empty or null system, which is denoted in the same way as the null vector, by 0.

The second elementary operation also has an inverse: this is the *generation of the null vector,* that is, passing from the system $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ to the system $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{v}, -\mathbf{v})$, where $\mathbf{v}$ is an arbitrary vector.

We say that two systems of vectors $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k)$ and $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_l)$ are *equivalent* if each can be reduced to the other by a finite sequence of elementary operations.

From the point of view of mechanics, it is clear why we are interested in elementary operations and equivalent systems. Indeed, two equivalent systems of forces act identically on a rigid body. You can convince yourself of this fact not only experimentally, but also by simply thinking about the physical meaning of the elementary operations. By reducing complex systems of vectors to simple ones, it becomes easier to understand the resulting action of a given complex system on a rigid body.

Problem 1. How do these elementary operations work when applied to free vectors? To localized vectors?

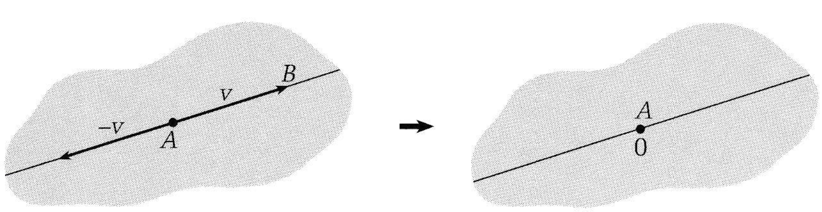
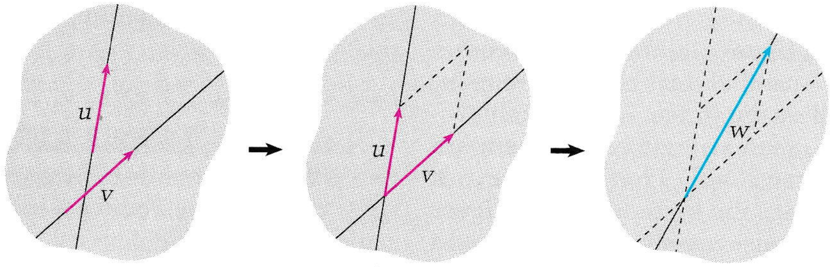
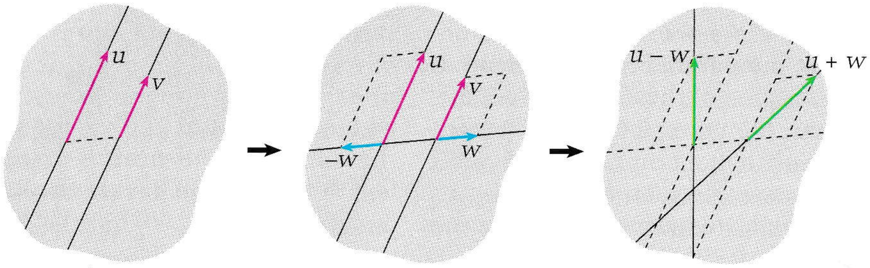


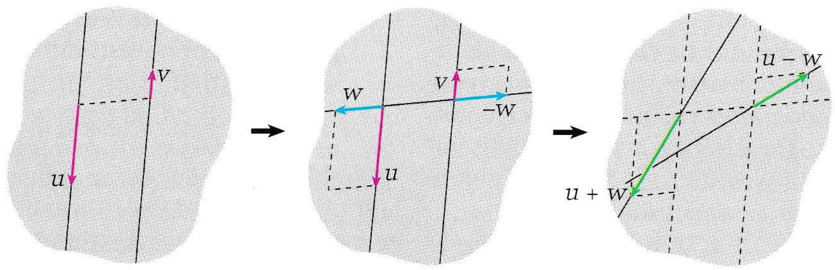
Figure 2



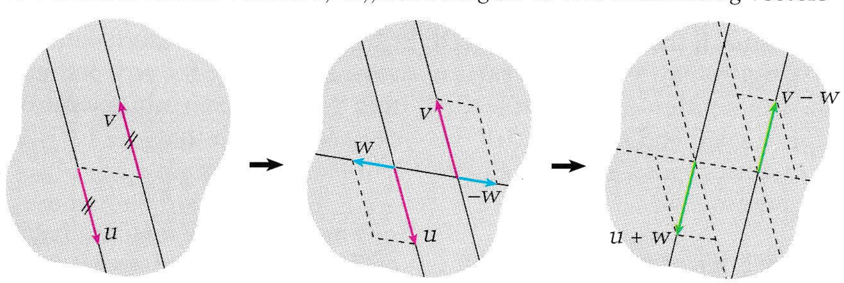
a) *two intersecting vectors to a single vector*



b) *two codirectional vectors to two intersecting vectors*



c) *two antidirectional vectors of different lengths to two intersecting vectors*



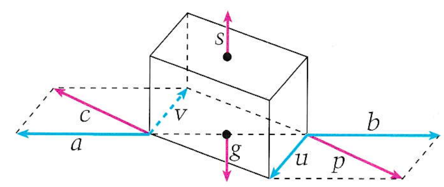
d) *two antidirectional vectors of equal length to another pair*

Figure 3. *The simplest transformations of planar systems of vectors.*

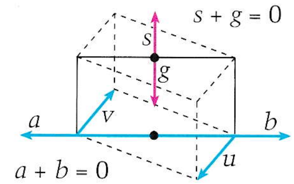
Simplest systems in the plane: pairs of vectors

Figure 3 shows how some simple systems of vectors in the plane can be reduced to other systems by elementary operations. We advise the reader to follow these operations carefully and to practice them in various cases.

Special attention should be paid to figure 3d. It shows systems of two opposite vectors of equal length with parallel lines of action. Such systems are called *pairs* for short. It can be



a



b

Figure 5

seen from the figure that a pair cannot be simplified; it can only be rotated (transformed into another pair).

Figure 4 demonstrates another series of elementary operations. Here, we make the initial one-vector system more complex, rather than simplify it: from the initial vector, we obtain a vector with another line of action and a pair. However, if we look at this figure in the reverse order (which makes sense because every elementary operation has an inverse), we obtain a simplification: the system consisting of a vector and a pair is reduced to a single vector. The transformation $(\boldsymbol{v}) \to (\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{y})$ is particularly important. We will use it a lot in later work.

Note that the methods used for planar systems can be applied in spatial problems as well. For example, in figure 5, we have modeled the efforts of Krylov's hapless creatures. The vector $\boldsymbol{s}$ models the force applied by the swan, $\boldsymbol{c}$ the crayfish, and $\boldsymbol{p}$ the pike, while $\boldsymbol{g}$ represents the force of gravity. Following Krylov, we have chosen $\boldsymbol{s} = -\boldsymbol{g}$ and $|\boldsymbol{c}| = |\boldsymbol{p}|$. Then, setting $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{v}$, $\boldsymbol{p} = \boldsymbol{b} + \boldsymbol{u}$ (figure 5a), where $\boldsymbol{b} = -\boldsymbol{a}$, we can reduce the system $(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{p}, \boldsymbol{g})$ to the planar system $(\boldsymbol{u}, \boldsymbol{v})$.

Moment of a vector and moment of a pair

We have already seen that a pair of vectors cannot always be reduced to a single sliding vector. However, a pair exhibits a useful vector characteristic.

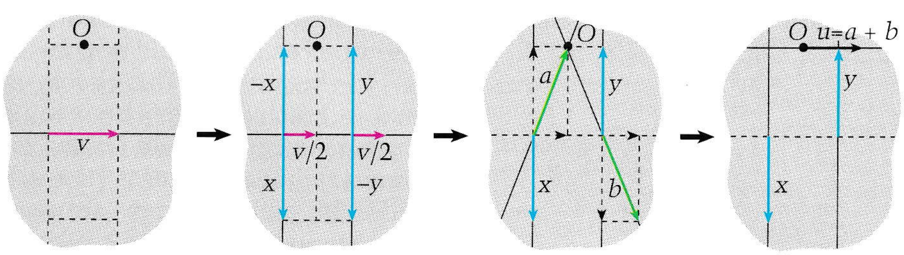


Figure 4. *Transformation of a single vector to a vector passing through a given point and a pair.*

Suppose we are given a vector $\boldsymbol{v} = \overline{AB} \neq 0$ and a point O not on its line of action. Then the *moment of vector* $\boldsymbol{v}$ *with respect to point* O is the *localized* vector OM, applied to the point. Its line of action is taken perpendicular to plane OAB, and its length $|OM|$ is the product of the length of $\boldsymbol{v}$ and the length of the perpendicular from point O to line AB. The direction of vector OM is chosen so that the direction of the rotation of $\boldsymbol{v}$ about O as observed from the point M is positive (i.e., counterclockwise; figure 6). If the point O lies on the line of action of $\boldsymbol{v}$ or if $\boldsymbol{v} = 0$, the moment is taken to be zero. The moment of a vector plays an important role in the analysis of rotations and has elegant applications in geometry.

Now suppose that a pair of vectors $\boldsymbol{v} = \overline{AB}$, $\boldsymbol{v}' = \overline{A'B'}$, and a point O are given. (Recall that when we talk about a *pair* of vectors, we mean that they act along parallel lines.) Let OM and OM' be the moments of $\boldsymbol{v}$ and $\boldsymbol{v}'$ with respect to O.

Problem 2. Show that the length and direction of the sum of the moments of vectors $\boldsymbol{v}$ and $\boldsymbol{v}'$ with respect to O are independent of the choice of this point. That is, show that this sum is actually a free vector.

This free vector is called the (*vector*) *moment* of the pair $(\boldsymbol{v}, \boldsymbol{v}')$.

Problem 3. If d is the distance between the parallel lines AB and $A'B'$, show that the length of the vector moment of the pair $(\boldsymbol{v}, \boldsymbol{v}')$ is $d \cdot |\boldsymbol{v}|$.

Problem 4. Show that a pair of vectors can be reduced to zero if and only if its vector moment is zero.

Reducing planar systems to a pair or a vector

We now prove the following remarkable theorem.

Any finite system of sliding vectors in a plane can be reduced either to a single vector or to a pair of vectors.

Proof. If the given system consists of a single vector, the theorem is proved. If the system contains vectors with intersecting lines of action, they can be simplified in pairs (figure 3a) until only vectors with parallel lines of action remain. If there are three or more such vectors, then at least two of them are codirectional. Therefore, we can simplify the system further (figure 3b). Thus, transformations 3a and 3b make it possible to reduce the proof of the theorem to the case of two vectors with opposite directions, but parallel lines of action. If the lengths of these vectors are different, then transformation 3c with the subsequent transformation 3a yields a single vector. If the lengths are equal, we have a pair. Thus, the theorem is completely proved.

Note that we not only proved the theorem, but also gave an effective algorithm for finding an equivalent vector (or pair). Also, if we are given an arbitrary fixed point, we can perform the elementary operations in such a way as to reduce the given system to a pair and a vector with the line of action passing through the given point. Indeed, if the system is reduced to a pair, this is obviously true. If the system is reduced to a single vector, the operation shown in figure 4 transforms this vector to a pair and a vector with the line of action passing through the given point.

Calculations with systems of vectors. Bases

We have learned how to reduce planar systems of vectors to simpler systems. We can also express one system of vectors in terms of other systems. In this way, we can perform calculations on systems of sliding vectors that are similar to our calculations with free vectors.

To describe this calculations, we define the sum of two systems $N = (\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k)$ and $M = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_l)$

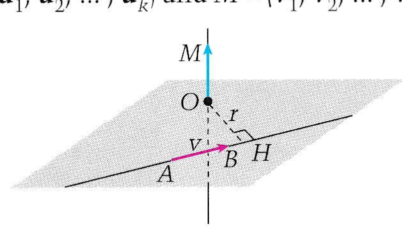


Figure 6

as their "free combination," i.e., as the system

$$L = N + M = (\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k, \boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_l).$$

Here, we interpret the equality of two systems of vectors to mean that they can each be reduced to the same system by elementary operations. The product of the system N and a number λ is defined as the new system

$$\lambda N = (\lambda \boldsymbol{u}_1, \lambda \boldsymbol{u}_2, \dots, \lambda \boldsymbol{u}_k),$$

where $\lambda \boldsymbol{u}$ denotes the vector with the same line of action as $\boldsymbol{u}$, of length $|\lambda| \cdot |\boldsymbol{u}|$, and having the same direction as $\boldsymbol{u}$ if $\lambda > 0$, but the opposite direction if $\lambda < 0$.

The system N is said to be expressed linearly in terms of systems $M_1, \dots, M_n$ if there exist numbers $\lambda_1, \lambda_2, \dots, \lambda_n$ such that

$$N = \lambda_1 M_1 + \lambda_2 M_2 + \dots + \lambda_n M_n.$$

Problem 5. Let O be an arbitrary point in the plane. Let $I = \boldsymbol{i}$ and $J = \boldsymbol{j}$ be two single-vector systems such that $\boldsymbol{i}, \boldsymbol{j} \neq 0$, $\boldsymbol{i}$ and $\boldsymbol{j}$ are non-collinear, and $\boldsymbol{i}$ and $\boldsymbol{j}$ are both applied at point O. Let K be any nonzero pair of vectors. Prove that for any planar system N there exist unique numbers α, β, and γ such that

$$N = \alpha i + \beta j + \gamma k$$

Those readers who are familiar with the concept of a vector space will see that this assertion states that I, J, and K form a basis for the set of planar system of vectors, and that therefore the *set of classes of equivalent sets of sliding vectors in the plane forms a vector space of dimension 3 with respect to the operations defined above.*

* * *

At this point, the authors had an argument about what applications of sliding vectors should be considered in this article. This theory is used in statics (truss theory), in structural mechanics, and in the kinematics and dynamics of rigid bodies. Its generalizations are used in some modern branches of geometry. However, in this article we give only one example from the kinematics of rigid bodies and present several problems at the end of the article.

Rotation of a rigid body

We are interested in the following kinematic problem: how can the motion of a rigid body rotating about an axis be described if this axis, in turn, rotates about another fixed axis? This problem will be solved if we give a method for finding the velocity of any point of the body with respect to a fixed frame of reference.

The solution can be elegantly formulated in terms of sliding vectors. Before presenting this solution, let us first discuss the various kinds of motion of a rigid body. We begin with the simplest examples.

During *uniform rectilinear motion,* a rigid body moves in a fixed direction with constant speed. For all points, the velocity is the same and is independent of time. All points move along parallel lines, and the motion is characterized by a single free vector $\vec{v}$ (the velocity).

During *uniform rotational motion,* a body moves with constant angular speed about a fixed axis. The velocity of all points on this axis is zero, and the velocity of any other point is perpendicular to the plane passing through this point and the axis. The magnitude of the velocity is proportional to the distance from the point to the axis. The points on the axis do not move at all, and all other points move along circles centered on the axis. This type of motion is characterized by a single sliding vector ω with the line of action identical to the axis. This vector ω is called the rotational velocity.

During *uniform spiral motion,* a body rotates uniformly about an axis (called the spiral axis) and simultaneously moves along it with constant speed. The velocity of every point is the sum of the vectors of the rotational and rectilinear motions (figure 7). The points of the body describe spiral lines, and only the points of the axis move along it. This type of motion is characterized by a pair of vectors $(\vec{v}, \omega)$: the free vector of the rectilinear motion $\vec{v}$ and the sliding vector of the angular velocity ω (figure 7).

Of course, the possible types of motion of a rigid body are not exhausted by these examples. For example, the vectors $\vec{v}$ and ω may depend on time. However, the following remarkable assertion is true (see problem 10): however complex the motion of a rigid body may be, the instantaneous distribution of velocities of its points coincides with one of the three types described above. By the way, note that rectilinear and rotational motion may be regarded as particular cases of spiral motion. Two examples of more complex motions are illustrated in figures 8 and 9.

To return to our kinematic problem, suppose that the moving axis M_tN_t rotates with constant angular speed ω_0 about a fixed axis AB, and that the body itself rotates about M_tN_t with constant (in magnitude) speed ω_1. We stress that the latter angular speed considered as a sliding vector varies: $\omega_1 = \omega_1(t)$, because its line of action, M_tN_t, changes. How can the resulting motion be described? This depends on the relative arrangement of AB and M_tN_t.

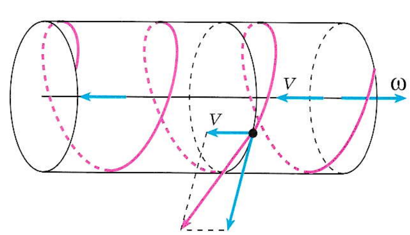


Figure 7

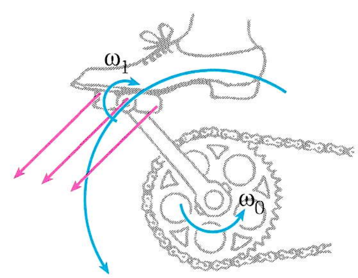


Figure 8

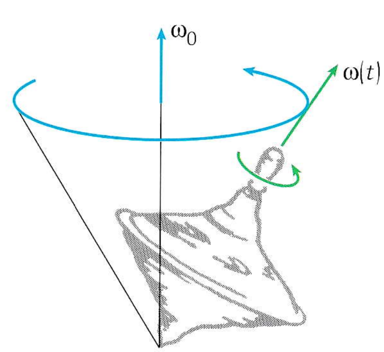


Figure 9

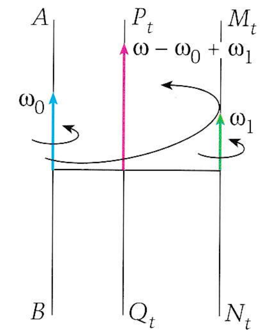


Figure 10

1. The axes coincide. In this case, the body rotates about the fixed axis $AB = M_tN_t$ with a constant velocity $\omega = \omega_0 + \omega_1$. An example is provided by the hour or minute hand of a clock at the North Pole.

2. The axes are parallel and $\omega_0 \neq -\omega_1$. At any instant of time t, the velocities of all points of the body are the same as if this body were uniformly rotating about the line P_tQ_t, the line of action of the vector $\omega = \omega_0 + \omega_1$ (here the plus sign denotes the sum of two sliding vectors, as in figures 3 and 10), with angular velocity ω. The body is said to have the instantaneous angular velocity ω and the instantaneous axis of rotation P_tQ_t. In the case under consideration, the instantaneous axis rotates about the fixed axis AB, remaining parallel to it.

3. The axes are parallel and $\omega_0 = -\omega_1$. In this case, the body executes translational motion with constant velocity $\vec{v}$ equal to the moment of the pair (ω_0, ω_1). An example is provided by the moving part of a bicycle pedal (figure 8).

4. The axes intersect. This case is similar to case 2: at any instant of time t, the body has an instantaneous axis of rotation and an instantaneous angular velocity $\omega = \omega_0 + \omega_1$. The difference is that the instantaneous axis describes a cone with vertex at the point of intersection of the axes, rather than a cylinder as in case 2. One such example is the precessional motion of a top (figure 9).

5. The axes are skew. In this case, the body has, so to speak, an "instantaneous spiral axis" (see problems 6 and 7).

We see that the solution to our kinematic problem can be elegantly formulated in terms of the addition of sliding vectors. Simple proofs of propositions 1–4 are left to the reader.

Various problems

1. Suppose that a system of sliding vectors $(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$ is given. Consider free vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ equal, respectively, to the given sliding vectors. We call the sum of these free vectors the resultant vector of the given system. The sum of the moments of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ with respect to a point O is called the resultant moment of the system with respect to this point. Prove that any two equivalent systems have identical resultant vectors and identical resultant moments with respect to the same point O.

2. Prove that any system of vectors lying in one plane is equivalent to three vectors directed along the sides of an arbitrary triangle in the same plane.

3. Prove that any planar system of vectors that are perpendicular to the sides of a convex n-gon at their midpoints is equivalent to zero if the lengths of these vectors are proportional to the corresponding sides and all the vectors are directed towards the interior of the polygon (or towards its exterior).

4. Formulate and prove a proposition for the tetrahedron, similar to that in the previous problem.

5*. Prove that any system of spatial vectors is equivalent to six vectors directed along the edges of an arbitrary tetrahedron.

6*. Prove that any system of spatial vectors is equivalent to a system consisting of a vector (passing through an arbitrary point) and a pair. Deduce from this fact that systems of sliding vectors in three-dimensional space (taking into account their possible equivalence) form a six-dimensional vector space).

Hint. First reduce the given system to three sliding vectors passing through three arbitrary points (by resolving each vector in terms of three directions), then apply the same technique to two vectors, one of which passes through the given point, and finally apply the construction from figure 4.

7*. A rigid body rotates with speed ω about an axis that, in turn, executes uniform rectilinear motion with velocity $\vec{v}$. Prove that (a) the body executes spiral motion if $\omega \parallel \vec{v}$; (b) the body has an instantaneous axis of rotation if $\vec{v} \perp \omega$; (c) the body has an instantaneous spiral axis if $\vec{v}$ is neither parallel nor perpendicular to ω (this means that all points of a line—the spiral axis—have identical velocities $\vec{v}$ directed along this axis, and the velocities of all other points are equal to the sum of $\vec{v}$ and the vector of the instantaneous speed of rotation of the point about the spiral axis).

8*. Prove that in the case of skew axes AB and M_tN_t in the above kinematic problem, the body has an instantaneous spiral axis parallel to the vector $\vec{\omega}_0 + \vec{\omega}_1$ and intersecting the common perpendicular to the axes AB and M_tN_t. Describe more precisely the position of the instantaneous spiral axis and the velocity along this axis.

Hint. Use problem 7.

9*. Describe the resultant motion of a body that executes instantaneous spiral motion with respect to an axis that, in turn, executes spiral motion with respect to a fixed axis.

10*. Prove that however complex the motion of a rigid body may be, the distribution of the velocities of its points at any instant of time is the same as if the body executed uniform (rectilinear, rotational, or spiral) motion.

11. The moment of the vector $\overline{AB}$ with respect to the point O is equal (in magnitude) to the doubled area of the triangle OAB. Use this fact to prove that the set of points in the interior of any convex polygon for which the sum of the distances to its sides (more precisely, to the lines containing the sides) is constant, is either a segment, or the entire polygon, or is empty.

"When, among partners, concord there is not,

Successful issues scarce are got ..."

Let us return to the fable cited at the beginning of this article. Suppose that a rectangular heavy cart is standing on a road. The gravitational force g acts on the cart; we may assume that this force is applied at the center of gravity of the cart. This force is compensated for by the reaction of the ground: the cart is not moving. Now three agents—the Swan, the Crayfish, and the Pike—begin pulling the cart in different directions with forces s, c, and p, respectively, as shown in figure 5. What will happen?

Having the theory at our disposal, we can answer this question. The system of forces $(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{p}, \boldsymbol{g})$, like any other system of vectors, can be reduced to a pair and a vector (see problem 6). We have already done this and found the resulting pair (u, v) in the horizontal plane. It is clear that the cart will rotate, remaining at the position where it is standing. Thus, we can confirm the words of the fable: "I know the cart remains there, yet."

HOW DO YOU FIGURE?

Challenges

Math

M286

Quad query. Quadrilateral $ABCD$ is inscribed in a circle. Let M be the point of intersection of its diagonals, and L be the midpoint of arc AD (which does not contain the other vertices of the quadrilateral). Prove that the distances from L to the centers of the circles inscribed in triangles ABM and CDM are equal.

M287

Up a creek. A mathematician was walking home along a creek upstream with a speed equal to one and a half times the speed of the current. He had a hat and a stick. He wanted to throw the stick into the stream, but accidentally threw his hat instead, and continued walking with the same speed. Several minutes later he noticed his mistake, threw the stick into the stream, and ran back with a speed equal to twice his walking speed. He caught up with the hat, picked it up (instantly), and continued walking upstream with his initial speed. Ten minutes after he caught up with the hat, the mathematician encountered the floating stick. How much earlier would he have arrived home if he didn't mix up the stick and the hat?

M288

Simply irreducible. Prove that all fractions of the form

$$\frac{101011\ldots10101}{110011\ldots10011}$$

are equal, provided that the numerator and the denominator contain four zeros and the same number of ones (the dots in the middle stand for a sequence of 1's). Find the irreducible fraction equal to these fractions.

M289

Smallest segment. A point M is given on the graph of the function $y = x^2$. A perpendicular is drawn to the tangent to this graph through the point M. This perpendicular cuts off a segment of the parabola. Find the minimum value of the area of this segment.

M290

Stuck in the middle. Two circles and an isosceles triangle are arranged as shown in figure 1. Find the altitude of the triangle drawn to its base if the sum of the circles' diameters is 2.

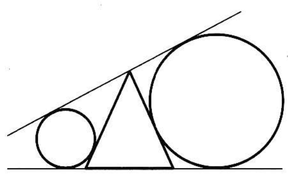


Figure 1

Physics

P286

Lunar perturbations. How does the Moon affect the Earth's orbit around the Sun? (A. Dozorov)

P287

Diving bell. A diving bell is an apparatus used for working under water. It is a thin-walled cylinder, which is sunk from a boat to the seabed with its base up. What should be the thickness of the bell's walls and base (that is, ceiling) to make it rest at the bottom of a pond at a depth of $H = 3$ m? The inner radius of the diving bell is $r = 1$ m, its height is $h = 2$ m, and the density of steel is $\rho_{st} = 7.8 \cdot 10^3$ kg/m^3.

P288

Inside a capacitor. A plate of thickness h is made of weakly conducting material with resistivity ρ. It is placed inside a parallel-plate capacitor, parallel to its plates. The capacitor is charged to voltage V_0. Find the maximal current that will flow via the plate after short-circuiting the capacitor. The area of each plate of the capacitor is S, and the distance between them is d, which is much smaller than the size of the plates. (V. Deryabkin)

P289

Flying plasmoid. Photographs of radiating spherical plasmoids (plasma clusters) that move with uniform deceleration until they stop completely look like bands of length l. The maximal width of the bands is $d \ll l$. The distance between the plasma and the objective of a camera is L, and the focal length of the objective is F. The minimal exposure time needed to obtain an image on the photograph is τ. Find the acceleration of a plasmoid. The objective is open until the plasmoids stop completely. (V. Sergievich)

P290

A spherical lens. A hemisphere of radius R made of transparent glass with refractive index $n = 2$ has a

CONTINUED ON PAGE 41

Principles of vortex theory

Inside the hydrodynamics of Helmholtz

by N. Zhukovsky

MECHANICS IS THE CHILD of the combined efforts of geometricians and analysts. It was not a rare event when the complicated analytic formulas were illustrated in a clear visual form through witty geometrical plots. Such interpretations encompassed the problem in its splendor and completeness and clarified many features that were overlooked in purely analytic studies. This was the case with the problem of the rotation of a solid body about its center of gravity. It was first solved by the great Leonard Euler in analytic form. However, the solution was buried under a mass of formulas until Louis Poinsot clarified it with simple and clear geometrical interpretations. Another example is the work of Hermann Helmholtz, who illuminated many dark corners in the problems of moving fluids.

Almost all the papers of Helmholtz on mechanics were focused on hydromechanical problems. It is not an exaggeration to say that modern hydrodynamics was developed mostly by Helmholtz. However, the most remarkable work of this scientist on hydrodynamics was published only in 1858, after a long period of 43 years, when the formulas that had the vortex conservation principle hidden somewhere inside their spacious structure were obtained by Augustin Cauchy. However, Cauchy treated his results only from the analytical viewpoint and did not foresee the huge number of questions that could be solved with the help of the corresponding geometrical interpretation of his inferences.

I shall try to explain in the simplest way the concept of a vortex as elaborated by Helmholtz. Imagine a cylindrical vessel of finite height (figure 1) and rather large base, which is filled with a fluid (a gas or a liquid). Suppose that this fluid moves in the following way: a central cylindrical column of some width rotates as a solid body about its axis, while other parts of the fluid circulate around this column with speeds that are inversely proportional to the distance from the axis of the column. Thus, the speed of the fluid grows in the direction toward the axis of the cylinder, and it coincides with the speed of the central column at its boundary.

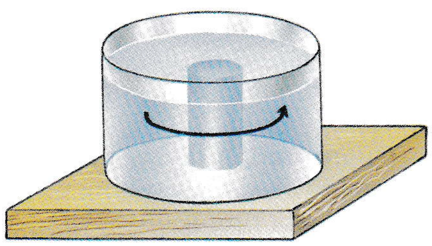


Figure 1

Such motion in fluids is called a vortex, and its characteristic column is referred to as the vortical filament. We define the intensity of a vortex as half the product of the fluid speed at the surface of the vortical filament and the circumference of the normal cross section of the filament. Twice the value of this product is called the circulation. In general, the circulation along a closed contour inside a moving fluid is equal to the product of the length of the contour and the mean tangential velocity along the contour.

Since in the moving liquid shown in figure 1 the speeds are inversely proportional to the radius, the circulations along all horizontal circles centered about the axis of the column (and enclosing it) are identical and equal to twice the intensity of the vortex.

In contrast, the circulation along a contour consisting of two arcs of concentric circles and two radial segments, and lying outside the filament (the contour *ABCD* in figure 2), is zero. This property can be gen-

Art by Pavel Chernusky

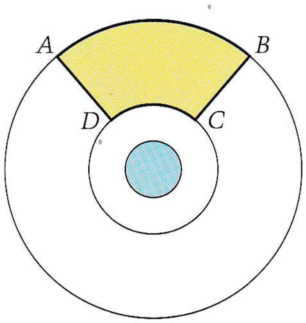


Figure 2

eralized: one can prove that the circulation along any closed contour enclosing the filament is equal to twice the intensity of the filament, while the circulation along any closed contour that does not enclose the filament is zero.

This property makes it possible to detect a vortical filament in a moving fluid. To this end, one must draw a closed contour and calculate the circulation along it. If this circulation is nonzero, the contour is threaded by a vortical filament. Now we contract the contour until the circulation changes. In this way we may detect the surface of the filament.

If our wide vessel contains only one vortex due to a straight vortical filament, it will remain motionless. However, if two such vortices are generated in the vessel, which circulate around parallel vortical filaments, the filaments will move. Figure 3 shows the top view of two vortical filaments of different intensities that rotate in the same direction. Since the vortex generated by the left filament spins the entire liquid mass clockwise around the axis of the filament, the right filament will acquire a velocity directed downward and perpendicular to the radius. For the same reason, the vortex generated by the right filament will impart to the left filament an upward velocity. As a result, both filaments rotate in the clockwise direction around a certain point. The location of this point can be determined if we place at the centers of the two filaments masses proportional to the intensity of the corresponding filaments and find their center of mass.

If the vortices circulate in opposite directions (figure 4), the vortical filaments rotate about an axis located at the side of the filament with the larger intensity, and this rotation will be in the same direction the stronger vortex rotates. If the intensities of the two vortices are equal, their common center of revolution is located at infinity, so that both filaments move forward in the direction normal to the line connecting their centers (figure 5).

Figure 6 shows the trajectories of three vortical filaments rotating both counterclockwise (filaments 1 and 2) and clockwise (filament 3). The indices mark the same time values for all the vortices.

Our concept of a straight vortical filament generated in a sufficiently wide cylindrical vessel can be generalized for vortical filaments generated in any mass of fluid. These filaments can also be detected by means of the circulation calculated along closed contours, as we did for a straight filament.

If vortices are generated in a frictionless liquid subject to forces which do not violate energy conservation, the following remarkable theorem holds: the circulation determined for any closed contour in such a liquid does not change when the particles forming the contour are displaced.

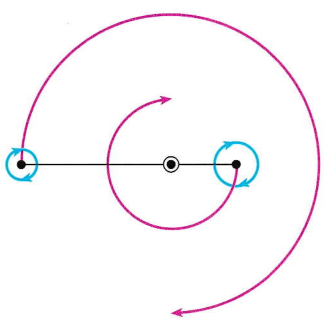


Figure 3

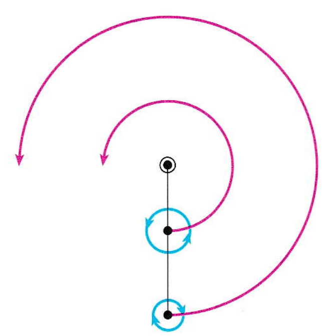


Figure 4

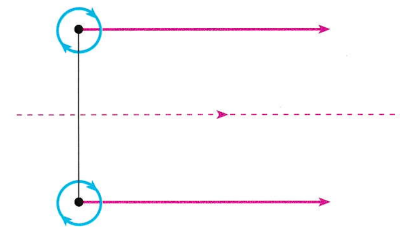


Figure 5

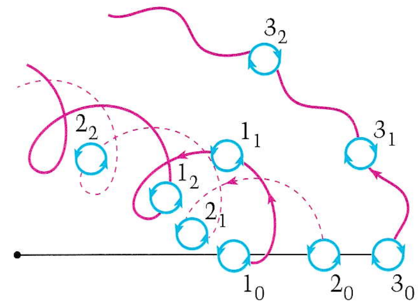


Figure 6

It follows from this theorem that the particles of the liquid forming the vortical filament will generate a filament of the same intensity during a displacement of the vortex. Another consequence of this theorem is that no new vortical filaments can arise in such a fluid. Indeed, if we look for a vortical filament with the help of circulation values calculated along closed contours, we shall obtain zero for all contours not threaded by the filament, and the same constant value for contours enclosing it. Therefore, we shall conclude that the chosen contours are pierced by a vortical filament of the same intensity.

Another consequence of the same theorem is that throughout its motion the vortical filament will either have its ends at the fluid boundary (at the walls of the vessel or at the free surface) or remain closed.

In fact, in order to detach itself from the walls of the vessel, the base of the vortex would have to shrink to zero size. Since the circulation along the circumference of the base must remain constant, such a contraction would require an infinite speed for the circulating liquid at the foot of the vortex.

The hydrodynamic pressure in a fluid decreases when the speed increases. Thus, when the base of the vortex at the wall of the vessel contracts, the increased speed will decrease the pressure there, so that the neighboring fluid will push the particles of the base of the vortex and prevent its detachment from the wall. It looks as if the vortical filament "sticks" to the walls of the vessel with its ends. If the end of the filament is located at the free surface, such "sticking" can be seen by the "crater" formed at the free surface near the end of the filament.

If the ends of a vortical filament are not located at the boundaries of the liquid, they must be attached to each other, which means that the vortical filament must be closed: its ends are "stuck" to each other.

The simplest form of a closed vortical filament is a ring (torus), as shown in figure 7. All the particles of the fluid lying outside the ring move along closed curves threading the ring in such a way that the circulation values along all these trajectories are identical and equal to the circulation along the contour of the normal cross section of the ring. If we "enter" into the ring, we obtain various values for the circulation along the trajectories of its particles. The particles at the surface of the ring have the largest speeds. The speed decreases in the direction from this surface to the inner part of the ring and becomes zero at some axial line. The speed also decreases with the distance from the ring in the adjacent mass of fluid. The speed is inversely proportional to the cube of the distance from the ring for particles at large distances from the ring.

As we have seen, two parallel straight vortical filaments of equal intensity, rotating in a fluid in opposite directions, run along the line normal to the plane in which both filaments lie. For the same reason, the vortical ring will not stay still, but will run in the direction normal to the plane of the ring to the side in which fluid runs out from the ring.

Figure 7 shows that the particles of the fluid that move along the upper closed trajectories will push the lower edge of the ring to the right.

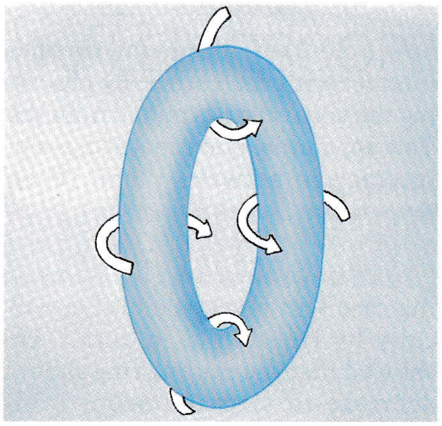


Figure 7

Similarly, the particles of the fluid moving along the lower closed trajectories will push the upper edge of the ring in the same direction. Therefore, the entire ring will move uniformly to the right and will carry the part of the fluid rotating around it. This motion is more rapid for smaller rings with stronger intensity.

As we mentioned above, vortical filaments generated inside an ideal fluid should always be preserved, and no new filaments may appear in it. However, in nature we often see the birth and death of vortices. The point is that water and air are somewhat viscous substances, and so our theoretical reasoning must be slightly modified to describe the behavior of real vortices. On the one hand, vortices can be generated predominantly in the regions where two layers of fluid move with different speeds and slide on each other; on the other hand, the generated vortices have only a limited lifetime and gradually decay.

Generation of straight vortices in liquids was demonstrated by Helmholtz in a beautiful experiment described in his lecture on vortical tempests. Here we repeat this demonstration.

A small hole at the bottom of a cylindrical vessel (figure 8) filled with water is plugged with a stopper. A jet of air is blown via a tube to one side of the free surface of the water, thereby spinning the water slowly. After opening the vessel, water begins to pour out of the hole, moving from the wall of the cylinder to its axis. Particles of water move along gradually decreasing circles centered at the axis of the cylinder. Since the circulation along these circles is constant, the decrease in radius is accompanied by an increase in speed of the particles. In approaching the axis of the cylinder, the speed of rotation becomes greater until a vortex with a characteristic deep crater can be observed clearly.

A vortex can also be produced by means of a rapidly rotating disk as shown in figure 9. A vertical axle with a small disk at its end is inserted through the bottom of a glass cylinder. The cylinder is filled with water, and oil is poured on top. When the disk is set in motion, it gradually spins the water and generates a vortical filament in it. The vortex is clearly seen by the crater at the oil-water boundary. The crater is filled with oil, which goes downward to the disk as a falling oil-spout. At the instant when the oil contacts the disk, the entire mass of the oil is dispersed in the water.

Straight vortices in air can be produced and visualized by a very interesting method. Air over water is spun by means of a rapidly rotating fan placed at some height over the surface of the water (figure 10). The

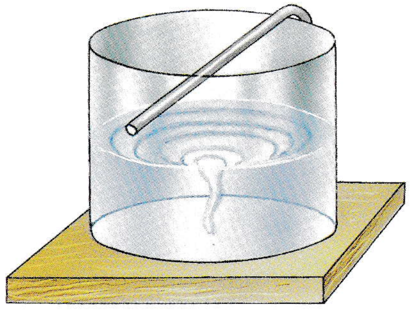


Figure 8

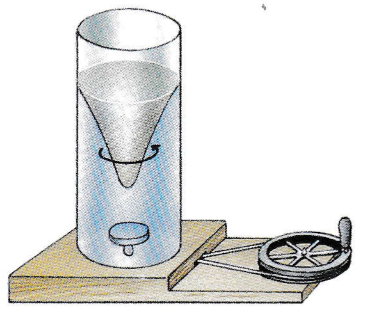


Figure 9

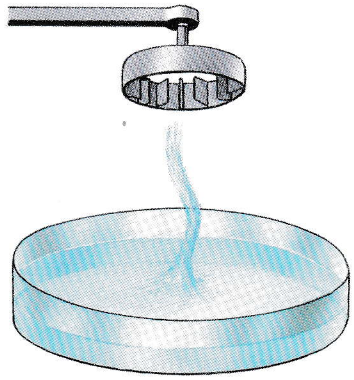


Figure 10

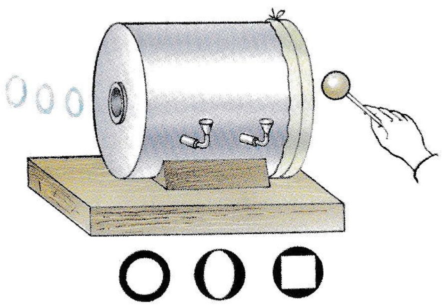


Figure 11

invisible air vortex catches water into its axial filament and lifts it as a rising waterspout up to the fan.

Vortical rings in air can be demonstrated with the help of a Tait device. This consists of a box (figure 11) whose back panel is closed by a piece of leather, while the front panel has an orifice with a sharp edge. The shape of the hole can be made round, elliptical, quadrangular, and so on, by using special fittings.

To visualize the vortical rings, two jars are placed in the box: one is filled with hydrochloric acid, and the other with liquid ammonia. The vapors of these chemicals produce a thick fog composed of ammonium chloride particles. Striking with a fist or a hammer on the stretched leather, we quickly push a mass of air with ammonium chloride fog out of the box. This mass passes through the motionless air and causes the air to execute vortical motion. The air from the Tait machine wraps itself into a vortical ring that is clearly seen by the fog that fills it. Evidently, the air near the ring will rotate in such a way that the vortical ring moves away from the orifice of the apparatus.

Previously, we saw (figures 3-6) how a number of straight vortices interact with each other. By observing the vortical rings puffed from the Tait apparatus, one can see cases of interaction between them. When rings contact each other with their sides, they are repelled. One ring can pass through another. This interesting case was investigated theoretically by Helmholtz in detail. He showed that the trailing ring should decrease in size and acquire increased speed. In contrast, the leading ring becomes larger in size and decelerates. These changes continue until the trailing ring passes through the leading ring. After this, the rings interchange roles: the trailing ring again catches up and passes through the leading ring. Unfortunately, this play of two rings can be observed only rarely, when their sizes and intensities are properly matched.

We can prove that a vortical ring carries rapidly rotating air by directing the ring to a burning candle. A candle placed at a large distance from the Tait apparatus is blown out whenever its flame is caught by a ring. In my youth I tried to find out why a percussion cap shot from a gun can blow out a burning candle at a great distance. Now I know the reason: a pistol fires not only a bullet, but also a vortical ring of air. Such a ring can travel a great distance without decay.

Previously, we puffed the rings from a round orifice. The rings may be formed by elliptical or square holes. However, such rings do not retain their shape and oscillate, trying to assume the "correct" round shape, which is a single stable form of a closed vortical filament.

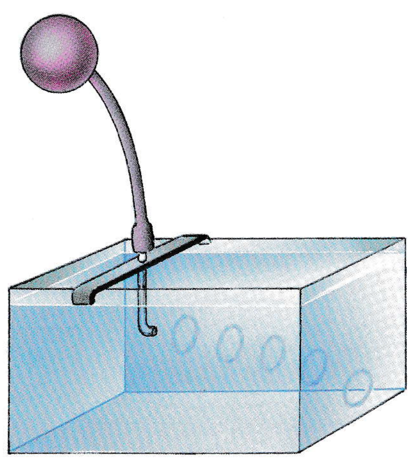


Figure 12

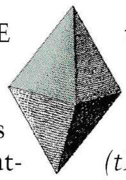
Now let us consider the influence of material bodies on the vortical rings. Solid bodies placed at the side of a moving ring repel it. However, if a ring runs into a plane parallel to its own plane, it will become greater and greater, as if spreading around the obstacle. If we put a knife in the way of the ring so that the plane of its blade contains the ring's axis, the ring will be cut into two half-rings, whose ends will slide on the blade. When these half-rings pass the knife, their ends are "glued" together to restore the ring.

In addition to smoke rings in air, one can observe air-traced rings in water. This interesting phenomenon, which at first glance seems paradoxical, has a simple explanation. The point is that the pressure at the axis of a vortical ring is greatly reduced as a result of the centrifugal force. If we introduce a few air bubbles into water at the instant of generation of a vortical ring, they quickly get to that part of the liquid where the pressure is minimal, that is, to the axis of the ring. The air bubbles will be retained there all the time, as long as the vortical ring travels in water, despite the fact that air is 800 times less dense than water.

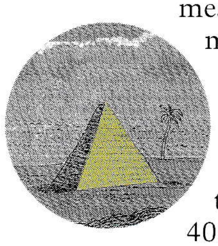
We describe here a device to generate air rings in water, which are visualized with the help of air bubbles. It consists of a large glass bath (figure 12) filled with water. A wide glass tube bent at a right angle is sunk into the bath. The upper end of the tube is held above the water surface. It is connected to a rubber ball, which can be pressed to push air into the tube, thereby expelling water from it. By rapidly squeezing the rubber ball, we expel water from the horizontal elbow and displace air to the very flange of the tube. In this way we not only expel water from the tube, but also release a small amount of air into the bath. After leaving the tube, the horizontal water column is curled into a vertical

CONTINUED ON PAGE 50

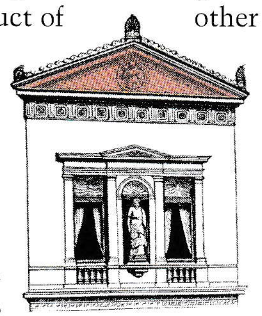
About the triangle



THE TRIANGLE IS THE simplest polygon—it has three vertices and three sides. The study of triangles gave rise to a branch of mathematics—trigonometry—in which the metric properties of the triangle are expressed in terms of functions of its angles. The development of this science was stimulated by practical demands. Trigonometry was used to measure parcels of land, to make maps, and to design machines.



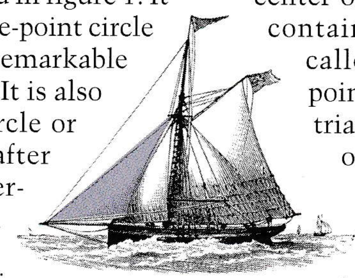
The earliest references to the triangle can be found in Egyptian papyri more than 4000 years old. For example, the Egyptians knew an approximate formula for calculating the area of an isosceles triangle: the area was found as the product of half the base and the lateral side. This formula gives a good approximation if the angle opposite the base is small.



In ancient Greece 2000 years later, considerable progress in the study of the triangle's properties was made. It suffices to recall the Pythagorean theorem and Hero's formula.

After a long period of cultural decline, the Renaissance began in the 15th century. Numerous studies of the triangle appeared, especially in the 18th century. These studies made up a large part of plane geometry, called the *new geometry of the triangle.*

Here is one remarkable theorem proved by Euler: *the midpoints of the triangle's sides, the feet of its altitudes, and the midpoints of the segments connecting its vertices with the orthocenter (the point of* intersection *of the triangle's altitudes) lie on a circle.* This circle is depicted in figure 1. It is often called the nine-point circle (in view of the nine remarkable points that lie on it). It is also known as Euler's circle or Feuerbach's circle, after the 19th-century German mathematician K. Feuerbach (the brother of the prominent philosopher L. Feuerbach), who proved that this circle is tangent to the inscribed circle of the triangle and all its escribed circles (that is, the circles tangent to a side of the triangle and the extensions of the other two sides; see figure 2).



Let us denote by H the orthocenter of a triangle ABC. It turns out that triangles ABC, ABH, BCH, and CAH have the same nine-point circle. Therefore, the nine-point circle is tangent to the 16 circles that are inscribed or escribed for triangles ABC, ABH, BCH, and CAH.

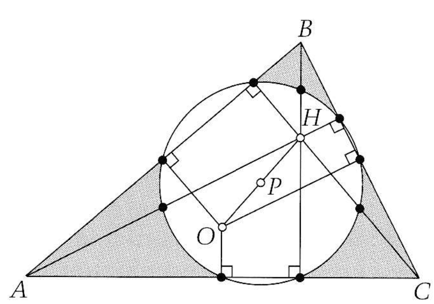


Figure 1

The radius of the nine-point circle is equal to half the radius of the circumscribed circle, and its center lies at the midpoint of the segment connecting the center of the circumscribed circle with the orthocenter of the triangle. The line containing this segment is called the *Euler line*. The point of intersection of the triangle's medians also lies on this line.

It is well known that Napoleon devoted part of his time to mathematics. The following elegant theorem is attributed to him: *the centers of the equilateral triangles constructed externally on the sides of any triangle form another equilateral triangle.* This equilateral triangle is called the *external Napoleon triangle.* An *internal Napoleon triangle* can be constructed similarly.

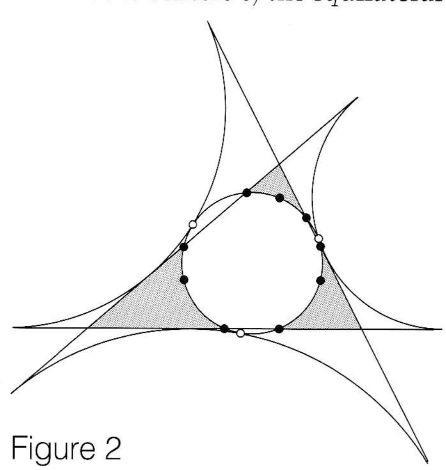


Figure 2

Even in the 20th century, there were some discoveries left to be made in the geometry of the triangle. In 1904, the American mathematician F. Morley proved that if

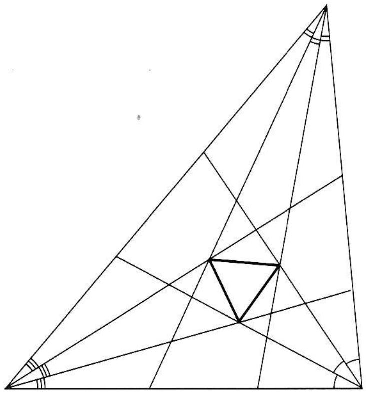
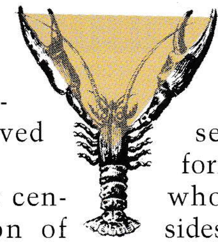
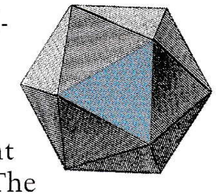


Figure 3

the angles of any triangle are trisected, then the points of intersection of adjacent trisectors form an equilateral triangle (figure 3). Ancient Greek mathematicians could have proved this theorem. Most probably, they didn't discover this remarkable fact because they considered only constructions that could be made using a compass and a straightedge. Angle trisectors, however, cannot be constructed using only these instruments (this wasn't proved until the 19th century).



In the middle of our century, a generalization of Morley's theorem was obtained. For every angle, we can consider three types of trisectors. The first kind was considered above. The second kind includes trisectors of angles adjacent to the given one. The third kind includes trisectors of the angle whose measures, together with that of the original angle, add up to 360 degrees. The points of intersection of these 18 trisectors form 18 (!) equilateral triangles, whose sides are parallel to the sides of the basic Morley triangle. A special report on this subject was presented at the Moscow Mathematical Congress in 1966.



AT THE BLACKBOARD I

An Olympian effort

by V. Tikhomirov

EACH GENERATION OF STUDENTS at the mathematical department of Moscow University has had its Olympic[1] leader. My generation also had such a leader.

The members of his family said that even before he could speak properly, he would tell people he wanted to be an *"alithmetician."* He started taking part in mathematical Olympiads when he was in the fifth grade, despite the fact that the contests were open only to students in the eighth grade and above. More incredible is that he always solved the problems and was among the winners.

[1]We mean Mathematical Olympiads, of course.

At that time, mathematical Olympiads were organized by the Moscow Mathematical Society, Moscow University, and the Moscow Department of Education. The winners were given books. The first-prize winners were given a large pile of books, the second-prize winners were given a smaller pile, the third-prize winners a still smaller pile, and the winners with honorable mention were awarded a single book, which was nevertheless nice. All the books were signed by renowned mathematicians, and the ceremonial presentation was held in the main auditorium of the mathematics department. According to legend, the hero of my story was once given a pile of books taller than himself. Be that as it may, he was always the winner. I believe that his achievements were never exceeded.

Later, my hero and I entered the mathematical department of Moscow University, then worked there and became friends. My friend became a prominent mathematician and could still solve any Olympiad-type problem.

Once, in the mid-1970s, a professor who had to make up problems for the entrance exams left for a long business trip abroad. He started the work but didn't finish it. The job was urgent (it was May already, and the exam was to take place in July), and I was asked to finish it. The list of people who were supposed to help me included the name of my friend, and I agreed to do the job.

Easy problems were already made up, but I lacked more difficult ones and, in particular, the so-called "nonstandard" problems. Each set of problems had to include a nonstandard problem.

You may ask what a nonstandard problem is. I also asked this question and was given the following explanation.

There exists an old and unsolved problem: *Is it possible to test talent?*

Athletes think that they have solved this problem. Once I observed how children were tested as to whether or not they could try out for gymnastics. A girl came up to the woman who was responsible for testing. The woman grasped the poor girl's buttocks and said: "The buns are too fat, doesn't suit." The procedure took only a second. I was delighted by its simplicity. However, we still don't know whether there exists a simple procedure for testing talent for mathematics. It is unclear whether it is possible, as A.N. Kolmogorov wrote, taking into account the particular atmosphere of the entrance exam used to assess a student's prospects in a certain branch of science.

Among other things, nonstandard problems were intended to test whether or not a student has a talent for mathematics. Such a problem has to include a difficult element that was impossible to overcome by standard methods. It was assumed that only talented students were able to solve this problem.

The nonstandard problem traditionally occupied the last, fifth place in the list of problems. Everybody knew that this problem was very difficult, and only ambitious students tried to solve it. As far as I know, the information on who solved the nonstandard problem was not made public, so I don't know who of the really talented mathematicians solved such a problem at the entrance exam.

Thus, among other things, I had to find a nonstandard problem, and I asked the hero of this story to do it. Naturally, he asked me what a nonstandard problem was. Instead of giving him lengthy explanations, I showed him a list of problems posed on the entrance exam of the previous year and asked him to make up a problem similar to the fifth problem in this list (however, I didn't solve it myself). We agreed that he would phone me two or three days later.

However, he didn't phone me even a week later. I was pressed to finish the work, and thus phoned him myself. With irritation I asked why he didn't contact me. Melancholically, he answered that he couldn't solve my problem. Here it is.

Find all pairs (x, y) of real numbers satisfying the conditions

$$4y^2 - 2x^2 = \sqrt{2(x+2y)^2 - (x+2y)^4},$$

$$x^4 + 2 \le 4y(x^2 - 1).$$

I think that my legendary friend was cunning. I don't believe that there is an Olympiad-type problem that he couldn't solve. Later, I solved this problem myself, which took several hours (but not days). I think my friend just didn't want to solve the problem, maybe because he considered it insufficiently elegant, or for some other reason. Be that as it may, he didn't pass the test.

What is your opinion? Do you think this is an adequate problem for testing talent? Try to solve it.

By the way, my friend and I failed to make up a nonstandard problem that year. The job was done by other people.

In case you don't want to ponder that problem, here is my solution (maybe not the best one).

Denote $(x + 2y)^2$ by z. The function $2z - z^2$ attains its maximum at $z = 1$. Thus, we find from the first equation that

$$4y^2 - 2x^2 = \sqrt{2z - z^2} \le 1. \quad (1)$$

The second inequality can be written as

$$x^4 - 4yx^2 + 4y + 2 \le 0,$$

or

$$x^4 - 4yx^2 + 4y^2 \le 4y^2 - 4y - 2,$$

or

$$(x^2 - 2y)^2 \le 4y^2 - 4y - 2.$$

This implies that

$$x^2 \le 2y + \sqrt{4y^2 - 4y - 2}. \quad (2)$$

Therefore,

$$4y^2 \le 2x^2 + 1 \le 4y + 2\sqrt{4y^2 - 4y - 2} + 1,$$

which is equivalent to the inequality

$$4y^2 - 4y - 2 - 2\sqrt{4y^2 - 4y - 2} \le 0.$$

This implies the inequality

$$\left(\sqrt{4y^2 - 4y - 2} - 1\right)^2 \le 0.$$

Thus, $4y^2 - 4y - 2 = 1$; therefore, $y = -1/2$ or $y = 3/2$. In the first case, the second condition of the problem (the inequality) entails $x^4 + 2x^2 <= 0$; that is, $x = 0$.

In the second case, we have $x^4 - 6x^2 + 8 \le 0$. This implies that $x^2 \le 4$. Substituting into the first condition of the problem (the equation), we find that

$$9 = 2x^2 + \sqrt{2(x+3)^2 - (x+3)^4};$$

but we have already observed that the function under the radical sign has its maximum when $x + 3 = 1$. Hence

$$9 \le 2x^2 + 1;$$

that is, $x^2 \ge 4$.

As a result, we arrive at the following conclusion: *there exist only two pairs of numbers satisfying the conditions of the problem:* (0, –1/2) *and* (–2, 3/2).

If you want to know my opinion about testing talent, I doubt that it is possible to make up a short test for finding out whether a student will be able to become a scientist or not. What's your opinion? ◘

Magnetic vee

by Larry D. Kirkpatrick and Arthur Eisenkraft

> *The secret of magnetism, now explain that to me! There is no greater secret, except love and hate.*
> *—Johann Wolfgang von Goethe*

ONE OF THE GOALS OF modern day theoretical physics is to reduce all forces in Nature to manifestations of a single force. In fact, the 1999 Nobel Prize in Physics was awarded to the Dutch physicists Gerardus 't Hooft and Martinus J. G. Veltman for their contributions in unifying the electromagnetic interaction and the weak interaction. Their theoretical methods have also been applied in attempts to combine this electroweak interaction with the strong interaction to form a grand unified theory. The biggest remaining problem is how to include the gravitational interaction in a "theory of everything."

During the early part of the 19th century, much effort went into establishing connections between electric forces and magnetic forces. Although there had been much discussion of possible connections, the first connection was established in 1819 during a classroom demonstration when the Danish scientist Hans Christian Oersted discovered that a current carrying wire deflected a compass needle. Furthermore, he discovered that the compass needle pointed at right angles to the current and that the compass needle pointed in the opposite direction when the current was reversed.

Within one week of the announcement of Oersted's discovery in 1820, the French physicist André Ampère formulated the right-hand rule: if you grasp a wire with your right hand so that your thumb points in the direction of the current, the compass points in the direction of the fingers. In modern language, the magnetic field lines are circles around the wire and your fingers point in the direction of the magnetic field.

A short time later, Ampère developed a formula for calculating the magnetic force between current-carrying wires. He also made the suggestion that all magnetic fields are due to currents, including those at the atomic level. Note that this occurred three-quarters of a century before the discovery of the electron and the publication of Niels Bohr's theory of the hydrogen atom a decade later.

During this time, Jean Baptiste Biot and Félix Savart obtained a quantitative expression giving the contribution $d\mathbf{B}$ to the magnetic field at a point P due to an element of current $Id\mathbf{s}$. The full expression, now known as the Biot-Savart law, is

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{Id\mathbf{s}\times\hat{\mathbf{r}}}{r^2},$$

where $\hat{\mathbf{r}}$ is a unit vector directed from the current element to the point P and μ_0 is the permeability of free space with the value $4\pi \cdot 10^{-7}$ T·m/A.

To find the total magnetic field created at point P by a current of finite length, we must sum up the contributions from all of the current elements:

$$\mathbf{B} = \frac{\mu_0}{4\pi}\int\frac{Id\mathbf{s}\times\hat{\mathbf{r}}}{r^2}.$$

We must be very careful when evaluating this integral as the integrand is a vector quantity and we must take its direction into account.

Notice the similarity of the Biot-Savart law with Coulomb's law. Here a current element produces a magnetic field that varies as the inverse-square of the distance from the current element. In Coulomb's law, a point charge produces an electric field that varies as the inverse-square of the distance from the point charge.

However, the directions of the two fields are very different. In Coulomb's law the electric field is radial; it points toward or away from the point charge. In the Biot-Savart law, the magnetic field is perpendicular to both the current element and the radius vector. The magnetic field points out of the plane determined by the current element and the point P.

Art by Tomas Bunk


THEORY OF EVERYTHING
ADD WATER
BIG
BANG
MAGNETISM
GRAVITY
TIME
ANTI
MATTER
LOVE AND
HATE
SINGLE
FORCE
STEW.
UNIVERSE

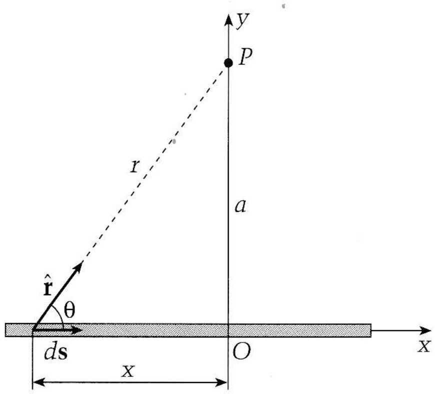


Figure 1

Let's apply the Biot-Savart law to find the magnetic field generated by a thin, straight wire carrying a constant current I. Let's set up the geometry as shown in figure 1. The wire is along the x-axis with the current in the positive x-direction. The point P is along the positive y-axis at a distance a from the origin O.

Using the right-hand rule, we see that each current element produces a contribution to the magnetic field that points out of the page. Therefore, the total magnetic field points directly out of the page and we only need calculate its magnitude. This means that we can replace $d\mathbf{s} \times \hat{\mathbf{r}}$ with $dx \sin\theta$, where θ is the angle between the direction of the current element and the direction to point P as shown in figure 1.

Before we go on, we note that we have more variables than we need. If we choose a given current element, the values of r, x, and θ are all specified. Therefore, we must express two of these three variables in terms of the third variable before we carry out the integration. Let's choose to express everything in terms of θ. We then have

$$r = \frac{a}{\sin\theta}$$

and

$$x = \frac{-a}{\tan\theta}.$$

Taking the derivative of the last expression, we obtain

$$dx = \frac{a\,d\theta}{\sin^2\theta}.$$

After making these substitutions, we are left with calculating the integral

$$B = \frac{\mu_0 I}{4\pi a}\int_{\theta_1}^{\theta_2} \sin\theta\, d\theta$$
$$= \frac{\mu_0 I}{4\pi a}\left(\cos\theta_1 - \cos\theta_2\right).$$

The angles are defined in figure 2.

If we look at the special case of an infinitely long, straight wire, $\theta_1 = 0$ and $\theta_2 = \pi$, and the magnetic field is given by

$$B = \frac{\mu_0 I}{2\pi a}.$$

Our contest problem is based on a problem that was given at the International Physics Olympiad that was held in Padua, Italy, last summer. A very long, thin, straight wire, carrying a constant current I, is bent to form a "V" of half-angle α as shown in figure 3.

A. What are the directions of the magnetic field at points P and P^*?

B. What is the magnitude of the magnetic field at point P?

C. What is the magnitude of the magnetic field at point P^*?

Please send your solutions to *Quantum*, 1840 Wilson Boulevard, Arlington, VA 22201-3000 within a month of receipt of this issue. The best solutions will be noted in this space.

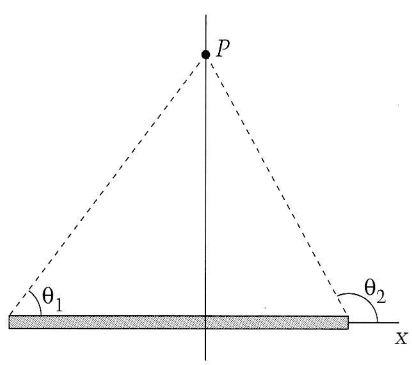


Figure 2

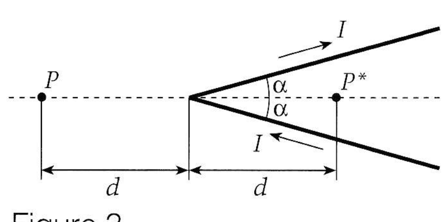


Figure 3

Breaking up is hard to do

In the September/October issue, we asked a series of questions related to the energy released during a fission reaction.

Part A asked readers to calculate the mass defect in the reaction

$${}^{235}_{92}\mathrm{U} + {}^{1}_{0}\mathrm{n} \rightarrow {}^{140}_{54}\mathrm{Xe} + {}^{94}_{38}\mathrm{Sr} + 2\ {}^{1}_{0}\mathrm{n},$$

where the Xe rapidly decays into ${}^{140}_{58}\mathrm{Ce}$ and the Sr into ${}^{94}_{40}\mathrm{Zr}$ with the emission of electrons of negligible mass. Given the following masses:

${}^{235}_{92}\mathrm{U}$	235.004 u
${}^{1}_{0}\mathrm{n}$	1.009 u
${}^{140}_{58}\mathrm{Ce}$	139.905 u
${}^{94}_{40}\mathrm{Zr}$	93.906 u,

we find that the initial and final masses are 236.013 u and 235.829 u, respectively, for a change in mass of 0.184 u and a corresponding energy release of 171 MeV.

Part B asked readers to follow the argument used by Frisch and Meitner and Joliot (independently) where the change in electrostatic potential energy provides the energy of the fission products. The radii of the Ce and Zr nuclei can be calculated using the approximate equation,

$$R = KA^{1/3},$$

where $K = 1.0 \cdot 10^{-15}$ m.

$$R_{\mathrm{Ce}} = 5.19 \cdot 10^{-15}\ \mathrm{m}$$

$$R_{\mathrm{Zr}} = 4.55 \cdot 10^{-15}\ \mathrm{m}$$

The distance between the centers of the two fragments is equal to the sum of their radii, $R = 9.74 \cdot 10^{-15}$ m.

We can now find the electrostatic potential energy:

$$U = \frac{-kq_1q_2}{R}$$
$$= \frac{-\left(9\cdot 10^9\ \frac{\mathrm{N}\cdot\mathrm{m}^2}{\mathrm{C}^2}\right)(58)(40)\left(1.6\cdot 10^{-19}\ \mathrm{C}\right)^2}{9.74\cdot 10^{-15}\ \mathrm{m}}$$
$$= 5.49 \cdot 10^{-11}\ \mathrm{J}$$
$$= 343\ \mathrm{MeV}.$$

The values for the energy calculated in parts A and B are relatively close.

Part C asked readers to show that the energy released is greatest for the rare case of a symmetric fission. Weizsacker's semi-empirical formula for the binding energy is

$$B=\left[15.753A-17.804A^{2/3}-0.7103\frac{Z^2}{A^{1/3}}-94.77\frac{\left(\frac{1}{2}A-Z\right)^2}{A}\right]\text{MeV.}$$

We built a spreadsheet to calculate the binding energies of all possible pairs of daughter nuclei. In doing this, we needed to make a decision on how to divide the neutrons between the daughter nuclei.

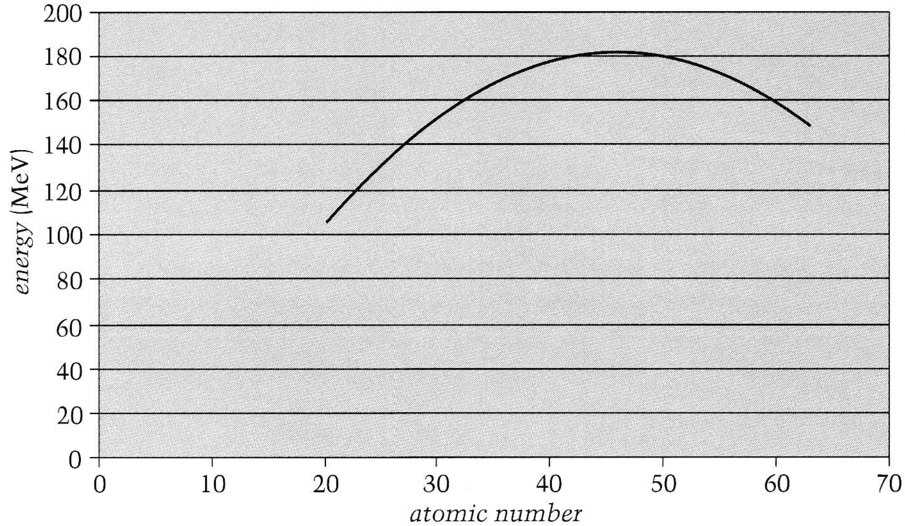


Figure 4

We chose to divide them so that the ratio of nucleons was the same as the ratio of protons. For example, if the daughter nuclei had 30 and 62 protons, the number of nucleons A was taken to be (30/92) of 236 and (62/92) of 236, respectively.

After completing the spreadsheet, we graphed the energy released versus the atomic number of one of the daughter nuclei (figure 4). The energy released is greatest for the symmetric fission with each daughter nucleus having 46 protons. ◘

Experiments of Frank and Hertz

by A. Levashov

THE BEGINNING OF THE 20th century was an exceptional period in the history of physics. This was a time of revision of the old and venerable basic concepts of the structure of the material world. In a number of cases this was not a revision, but an outright revolution, which made obsolete the good old laws of classical mechanics and electrodynamics.

In 1900 Max Planck created the quantum theory of electromagnetic radiation. In 1905 Albert Einstein formulated the principles of the special theory of relativity. During the same period, science took a close look at the problem of the physical nature and internal structure of the atom. In 1911 Ernest Rutherford published the results of his experiments, the cornerstones of the nuclear model of the atom, which contradicted classical physics. In 1913 Niels Bohr formulated quantum postulates that explained Rutherford's atomic model and the character of atomic spectra, but was at odds with both classical mechanics and electrodynamics.

Each of these discoveries opened up new vistas in our knowledge of the material world. The development of new theoretical ideas is always supported by a sound experimental foundation. Therefore, experimental tests of new hypotheses play an exceptionally important role in science.

The work of the German physicists James Frank and Gustav Hertz was practically the first experimental confirmation of Bohr's postulates. Historically, their experiments continued the work of the German physicist Philip Lenard, who tried to determine experimentally the ionization energy of atoms long before the advent of Rutherford's model. In 1902 Lenard carried out an interesting experiment. The main part of his setup was a glass tube with three soldered electrodes. In fact, it was a prototype of a triode, which was the cornerstone of electronics in the first half of the 20th century. The tube was filled with a gas whose ionization energy was to be measured. The gas pressure in the tube was about 0.01 mm Hg. The setup is shown schematically in figure 1. The voltage V_1 applied between the cathode C and the grid G by battery 1 can be changed with the variable resistor R_1. The voltage V_2 applied between the anode A and the grid G by battery 2 can be changed with the variable resistor R_2.

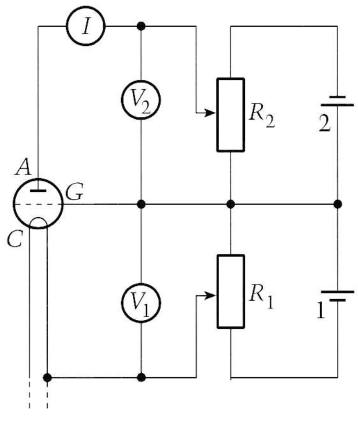
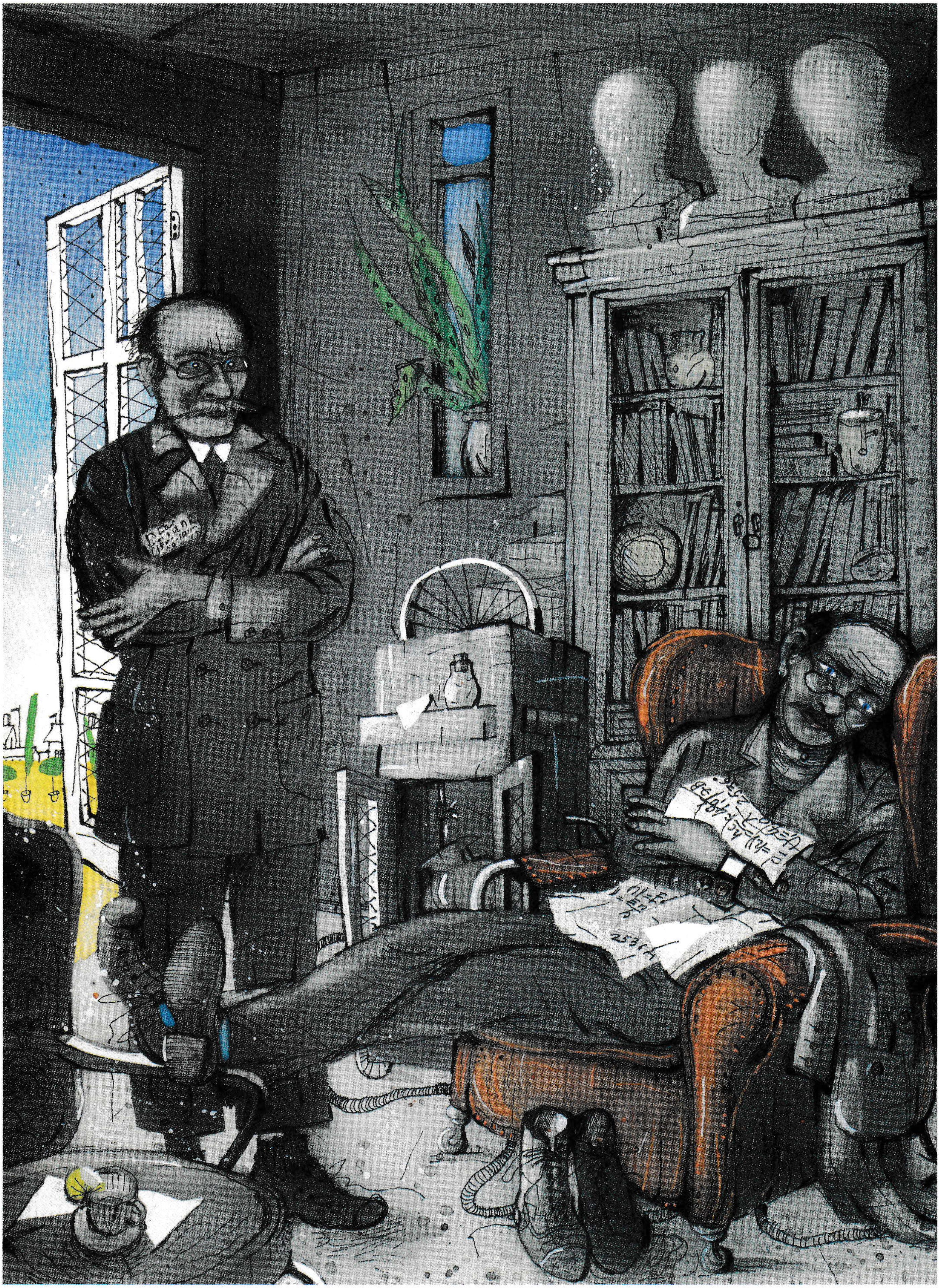


Figure 1

The idea of the experiment was very simple. Electrons emitted from the cathode are accelerated by an electric field in the C-G space and rush to the grid. Therefore, the voltage V_1 is called the "accelerating voltage." Passing through the grid, the electrons enter the G-A space. Here the electric field has the opposite direction, and it decelerates the electrons. Accordingly, the voltage V_2 is called the "decelerating voltage." Clearly, depending on the relation between V_1 and V_2, the electrons either arrive at the anode ($V_1 > V_2$) and produce an electric current I recorded by an ammeter, or they are decelerated in the G-A space ($V_2 > V_1$) and are captured by the grid to be returned to battery 1 in the grid circuit. In Lenard's experiments the decelerating voltage was always higher than the accelerating one, so the emitted electrons could not reach the anode.

In starting his experiments to determine the atomic ionization energy, Lenard proceeded from the following assumptions about the interactions of electrons with atoms. In the C-G space the emitted electrons collide with the atoms of the gas. These collisions are of two types—elastic and inelastic. The elastic collisions do not change the kinetic energy of the electrons appreciably. Indeed, because the mass of an atom is thousands of times larger than the mass of an electron, an elastic collision changes only the direction of the electron velocity, but not its magnitude. Thus, the ac-

Art by Ekaterina Silina



celerated electrons follow zigzag paths to the grid. Near the grid, the electron energy reaches its maximum value determined by the conservation of energy

$$\frac{mv^2}{2} = eV_1,$$

where m is the electron mass and e is its charge.

As the accelerating voltage V_1 is increased, the energy of the electrons increases. Lenard assumed that when the maximal energy of electrons near the grid was equal to or somewhat greater than the ionization energy of the gas, the collisions of electrons with the atoms would become inelastic. In this case, an electron transfers essentially all its kinetic energy to the atom, and this energy initiates the internal atomic processes that lead to ionization of the atom. As a result, secondary charged particles appear: electrons and positively charged ions (that is, atoms deprived of at least one electron). All the electrons in the gas (both secondary and primary—those emitted by the cathode) finally travel to the grid circuit. In contrast, the positive ions that penetrated into the *G-A* space move to the anode under the influence of the accelerating electric field. Therefore, the current that arises in the anode circuit is caused by the positive ions.

Thus, according to Lenard, the appearance of an electric current in the anode circuit should indicate the beginning of ionization. The ionization energy is eV_i, where V_i is the value of V_1 at which the anode current arises. Although certain values of the accelerating voltage V_1 did produce an anode current, Lenard could not obtain reliable data, owing to technical imperfections in his experimental setup.

It is important to note that Lenard proceeded from incorrect assumptions about the interaction of electrons with atoms because he could not know about the existence of discrete energy levels in atoms. For this reason, Lenard's experiment, although very interesting, was not a landmark in the history of physics. A refined version of Lenard's method was subsequently used by Frank and Hertz in their classic experiments.

James Frank and Gustav Hertz began their joint work at Berlin University in 1911. At that time they studied the passage of an electric current in gases. Naturally, they were interested in the value of the ionization energy of the atoms. To determine it, they first used Lenard's method, but later improved it considerably.

First of all, Frank and Hertz modified the tube. The distance between the cathode and the grid was made far greater than the distance between the grid and the anode. The gas (or vapor) pressure in the tube was increased to 1 mm Hg. Thus, the electrons emitted by the cathode underwent multiple collisions with gaseous atoms. The decelerating voltage was decreased and made constant (in the range of 0.5–2.0 V). Therefore, a current appeared in the anode circuit when the accelerating voltage V_1 was larger than the decelerating voltage V_2.

Frank and Hertz, like Lenard, wanted to measure the ionization energy as a first step. The result of one of their experiments with a tube filled with mercury vapor is shown in figure 2. The curve in this figure shows how the anode current varies with increasing accelerating voltage (the current is measured in arbitrary units). Our attention is drawn to the curious shape of the plot: the current rises and falls "periodically" at regular intervals of the voltage V_1. The current peaks correspond to voltages 4.9 V, $2 \cdot 4.9$ V, and $3 \cdot 4.9$ V.

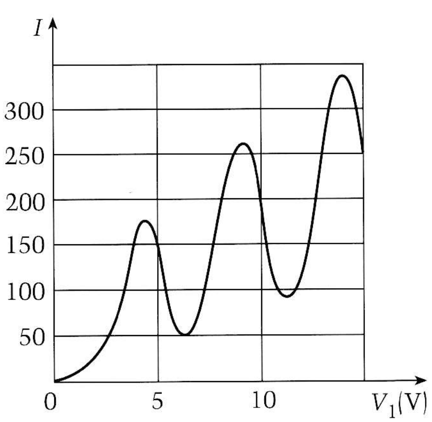


Figure 2

Frank and Hertz interpreted these data in the following way. When the accelerating voltage is increased, the anode current grows (specifically, from the instant when V_1 becomes larger than V_2). When V_1 is equal to or slightly larger than 4.9 V, an accelerated electron acquires an energy of 4.9 eV along its way to the grid, so that collisions in the vicinity of the grid become inelastic. As a result, an atom is ionized, and the decelerated electron is trapped by the grid. Therefore, the anode current drops drastically.

With a further increase of the accelerating voltage, an electron acquires the critical energy 4.9 eV before arriving at the grid, so that the region of inelastic collisions is shifted toward the cathode. After a collision, an electron loses energy, but it continues its accelerated motion (from zero velocity) to the grid. If it is accelerated strongly enough to acquire an energy larger than eV_2, it will overcome the decelerating effect of the voltage V_2 and arrive at the anode. In the plot, this process is shown by the rising segment of the anode current after the first decline.

When V_1 becomes so large that an electron that has lost 4.9 eV of energy in the first inelastic collision manages to acquire the same critical energy of 4.9 eV along its way to the grid, its collision with a second atom is inelastic, and it no longer reaches the anode, but is captured by the grid. The second inelastic collision occurs near the grid, which traps the decelerated electron. As a result, the anode current drops for the second time. In the plot, this process is shown by the decline after the second peak.

The third maximum indicates that the electrons undergo three inelastic collisions on their way to the grid.

On the basis of these results, Frank and Hertz concluded that the ionization energy of mercury is 4.9 eV, which is equal to the energy gained by the electrons between two successive inelastic collisions with atoms.

Although Frank and Hertz were confident of the value 4.9 eV that they obtained for the ionization energy of mercury, they decided to test their results in another way. They knew that mercury vapor strongly absorbs ultraviolet radiation at a wavelength $\lambda = 2536$ Å. According to Planck's theory, this radiation corresponds to the energy

$$E = h\nu = \frac{hc}{\lambda} \approx 4{,}84 \text{ eV}.$$

This value is remarkably close to the value obtained for the ionization energy. Is this coincidence accidental? Frank and Hertz decided to check and see whether radiation with the wavelength $\lambda = 2536$ Å is generated in gaseous mercury at the voltage $V_1 = 4.9$ V.

The results were spectacular. While the voltage V between the grid and the cathode was less than 4.9 V, the grid current grew with V but no radiation appeared. However, when this voltage reached a value of about 4.9 V, the current dropped sharply and, at the same time, a spectrograph detected radiation at a wavelength of precisely 2536 Å!

On the basis of these data, Frank and Hertz concluded that in most cases of inelastic collisions the electron energy is expended on atomic ionization, but sometimes it induces radiation. These data were explained correctly four years later, when the theory of Niels Bohr began to win recognition. In essence, the experiments of Frank and Hertz were the first direct experimental confirmation of this theory, although Niels Bohr did not realize it at the time.

According to Bohr's first postulate, an atom can exist only in a number of states in which its energy has certain discrete values. In Frank and Hertz's experiments this postulate was supported by the fact that the sharp decrease in the anode current occurred at values of V_1 which were multiples of the same value 4.9 V. This means that in inelastic collisions a mercury atom absorbs energy in fixed amounts: it cannot "swallow" an energy less than 4.9 eV. In other words, an atom changes its energy in jumps or, simply, by quanta. If the energy of a mercury atom is E_0 in the ground state with the lowest energy, the energy of its first excited state will be $E_1 = E_0 + 4.9$ eV.

Bohr's second postulate was also supported in the experiments of Frank and Hertz. It says that when an atom makes a transition from a state with larger energy E_{n+1} to a state with smaller energy E_n, a photon is radiated, and its energy $h\nu$ (where ν is the radiation frequency) is determined by the equation $h\nu = E_{n+1} - E_n$, whence $\nu = (E_{n+1} - E_n)/h$. Now it is evident why Bohr's second postulate is also called "the rule of frequencies."

The experiments of Frank and Hertz with a quartz tube filled with mercury vapor confirmed this postulate: having absorbed an energy of 4.9 eV in the course of an inelastic collision, a mercury atom goes from the ground state E_0 to the first (excited) state with energy $E_0 + 4.9$ eV. The inverse transition to the ground state is accompanied by radiation of a photon with energy 4.9 eV, which is observed as ultraviolet radiation at the wavelength 2536 Å.

Thus, Frank and Hertz observed not ionization, but excitation of mercury atoms. What they thought was the ionization energy was in reality the energy of the first excited state.

It was impossible to measure the energy of the higher excited states with Frank and Hertz's setup. After a collision, an electron had no chance to acquire an energy greater than 4.9 eV, because the number of "obstacles" (atoms) was too large. Therefore, the acceleration of the electron was interrupted by too many colliding atoms absorbing the electron's energy in numerous inelastic collisions. This obstacle must be removed to measure the energy of the higher excited states of an atom.

Later, Hertz modified the experiment. He separated the region in which the electrons collided with atoms. This allowed an electron to acquire energy larger than 4.9 eV because it did not meet obstacles in the accelerating region. The experiments showed that when the electron's energy reached 9.8 eV, its collision with a mercury atom became inelastic. Having absorbed an energy of 9.8 eV, mercury atoms undergo a transition to the second excited state. Thus, the modified method revealed the higher energy states of atoms.

In 1925, James Frank and Gustav Hertz became Nobel Prize laureates for the discovery of the laws describing collisions between electrons and atoms.

Quantum on atoms and collisions:

M. Digilov, "A Strange Box and a Stubborn Brit," March/April 1991, pp. 26–27.

S. R. Filonovich, "The Power of Likeness," September/October 1991, pp. 23–27.

A. Korzhuyev, "Bohr's Quantum Leap," January/February 1999, pp. 42–46.

CONTINUED FROM PAGE 25

symmetric spherical cavity. The thickness of glass along the line passing through the centers of both spheres is $R/2$ (figure 2). A point

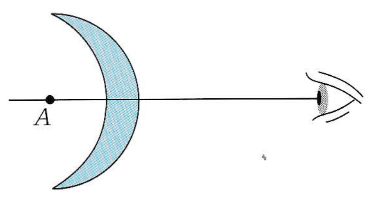


Figure 2

source of light is located at the center of the outer spherical surface (point A). Where will this source be seen by an observer whose eye is far from the lens along the line connecting the centers? (A. Zilberman).

SOLUTIONS ON PAGE 50

AT THE BLACKBOARD II

Inequalities become equalities

by A. Egorov

THE COMMUNICATION BY TIKHOMIROV (pp. 32–33) describes a rather funny situation: a first-rate mathematician could not quickly solve a problem for an entrance exam. This was in 1973. Another remarkable mathematician told me that he learned about this very problem the day he was leaving Moscow for Vladivostok. He attempted to solve it during his 8-hour flight, but couldn't find the solution.

It must be said that mathematicians are rarely interested in examination problems. So it is all the more significant that this young mathematician got interested in the problem. There must be something to it. Unfortunately, entrance-exam problems are rarely interesting, especially in the last 15 years.

In this article we discuss problems that are similar to the very difficult problem mentioned above. They are linked by several themes, which we will point out as we go along.

Problem 1. *Solve the equation*

$$\sqrt{x-1}+\sqrt{3-x}=x^2-4x+6.$$

Solution. An attempt to find a straightforward solution, i.e., eliminating the radicals by squaring both sides of the equation, leads to an eighth-degree equation, which is very difficult to solve. Let us try to use the following simple fact.

The left-hand side of the equation,

$$y=\sqrt{x-1}+\sqrt{3-x}$$

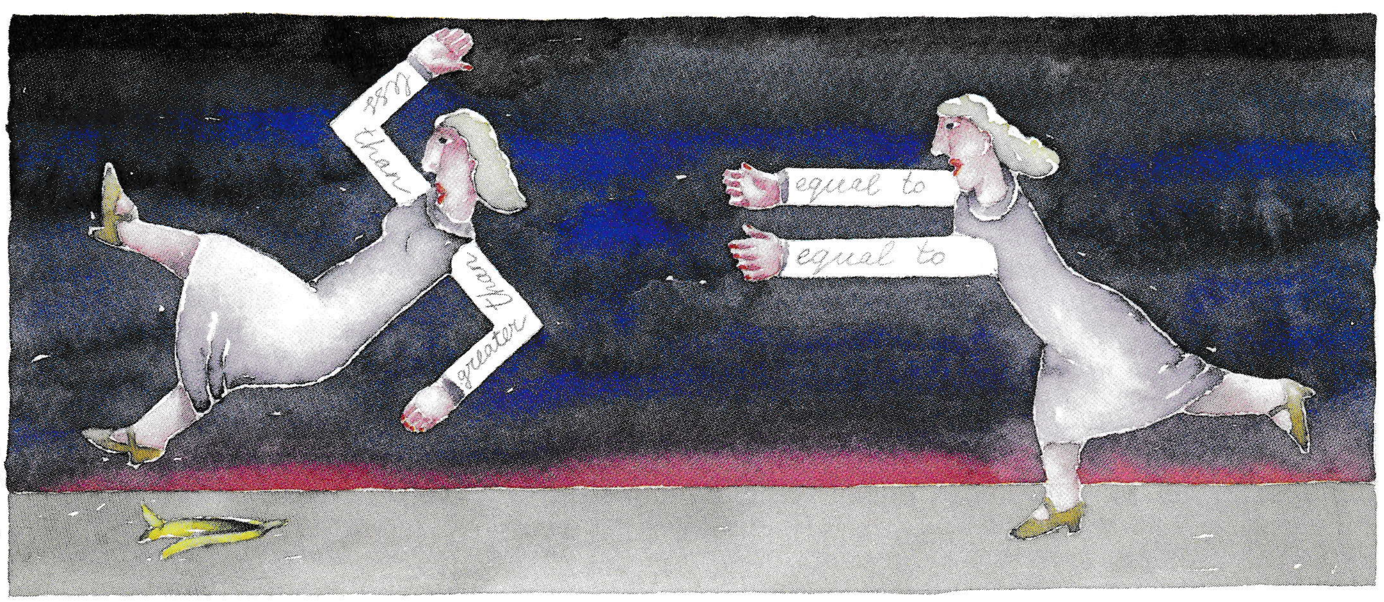
is defined for $1 \le x \le 3$, and its graph is symmetric about the vertical line $x = 2$. It seems likely that the point $x = 2$ has some special property. What is it? In fact, the left-hand side attains its maximum at this point, as we will now prove. Consider the square of the left-hand side:

$$y^2=\left(\sqrt{x-1}+\sqrt{3-x}\right)^2=2+2\sqrt{(x-1)(3-x)}.$$

The maximum is attained at the point where the radicand attains its maximum, i.e., at $x = 2$. Indeed,

$$(x-1)(3-x)=-3+4x-x^2=1-(x-2)^2.$$

Thus, the left-hand side is not greater than 2, and is equal to 2 only for $x = 2$. However, the right-hand side is not less than 2, since



Art by Sergey Ivanov

$$x^2 - 4x + 6 = (x-2)^2 + 2,$$

and it is equal to 2 only for $x = 2$. Thus, the problem is solved.

Answer. $x = 2$.

In this problem, we dealt with two functions, one of which attains its maximum at $x = 2$, while the other attains its minimum at this point, and the maximum coincides with the minimum. In this article, we discuss problems of this type. Consider the following classic example.

Problem 2. *Solve the equation*

$$\sin^5 x + \cos^3 x = 1.$$

Solution. All attempts to solve this problem by conventional methods are bound to fail. Let us try to guess the solution. The solutions for which either $\sin x = 1$ or $\cos x = 1$ are evident: these are

$$x = \frac{\pi}{2} + 2\pi k \text{ and } x = 2\pi k,\ k \in \mathbf{Z}.$$

We prove that the equation has no other solutions. For this purpose, it is sufficient to prove that, for other x, the left-hand side of the equation is less than 1.

Suppose that $\sin x \ne 0, 1$ and $\cos x \ne 0, 1$. Then, $\sin^5 x < \sin^2 x$ and $\cos^3 x < \cos^2 x$. Thus,

$$\sin^5 x + \cos^3 x < \sin^2 x + \cos^2 x = 1.$$

For $\sin x = 0$ and $\cos x = 0$, we obtain the solutions that we have already guessed.

Answer. $\frac{\pi}{2} + 2k\pi,\ 2k\pi,\ k \in \mathbf{Z}$.

The next problem is similar to problem 2.

Problem 3. *Solve the system of equations (n is a positive integer)*

$$\begin{cases} x + y = 1, \\ x^{2n} + y^{2n} = 1. \end{cases}$$

Solution. The second equation implies that $|x| \le 1$ and $|y| \le 1$. Then the first equation implies that $x \ge 0$ and $y \ge 0$. If $0 < x < 1$ and $0 < y < 1$, then $x^{2n} + y^{2n} < x + y = 1$. Thus, we immediately obtain the solution.

Answer. $(1, 0)$, $(0, 1)$.

Here is another example.

Problem 4. *Solve the inequality*

$$-x - y^2 - \sqrt{x - y^2 - 1} \ge -1.$$

Solution. Since the radicand cannot be negative, we see that $x \ge y^2 + 1$. That is, $x > 1$ for $y \ne 0$. Then $-x < -1$, which implies that the left-hand side of the inequality is less than its right-hand side. Thus, $y = 0$, $x = 1$.

Answer. $(1, 0)$.

Now let us consider a system of equations.

Problem 5. *Solve the system of equations*

$$\begin{cases} x + y + z = 2, \\ 2xy - z^2 = 4. \end{cases}$$

Solution. Here we have two equations and three unknowns. Let us express y in terms of z and x, and substitute it into the second equation:

$$y = 2 - (x + z),$$

$$2(2 - (x + z))x - z^2 = 4,$$

$$4x - 2x^2 - 2xz - z^2 = 4.$$

Therefore, $(x - 2)^2 + (x + z)^2 = 0$.

This implies that $x = 2$, $z = -x = -2$, and $y = 2$.

Answer. $(2, 2, -2)$.

Here is another problem where the number of unknowns exceeds the number of equations.

Problem 6. Solve the equation

$$2(x^4 - 2x^2 + 3)(y^4 - 3y^2 + 4) = 7.$$

Solution. Each of the factors on the left-hand side is a quadratic trinomial: the first one is quadratic with respect to x^2 and the second one with respect to y^2. Therefore, the factors are minimal for $x^2 = 1$ and $y^2 = 3/2$, respectively. Substitution shows that both these minimal values are positive. This means that both factors are positive for all x and y, so the product attains its minimum when each of the factors is minimal. A quick computation shows that the minimum of the product is 7.

Answer. The problem has four solutions:

$$\left(\pm 1; \pm\sqrt{3/2}\right).$$

Problem 7. *Solve the system of inequalities*

$$\begin{cases} x^2 - 6x + 6y \le 0, \\ y^2 - 2xy + 9 \le 0. \end{cases}$$

Solution. We add the inequalities to obtain

$$(x-y)^2 - 6(x-y) + 9 \le 0,$$

or

$$(x - y - 3)^2 \le 0.$$

The last inequality implies that $x - y = 3$. Also, all the inequalities must become equalities (if any of the inequalities of the system is strict, their sum is also a strict inequality, which is impossible, since a square cannot be negative). Thus, $y = x - 3$ and $x^2 - 6x + 6y = 0$. Therefore,

$$x = \pm 3\sqrt{2} \text{ and } y = -3 \pm 3\sqrt{2}.$$

Answer. $\left(3\sqrt{2}, 3\sqrt{2} - 3\right)$, $\left(-3\sqrt{2}, -3\sqrt{2} - 3\right)$.

Problem 8. *Solve the system*

$$\begin{cases} \sqrt{\frac{1}{2}(x-y)^2 - (x-y)^4} = y^2 - 2x^2, \\ y \ge 4x^4 + 4yx^2 + \frac{1}{2}. \end{cases}$$

Solution. We define $t = (x - y)^2$. Then the left-hand side of the first equation can be written as

$$f(t) = \sqrt{\frac{1}{2}t - t^2}.$$

The radicand is a quadratic trinomial that attains its maximum at $t = 1/4$. The maximum itself is

$$\max f(t) = \sqrt{\frac{1}{2} \cdot \frac{1}{4} - \frac{1}{16}} = \frac{1}{4}.$$

Now we replace the first equation by the inequality (this is the decisive step in the solution)

$$y^2 - 2x^2 \le \frac{1}{4}.$$

We add this inequality to the second inequality in the system to obtain

$$y + \frac{1}{4} \ge 4x^4 + 4yx^2 + \frac{1}{2} + y^2 - 2x^2.$$

This inequality can be written as

$$0 \ge \left(2x^2 + y\right)^2 - \left(2x^2 + y\right) + \frac{1}{4}$$

or

$$\left(2x^2 + y - \frac{1}{2}\right)^2 \le 0.$$

This implies, first, that $2x^2 + y - 1/2 = 0$ and, second, that all the above inequalities are in fact equalities. Taking into account the fact that $t = (x - y)^2 = 1/4$, we arrive at the following system of equations:

$$\begin{cases} y^2 - 2x^2 = \frac{1}{4}, \\ 2x^2 + y = \frac{1}{2}, \\ (x - y)^2 = \frac{1}{4}. \end{cases}$$

This system can be solved, for example, by solving the second equation for y, substituting in the third equation, then checking that the first equation is satisfied.

Answer. $(0, 1/2)$, $(-1, -3/2)$.

Here is a similar problem with a parameter.

Problem 9. *Find the values of a for which the following system of equations has a unique solution:*

$$\begin{cases} x \ge (y - a)^2, \\ y \ge (x - a)^2. \end{cases}$$

Solution. It is easily seen that if (x_0, y_0) is a solution of the given system, then (y_0, x_0) is also a solution. Therefore, if the solution is unique, then $x_0 = y_0$. Therefore, the inequality

$$x \ge (x - a)^2,$$

and the equivalent inequality,

$$x^2 - (2a + 1)x + a^2 \le 0,$$

must also have a unique solution. Thus, the discriminant of the quadratic trinomial on the left-hand side must be zero:

$$(2a + 1)^2 - 4a^2 = 0.$$

Thus, $a = -1/4$.

It remains to verify that for $a = -1/4$ the given system has a unique solution. We substitute $a = -1/4$ in the system and add the inequalities to obtain the inequality

$$x + y \ge y^2 + \frac{y}{2} + \frac{1}{16} + x^2 + \frac{x}{2} + \frac{1}{16},$$

which reduces to

$$\left(y - \frac{1}{4}\right)^2 + \left(x - \frac{1}{4}\right)^2 \le 0.$$

Thus, $x = y = 1/4$ is the unique solution of the system.

Answer. $a = -1/4$.

The following two problems require us to estimate maxima and minima of certain trigonometric expressions. We will make repeated use of the following algebraic inequality:

If $A > 0$, then

$$A + 1/A \ge 2, \tag{1}$$

with equality only when $A = 1$.

Proof. We have

$$\left(\sqrt{A} - \sqrt{\frac{1}{A}}\right)^2 = A + \frac{1}{A} - 2\sqrt{A\frac{1}{A}}$$
$$= A + \frac{1}{A} - 2 \ge 0.$$

This implies the result[1].

We have

$$A + \frac{1}{A} \ge 2\sqrt{A\frac{1}{A}} = 2,$$

with equality only when $A = 1/A$; that is, when $A = 1$.

Problem 10. *Solve the system of equations*

$$\begin{cases} \tan^2 x + \cot^2 x = 2\sin^2 y, \\ \sin^2 y + \cos^2 z = 1. \end{cases}$$

Solution. Let $\tan^2 x = t$. Then $t > 0$. The left-hand side of the first equation is $t^2 + 1/t^2$ and so (result (1)) it is not less than 2. The right-hand side is not greater than 2.

[1]This inequality is in fact a special case of the arithmetic/geometric mean inequality for two variables.

Therefore,

$$\tan^2 x = 1,\ \sin^2 y = 1,\ \cos^2 z = 0.$$

Answer.

$$\left(\pm\frac{\pi}{4}+\pi k, \frac{\pi}{2}(2l+1), \frac{\pi}{2}(2m+1)\right),\ k,l,m \in \boldsymbol{Z}.$$

Problem 11. *Solve the equation*

$$\tan^4 x+\tan^4 y+2\cot^2 x\cot^2 y = 3+\sin^2(x+y).$$

Solution. Let $a = \tan^2 x$ and $b = \tan^2 y$. Then the left-hand side of the equation can be written as

$$a^2+b^2+\frac{2}{ab}=(a-b)^2+2ab+\frac{2}{ab}\geq(a-b)^2+4.$$

(We have used the inequality (1) with $A = ab$.) Therefore, the left-hand side is not less than 4, and is equal to 4 if $a = b$ and $ab = 1$. However, the right-hand side is not greater than 4, and is equal to 4 only if $\sin^2(x + y) = 1$. Thus, it remains to solve the system

$$\sin^2(x+y) = 1,\ \tan^2 x = \tan^2 y,\ \tan^2 x\tan^2 y = 1.$$

Answer.

$$\left(\frac{\pi}{4}+\frac{\pi n}{2}; \frac{\pi}{4}+\frac{\pi m}{2}\right),$$

where $m + n$ is even and $m,n \in \boldsymbol{Z}$.

Some problems require a transformation of the expression $f(x) = a\sin x + b\cos x$ using an auxiliary angle. Let us recall how this is done. We write $f(x)$ as

$$f(x)=\sqrt{a^2+b^2}\times\left(\frac{a}{\sqrt{a^2+b^2}}\sin x+\frac{b}{\sqrt{a^2+b^2}}\cos x\right).$$

Consider the point with coordinates

$$\left(\frac{a}{\sqrt{a^2+b^2}}; \frac{b}{\sqrt{a^2+b^2}}\right).$$

This point lies on the unit circle; therefore, there exists a unique angle $0 \leq \phi \leq 2\pi$ such that

$$\cos\phi = a/\sqrt{a^2+b^2} \text{ and } \sin\phi = b/\sqrt{a^2+b^2}.$$

Therefore,

$$\begin{aligned} f(x) &= \sqrt{a^2+b^2}(\cos\varphi\sin x+\sin\varphi\cos x) \\ &= \sqrt{a^2+b^2}\sin(x+\varphi). \end{aligned} \tag{2}$$

This immediately implies the estimate

$$|a\sin x+b\cos x| \leq \sqrt{a^2+b^2}, \tag{3}$$

and the equality holds for x such that $\sin(x + \phi) = 1$ or $\sin(x + \phi) = -1$, i.e., for $x = -\phi \pm \pi/2 + 2k\pi$.

Here is a problem that uses this method.

Problem 12. *Solve the equation*

$$\sin 3x-2\sin 18x\sin x = 3\sqrt{2}-\cos 3x+2\cos x.$$

Solution. We rewrite the given equation as

$$\sin 3x+\cos 3x-2(\sin 18x\sin x+\cos x) = 3\sqrt{2}.$$

By virtue of (3), we have

$$|\sin 3x+\cos 3x| \leq \sqrt{2}$$

(here $a = 1$, $b = 1$, and $\phi = \pi/4$).
Similarly,

$$2|\sin 18x\sin x+\cos x| \leq 2\sqrt{1+\sin^2 18x} \leq 2\sqrt{2}$$

(here $a = \sin 18x$ and $b = 1$, and we have used the fact that $\sin 18x \leq 1$).

These inequalities imply that the modulus of the left-hand side is not greater than $3\sqrt{2}$, and it can be equal to $3\sqrt{2}$ only for $\sin^2 18x = 1$, i.e., if $\sin 18x = 1$ or $\sin 18x = -1$.

For the left-hand side to be equal to the right-hand side, it is necessary and sufficient that

$$\begin{cases} \sin 3x+\cos 3x = \sqrt{2}, \\ \sin 18x = 1, \\ \sin x+\cos x = \sqrt{2}, \end{cases}$$

or

$$\begin{cases} \sin 3x+\cos 3x = \sqrt{2}, \\ \sin 18x = -1, \\ -\sin x+\cos x = -\sqrt{2}. \end{cases}$$

Let us solve the first system. The third equation is the simplest one. For this reason, we solve it first, and then check which of the solutions obtained satisfy the other two equations.

To solve the third equation, we use (2), and find that

$$\sin x+\cos x = \sqrt{2}\sin\left(x+\frac{\pi}{4}\right) = \sqrt{2}.$$

We have

$$\sin\left(x+\frac{\pi}{4}\right) = 1,$$

and, therefore,

$$x = \frac{\pi}{4}+2\pi n.$$

For these x,

$$\sin 3x+\cos x = \sin\frac{3\pi}{4}+\cos\frac{3\pi}{4} = 0.$$

Thus, the first system has no solutions.

For the second system, we similarly have

$$\sin\left(x-\frac{\pi}{4}\right)=1,$$

and, therefore,

$$x=\frac{3\pi}{4}+2\pi n.$$

For these x,

$$\sin 3x+\cos 3x=\sin\frac{9\pi}{4}+\cos\frac{9\pi}{4}=\sqrt{2}$$

and

$$\sin 18x=\sin\frac{27\pi}{2}=-1;$$

that is, these x satisfy the second equation and, therefore, the original equation.

Answer.

$$\frac{3\pi}{4}+2\pi n,\ n\in\mathbf{Z}.$$

Finally, we consider one more problem.

Problem 13. *Solve the equation*

$$\left(\sqrt{3-\tan^2\frac{3}{2}\pi x}\right)\sin\pi x-\cos\pi x=2.$$

Solution. Let

$$A=\sqrt{3-\tan^2\frac{3\pi x}{2}}.$$

We estimate the left-hand side of the equation:

$$\left|A\sin\pi x-\cos\pi x\right|\le\sqrt{A^2+1}=\sqrt{4-\tan^2\frac{3\pi x}{2}}.$$

We see that it does not exceed 2, and is equal to 2 only if

$$\tan^2\frac{3\pi x}{2}=0,$$

i.e., for

$$x=\frac{2n}{3},\ n\in\mathbf{Z}. \qquad (4)$$

For these x, the equation can be written as

$$\sqrt{3}\sin\pi x-\cos\pi x=2$$

(we intentionally substitute for x in some places, but not in others). We solve the last equation, to obtain

$$\sin\left(\pi x-\frac{\pi}{6}\right)=1$$

or

$$x=\frac{2}{3}+2m,\ m\in\mathbf{Z}.$$

Using (4), we obtain

$$\frac{2n}{3}=\frac{2}{3}+2m,$$

or

$$n=3m+1.$$

Thus, $x=\frac{2}{3}(3m+1)$, where $m\in\mathbf{Z}$.

Answer.

$$\frac{2}{3}+2m,\ m\in\mathbf{Z}.$$

In conclusion, we offer you the following exercises.

Exercises

1. Solve the following equations and systems:

(a) $\sqrt{x+2}+\sqrt{6-x}=3x^2-12x+16$;

(b) $\sin^{13}2x+\cos^{12}2x=1$;

(c) $\sin x+\sin 9x=2$;

(d) $\begin{cases}x+y+z=\sqrt{3},\\ x^2+y^2+z^2=1;\end{cases}$

(e) $\left(x^2-2x+3\right)\left(y^2+6y+12\right)=6$;

(f) $2^{|x|}-\cos y+\log_{10}(1+x^2+|y|)=0$;

(g) $\left(\sin^2x+\dfrac{1}{\sin^2x}\right)^2+\left(\cos^2x+\dfrac{1}{\cos^2x}\right)^2=12+\dfrac{1}{2}\sin y$.

2. Find all a for which the following systems have a unique solution:

(a) $\begin{cases}y\ge x^2+2a,\\ x\ge y^2+2a;\end{cases}$ (b) $\begin{cases}x+y+z=z,\\ x^2+y^2=z.\end{cases}$

3. Find all pairs of numbers that satisfy the following conditions:

(a) $\begin{cases}y^6+y^3+2x^2=\sqrt{xy-x^2y^2},\\ 4xy^3+y^3+\dfrac{1}{2}\ge 2x^2+\sqrt{1+(2x-y)^2};\end{cases}$

(b) $\begin{cases}\sqrt{2x^2y^2-x^4y^4}=y^6+x^2(1-x),\\ \sqrt{1+(x+y)^2}+x\left(2y^3+x^3\right)\le 0.\end{cases}$

4. Solve the equations

(a) $2\sqrt{3}\sin 5x-\sqrt{3}\sin x$

$=\cos 24x\cos x+2\cos 5x-6\left(\dfrac{\pi}{3}+2\pi n\right)$;

(b) $\sqrt{1-\cot^2 2\pi x}\cos\pi x+\sin\pi x=\sqrt{2}$.

5. For every value of b, solve the equation

$3\cos x\sin b-\sin x\cos b-4\cos b=3\sqrt{3}$.

Investigating Air

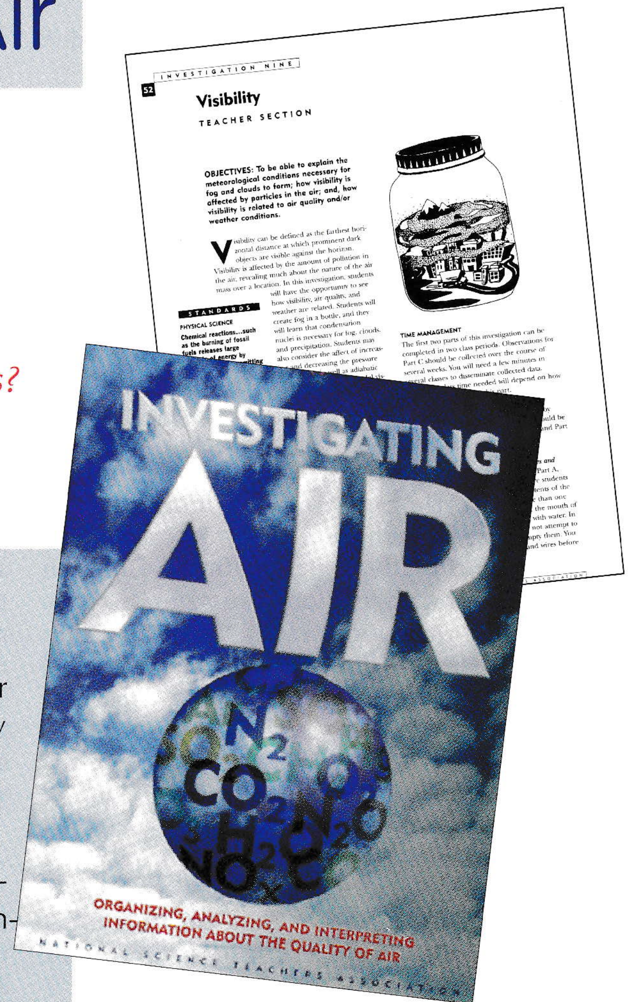
What is air made of?

How does air work?

Do humans impact air quality?

How does air quality impact humans?

Do we have a right to clean air?



Investigating Air provides the investigations and information that lead students to the answers to these and other questions. Students use data generated by researchers, and generate their own data, in situations analogous to real air quality issues—increases in atmospheric CO_2, acid deposition, ozone issues, and visibility. This resource culminates in a decision-making investigation that builds on the science and equity knowledge students have developed.
(grades 9–12, 1998, 78 pp.)

Produced by the National Science Teachers Association with support from the US Environmental Protection Agency.

#PB148X **$15.95**

Table of Contents

To order, call
1-800-722-NSTA

NSTA Pub Sales, 1840 Wilson Blvd.
Arlington, VA 22201-3000
www.nsta.org/scistore

P999Q2

A watery view and Waterloo

by A. Stasenko

JUST ABOUT ANY FISH KNOWS SOMETHING about waves and beaches. But that's not the point. We want to settle a matter of principle—why waves roll onto a beach parallel to the coastline. In doing so, they disgracefully neglect both the direction of the wind and their own previous course in the open sea. In other words, we are interested in why the lines perpendicular to the crests of the waves (the wave rays) curve in such a way that they come ashore at right angles to the shoreline.

Many things in physics are related to each other. This particular problem is directly related to the phenomenon of refraction of light rays at the boundary between two media—for example, when light travels from air to water. This is a simple process: in air light propagates with speed v_1, while in water it travels with a smaller speed v_2 ($v_2 < v_1$). The ratio of these speeds is the index of refraction of water relative to air:

$$n = \frac{n_2}{n_1} = \frac{v_1}{v_2} > 1$$

Here n_1 and n_2 are the absolute indices of refraction of air and water relative to the vacuum.

Snell's law of refraction says that

$$n_2 \sin\alpha_2 = n_1 \sin\alpha_1, \qquad (1)$$

where the α_i are the angles between the rays and the normal to the air-water interface. Therefore, when light travels to a more optically dense medium ($n > 1$), the refracted ray is closer to the normal than the incident ray: $\alpha_2 < \alpha_1$. Snell's law of refraction is universal: it is independent of the nature of the waves.

First of all, let's find the speed of the waves near a beach. What does it depend on? If we produce a "hump" on the surface of still water in the sea, a lake, or a pond of depth h (figure 1) and then release it, the hump will move downward under the effect of the gravitational force, which is proportional to the acceleration due to gravity. When the hump reaches the initial (unperturbed) position, it will not stop because of its inertia. Therefore, a cavity will appear instead of the hump, which generates a circular crest. As a result, a propagating water wave is generated.

What quantities (and corresponding dimensions) can affect the speed of water waves? We know that the waves that roll onto a beach are not small ripple waves, so we neglect surface tension. The size of sea waves is comparable to the depth of the water—at least in the beach area. Thus, two physical parameters should play a decisive role in determining the speed of sea waves—the depth h (m) and the acceleration due to gravity g (m/sec^2). Dimensional analysis yields the following formula:

$$v \sim \sqrt{gh} \qquad (2)$$

Thus, the waves travel slower in shallow regions (this is true only for rather shallow water, where the speed does not depend on the wavelength). When a wave approaches the shore, it enters less deep regions, so its speed decreases. By analogy with optics, we may say that the waves travel into a more and more "optically" dense medium.

By the way, we can model such a gradual change of optical density in a glass of water, where the index of refraction grows smoothly with increasing depth. To do this, we prepare a very concentrated solution of table salt, pour it into a glass, and then add less and less concentrated salt solutions. We must pour them layer by layer and finish with pure water. The composite optical medium is ready! A teaspoon in this glass will look not like a broken object, but like a smoothly curved object.

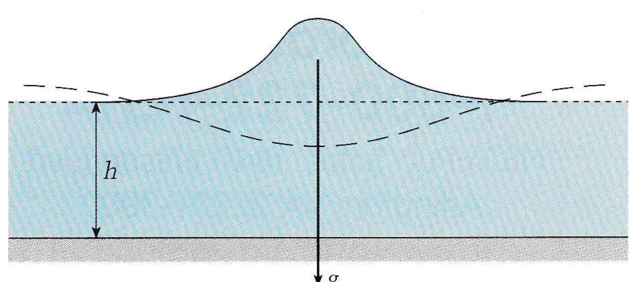


Figure 1

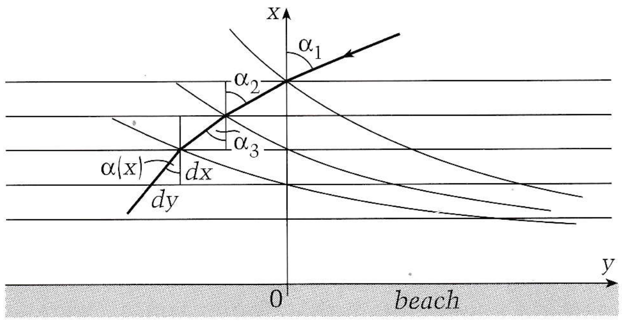


Figure 2

Let's divide the sea's surface into bands of width dx parallel to the coastline (figure 2). Each band is characterized by its own depth h, speed of wave propagation, index of refraction *(inversely proportional to this speed)*, and the angle between the wave ray and the axis x normal to the shore. Now Snell's law (2) for all layers can be written as

$$\frac{\sin\alpha_1}{\sqrt{h_1}} = \frac{\sin\alpha_2}{\sqrt{h_2}} = \ldots = \frac{\sin\alpha(x)}{\sqrt{h(x)}} = \text{const}\,. \qquad (3)$$

Even this initial equation shows that if the depth approaches zero, the angle between the ray and the normal also tends to zero. This feature explains why the waves approach the shore at right angles. However, we may proceed a little further and obtain the equation of the ray itself. Figure 2 yields:

$$\frac{dy}{dx} = \tan\alpha(x) = \frac{\sin\alpha(x)}{\cos\alpha(x)} = \frac{\sin\alpha(x)}{\sqrt{1-\sin^2\alpha(x)}}.$$

With the help of equation (3) we may obtain the relationship between the inclination of the wave to the normal at any distance x from the shore and its value α_1 at any fixed distance x_1:

$$\frac{dy}{dx} = \frac{\sqrt{h/h_1}\sin\alpha_1}{\sqrt{1-(h/h_1)\sin^2\alpha_1}}. \qquad (4)$$

We need only choose a particular profile of the seabed, that is, a dependence of its depth on the distance to the shore, $h(x)$—and you are welcome to integrate equation (4).

For example, let's take the quadratic law $h = h_1(x/x_1)^2$. This yields

$$\int_{y_1}^{y} dy = y - y_1 = \int_{x_1}^{x} \frac{(x/x_1)\sin\alpha_1}{\sqrt{1-(x/x_1)^2\sin^2\alpha_1}}\,dx.$$

At this stage we may want to ask mathematicians what to do with this monster. However, courage overcomes all obstacles—let's try to cope with this integral ourselves. First, we make a change of variable:

$$1-\left(\frac{x}{x_1}\right)^2\sin^2\alpha_1 = \gamma.$$

By differentiating this equation we get

$$2\frac{dx}{x_1}\frac{x}{x_1}\sin^2\alpha_1 = -d\gamma.$$

Substituting this into the integrand, we obtain

$$\frac{y-y_1}{x_1/\sin\alpha_1} = -\frac{1}{2}\int_{\gamma_1}^{\gamma}\frac{d\gamma}{\sqrt{\gamma}} = -\sqrt{\gamma}\Big|_{\gamma_1}^{\gamma}$$

$$= \sqrt{1-\sin^2\alpha_1} - \sqrt{1-\left(\frac{x}{x_1}\right)^2\sin^2\alpha_1}\,.$$

This clumsy formula is the desired ray equation $y(x)$. What kind of curve is this? We transfer the first term on the right-hand side to the left and then square both sides. By leaving only unity on the right-hand side we get

$$\left(\frac{y-y_1}{x_1/\sin\alpha_1} - \cos\alpha_1\right)^2 + \left(\frac{x}{x_1/\sin\alpha_1}\right)^2 = 1.$$

Isn't this familiar? Yes, this is the equation of a circle! Of course, another profile of the sea bed $h(x)$ would generate some other ray equation.

Let's proceed further. Snell's law of refraction (1) explains many other natural phenomena, including the *atmospheric refraction* of solar rays (figure 3). Since the density of the atmosphere grows in approaching the Earth's surface, the index of refraction in the air decreases at higher altitudes. Therefore, the solar rays curve in such a way that an observer on the Earth sees the Sun during some period after the geometrical sunset or before the geometrical sunrise. As a result, the period of daylight is prolonged for several minutes, a very regrettable phenomenon for nocturnal lovers. In contrast, the atmospheric refraction curtails the very long polar nights at high latitudes, which is a very useful result of good old Snell's law.

The same equation of refraction explains mirages in the desert. The hot sand warms the adjacent layer of air, so that the index of refraction of the upper layer becomes larger than that of the bottom layer. As a result, the rays emerging, say, from point A (figure 4) are de-

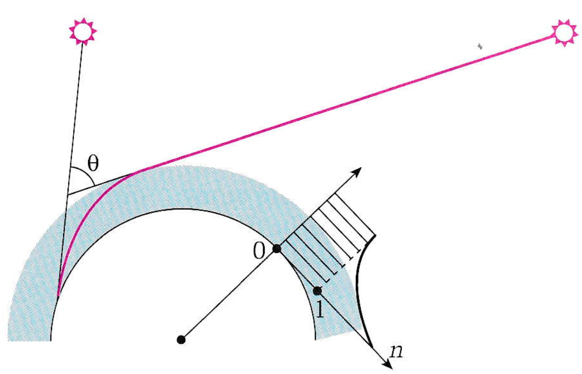


Figure 3

flected upwards, so that a tired traveler treats the point A^* as a reflection in a desert lake, which is unfortunately absent in reality.

Let's revisit the fish's eyeview of the world. The index of refraction of its environmental media (water) is significantly larger than unity ($n \cong 4/3$). Therefore, to focus the rays, the index of refraction of the fish's eye must be even greater. Should Nature make the eyes of fishes of flint or crown glass? This is indeed a difficult problem, and how Nature solved it should be considered in a special paper. Here we will just recall that the famous physicist Sir James Clerk Maxwell invented such an apparatus and called it a "fish eye." It is an unbounded refracting medium with index of refraction depending only on the distance (r) to a fixed point:

$$n(r) = \frac{n_0}{1+(r/a)^2}.$$

Here n_0 and a are constants. It can be shown that in such a medium a light ray moves along a circle, independently of the point of emergence and the initial direction.

At first glance, all this is optics. However, the same phenomena take place in acoustics, because equation (1) describes a general property of the rays for waves of any nature, which curve to the region where the speed of the wave decreases. On a sunny day the sand on the beach is strongly heated, so that the speed of sound in the adjacent hot layer of air is larger than in the upper layer (the speed of sound is proportional to the square root of the temperature). As in the case of optical illusions, the "sound rays" go upward, so that voices are muffled on the hot beach. In contrast, in the evening the soil cools down before the warm upper air, and the opposite phenomenon takes place: the upward sounds curve downward, so the sound of the "evening bells" really does carry across the fields at night.

However, while evening songs propagate far into the distance, they might not be heard nearby. During the famous battles of yore, full of sound and fury, sometimes

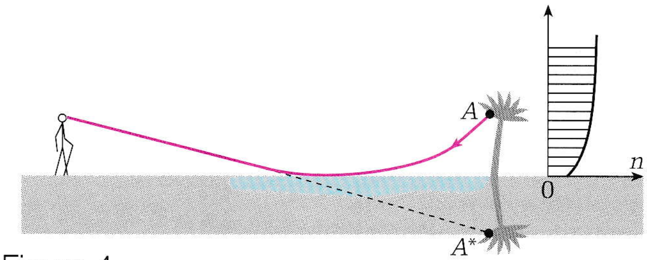


Figure 4

atmospherics were such that "quiet zones" were created. (Of course, surrounding terrain clouds and other factors described in textbooks on acoustics also contribute to this phenomenon.)

For example, the noise of the battle of Waterloo was heard at very great distances, but not nearby, where Napoleon's general Emmanuel Grouchy waited with his troops. Alas, the general did not come to save his Emperor. Thus, an acoustic event changed the course of European history. It certainly pays to learn physics—even if you're a general.

Quantum on waves, refraction, and optical illusions:

A. Eisenkraft and L. D. Kirkpatrick, "A Snail That Moves Like Light" (Physics contest), September/October 1991, pp. 28–29, and March/April 1992, p. 50.

A. Varlamov and A. Malyarovsky, "The Oceanic Phone Booth," May/June 1993, pp. 37–39.

I. Vorobyov, "The Bounding Main," May/June 1994, pp. 20-25.

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

A. Eisenkraft and L. D. Kirkpatrick, "Sea Sounds," March/April 1996, pp. 34–46, and September/October 1996, pp. 36–37.

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

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

L. A. Ostrovsky, "Sea Waves," November/December 1998, pp. 20–24.

CONTINUED FROM PAGE 30

ring. A few bubbles of air from the tube will be split into smaller bubbles that take their places at the axis of the ring. These bubbles will clearly reveal the water vortex traveling in water: it will be traced by the sparking beads. Running along the entire length of the bath, the rings hit the opposite wall, expand, and finally disintegrate.

Using this device, we can clearly demonstrate reflection of the rings from the free water surface. To this end, we need only turn the lower flange of the tube slightly upward. When a ring arrives at the water surface, it is reflected from it. Note that the angles of incidence and reflection are equal.

Since real liquids are viscous and rub against the walls of the vessels during their motion, they are constantly filled with generated vortical filaments. Helmholtz showed that a liquid mass set in any imagined motion could be considered as continuously filled with vortical filaments. He also developed mathematical and physical tools to study the motion of these filaments.

Quantum on vortices:

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

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.

AT THE BLACKBOARD III

An old algorithm

by Y. Solovyov

NOWADAYS COMPUTATIONS are mainly done with the help of calculators or computers. Nevertheless, it is still worthwhile to consider some of the many methods that have been developed since ancient times to facilitate doing calculations by hand. Everyone knows algorithms for addition, subtraction, multiplication, and division. In this paper we present an algorithm for taking square roots. This algorithm, which has been known since the fifteenth century, is not much more difficult than ordinary long division. It can be mastered by working just a few examples. Let us first consider the case when the number is a perfect square. For example, let us find the root of the number 294849.

We start by partitioning the digits of this number into groups of two, starting from the right.

$$\sqrt{29\ 48\ 49}$$

The groups are numbered (first, second, ...) in order of decreasing significance, from left to right. The total number of groups is equal to the number of digits that will appear in the square root.

The first digit of the square root is the square root of the largest perfect square that is equal to or smaller than the first group of digits. In our case this is 5. Now square this digit, subtract this squared value from the first group, and "bring down" the two digits of the next group, writing them to the right of the remainder. If the remainder is zero, this next group will stand alone. In our example we obtain the number 448.

$$\begin{array}{l} \ \ 5 \\ \sqrt{29\ 48\ 49} \\ -25 \\ \ \ \ 4\ 48 \end{array}$$

Now compute the second digit. Draw a vertical line to the left of the number obtained (448). Multiply the first digit obtained by two ($2 \times 5 = 10$),

$$\begin{array}{rl} & \ \ 5 \\ & \sqrt{29\ 48\ 49} \\ & -25 \\ 10 & \vert\ \ 4\ 48 \end{array}$$

and write the product to the left of the line you just drew, leaving space for one digit to the right of this product (the rightmost digit of the product is in the "tens" place). The digit that goes in the "ones" place is the largest digit $\underline{a}$ such that the difference $448 - \overline{10a} \cdot a$ is nonnegative (i.e., positive or zero).[1] By trial and error, we see that in our case this digit is 4, and we write this as the second digit of the result.

$$\begin{array}{rl} & \ \ 5\ \ 4 \\ & \sqrt{29\ 48\ 49} \\ & -25 \\ 104 & \vert\ \ 4\ 48 \end{array}$$

Now multiply 104 by 4 and write down the result to the right of the vertical line. Calculate the difference $448 - 416 = 32$ and write down the next group to the right of the remainder, to obtain 3249.

$$\begin{array}{rl} & \ \ 5\ \ 4 \\ & \sqrt{29\ 48\ 49} \\ & -25 \\ \times\ 104 & \vert -4\ 48 \\ 4 & \vert\ \ \ 4\ 16 \\ & \vert\ \ \ \ 32\ 49 \end{array}$$

The third digit of the result can be found in much the same way as the second digit: double the answer obtained so far (multiply 54 by 2), write down the product, 108, to the left of the vertical line, and find the largest digit b such that the difference $3249 - \overline{108b} \cdot b$ is nonnegative. In our example we find $b = 3$, and we write this as the third digit of the result.

$$\begin{array}{rl} & \ \ 5\ \ 4\ \ 3 \\ & \sqrt{29\ 48\ 49} \\ & -25 \\ \times\ 104 & \vert -4\ 48 \\ 4 & \vert\ \ \ 4\ 16 \\ 108 & \vert\ \ \ \ 32\ 49 \end{array}$$

We now multiply 1083 by 3, write down the product to the right of the vertical line, and then subtract it from 3249.

$$\begin{array}{rl} & \ \ 5\ \ 4\ \ 3 \\ & \sqrt{29\ 48\ 49} = 543 \\ & -25 \\ \times\ 104 & \vert -4\ 48 \\ 4 & \vert\ \ \ 4\ 16 \\ \times\ 1083 & \vert\ -32\ 49 \\ 3 & \vert\ \ \ \ 32\ 49 \\ & \qquad\ \ 0 \end{array}$$

Since the remainder is zero, the calculation is completed.

[1] Here, as usual, $\overline{10a}$ denotes the integer with hundreds digit 1, tens digit 0, and units digit a.

Now, looking at the calculations below, try to repeat all the necessary steps and find the roots $\sqrt{212521}$ and $\sqrt{165649}$.

```
 √21 25 21 = 461
 -16
 ×86 | -5 25
   6 |  5 16
×921 |   -921
   1 |    921
            0
```

```
 √16 56 49 = 407
 -16
 ×80 | -56
   0 |  00
×807 | -56 49
   7 |  56 49
            0
```

So far, we have considered only numbers with an even number of digits. If the number of digits in the given number is odd, the leftmost group will consist of a single digit.

```
 √1 46 41 = 121
 -1
 ×22 | -46
   8 |  44
×241 |  -2 41
   1 |   2 41
            0
```

```
 √5 33 61 = 231
 -4
 ×43 | -1 33
   3 |  1 29
×461 |  -4 61
   1 |   4 61
            0
```

What is to be done if the given number is not a perfect square? The algorithm does not change in this case, but the number itself must be treated in a certain way. Let N be a number (not necessarily an integer) written as a decimal fraction; suppose that we are required to calculate its square root to an accuracy of $1/10^m$, i.e., to m decimal places. As before, we partition the digits of the integer part of the number into groups of two digits each, starting from the right, while the fractional part is partitioned into similar groups starting from the left (in other words, start from the decimal point and work in both directions). If the number of digits in the integer part is odd, the leftmost group consists of a single digit; if the number of digits in the fractional part is odd, we write down an extra zero to the right of the last digit. If the number of groups in the fractional part is greater than m, we eliminate the extra groups from the right; if it is less than m, we fill in the missing groups with zeros. Now everything is ready for our algorithm to be applied. Here we demonstrate the calculation of the square roots $\sqrt{2}$ and $\sqrt{12.5}$ to an accuracy of 0.001.

```
 √2 = √2.00 00 00 = 1.414
      -1
  ×24 | -1 00
     4 |    96
  ×281 |   -4 00
     1 |    2 81                  (*)
 ×2824 |   -1 19 00
     4 |    1 12 96
                6 04
```

```
 √12.5 = √12.50 00 00 = 3.535
         -9
  ×65 | -3 50
     5 |  3 25
  ×703 |  -25 00
     3 |   21 09
 ×7065 |   -3 91 00
     5 |    3 53 25
              37 75
```

To check whether you fully understand the algorithm, calculate the following roots to an accuracy of 0.001: $\sqrt{18769}$, $\sqrt{24336}$, $\sqrt{35721}$, $\sqrt{232324}$, $\sqrt{243049}$, $\sqrt{104.2441}$, $\sqrt{1867.1041}$, $\sqrt{7}$, and $\sqrt{10}$.

In this paper we will not prove the validity of the algorithm described. Note that the calculation of each new digit of the square root requires increased computational effort; for this reason, this algorithm should be used only when the required accuracy does not exceed three or four significant digits (an accuracy sufficient for most practical calculations).

The accuracy of a calculation can be improved by using the following theorem.

Suppose that we have calculated n significant digits of the square root. Subtract the square of the value found for the root from the number whose square root you are taking, and divide the difference by twice the value found for the root; the result gives the next $n - 1$ digits.

Let us prove this proposition. First, suppose that the integer part of the given number a consists of n groups of two digits. Let the first n digits of the root give a number c. Then $\sqrt{a} = c + x$, where x is the fractional part to be found. Therefore,

$$a = c^2 + 2cx + x^2,$$

$$\frac{a - c^2}{2c} = x + \frac{x^2}{2c}.$$

The difference $a - c^2$ is the remainder obtained after the calculation of the first n digits of the root. The quotient

$$\frac{a - c^2}{2c}$$

is the number in the proposition (the next $n - 1$ digits). Hence,

$$x = \frac{a - c^2}{2c} - \frac{x^2}{2c}.$$

Setting

$$x \cong \frac{a - c^2}{2c},$$

we make an error

$$\frac{x^2}{2c}.$$

Since $x < 1$ and $a \geq 10^{n-1}$, we have

$$\frac{x^2}{2c} < \frac{1}{2 \cdot 10^{n-1}},$$

which gives the necessary estimate. If the decimal point in the number a is to be moved an even number of decimal places from where it was when we calculated the root, we get the new answer by moving the decimal point in the old root by half that number of decimal places in the same direction (i.e., if a is multiplied or divided by some power of 10 with an even exponent, the root found previously must be multiplied or divided by 10 to the power of one-half that exponent).

Consider an example. Calculate $\sqrt{2}$ to six decimal places. We have already calculated $\sqrt{2}$ out to three decimal places [see (*)]. Dividing the remainder by twice the value obtained for the square root, we obtain the next three digits.

$$\sqrt{2} = \sqrt{2.00\,00\,00} = 1.414213$$

```
        −1
  ×24 | −1 00
    4 |    96
 ×281 |  −4 00
    1 |   2 81
×2824 | −1 19 00
    4 |   1 12 96
            6 04 | 28 28
           −6040 | 0.213
            5656
           −3840
            2828
          −10120
            8484
            1636
```

Certainly nowadays, when every primary school pupil has a calculator, it is not easy to delight a student with an algorithm for calculating square roots by hand. However, we can still hope. The well-known American mathematician Ron Graham recalled that his first interest in mathematics was invoked by this very algorithm for the calculation of square roots. He wondered whether it was possible to find a similar algorithm for the calculation of cube roots and roots of higher orders. He found such an algorithm, and since then mathematics has always been part of his life.

The Death of a Star

(Part 1)

by David Arns

What happens to a star whose life is drawing to an end?
These luminaries shone when in their prime,
But as the final curtain nears, which attributes attend?
Do they waste away and shrivel?
Or get rude and most uncivil?
Or just disappear with style and chic sublime?

It depends, of course, upon the kind of star that you might choose:
The Hollywood variety, perchance?
And all of their shenanigans that always make the news?
Or would you choose the kind
with which heaven's void is lined:
And are scattered through the cosmos' vast expanse?

The subject of this magazine demands the latter choice
(A checkout-counter tabloid this is not);
And so I'll choose the stars that make astronomers rejoice:
Those spheres of white-hot fusing
Hydrogen, which they keep using
To maintain their normal size and keep them hot.

So what happens to these stars when all their hydrogen's consumed?
Well, they're all so far away, it's hard to tell,
But according to the theories, even when a star is doomed
And its life is nearly through,
Then the next thing it will do
Depends upon its gravitational well.

See, a little star has insufficient mass and gravitation
To make a big display when once it dies;
Its fuel used up, it undergoes a gradual transformation
Where it cools, and then grows dark,
Leaving not a single spark
Of the splendor that it had ere its demise.

A star of somewhat larger mass and gravity will die
With significantly more than just a "pop."
Its gravity will cause some different physics to apply:
When its hydrogen is gone,
Then the helium's light lives on,
Since its gravity's compression heats it up.

So the helium starts to fuse, and a brand-new lease on life
Is now granted to this star of medium mass.
But then the helium's ash (that's carbon) in the stellar core gets rife
And the star again compresses
Under gravitational stresses
And so iron, fused from carbon, comes to pass.

CONTINUED ON PAGE 56

Boiling liquid

by A. Borovoi

HOW DOES A BUBBLE CHAMber work? During the early 1950s, Donald Arthur Glaser, a young scientist at the University of Michigan, invented an original device to record elementary particles. His apparatus became known as a bubble chamber.

Figure 1 shows a modern bubble chamber. It is a rather large setup that occupies a spacious hall with a number of floors equipped with sophisticated devices and controlled by powerful computers. Many bubble chambers have names, such as Ludmila, Mirabelle, and Gargamelle.

However, the first models of a bubble chamber made by Glaser were in no way "impressive" (figure 2). Their main part was a glass ampoule filled with ether. Its volume was only a few cubic centimeters. The liquid was heated and compressed to 20 atm (1 atm $\cong 10^5$ Pa). A simple contrivance could rapidly relieve the pressure. According to thermodynamics, this maneuver results in boiling of the liquid.

However, boiling cannot start immediately: there is some latent waiting period before it starts. If a charged particle encounters such a superheated (with respect to the boiling point) liquid during this period, it will produce a track of ions along its path, marked by bubbles of vapor. One can photograph the track and raise the pressure again. The bubbles collapse under the raised pressure, and the chamber is ready for a new measurement.

Bubble chambers have been extremely useful for studying the physics of elementary particles. Figure 3 shows a photograph taken in a bubble chamber. *Why did the bubbles appear precisely along the path of the particle?*

Let's recall boiling

Textbooks say that boiling is the process of formation of bubbles throughout the volume of a liquid, which grow and rise to the surface. Evidently, the pressure inside the bubbles is greater than the pressure in the liquid—otherwise they would collapse. Everybody knows that under normal atmospheric pressure water boils at 100°C. Precisely at this temperature the pressure of saturated water vapor becomes sufficient to produce bubbles, so that water begins to boil. However, is it precisely so, and does a liquid always boil under such conditions?

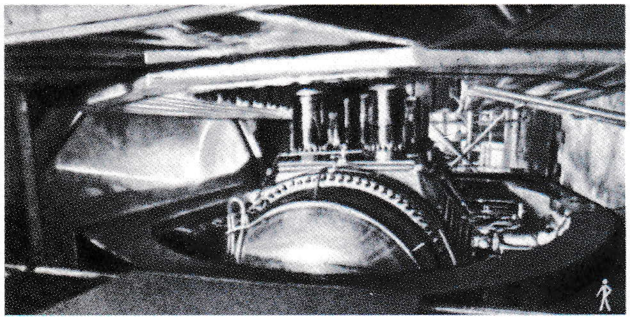


Figure 1. *The hydrogen bubble chamber "Mirabelle" is a key instrument at the High Energy Physics Research Institute.*

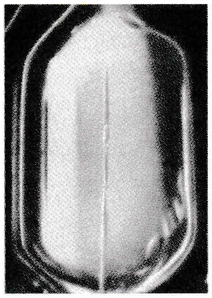


Figure 2. *The track of a charged elementary particle recorded in the first bubble chamber of Glaser.*

Let's take two test tubes. We wash one of them very carefully. This test tube must not have any scratches or foreign inclusions or particles. We fill it with distilled water (about 10 cm³). The other test tube should not be particularly clean. We fill it with the same amount of tap water and drop a piece of chalk into it. Both test tubes are heated under identical conditions, avoiding direct contact with the flame. It turns out that the water boils quite differently in these two test tubes (figure 4).

In the test tube with tap water the boiling starts earlier and is a quiescent and continuous process. As a rule, the bubbles are produced on the surface of the chalk. In contrast, in the test tube with distilled water, boiling starts later (at a higher temperature) and proceeds irregularly. Large bubbles appear now and then, and their generation and collapse are accompanied by loud crackles.

Using twice-distilled water, it is possible to purify water and glassware (retorts, flasks, and the test tubes) so carefully that boiling does not start until 140°C. Such water is said to be superheated. It is very dangerous: if a grain is thrown into the

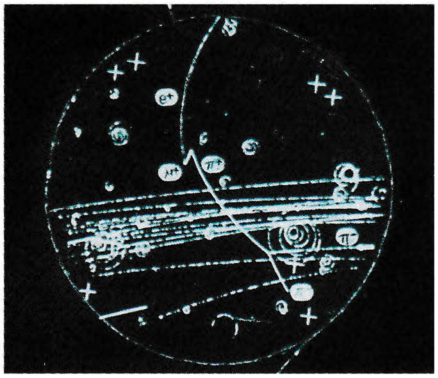


Figure 3. *A photograph of processes of interaction and conversion of elementary particles in a liquid-hydrogen bubble chamber.*

water, the water will burst into boiling. The reason is an extremely high rate of production of bubbles.

Chemists are well aware of this dangerous property of liquids. In order to guard against bursts and ensure a quiescent and homogeneous boiling process, they place so-called boiling chips (pieces of glass and porcelain tubes, or marble fragments) into the vessel with the boiling liquid.

These remarkable features of boiling are explained by the liquid's surface tension. The surface of a liquid can be imagined as a stretched elastic film. Such a film "wants" to contract, so the surface tension tries to crush a newly formed bubble. The smaller the bubble's radius, the higher the additional pressure developed by the surface tension[1], which nips the boiling in the bud. This is the reason why a pure homogeneous liquid can be superheated. At the same time, any heterogeneity in the liquid itself or foreign bodies facilitate boiling.

In particular, such heterogeneous zones are produced along the trajectory of a charged particle in a bubble chamber, so that the bubbles form along the particle's path. Now it is clear why the homogeneity of the ether and the purity and smoothness of the ampoule are the most important conditions for the operation of the Glaser bubble chamber.

Some interesting observations on bubbles in a liquid

Such observations can be carried out with ordinary aerated water, which sometimes behaves very much like a superheated liquid. This raises the following question: *Why and under what conditions can aerated water be a model for a superheated liquid?*

For our experiments we need a bottle of lemonade, a glass, a teaspoon, sugar, and a piece of chocolate (in any case, one can be confident about the pleasant side effects of experiments with such test objects). We begin our study with a rapid opening of the bottle with aerated water. Instantaneously, a light puff of smoke appears above the neck of the bottle. Why does this happen?

We now pour the water into the glass and wait until the foam dissipates and only individual bubbles rise in the water. We put a close-fitting lid on this glass and tighten it. After a while, the bubbling stops. If we open the glass, bubbles appear again. *Why?*

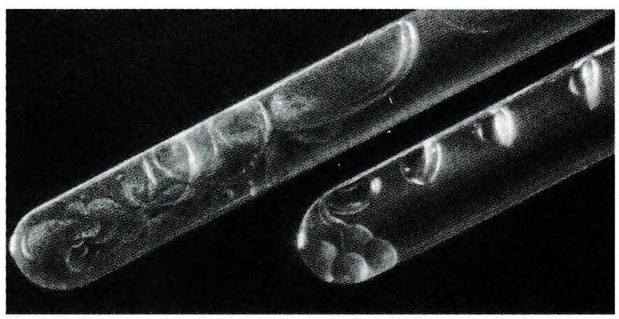


Figure 4. *The process of boiling proceeds quite differently in test tubes with distilled and tap water.*

We throw a pinch of sugar into the glass. The "boiling" of the aerated water is intensified, and foam appears on its surface again. Evidently, there was still a lot of carbon dioxide in the water. *What prevented its release previously, and why did the granulated sugar induce it?* There are other ways to induce the formation of foam—for example, by stirring the water or by pouring it from one glass into another.

A spectacular demonstration can be performed by throwing a small piece of chocolate or a berry into a glass of aerated water. Since chocolate is more dense than water, it will quickly sink to the bottom. Here it is "overgrown" with bubbles. Like buoys, they lift the sunken piece of chocolate. When it reaches the surface, the bubbles escape to the air, and the piece of chocolate sinks again. Sometimes this sink-rise cycle will be repeated a dozen times.

How can these questions be answered?

In essence, all these questions can be reduced to the first problem—why is aerated water similar to a superheated liquid?

Perhaps you know how to carbonate water at home. It can be done by placing a small balloon filled with carbon dioxide (CO_2) over the mouth of a test tube filled with water. To accelerate the dissolving of the carbon dioxide, the test tube should be shaken vigorously. The gas continues to dissolve until equilibrium is reached. If this gas does not react chemically with water, its steady-state content (that is, its density and concentration) in water is proportional to its pressure at the surface.[2] There are other recipes for preparing fizzy water at home with the help of lemon juice or soda water, etc.

If we quickly open a bottle of aerated water, the pressurized gas located above the surface expands rapidly and cools off. The water vapor released with the gas will condense and form a small misty cloud. After the pressure on the liquid drops ap-

[1]The additional pressure applied to a bubble can be evaluated by the formula $P = 2\sigma/r$, where σ is the coefficient of surface tension and r is the radius of the bubble. The same pressure is applied to the liquid in a capillary directly under the convex hemispherical surface (meniscus).

[2]In the case of carbon dioxide and water, the process is somewhat more complicated because these substances do react with each other. However, the resulting acid is not stable and rapidly disintegrates.

preciably, the dissolved carbon dioxide is released from the water in two ways: through the gas-water interface and into the bubbles formed in the water.

Immediately after opening the bottle, when the equilibrium is disturbed most strongly, and when the water contains a great surplus of carbon dioxide, bubbles are produced easily and in great numbers. Thus, foam appears on the surface of the water. Gradually, the foam abates, although there is some "extra" gas left in the water. However, the formation of bubbles is now impeded because the concentration of the gas is considerably reduced, and therefore the chances are very low that the gas molecules will come together to produce a bubble. Bubbles appear and grow only at the heterogeneous places in the water, and this process looks very much like the boiling of superheated water. Perhaps it will not be difficult for our readers to explain all the following experiments.

An absolutely unexpected application of aerated water

Our article began with a story about the bubble chamber. We want to finish it with another story, which also deals with an application of bubbles—and again in the physics of elementary particles.

Several years ago, Soviet physicists investigating the properties of neutrinos were faced with an exceptionally difficult problem. It was necessary to extract a few atoms of gas formed in a large volume of liquid (hundreds of liters) by collisions with neutrinos. This gas (neon-23) is radioactive and highly unstable: its atoms decay in less than one minute. Therefore, it was necessary to find 5–10 neon atoms among 10^{28} atoms of liquid, extract them, transport them to a counter, and finally count them. At first glance, the problem seemed insoluble.

Nevertheless, the problem was solved with the help of a carbonated liquid. Prior to irradiation, the liquid was saturated with carbon dioxide. Immediately after the irradiation, the vessel was opened and the liquid was vigorously stirred with special blades. As a result, turbulent boiling was induced, which increased the area of the gas-liquid boundary by tens of thousands of times, thereby drastically raising the rate of escape of neon atoms from the liquid. The evaporating carbon dioxide carried them away from the vessel with the boiling liquid. The mixture of neon and carbon dioxide was transported to another vessel, where the carbon dioxide was absorbed by an alkali solution. Then the neon atoms were collected in a special test tube, which was quickly placed into the counter. The entire procedure took less than 20 sec. Thus, aerated liquids are not only pleasant, but useful as well.

Quantum on boiling, vapors, and elementary particles:

S. L. Glashow, "The Elementary Particles," September/October 1990, pp. 49–51.

M. Digilov, "A Strange Box and a Stubborn Brit", March/April 1991, pp. 26–27.

A. Buzdin and V. Sorokin, "Double, Double Toil and Trouble," May/June 1992, pp. 52–53.

C. Waltham, "The Omnipresent and Omnipotent Neutrino," July/August 1993, pp. 10–15.

A. Eisenkraft and L. D. Kirkpatrick, "Focusing Fields," January/February 1996, pp. 30–31.

H. D. Schreiber, "The Name Game of the Elements," September/October 1996, pp. 24–30.

T. Polyakova, V. Zabolotsky, and O. Tsyganenko, "Stirring Up Bubbles," March/April 1997, pp. 52–55.

I. Mazin, "Physics in the Kitchen," September/October 1997, pp. 54–56.

"Do You Really Know Vapors?", Kaleidoscope, September/October 1998, pp. 32–33.

S. Krotov and A. Chernoutsan, "Cold Boiling," January/February 1999, p. 33.

CONTINUED FROM PAGE 53

Well, this iron core is stubborn, and resists attempts to fuse
Into anything more dense than iron at all,
But the grav'ty of this star of medium mass is loathe to lose
When it fights atomic forces
So its tendency, of course, is
Just to squeeze it all into a neutron ball.

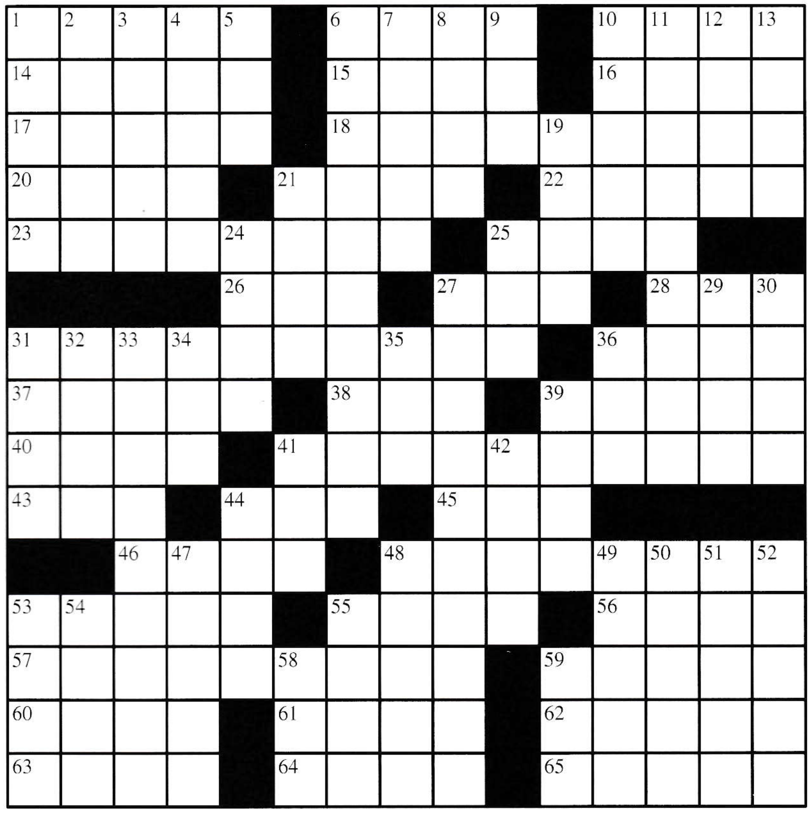
This "neutron star," as it is called, is just exactly that:
The pressure got so high in its collapse,
That electron shells of atoms, in their white-hot habitat,
Just cave in, and burst asunder,
Nuclear forces knuckling under,
And neutrons all crowd in and fill the gaps.

Now, as you can well imagine, since an atom's mostly space,
A neutron star is dense beyond belief.
Just try to grab a teaspoonful and take it anyplace:
The hundred million tons in weight
Would crush your spine to such a state
You'd hardly be much more than bas-relief.

So now, you ask, what happens when a star's so very large
Even neutrons can't resist gravitic squeeze?
Ah, well, that's another duty I will faithfully discharge,
In an issue not yet printed
(As he shrewdly, subtly hinted),
Assuming that the publisher agrees!

Criss × cross science

by David R. Martin



ACROSS

1 Spirit
6 Surgeons William James, and Charles Horace ___
10 Coup d'___
14 Hungarian composer Erno ___
15 Stravinsky
16 Heraldic border
17 "The Phantom of the ___"
18 ___ coordinate system
20 Forearm bone
21 Type of cheese
22 Submerged
23 Elementary particle
25 ___ cava (large vein)
26 Point
27 1968 Physiol. Nobelist ___ Khorana
28 Calculator display
31 .3937 inches
36 Of aircraft
37 Help
38 Titanium carbide
39 Astronomer ___ Cannon (1863–1941)
40 Focusing device
41 3.937×10^{-11} mil
43 Manganese oxide
44 It's mostly Nitrogen
45 Japanese sash
46 Speck of dust
48 Like some alleles
53 One cubic meter
55 Fencing sword
56 Singer ___ Horne
57 Trigonometric function
59 Pennies
60 Gebang palm fiber
61 Reverse curve
62 Editor's mark
63 Pressure units
64 An Esker
65 Carpel stalk

DOWN

1 Periodic table division
2 Simple: comb. form
3 Unlocks
4 Near Eastern inn
5 Oolong, e.g.
6 .03937 mil
7 Over
8 "In days of ___"
9 Scrap of food
10 Microscopy dye
11 Able to displace 3 H atoms
12 Wings
13 Be inclined
19 Dammed German river
21 Edge
24 Before bone or gate
25 Reactive power unit
27 3937 inches
29 Cleveland's lake
30 Active person
31 Peaceful
32 Type of parity
33 3.937×10^{-5} mil
34 "___ the season ..."
35 Actor ___ Robbins
36 Sue ___ Langdon
39 Moslem ruler
41 For shame!
42 Woodwind
44 Certain asteroid
47 Type of exams
48 Asphyxia
49 Double fold
50 Inductance unit
51 Chip maker
52 Social insect division
53 Wound cover
54 Roman garment
55 Brain graphs
58 Tibetan gazelle
59 Units of volume: abbr.

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE JANUARY/FEBRUARY PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M286

Let I, J be the centers of the circles inscribed in triangles ABM and CDM, respectively (see figure 1). First we note that I, M, and J are collinear.

Indeed, I is the intersection point of the angle bisectors of triangle ABM, so $\angle IMB = \angle IMA$. Similarly, $\angle CMJ = \angle DMJ$. And $\angle BMC = \angle AMD$ (they are vertical angles) so $\angle IMB + \angle BMC + \angle CMJ = \angle IMA + \angle AMD + \angle DMJ$. This implies that $\angle IMJ = 180°$, so I, M, and J are collinear.

Next we prove that triangle IJL is isosceles. $\angle LIM$ is an exterior angle of triangle IBM, so $\angle LIM = \angle IMB + \angle IBM$. Similarly (using triangle CMJ), we have $\angle LJM = \angle CMJ = \angle MCJ$. But $\angle IBM = \angle DMJ = \angle CMJ$, and $\angle IBM = \angle JCM$ (they intercept equal arcs along the circle). Hence $\angle LIJ = \angle LJI$, which proves our assertion.

M287

The delay time starts accumulating when the mathematician drops his stick into the water. We could calculate it using straightforward algebraic methods, but the calculation is made simpler if we use one of the physicist's prize tricks. Let us consider all the events from the point of view of someone floating down the river. That is, let us take a frame of reference moving with the river. From this point of view, neither the hat nor the stick move at all, but the mathematician (and his house!) are moving upstream, as if on a conveyor belt.

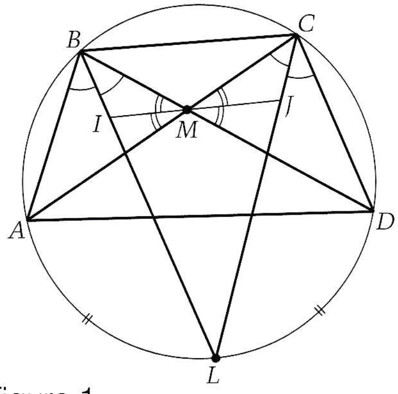


Figure 1

We can divide the time lost into three parts: (a) the time lost walking downstream from the stick back to the hat, (b) the time lost walking upstream from the hat up to the stick, and (c) the time lost because the house has moved further on the conveyor belt during all these maneuvres.

Suppose the speed of the current is v. Then the mathematician walks upstream at a speed of $1.5v$, to which we must add the speed of the conveyor belt. So, from our point of view, he walks upstream at a rate of $2.5v$. When he turns back to get his hat, he runs at a speed of $3v$, from which we must subtract the speed of the river (he runs downstream), so his speed is $2v$.

For part (b), we know that at this rate it takes him 10 minutes, walking upstream, to cover the distance between the stick and the hat.

For part (a), the distance traveled is the same as part (b) (since the hat and stick are stationary). How long will this take at the rate of $2.5v$? We have $(10)(2.5v) = (t)(2v)$, which gives $t = 12.5$ minutes. So parts (a) and (b) account for 22.5 minutes of lost time.

For part (c), we note that the house 'traveled' $22.5v$ meters in the time taken by procedures (a) and (b). The mathematician is traveling $1.5v$ faster than the house, so he makes up this distance in $22.5/1.5 = 25$ minutes.

Altogether, the mathematician has lost 37.5 minutes.

The reader may find it amusing to work this all out by conventional algebra. The key insight, capture by the physicist's trick, is that the distance between the hat and the stick, once they are in the river, does not change. (Solution by Boris Korsunsky.)

M288

The solution will follow if we prove that both the numerator and denominator are divisible by 111...111, the number consisting entirely of 1's, and containing as many 1's as either the numerator or the denominator of the given fraction. To simplify the solution, we take the case when the numerator and denominator each contain ten 1's. The general case will be clear from this discussion and is left to the reader.

In this case, the denominator is 11001111110011 and can be written as (we group the digits to make the numbers easier to read):

$$\begin{array}{r} 11\ 111\ 111\ 110\ 000 \\ +\quad 1\ 111\ 111\ 111 \\ -\ 111\ 111\ 111\ 100. \end{array}$$

(The reader is invited to check this.) This sum is equal to:

$(1\ 111\ 111\ 111)(10\ 000 + 1 - 100) = (1111111111)(9901)$.

The numerator can be treated similarly and is equal to:

$$\begin{array}{r} 11\ 111\ 111\ 110\ 000 \\ +\ 111\ 111\ 111\ 100 \\ +\quad 1\ 111\ 111\ 111 \\ -1\ 111\ 111\ 111\ 100 \\ -\quad 11\ 111\ 111\ 110, \end{array}$$

which can be written as

$$(1\ 111\ 111\ 111)(9091).$$

Hence the original fraction is equal to 9091/9901.

M289

Let $M(t, t^2)$, with $t > 0$, be a point on the parabola $y = x^2$ (see figure 2). The slope of the tangent at this point is $2t$, the slope of the perpendicular (normal) is $-1/2t$, and the equation of the normal is

$$y = -\frac{1}{2t}x + t^2 + \frac{1}{2}.$$

Let us find the second point of intersection of this normal N with the parabola $y = x^2$. We have the equation

$$x^2 = -\frac{1}{2t}x + t^2 + \frac{1}{2}.$$

One root of this equation corresponds to the point M: $x_1 = t$. The sum of the roots is $1/2t$, so the second root is

$$x_2 = -t - \frac{1}{2t}.$$

The area of the segment is

$$S = \int_{-t-\frac{1}{2t}}^{t} \left(-\frac{1}{2t}x + t^2 + \frac{1}{2} - x^2\right)dx$$

$$= -\frac{1}{2t}\cdot\frac{x^2}{2} + \left(t^2 + \frac{1}{2}\right)x - \frac{x^3}{3}\Bigg|_{-t-\frac{1}{2t}}^{t}$$

$$= \frac{4}{3}t^3 + t + \frac{1}{4t} + \frac{1}{48t^3} = \frac{4}{3}\left(t + \frac{1}{4t}\right)^3.$$

This function of t attains its minimum at the point $t = 1/2$, and the minimal value itself is 4/3.

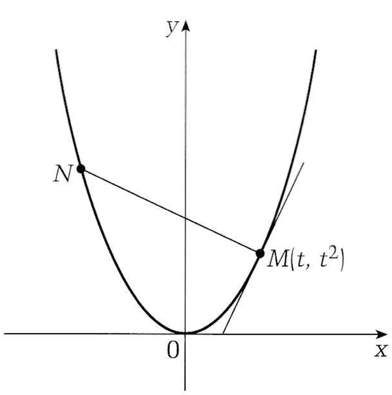


Figure 2

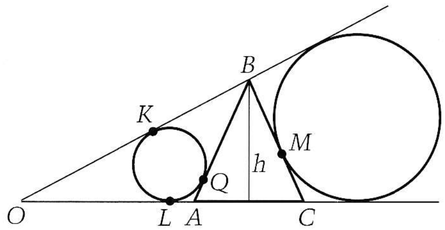


Figure 3

M290

In figure 3, $AB = BC = a$, $AC = 2b$, and let h be the triangle's altitude. Using the fact that the tangents drawn to a circle from the same point are equal, we find that $AL = AQ$, $BQ = BK$, and $OL = OK$. Therefore, $OL + BQ = OK + KB = OB$. Thus, $(AO - AL) + (AB - AQ) = OB$. Since $AL = AQ$, we have

$$AQ = \frac{1}{2}(AO + AB - OB).$$

In the same way, we find that

$$CM = \frac{1}{2}(BO + CB - OC).$$

Then

$$AQ + CM = \frac{1}{2}(AO + AB + BC - OC)$$

$$= \frac{1}{2}(2AB - AC) = a - b.$$

Thus, the sum $AQ + CM$ is independent of how the line passes through vertex B. If r and R are the radii, then

$$\frac{r}{AQ} = \frac{R}{CM} = \cot\frac{\alpha}{2},$$

where α is the internal angle at the base of the isosceles triangle ABC; therefore,

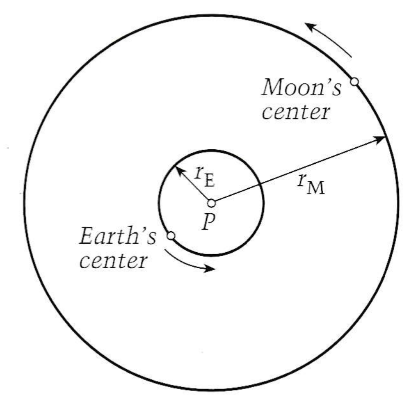


Figure 4

$$\frac{r + R}{AQ + CM} = \cot\frac{\alpha}{2}$$

or

$$\frac{r + R}{a - b} = \cot\frac{\alpha}{2} = \frac{\sin\alpha}{1 - \cos\alpha} = \frac{h}{a - b},$$

whence $r + R = h$. Thus, $h = 1$.

Physics

P286

The mutual gravitational attraction between the Earth and the Moon results in their revolution around some point P (figure 4). The Moon's center of gravity (CG) revolves along a circle of radius $r_M \cong 380{,}000$ km, while the corresponding value for the Earth's CG is $r_E \cong 4{,}700$ km. The Earth-Moon system revolves as a whole about the Sun under the action of solar gravitation. The point P (the system's CG) revolves about the Sun along a circle of radius $R_0 \cong 150 \cdot 10^6$ km.

Owing to the Moon's revolution about the Earth, the CG of our planet is located first on one side of the point P, and then on the other side of this point. Therefore, the Earth does not simply revolve about the Sun along the circular orbit of radius R_0, but it also oscillates about this circular orbit. The maximal distance of the Earth's center from the Sun's center is $R_0 + r_E$, while the minimal distance is $R_0 - r_E$.

The Moon makes a full revolution about the Earth in $T_M = 27$ days, 7 hours, 43 minutes, and 11 seconds (27.322 days). At the same time, the Earth's center travels through a distance $S = 2\pi r_E \cong 29{,}500$ km. This motion is not rapid: its linear speed $v = S/T_M \cong 12.5$ m/sec is a little higher than the fastest human speed ($\cong$10 m/sec). In contrast, the linear velocity of the orbital motion of the point P about the Sun is much greater: $V \cong 30$ km/sec.

The segment of the trajectory traversed by the point P during a small interval of time can be approximated by a straight line. Indeed, the daily rotation of the radius vector of the point P is $\alpha \cong 2\pi/365 \cong 0.99°$, so

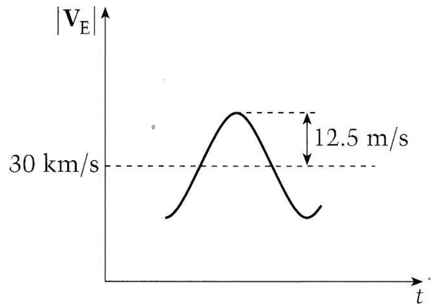


Figure 5

that the corresponding arc of the trajectory is practically a straight line segment. In other words, the velocity of the point P maintains the same direction during a small interval of time. Therefore, in this interval the projection of the velocity v onto the direction of the basic trajectory can be written as

$$v_0 = v \sin(\omega t + \phi_0),$$

where $\omega = v/r_E$ is the angular frequency of rotation, while the initial phase ϕ_0 depends only on the choice of the origin for the time variable t. During a small interval of time, the speed of the Earth's CG relative to the Sun also changes according to a sinusoidal law (figure 5):

$$|\mathbf{V}_E| = V + v_0 = V + v \sin(\omega t + \phi_0).$$

Thus, during a small time interval, the trajectory of the Earth's revolution about the Sun is a fragment of a sinusoid. The entire trajectory of the Earth during a full revolution can be represented as successive fragments of a sinusoid "bent" into a circle. The number of "swings" performed by the Earth near the basic trajectory during a year is 1 year/$T_M \cong 13.5$. Therefore,

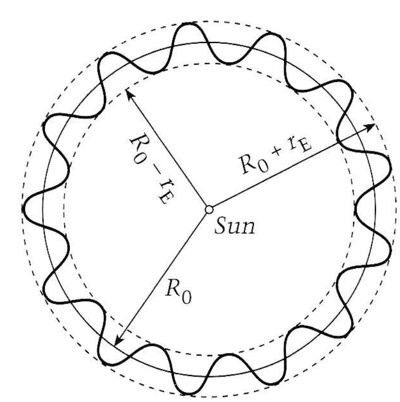


Figure 6

the Earth's CG will not arrive at the same position after traveling for one year.

We see that the Earth does not undergo repeated motion. Strictly speaking, there is no such value as a "period" for the Earth's revolution around the Sun. From the mathematical viewpoint, only the motion of the combined Earth-Moon system about the Sun is periodic. The trajectory of the Earth's CG is shown schematically in figure 6.

P287

When the diving bell just makes contact with the bottom of the pond, there is a thin layer of water separating it from the solid surface (figure 7). However thin, this layer of water provides two keys to solving the problem. First, the thin layer of water immediately determines the pressure in the water inside the bell. Second, it clears the way to apply Archimedes' principle, since the bell makes contact with only water.

Let us start with the latter approach. For the sake of simplicity, we assume that the thicknesses Δ of the walls and ceiling of the bell are equal and small in comparison with its radius and height. The weight of the water displaced by the bell is determined by the volume of the air layer ($2\pi r^2 x$), the side walls ($2\pi rh\Delta$), and the ceiling ($\pi(r+\Delta)^2\Delta \cong \pi r^2\Delta$). The necessary width Δ can be found from the condition that the bell's weight is greater than or equal to the buoyant force:

$$\rho_w\left(\pi r^2 x + 2\pi rh\Delta + \pi r^2\Delta\right)g$$
$$\le \rho_{st}\left(\pi r^2\Delta + 2\pi rh\Delta\right)g, \qquad (*)$$

where $\rho_w = 10^3$ kg/m^3 is the density of water, and g is the acceleration due to gravity.

Now let us find the thickness x of the layer of air left at the ceiling at the instant when the bell "lands" at the bottom of the pond. The air in the bell is compressed by a pressure $\rho_w g(H + H_0 - (h - x))$, and $\rho_w g H_0$ is the external atmospheric pressure ($H_0 = 10.3$ m). According to Boyle's law,

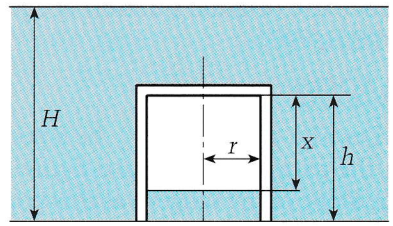


Figure 7

$$\rho_w g H_0 h\pi r^2 = \rho_w g(H_0 + H - (h - x))x\pi r^2,$$

whence

$$x^2 + (H_0 + H - h)x - H_0 h = 0,$$

and

$$x \cong 1.6 \text{ m}.$$

Finally, the inequality (*) yields the required value for the thickness of the wall:

$$\Delta = \frac{\rho_w r^2 x}{(\rho_{st} - \rho_w)\left(r^2 + 2rh\right)} \approx 4.6 \cdot 10^{-2} \text{ m}.$$

Another way to find the maximum value of the buoyant force is to use Pascal's law (this is the point where the thin film helps us again). The air inside the bell acts on the ceiling of area πr^2 with a force $\rho_w g(H_0 + H - (h - x))\pi r^2$. The water pressure applied via the thin film to the flange surface of the bell at the bottom of the pond is $\rho_w g(H_0 + H)2\pi r\Delta$. The ceiling of the bell with area $\pi(r + \Delta)^2 \cong \pi r^2 + 2\pi r\Delta$ is pressed downward by a force $\rho_w g(H_0 + H - h - \Delta)(\pi r^2 + 2\pi r\Delta)$. The buoyant force is equal to the difference of these upward and downward forces:

$$\rho_w g(H_0 + H - (h - x))\pi r^2 + \rho_w g(H_0 + H)2\pi r\Delta - \rho_w g(H_0 + H - h - \Delta)(\pi r^2 + 2\pi r\Delta) \cong \rho_w g(\pi r^2 x + (2\pi rh + \pi r^2)\Delta),$$

which agrees with the left-hand side of (*).

P288

If the conducting plate is shifted inside the capacitor without rotation (that is, if this plate is oriented

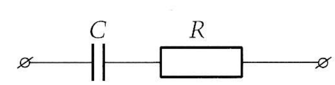


Figure 8

always parallel to the plates of the capacitor), the electric field inside the capacitor does not change. Let us shift it close to one of the plates. Figure 8 shows the equivalent scheme of this new arrangement. In this scheme the area of a plate is S, and the distance between the plates is $d - h$. The capacitor is charged to voltage V_0. The value of the series resistance is $R = \rho h/S$. The maximal current will flow in the first instant after closing the circuit:

$$I = \frac{V_0}{R} = \frac{V_0 S}{\rho h}.$$

P289

Evidently, the maximal width of the band is determined by the plasmoid diameter D. The magnification of the camera is

$$\Gamma = \left|\frac{F}{F-L}\right|,$$

so that $D = d/\Gamma$. To obtain an image on the film, the velocity $\mathbf{v}$ of a plasmoid must be sufficiently small for it to move a distance less than D during the time τ. If a time t elapses from the instant at which the image appears to the complete stop of the plasmoid, the plasmoid velocity at the instant when the image appears is $|\mathbf{v}| = |\mathbf{a}|t$, which satisfies the condition

$$\frac{D}{|\mathbf{a}|t} = \frac{d}{\Gamma|\mathbf{a}|t} = \tau. \qquad (1)$$

The plasmoid travels a distance

$$L = \frac{l}{\Gamma} = \frac{|\mathbf{a}|t^2}{2} \qquad (2)$$

before it stops completely.

Equations (1) and (2) yield the acceleration of the plasmoid:

$$|\mathbf{a}| = \frac{d^2}{2\Gamma l\tau^2} = \frac{d^2}{2l\tau^2}\left|\frac{F-L}{F}\right|.$$

P290

As a first step, let's find the radius of curvature of the hollow [it is the radius of the inner spherical surface (figure 9)]. From the right triangle, we obtain

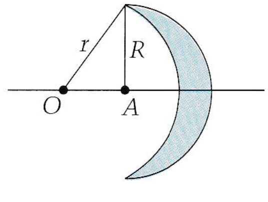


Figure 9

$$r^2 = R^2 + (r - R/2)^2,$$

whence

$$r = 1.25R.$$

Now let's draw a ray emitted by the light source (figure 10). For the sake of convenience, we draw the ray impinging on the spherical surfaces at a fairly large angle (otherwise we will not discern the details). In reality, the image is formed by rays which travel at very small angles to the principal optical axis (POA), because the pupil of an observer's eye is narrow. Moreover, according to the conditions of the problem, the eye is located at a large distance from the lens. Therefore, in this problem we use the conventional approximations for small-angle trigonometry: the values of the sine and tangent are equal to the angles themselves, provided that the latter are measured in radians.

Consider the path of a ray emitted at an angle α to the POA. It intersects the inner spherical surface at a distance $0.5R\alpha$ from the POA. Now draw a normal from the center of the inner spherical surface O to this intersection. This ray forms an angle β with the POA, which is related to the angle α by

$$0.5R\alpha = r\beta = 1.25R\beta,$$

$$\beta = 0.4\alpha.$$

The angle between this incident ray and the normal line is 0.6α, so after refraction at the surface of the glass with $n = 2$ it will form an angle 0.3α with the normal. At the same time, this ray forms an angle $0.4\alpha + 0.3\alpha = 0.7\alpha$ with the POA. The ray will hit the inner side of the outer spherical surface at a distance $0.5R\alpha + 0.5R \cdot 0.7\alpha = 0.85R\alpha$ from the POA. Let's draw the normal to the spherical surface at this point (it is

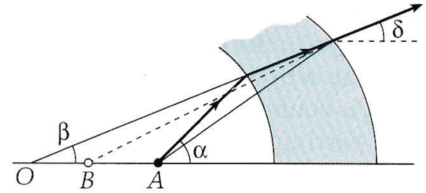


Figure 10

the radius drawn from point A, the center of this surface). The angle between this radius and the POA is $\gamma = 0.85R\alpha/R = 0.85\alpha$, so that the angle of incidence is $\gamma - 0.7\alpha = 0.15\alpha$. After refraction the angle is doubled and becomes 0.3α. The refracted ray makes an angle $\delta = \gamma - 0.3\alpha = 0.55\alpha$ with the POA. The continuation of this ray intersects the POA at a point B at distance $L = 0.85R\alpha/0.55\alpha = 17R/11 \cong 1.55R$ from the point of emergence of the ray (for small angles, this is the point of intersection of the outer spherical surface with the POA). Since we chose a small arbitrary angle of incidence for a ray traveling from the source to the lens, the position of this point does not depend on the value of this angle (after refraction a narrow beam of light appears to be emitted from this point). Therefore, we have found the location of the image observed by an eye.

Brainteasers

B286

The 500,000th place is occupied by the digit 5. Indeed, to write the numbers 1 through 99,999 we need $9 + 2 \cdot 90 + 3 \cdot 900 + 4 \cdot 9000 + 5 \cdot 90000 = 488{,}889$ digits. It remains to find the digit number $500{,}000 - 488{,}889 = 11{,}111$ among the 6-digit numbers. We notice that $11{,}111 = 6 \cdot 1851 + 5$. Therefore, the desired digit is the fifth one in the numeral 101,851. Writing down all these numbers, at the rate of one per second, would take almost six days!

B287

Denote the number of students who participated in only one event by $6x$. Then $3x$ students participated in three events and $2x$ students in two events. If we add all the numbers given in the problem, then each student who participated in one event is counted once, every student who participated in two events is counted twice, and everyone who participated in three events is counted three times. Thus, we have the equation $100 + 50 + 48 = 6x + 6x + 6x$, which yields $x = 11$. From this, we find that the number of participants is $6x + 3x + 2x = 11x = 121$.

B288

The hot plate with the most power will cook the cutlets the fastest. The current through the upper two hot plates is given by Ohm's law:

$$I_u = \frac{V}{R_u}.$$

Because the current through both hot plates is the same, the right-hand hot plate has the greater power,

$$P_u = I_u^2 R_{ur} = \frac{V^2}{R_u^2} R_{ur}$$

$$= \frac{V^2}{(30\ \Omega)^2}(20\ \Omega) = \frac{2V^2}{90\ \Omega}.$$

The same calculation for the bottom pair of hot plates yields

$$P_l = \frac{2V^2}{245\ \Omega}.$$

Therefore, the hot plate with the 20-Ω resistance will cook the cutlets the fastest.

B289

We assume that the steps in each staircase are identical (an assumption forced on most architects by their lawyers, because if the steps are uneven people tend to trip and fall). Since the staircases have the same slope, each step in one staircase is the same length and height as each step in the other staircase (the ratio of the height to the length will be the common slope). Since the towers are the same height, the two staircases must have the same number of steps. Therefore they must be the same length. Note that we did not use the fact that the towers were circular. The solution holds for any two staircases for which the word 'slope' is meaningful and which have the same slope.

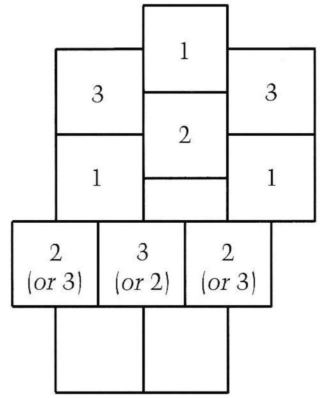


Figure 11

B290

One such arrangement of squares is shown in the figure 11. Let's prove that this set of squares satisfies the condition of the problem. Assume the opposite. If we paint the two upper middle squares in colors 1 and 2 (as shown in figure 11), we automatically obtain the coloring for four more squares (the two on the left and the two on the right of the first two we've colored). Now we see that the next layer of squares can be painted only in colors 2 and 3. However we choose these colors, we cannot paint the two bottom squares as required.

Bulletin Board

A not-so-smooth ascent for some

There were a few missteps taken by some entrants trying to solve this month's CyberTeaser involving two cylindrical staircases. However, many more were able to overcome the treacherous ascent the problem posed. The following were the first ten correct solutions to the problem:

Helio Waldman (Campinas, Brazil)
John E. Beam (Bellaire, Texas)
Jerold Lewandowski (Troy, New York)
Rick Armstrong (St. Louis, Missouri)
Vincze Zsombor (Szeged, Hungary)
Maxim Bachmutsky (Kfar-Saba, Israel)
Theo Koupelis (Wausau, Wisconsin)
Bruno Konder (Rio de Janeiro, Brazil)
Mirella Murad (Curitiba, Brazil)
Michael H. Brill (Morrisville, Pennsylvania)

Our congratulations to this month's winner, who will receive a copy of this issue of *Quantum* and the coveted *Quantum* button. Everyone who submitted a correct answer (up to the time the answer is posted on the Web) is entered into a drawing for a copy of *Quantum Quandaries*, a collection of 100 *Quantum* brainteasers. Our thanks to everyone who submitted an answer—right or wrong. The new CyberTeaser is waiting for you at http://www.nsta.org/quantum.

A summer of science

The Weizmann Institute of Science in Israel is opening its doors and laboratories to gifted high school seniors who are graduating this June for a one-month summer program at its Rehovot campus. Since its establishment in 1969, each summer the Institute has brought together 75 talented science students from around the world to experience the challenges and rewards of scientific research.

American participants in the Institute receive a full scholarship, including travel to and from Israel. For information and an application, contact Debbie Calise, American Committee for the Weizmann Institute of Science, 130 East 59th Street, New York, NY 10022; telephone: (212) 895-7906; email: debbie@acwis.org.

Langton's ant

by Don Piele

MANY OF YOU HAVE HEARD ABOUT, OR perhaps played with, John Conway's "Game of Life." It first appeared in Martin Gardner's "Mathematical Recreations" column in *Scientific American* in October, 1970. When I read this column, I was astounded at the variety of life forms that could evolve from such a simple set of rules. Furthermore, it seemed impossible to predict what would evolve knowing the rules and the life form at the first generation. Such is the nature of the world of Cellular Automata. In this column we will explore a small corner of this world through the eyes of Langton's Ant, created by Chris Langton of the Sante Fe Institute.

The rules

Imagine that you are an ant in the middle of a large square array of cells—such as a checker board. Imagine that all the cells are colored white. You have a black and a white magic marker and you are facing North. Your instructions are as follows:

1. Move forward one cell;
2. If the cell you are on is colored white, color it black and turn right 90 degrees;
3. If the cell you are on is black, color it white and turn left 90 degrees;
4. Repeat steps 1-3.

Can you predict what pattern of black cells will evolve on the board? Will any identifiable pattern evolve or will it just appear as a random array of black dots? To find the answers to these questions, let's write the necessary algorithms and build a program in *Mathematica* to carry it out.

Translating a set of rules that we understand in the real world into a set of procedures a computer can carry out is the art of programming. Programming requires, first, that you know the data structures available to you in your computer language and, second, how to build algorithms that solve problems within those structures.

Let's start by naming the board **antland** and defining it as a square matrix of cells each set initially to 0 for white.

```
size = 10;

antland = Table[0, {size}, {size}];
```

```
MatrixForm[antland]
```

$$\begin{pmatrix} 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0&0 \end{pmatrix}$$

Let's put the ant initially set at $\{x, y\}$ near the center of antland.

$$\mathbf{x = y = Floor}\left[\frac{\mathbf{size}}{\mathbf{2}}\right]$$

The state of each position x, y is the value of antland[[x, y]]. Of course, initially antland[[x, y]] = 0 for all x, y.

Changing states

If the ant lands on white (0), it must be colored black (1) and vice versa. Notice if we take {0, 1} and add 1 to each position we get {1, 2}. Now, if we take the result mod 2, we end up with {1, 0}, a reversal of values. Thus we have our first algorithm for changing states of a square the ant is on: Add 1 to the current state of the cell and compute the value mod 2.

```
antland[[x, y]] = Mod[antland[[x, y]] + 1, 2]
```

Enter the above command several times and watch the state switch from 0 to 1 and back to 0.

Changing directions

The direction the ant is facing determines where it will be when it moves forward one cell. We number the four directions as follows: North = 0, East = 1, South = 2, and West = 3. Initially, we have the ant facing North or direction = 0. Let's examine how the directions change after a right turn. Start with {N, E, S, W} = {0, 1, 2, 3} and do a right turn. Clearly N -> E, E -> S, S -> W, and W -> N or {0, 1, 2, 3} -> {1, 2, 3, 0}. This is exactly the same result you would get if you added one to {0, 1, 2, 3} to get {1, 2, 3, 4} and took the results mod 4, {1, 2, 3, 0}. It is easy to see that a left turn is the result of subtracting 1 and computing the result mod 4. Assuming **dir** is the existing direction number, the following commands, **rightTurn** and **leftTurn**, change the direction **dir** in the correct way.

```
rightTurn := dir = Mod[dir + 1, 4]
```

```
leftTurn := dir = Mod[dir - 1, 4]
```

```
dir = {0, 1, 2, 3};
```

```
rightTurn; dir
```

```
{1, 2, 3, 0}
```

Moving forward

Next we need to work out the movement forward algorithm. Depending upon the direction we are facing, a move forward from {x,y} has a different result. Here are the transformations for each direction: North $\{x, y\} \to \{x, y+1\}$, East $\{x, y\} \to \{x+1, y\}$, South $\{x, y\} \to \{x, y-1\}$, West $\{x, y\} \to \{x-1, y\}$. In *Mathematica* we use the **Switch** command to change the position $\{x, y\}$ depending on the direction **dir** the ant is facing. Note that **x++** increases x by 1 and **x--** decreases x by 1. If **dir** = 0, then **y++**; if **dir** = 1 then **x++**; if **dir** = 2 then **y--**; if **dir** = 3 then **x--**. This is expressed in the command

```
Switch[dir, 0, y++, 1, x++, 2, y--, 3, x--];
```

Note: It is true that an array in *Mathematica* starts with {1, 1} in the Northwest corner of the array and not in the usual Southwest corner. But for purposes of viewing the pattern we will use a Graphics command that takes this into account. Let's assume the usual x, y coordinate system where going North increases y and going South decreases y.

Putting the pieces together

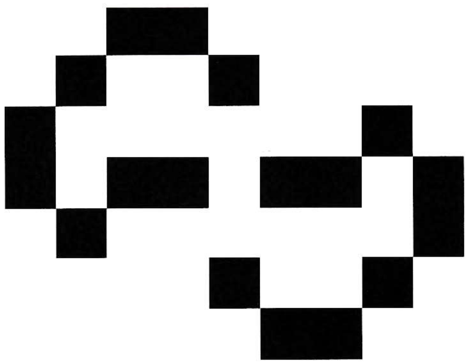
Let **size** be the dimension of **antland** on each side. Let's agree that **dir** always begins facing North. Let **tourLength** be the total number of steps the ant will take, and let **moves** be the current number of moves taken. Continue making moves as long as the ant is still on the board and the number of moves is less than the tourLength. Here is a module called *LangtonsAnt* that puts all the pieces together.

```
<< Graphics`Colors`
LangtonsAnt[tourLength_, size_] :=
 Module[{dir = 0, moves = 0, x = y =
Floor[size/2]},
      antland = Table[0, {size}, {size}];
      rightTurn := dir = Mod[dir + 1, 4];
      leftTurn := dir = Mod[dir - 1, 4];
      While[moves < tourLength && 1 <= x
<= size && 1 <= y <= size,
         antland[[x,y]] =
Mod[antland[[x,y]] + 1, 2];
         If[antland[[x,y]] == 1, leftTurn,
rightTurn];
         Switch[dir, 0, y++, 1, x++, 2, y--,
3, x--];
      moves++];

Show[Graphics[RasterArray[antland] /.
 {0 -> White, 1 -> Black}], AspectRatio ->
Automatic]]
```

Here is the final state of antLand after 96 moves.

```
LangtonsAnt[96, 11]
```



Big moves

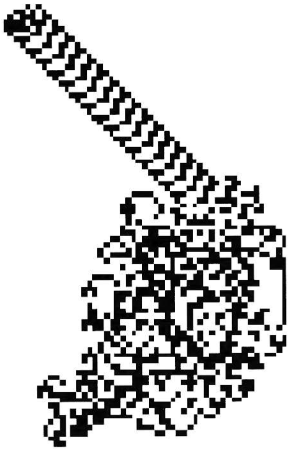
After 10,000 moves of Langton's Ant on an array of size 100, nothing interesting seems to be happening.

```
LangtonsAnt[10000, 100]
```

But going a bit further we see something developing.

```
LangtonsAnt[11000, 100]
```

```
LangtonsAnt[12000, 100]
```

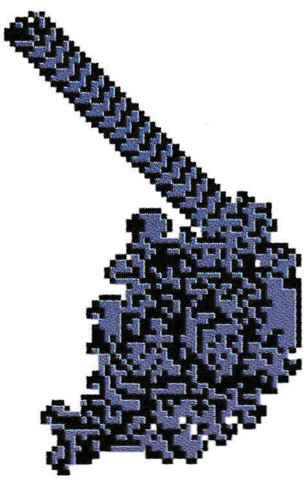


The ant has now settled into a pattern, building a highway to infinity and beyond. Could you have predicted that? That's the nature and the beauty of cellular automata.

Your turn

1. As a first exploration, see if you can make a one line change to the LangtonsAnt program in order to be able to view all the cells the ant visits. As it now stands, you see only the cells the ant visits an odd number of times. Color the even number of times blue, the odd black, and unvisited squares white. It will look like:

```
LangtonsAnt1[12000, 100]
```



2. Modify the rules in problem 1 so that once an ant steps on a blue cell, it does not change color and it moves forward on the next step instead of turning. The results are completely different and unexpected.
3. Explore making up your own rules and implementing them. I'm interested in any unusual patterns you discover.

2000 USA Computing Olympiad

Over 150 students from 31 countries took part in the USACO Fall Internet competition in November, 1999. In this programming competition, the problems are sent out via email to students who subscribe to the USACO listserv at majordomo@delos.com.

The top four students in the Senior Division of the Fall Internet competition were: Percy Liang, USA; Omid Etesami and Siamak Tazari, Iran; and Jing Xu, China. All four students had perfect scores on all four problems.

Three Internet competitions are held each year in November, January, and March. The USACO National competition will be held at local high schools in the United States on April 12, 2000. To be included in any of these competitions, send an email to me at piele@uwp.edu, or visit our website at www.usaco.org.

Finally

Waiting for two months to see a solution is not necessary today thanks to the Internet.Therefore, all solutions to the problems presented in this column are available at the Informatics website: http://www.uwp.edu/academic/mathematics/usaco/informatics/. Send your solutions to me at piele@uwp.edu.

To participate fully in this column, you will need to have access to Mathematica. Readers who are students in any school or college may purchase the student edition of Mathematica. For details go to http:// www.wolfram.com/products/student/.

Index of Advertisers