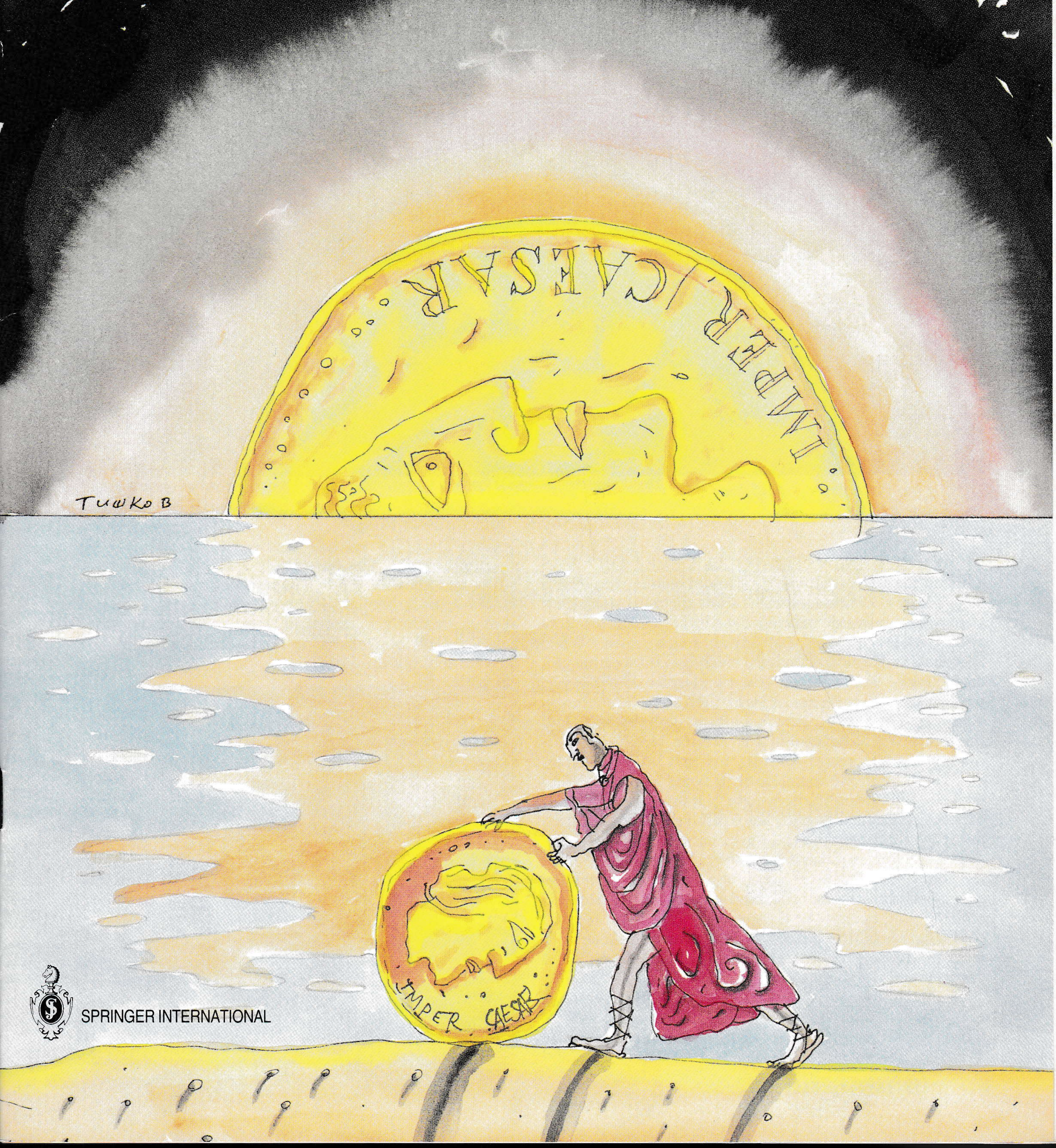
QUANTUM

MAY/JUNE 1994

$5.00



SPRINGER INTERNATIONAL

Photo by Tom Munk, courtesy of University Relations, Southern Illinois University at Carbondale (SIUC)

Geodesic Dome by R. Buckminster Fuller

IF ALL BUCKMINSTER FULLER EVER DID WAS invent the geodesic dome, his place in history would have been assured. It is considered by some the most significant structural innovation of the 20th century. It encloses more space with less material than any alternative form. "When I invented and developed my first clear-span, all-weather geodesic dome," Fuller wrote, "the two largest domes in the world were both in Rome and were each about 50 meters in diameter. They are St. Peter's, built around A.D. 1500, and the Pantheon, built around A.D. 1. Each weighs approximately 15,000 tons. In contrast, my first 50-meter-diameter geodesic dome installed in Hawaii weighs only 15 tons—one thousandth the weight of its masonry counterpart. An earthquake would tumble both the Roman domes, but it would leave the geodesic unharmed."

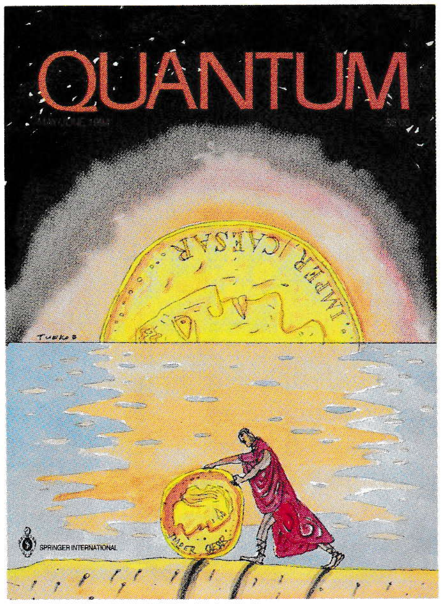
Fuller coined the word "tensegrity" for the continuous tension/discontinuous compression structural system he developed from an idea he learned from a sculptor. He soon realized that, because of the greater efficiency of tension compared to compression, very large domes could be built with his tensegrity trusses. He calculated that a 3-km dome would weigh only 4,000 tons.

In the photograph above, Fuller is standing in front of the 76-m dome that housed the US pavilion at the world's fair in Montreal in 1967. Not all geodesic domes are quite so monumental. The photo below shows Buckminster Fuller and his wife, Anne Hewlett Fuller, in their geodesic home in Carbondale, Illinois. Turn to page 8 for a look at even smaller geodesic structures: hollow carbon molecules called "buckyballs" in honor of this wide-ranging, forward-looking thinker.

Photo by Photographic Services, courtesy of University Relations, Southern Illinois University at Carbondale (SIUC)

QUANTUM

MAY/JUNE 1994 VOLUME 4, NUMBER 5



Cover art by Leonid Tishkov

Some of us can recall, with a twinge of shame, a time when it was oh-so-easy to make a bit of money. All you had to do was strike a little deal with a younger kid: "Here, I'll trade you this nice, big nickel for that little dime of yours." It was a nice scam—for a while, at least. Either the kid got wise, or someone older intervened—with dire consequences for the clever currency trader.

Who can blame the four-year-old for thinking a nickel is worth more than a dime? Only an adult would think of making a bigger coin less valuable than a smaller one. It all but invites fraud in the seven-year-old mind. Of course, grown-ups aren't immune to psychological mishaps when it comes to money. In fact, the noble Roman on our cover has fallen prey to a cunning emperor. Just how cunning? You be the judge. The story begins on page 16.

FEATURES

DEPARTMENTS

PUBLISHER'S PAGE

Numbers in our genes

The role of quantification in molecular biology

IN A RECENT EDITORIAL IN the British journal *Nature* (Vol. 368, 10 March 1994), John Maddox makes an important point regarding the present descriptive character of molecular biology. Maddox observes that in the human genome projects, the main goal is to list the genes and "to specify their nucleotide sequences." The projects also try to specify "the sequences of regions of DNA that hold the genes together in the chromosome."

These activities are, first of all, ones of identifying structure and giving names to that structure. Second, researchers make connections between these structures and other structures or characteristics of the organism that are inherited through the gene. Still, these processes are inherently descriptive. They do not rely on *quantitative* relationships. As description, we can tell what the results will be, and we can give names to genes and sequences of regions of DNA. But we cannot make quantitative predictions. Furthermore, qualitative predictions that fail to consider underlying quantitative variables may well be wrong.

Most important, regardless of how well descriptive molecular biology may tell *what* happens, it does not tell *why* it happens. This can only occur if we can apply laws of science that are quantitative to the situation. Maddox uses as an example a virus that infects *E. coli*, called bacteriophage λ, and a *repressor* protein produced by one of the viral genes. A second gene, called *Cro*, will bind to the same DNA site if there is no repressor protein present and repress the activity of the repressor. This is called a switching mechanism. In examining this situation, researchers have found that there are only about 100 or so free molecules in the cell under consideration. This means that there are fluctuations of macroscopic laws of equilibrium thermodynamics that are not accounted for. Maddox speculates that the genetic switch may in fact be a kinetic phenomenon, and that "the energetic implications of what appears to be equilibrium constants may be spurious."

As Maddox points out, "The naming of parts does not in itself yield understanding." He goes on to list the problems yet to be resolved: "how the molecules of *repressor* fold into their characteristic dumbbell shape, why dimers are so much more stable than monomers, and how an alpha helix in the amino-component interacts with DNA at the binding sites. To be more precise, *what* happens has been determined by elegant genetic experiment; *why* that, not something else, happens remains to be discovered."

In this brief overview I haven't done justice to Maddox's editorial. If you have read some molecular biology, I recommend that you read the essay in its entirety. Here, I simply wanted to give a sketch of his ideas to suggest the importance of two aspects of science that aren't often associated with molecular biology. One is the use of quantity, symbols, and equations; the other is their application in fundamental laws of science, such as thermodynamics and kinetic theory.

I have noticed a tendency to downplay the "hard sciences" on the grounds that the really exciting areas of research are in molecular biology. Yet even if that is where the "excitement" is, you will come to a dead end if you fail to utilize physics and chemistry. You can't get to the heart of natural phenomena if you don't understand the basic laws and principles of science that underlie those phenomena.

Readers of *Quantum* magazine continually see the unusual and exciting ways in which such fundamentals lead to a more profound understanding of varied phenomena in biology as well as in other areas of science. Recall, for instance, "Mathematics in Living Organisms" (November/December 1992), where cats are shown to be handy with logarithms; or "Trees Worthy of Paul Bunyan" (November/December 1993), where physical processes affecting plant growth are explored. Regardless of your interests—but especially if you intend to go into the life sciences—continue to study physics and chemistry and acquire the mathematical tools needed for research in all the sciences. We hope *Quantum* will help keep you on that productive path.

—Bill G. Aldridge

Be a factor in the QUANTUM equation!

Have you written an article that you think belongs in *Quantum*? Do you have an unusual topic that students would find fun and challenging? Do you know of anyone who would make a great *Quantum* author? Write to us and we'll send you the editorial guidelines for prospective *Quantum* contributors. Scientists and teachers in any country are invited to submit material, but it must be written in colloquial English and at a level appropriate for *Quantum*'s predominantly student readership.

Send your inquiries to:

Managing Editor
Quantum
1840 Wilson Boulevard
Arlington, VA 22201-3000

What's happening?

Summer study ... competitions ... new books ... ongoing activities ... clubs and associations ... free samples ... contests ... whatever it is, if you think it's of interest to *Quantum* readers, let us know about it! Help us fill Happenings and the Bulletin Board with short news items, firsthand reports, and announcements of upcoming events.

What's on your mind?

Write to us! We want to know what you think of *Quantum*. What do you like the most? What would you like to see more of? And, yes—what *don't* you like about *Quantum*? We want to make it even better, but we need your help.

What's our address?

Quantum
National Science Teachers Assoc.
1840 Wilson Boulevard
Arlington, VA 22201-3000

Be a factor in the QUANTUM equation!

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 National Science Teachers Association is an organization of science education professionals and has as its purpose the stimulation, improvement, and coordination of science teaching and learning.

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York Inc. Volume 4 (6 issues) will be published in 1993–1994. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau of the Russian Academy of Sciences, as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington, VA 22201, telephone (703) 243-7100. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010-7858. Supervising Production Editor: Madeline R. Kraner.

Advertising:

Advertising Representatives: (Washington) Paul Kuntzler (202) 328-5800; (New York) Patrick Ferencz (212) 460-1575; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, FRG, telephone (0) 30-82 07-1, telex 185 411.

Second class postage paid at New York, NY, and additional mailing ofices. **Postmaster:** send address changes to: *Quantum* Springer-Verlag New York, Inc., Journal Fulfillment Services Department, P. O. Box 2485, Secaucus, NJ 07096-2485.

Subscription Information: North America

Student rate: $15; Personal rate (nonstudent): $20; Institutional rate: $34; Single Issue Price: $5. 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. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus, NJ 07096-2485; or call 1-800-SPRINGER (in New York, call [201] 348-4033).

All Countries Outside North America

Subscription rates in U.S. currency as above (all rates calculated in DM at the exchange rate current at the time of purchase) plus postage and handling. SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Air mail delivery to all other countries is available upon request. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, FRG.

Art by Sergey Ivanov

Three paths to Mt. Fermat-Euler

Let Lagrange, Zagier, and Minkowski be your guides

by Vladimir Tikhomirov

LOOK AT THE FIRST FEW prime numbers greater than 2: 3, 5, 7, 11, 13, 17, 19, ... The numbers 5, 13, and 17 can be represented as the sums of two squares—

$$5 = 1^2 + 2^2,$$
$$13 = 2^2 + 3^2,$$
$$17 = 1^2 + 4^2$$

—while the other four numbers (3, 7, 11, 19) cannot. (Check it yourself!) Is there any way to tell one sort of number from the other without a brute-force search through all possible breakdowns? And how can this difference be explained? The answer is given by the following theorem.

THEOREM. *A prime number greater than 2 is representable as the sum of two squares if and only if its remainder upon division by four is one.*

(Indeed, $5 = 4 \cdot 1 + 1$, $13 = 4 \cdot 3 + 1$, $17 = 4 \cdot 4 + 1$, whereas $3 = 4 \cdot 0 + 3$, $7 = 4 \cdot 1 + 3$, $11 = 4 \cdot 2 + 3$, ...)

Problem 1. Prove the "only if" part of the theorem: any prime (except 2) equal to the sum of two squares can be written as $4n + 1$ for some integer n.

Who first discovered this mathematical phenomenon? There is evidence that not long ago we could have celebrated the 350th anniversary of this remarkable result. On Christmas Day in 1640, the great Pierre Fermat (1601–1665) wrote a letter to the renowned Mersenne, a faithful friend of Descartes and the main intermediary in the correspondence of scientists of that time. He informed Mersenne that "any prime number that yields a remainder of one when divided by four is uniquely representable as the sum of two squares."[1] At that time there were no mathematical journals, so mathematicians exchanged information by mail. In general they simply announced their results and didn't include any proofs.

However, almost 20 years after writing to Mersenne, Fermat described his plan of attack in proving the theorem presented above. In a letter to Carcavy[2] sent in August 1659, he writes that his proof is based on the method of infinite descent. Starting from the assumption that the conclusion of the theorem is not valid for a certain prime of the form $4n + 1$, he proves that it must be wrong for some smaller number and proceeds all way down to the number 5, thus arriving at a contradiction (since the theorem is true for 5).

The first complete proofs were found by Leonhard Euler (1707–1783) between 1742 and 1747. Euler held Fermat in the highest esteem and, ceding priority to his predecessor, created a proof that elaborated the idea in Fermat's letter. Giving credit to both great scholars, we now call this statement the Fermat–Euler Theorem.

There is a feature inherent in almost any beautiful mathematical result (as well as almost any beautiful but forbidding summit): many paths lead to it. We can approach it from different sides, and all the paths give sheer delight to those who aren't afraid to take them.

The Fermat–Euler Theorem vividly displays this wonderful feature, and I'm going to demonstrate this below.

We'll ascend to this peak, discovered in the 17th century, in three different ways. One of them was

[1]In this article we won't touch on uniqueness, which was established long before Fermat. See the solution to problem M115 in this issue.—*Ed.*

[2]After his death, Mersenne's role as a scientific intermediary was played by the Royal Librarian, an amateur mathematician and Fermat's friend Pierre de Carcavy (d. 1684).—*Ed.*

Art by Sergey Ivanov

found in the 18th century, another in the 19th century, and the third only recently, in our own century.

Lagrange's proof

The first proof (with certain modifications) is given in almost every textbook on number theory. It's based on the following lemma.

WILSON'S LEMMA. *For any prime p the number $(p-1)!+1$ is divisible by p.*

In order not to digress for the proof of this auxiliary (but very useful) fact, I'll demonstrate its main idea using the prime number 13 as an example. For any integer from 2 through 11, let's find the factor whose product with this integer gives a remainder of 1 when divided by 13, and collect such pairs of factors in the factorization of $(13-1)!$ together:

$$(13-1)! = 12!$$
$$= (2\cdot 7)(3\cdot 9)(4\cdot 10)(5\cdot 8)(6\cdot 11)\cdot 12$$

(where $2\cdot 7 = 14 = 13+1$, $3\cdot 9 = 27 = 2\cdot 13+1$, $4\cdot 10 = 5\cdot 8 = 40 = 3\cdot 13+1$, $6\cdot 11 = 66 = 5\cdot 13+1$. It follows that the remainder of 12! upon division by 13 is 12—that is, $12!+1$ is divisible by 13. The general case is treated likewise.

Problem 2. Prove that for any prime p the integers $2, 3, \ldots, p-2$ can be paired so that the product of the numbers in each pair (a, b) gives a remainder of 1 when divided by p. (This is written as $ab \equiv 1 \pmod p$.) Use this fact to prove Wilson's lemma in the general case.

From Wilson's lemma we derive a corollary.

COROLLARY. *If the number $p = 4n+1$ is prime, then $[(2n)!]^2+1$ is divisible by p.*

To prove it, we rewrite $(p-1)!+1$ as $(4n)!+1 = 1\cdot 2\cdot\ldots\cdot 2n\cdot(2n+1)\cdot\ldots\cdot(4n)+1 = 1\cdot 2\cdot\ldots\cdot 2n\cdot(p-2n)\cdot(p-2n-1)\cdot\ldots\cdot(p-1)+1 \equiv (2n)!\cdot(-1)^{2n}(2n)!+1 \equiv [(2n)!]^2+1 \pmod p$, and note that the left side is divisible by p.

Denote $(2n)!$ by N. Then our corollary means that $N^2 \equiv -1 \pmod p$. Now we have to overcome the major difficulty.

Consider all the pairs (k, m) of nonnegative integers that are no larger than $\sqrt{p}$. The largest of the numbers k (or m) is greater than $\sqrt{p}-1$, so the number of such pairs is greater than $[(\sqrt{p}-1)+1]^2 = p$. Therefore (by the pigeonhole principle applied to the pairs as pigeons and the remainders modulo p as pigeonholes),[3] there are at least two different pairs (k_1, m_1) and (k_2, m_2) such that the remainders of k_1+Nm_1 and k_2+Nm_2 when divided by p are the same. Then $a+Nb$ (where $a = k_1-k_2$, $b = m_1-m_2$) is divisible by p. Note that $|a| < \sqrt{p}$ and $|b| < \sqrt{p}$.

Now, $a^2-N^2b^2 = (a-Nb)(a+Nb)$ is divisible by p, and since $N^2 \equiv -1 \pmod p$, a^2+b^2 is also divisible by p—that is, $a^2+b^2 = rp$ for some positive integer r ($r \neq 0$, because otherwise the two pairs above would be the same). It remains to note that $a^2+b^2 < 2p$, which means that $r = 1$ and $a^2+b^2 = p$, completing the proof.

Zagier's proof

Another proof, which is due to the contemporary mathematician D. Zagier, completely stunned me: the result seems to emerge miraculously, out of thin air.

Our goal in reproducing Zagier's proof will be to show that for any prime p of the form $p = 4n+1$ the equation in positive integers

$$x^2+4yz = p$$

has a solution (x, y, z) with $y = z$. This would yield the representation $p = x^2+4y^2 = x^2+(2y)^2$, which proves the theorem.

We'll prove the existence of such a solution in a rather bizarre way: by proving that the (obviously finite) *number of solutions to the above equation is odd.* How is the oddness connected to the existence of the solutions we need? All the solutions with $y \neq z$ can be arranged in pairs by swapping y and z: if (x, y, z) is a solution, then (x, z, y) is a solution as well. So the number of these solutions should be even, and the total number can be odd only if there's a solution with $y = z$.

The modern way to articulate this reasoning is to consider the transformation J of the set S of all positive integer triples satisfying our equation that swaps y and z—$J(x, y, z) = (x, z, y)$—and note that, first, it's an *involution*—that is, when applied twice it takes us back to the start; second, its *fixed points* $(x, y, z) = J(x, y, z)$ supply the required decomposition of p (since they are characterized by $y = z$); and third, the number of points that aren't fixed is even, because they can be arranged in pairs such that either element of each pair is the image of the other element. Of course, the last statement holds for *any* involution of *any* finite set.

And now let's consider the transformation B of triples (x, y, z) defined as follows: $B(x, y, z) = (x', y', z')$, where

(1) for $x < y-z$: $x' = x+2z$, $y' = z$, $z' = y-x-z$;
(2) for $y-z \le x \le 2y$: $x' = 2y-x$, $y' = y$, $z' = x-y+z$;
(3) for $x > 2y$: $x' = x-2y$, $y' = x-y+z$, $z' = y$.

Like J, this transformation considered on the set S is also an involution of S. First of all, it maps the set S into itself, because it preserves the value x^2+4yz. Indeed, take case 1, for instance:

$$x'^2+4y'z' = (x+2z)^2+4z(y-x-z)$$
$$= x^2+4xz+4z^2+4zy-4zx-4z^2$$
$$= x^2+4yz.$$

Verification in the other two cases is just as straightforward. Further, if $(x', y', z') = B(x, y, z)$, then $B(x', y', z') = (x, y, z)$. This is also verified by direct calculation. For instance, if $x < y-z$, we must apply the equations in case 1: they yield $x' = x+2z > 2z = 2y'$, so $(x'', y'', z'') = B(x', y', z')$ must be computed using the equations in case 3, and we get

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

Examination of the other two cases is left to the reader. After such verification, we conclude that B is an involution of S.

What about the fixed points of B?

[3]For an explanation of the pigeonhole (or Dirichlet) principle, see "Pigeons in Every Pigeonhole" in the January 1990 issue of *Quantum*.—*Ed.*

Looking at the definition, we see that in cases 1 and 3, $x' > x$ or $x' < x$, respectively, so a fixed point can arise only in case 2, which yields $x = x' = 2y - x$, or $x = y$.

Conversely, you can see at once that any triple of the form (x, x, z) is preserved under B. But only one of these triples belongs to the set S of positive integer solutions to our equation: if $p = x^2 + 4xz = x(x + 4z)$, then $x = 1$ (since p is a prime) and $z = n$ (recall that $p = 4n + 1$). Thus, the involution B of the set S has a unique fixed point $(1, 1, n)$, and therefore, as we've seen, S consists of an odd number of triples, which is what we set out to prove.

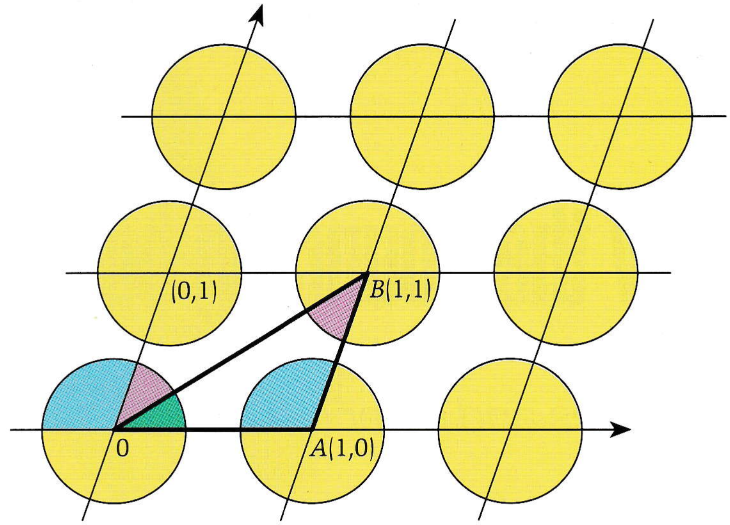


Figure 2

Minkowski's proof

The (slightly modified) proof by Hermann Minkowski (1864–1909), which I'm going to present now, staggers the imagination perhaps even more.

Minkowski's proof begins with a result that doesn't seem to have anything to do with the Fermat–Euler Theorem.

THEOREM. *Let a, b, and c be any integers satisfying $a > 0$ and $ac - b^2 = 1$. Then the equation $ax^2 + 2bxy + cy^2 = 1$ has an integer solution (x, y).*

Proof. The expression $ax^2 + 2bxy + cy^2 = 1$ can be viewed as the square of the distance from the origin O to the point P with coordinates (x, y) in a certain coordinate system (not necessarily rectangular). To construct such a system, draw the axes at the angle α defined by $\cos\alpha = b/\sqrt{ac}$ (this is possible because $ac > 0$ and $|b/\sqrt{ac}| < 1$, since $ac = b^2 + 1 > b^2$). Choose the units of scale on the x- and y-axes equal to $\sqrt{a}$ and $\sqrt{c}$, respectively (fig. 1). Then the square

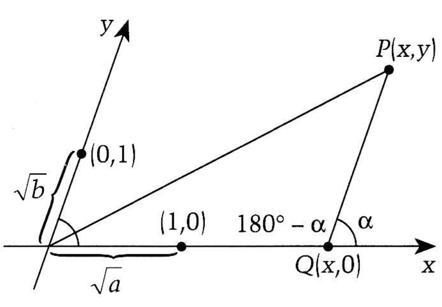


Figure 1

of the distance OP, $P = (x, y)$, can be found from the triangle OQP, where $Q = (x, 0)$: in this triangle, $OQ = |x|\sqrt{a}$, $QP = |y|\sqrt{c}$, and the angle at Q is α or $180° - \alpha$, depending on the signs of x and y. However, no matter what these signs are, the Cosine Law always yields

$$OP^2 = OQ^2 - 2OQ\cdot OP\cos(\angle Q) + OP^2 = ax^2 + 2bxy + cy^2.$$

The points with integer coordinates form the integer grid with respect to our coordinate system (fig. 2), and we have to prove that there is a node in the grid at a unit distance from the origin.

Let d be the smallest distance from the origin O to another node, and let (m, n) be a node at a distance d from O. Since the distance from (x, y) to (x_1, y_1) is equal to the distance from $(0, 0)$ to $(x_1 - x, y_1 - y)$, the distance between any two nodes is no smaller than d. Therefore, the circles of radius $d/2$ centered at all the nodes of our grid do not overlap: if two such circles, with centers A and B, have a common interior point C, then $AB < AC + CB < d/2 + d/2 = d$. As is clear from figure 2, the area covered by these circles in the triangle with vertices $O(0, 0)$, $A(1, 0)$, and $B(1, 1)$ is half the area of one circle–that is, $\pi d^2/8$. And this is only a part of the triangle's area, which equals

$$\frac{1}{2}OA\cdot OB\cdot\sin(\angle A) = \frac{1}{2}\sqrt{a}\sqrt{c}\sin(\angle A)$$
$$= \frac{1}{2}\sqrt{ac\left(1-\cos^2\alpha\right)}$$
$$= \frac{1}{2}\sqrt{ac-b^2}$$
$$= \frac{1}{2}.$$

So $\pi d^2/8 < 1/2$, or $d^2 < 4/\pi < 2$. Since d^2 is an integer ($d^2 = am^2 + 2bmn + cn^2$), we get $d = 1$, which proves Minkowski's theorem.

But what relationship does this marvelous theorem have to Fermat and Euler? The most direct!

By the corollary of Wilson's lemma proved above, we know that the number $N^2 + 1$, where $N = [(p - 1)/2]!$, is divisible by p, don't we? Well, now let's apply Minkowski's theorem to the numbers $a = p$, $b = N$, $c = (b^2 + 1)/a$. The theorem says that for certain integers m and n

$$1 = am^2 + 2bmn + cn^2,$$

so

$$p = a = a^2m^2 + 2abmn + (b^2 + 1)n^2 = (am + bn)^2 + n^2$$

—that is, p is the sum of two squares. And, once again, the theorem is proved!

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

Follow the bouncing buckyball

On fullerenes and other carbon structures

by Sergey Tikhodeyev

EVERYBODY KNOWS THAT carbon is one of the most common elements. But did you know that carbon atoms are a first-rate building material for constructing a wide variety of crystals and molecules? The record in the hardness department belongs to diamond, which is one of the crystalline forms of carbon. The complex organic molecules known as proteins—whose atomic "skeletons" are atoms of carbon and nitrogen—form the basis of all living things.

The great variety of atomic structures made of carbon is due to the fact that carbon—an element of group IV of the periodic table—has four electrons in its outer valence shell and can form valence bonds with four, three, or two neighboring atoms. If a carbon atom has four close neighbors, the resulting structure is three-dimensional. One example of such a structure is the diamond crystal, in which each carbon atom sits in the center of a regular tetrahedron whose corners are the neighboring carbon atoms.

If there are only two adjacent atoms, a one-dimensional linear structure appears—long polymer molecules are examples of this type. When there are three neighbors, the atomic structures include flat regions. For example, in the flat molecule of benzene C_6H_6 each carbon atom forms bonds with one hydrogen and two carbon atoms.

Another example of atomic structures where each carbon atom has three neighbors is graphite, the second natural form of carbon. Graphite is a layered substance whose structure is based on planes in which the atoms sit at the corners of regular hexagons, forming a kind of honeycomb. Actually, no other structure is possible when each carbon atom forms valence bonds with only three neighbors and all the atoms are arranged in the same way. Fortunately there is a mutual attraction between adjacent planes, which connects the carbon layers to form a crystal of graphite. These attractive forces (known as van der Waals forces, which decrease with distance as r^{-7}) are much weaker than the interaction between adjacent carbon atoms in the same layer. Thus, graphite isn't strong mechanically, and so it can be used to make pencil lead. The carbon planes themselves, however, are as strong as diamond.

The question arises: can we make something more interesting from carbon atoms than just a flat layer in a graphite crystal—say, a polyhedron? Since each carbon atom must have exactly three neighbors, the following geometrical problem arises: how to construct a polyhedron in which exactly three edges come together at each corner?

Here we'll make use of Euler's theorem: for any convex polyhedron, where C is the number of corners, F is the number of faces, and E is the number of edges,

$$C + F - E = 2. \qquad (1)$$

For more complicated polyhedrons equation (1) must be modified by introducing a new concept having to do with the number of "handles" in a polyhedron. For a torus $g = 1$, which means that it has one handle, while for a convex polyhedron g equals zero.

The generalized Euler theorem yields

$$C + F - E = 2 - 2g. \qquad (2)$$

It's surprisingly easy to prove equation (2). We need but note that each handle of a polyhedron and the polyhedron with its handles cut off satisfy equation (1). When we paste each handle back onto the polyhedron, the four glued faces disappear, while the difference $C - E$ doesn't change.

Now we have all we need to deduce the architectural rules for constructing polyhedrons out of carbon atoms. Suppose we want to construct a closed polyhedron with hexagonal faces only, and that there are n_6 of them. Since three faces come together at each corner, and because each edge belongs to two faces simultaneously, we get

Art by Pavel Chernusky

Art by Pavel Chernusky

$$C = \frac{6n_6}{3}, F = n_6, E = \frac{6n_6}{2}. \quad (3)$$

Substituting this equation into equation (2) results in $g = 1$. Thus, using hexagons only we can construct a polyhedron that is topologically equivalent to a torus.

If we want to construct more diverse structures, hexagonal "graphitelike" faces won't suffice. Suppose, in addition to n_6 hexagons, we have n_5 pentagons and n_7 heptagons. Repeating the reasoning above, we get

$$n_5 - n_7 = 6(2 - 2g) \quad (4)$$

So if we're interested in convex polyhedrons ($g = 0$) only, we can do without heptagons, but in that case we must add precisely 12 pentagons. (Incidentally, it was Euler himself who first noticed and proved this fact.) To construct more complicated structures, we need heptagons as well.

Thus, we have deduced the basic rules for constructing complicated three-dimensional structures from carbon atoms. Large structures similar to these exist in nature. For example, the skeletons of radiolarians—the simplest organisms among plankton—as well as many viruses are constructed in just this way.[1] These structures are also familiar in architecture. The geodesic domes of R. Buckminster Fuller spring to mind.[2] To drive home the fact that such structures aren't rarities, pick up an ordinary soccer ball, which is stitched together from 20 hexagons and 12 pentagons. But the question remains: is it possible to build such a structure out of carbon atoms?

Fullerenes and fullerites

In 1985 H. W. Kroto (Great Britain), J. Heath, S. O'Brien, R. Curl, and R. Smalley (United States) found that fairly stable molecules consisting of a large (32–90) and always even number of carbon atoms were formed when graphite was vaporized by a laser beam in a stream of helium. The most stable was the C_{60} molecule, and the discoverers thought that it had the form of a hollow soccer ball (fig. 1). In honor of "Bucky" Fuller the researchers called their molecule a *buckyball* (which soon evolved into the more stately *buckminsterfullerene*), and they named the whole class of C_{60}-like molecules *fullerenes*.

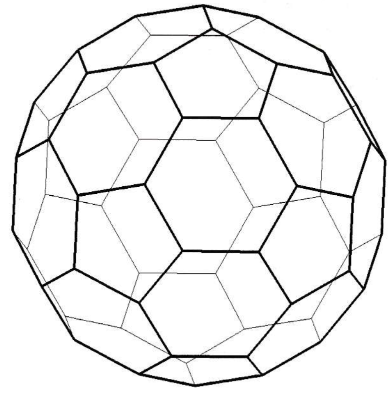


Figure 1

Three-dimensional structure of buckminsterfullerene C_{60} (buckyball). Carbon atoms sit at the corners of the polyhedron.

It turned out that fullerenes aren't all that rare: there are plenty of them in lampblack, gas soot, and other substances resulting from incomplete combustion. The problem is isolating a pure sample of such molecules—that is, obtaining a substance that contains almost nothing else.

In the molecule buckminsterfullerene C_{60}, the corners are carbon atoms and the edges are their valence bonds. There are no free valences in C_{60}, which explains its high chemical and physical stability. The C_{60} molecule is the most symmetrical and stable among fullerenes. The next (in order of increasing numbers of carbon atoms) stable molecule C_{70} has 25 hexagons and the same 12 pentagons. It is formed less often than C_{60}. The highly symmetrical molecules C_{240}, C_{540}, and C_{960}, which are thought to be stable as well, have not been found yet.

It's curious that there is nothing to prevent a C_{960} molecule from having a C_{540} molecule (or other fullerene) inside it. Such compound molecules have not been found either, though the term "matryoshka" has already been coined for them (after the Russian nesting dolls). It's also curious that almost 20 years before the discovery of hollow molecules, their existence was hypothesized by David E. H. Jones, who for many years wrote the famous Daedalus column in the journal *New Scientist*.[3]

At present the dimensions of the buckyball are well known: its radius is 0.3512 nm, the length of the short bond (which separates the hexagons) is 0.1388 nm, and that of the long bond (the same for hexagons and pentagons) is 0.1433 nm. The numbers are very similar to those for graphite. Quantum-mechanical calculations show that the valence electrons must be distributed more or less homogeneously in the spherical shell with a width of approximately 8 a.u. (1 a.u. [atomic unit] = 0.0529 nm, the Bohr radius). An electron-free cavity about 2 a.u. in radius is formed in the center of the buckyball. So the C_{60} molecule resembles a little empty cage.

The existence of the cavity inside the buckyball appears to have been proved experimentally by muon analysis. (The sensor in this method is muonium, which is something like the hydrogen atom, but instead of a proton it has a muon—an elementary particle with a charge of $+e$ and a mass of 200 m_e. The properties of muonium depend strongly on the electron density at its location. The researchers managed to place muonium inside a "fullerene cage" and show that the properties of free and captive muonium are virtually the same.

Free buckyball molecules attract one another with the same weak van der Waals forces that appear between the carbon layers in graphite. Because of this attraction, bucky-

[1]See *Growth and Form* by W. D'Arcy Thompson or the book cited in footnote 3.—*Ed.*

[2]See Gallery Q in this issue.—*Ed.*

[3]The ideas expressed by "Daedalus" in 1966 are not so much fantasy as scientific prediction that has come true. See *The Inventions of Daedalus* by David E. H. Jones (San Francisco: W. H. Freeman & Company, 1982).—*Ed.*

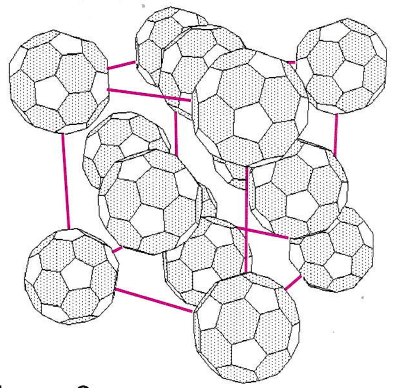


Figure 2

Structure of pure fullerite—the crystal formed from buckminsterfullerenes. One elementary cell of the crystal is shown. The buckyballs are situated in the corners and in the middle of the cube faces.

balls crystallize at room temperature into a fragile yellow-red crystal with a face-centered cubic lattice. The new crystal was named *fullerite*. The distance between adjacent buckyballs in this crystal at room temperature is 1.00 nm. Pure fullerite that contains nothing but buckyballs is a dielectric.

Figure 2 shows an elementary cell of fullerite. Buckyballs play the same role in a fullerite crystal as atoms in an ordinary crystal. Many characteristics of fullerites (for example, the electron spectrum) can be calculated with great accuracy by treating the buckyballs as if they were atoms and applying traditional methods of calculating the properties of crystals.

The new carbon molecules (fullerenes) and the crystals made from them (fullerites) are the third form of naturally occurring carbon—or in scientific terms, the third allotrope of carbon. The first two allotropes—diamond and graphite—have free bonds that grab stray atoms (for instance, hydrogen atoms). Such is not the case with fullerenes and fullerites, since they don't have any free bonds, so among carbon allotropes, they are the purest.

How to build a fullerene

By the mid-1980s, when fullerenes were being discovered, methods for experimentally producing so-called cluster molecules (consisting of a small number of identical atoms) had reached an advanced state of development. Usually the number of atoms in such clusters is rather arbitrary. However, the very first experiments with carbon produced a surprise: large carbon clusters with an odd number of atoms were never formed! At first this fact was explained by the formation of polymer chains of the type [–C≡C–]. H. W. Kroto and his colleagues were the first to provide a correct interpretation, though they couldn't perform a reliable structural analysis—they had too few fullerenes. Their explanation remained a hypothesis, and the fullerene an exotic toy for theoreticians, until the summer of 1990, when a revolutionary event occurred: a method of large-scale production of fullerenes was proposed.

The solution was found rather surprisingly by a group of American astrophysicists—specialists in the area of cosmic dust: W. Kretschmer, D. Huffman, and their students L. Lamb and C. Fostiropoulos. As far back as 1983 Kretschmer and Huffman had tried to experimentally reproduce the natural conditions needed for the formation of cosmic dust. To this end they evaporated graphite samples heated by an electric current in gaseous helium. After the discovery of buckminsterfullerene, the researchers decided to repeat their old experiments. To extract the spherical molecules they expected to produce, Kretschmer and Huffman took advantage of the old chemical rule: dissolve a substance in a similar one. They dissolved the lampblack (formed by carbon vaporization) in benzene, which also consists of closed molecules. A yellowish or reddish liquid was produced whose color depended on its concentration. Soon it was clear that the dissolved fraction of lampblack was composed of molecules of C_{60} (75%), C_{70} (23%), and even larger fullerenes (2%). After the benzene evaporated off, small fullerite crystals remained on the bottom of the cuvette! Analysis of these crystals produced the first reliable information about the shape and properties of fullerenes.

Later this method was perfected. It turned out that in order to obtain the fullerene-rich carbon soot, it was convenient to use an electric arc between carbon electrodes. When the monomolecular fractions were extracted from the fullerene solution, a purity of 99.99% was achieved. Yet the basic production stages remained the same: evaporating graphite electrodes in helium, then dissolving the soot in an organic solvent. So the price of the final product in this improved production process depended only on the cost of the electricity consumed: about 5 cents per gram of fullerenes!

How are fullerenes formed when graphite is evaporated, and why is an atmosphere of helium essential? As was mentioned above, graphite consists of flat layers of carbon hexagons. Fairly small carbon clusters seem to be formed initially during graphite vaporization in the electric arc. They are linear and have plenty of free bonds. In the cooling atmosphere of helium these clusters form graphite "fish scales" resembling scraps of graphite planes (see figure 3 on the next page). From the energy standpoint it's advantageous for these fish scales to change their shape (because they have free bonds at the edges, which are disadvantageous)—they form several pentagons instead of hexagons and bend because of it (the ends of the free bonds come together and thus lower their energy). Since it is energetically disadvantageous for two pentagons to be next to one another, the open ends must come together in the course of this evolution, the structure that emerges automatically is—the soccer ball! (The buckyball is a minimal fullerene, in which the pentagons have no common edges.) Thus, if a fullerene grows slowly enough, it must necessarily become a buckminsterfullerene. Under actual conditions, of course, the shell can close up before the ideal soccer-ball

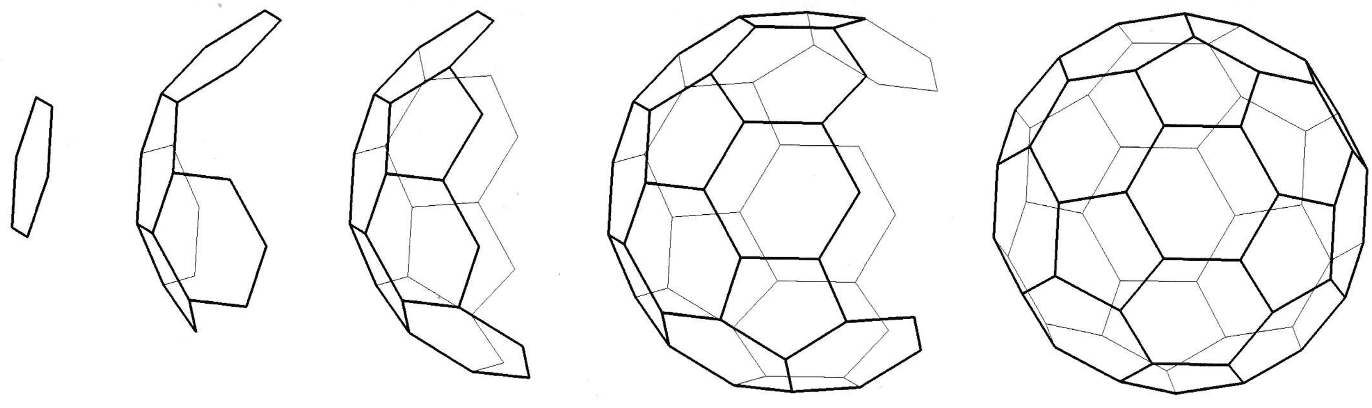


Figure 3

Consecutive stages of formation of buckminsterfullerene C60 during graphite vaporization in a helium atmosphere.

structure is formed, and then other structures arise.

It's not accidental that inert helium serves as the cooling bath when fullerenes are grown. ("Bath" here is a term of art, not figurative language.) This is due to the fact that helium doesn't saturate the free carbon bonds, which lets the carbon fish scales close in on themselves. If there were hydrogen atoms, for instance, in the combustion atmosphere, they could saturate some of the free bonds and destroy the symmetry of the curling fish scale. The opposite sides wouldn't be able to come together. As a result, structures that resemble shells would grow instead. It's interesting that this very process underlies the formation of carbon soot during incomplete combustion in ordinary air.

The smallest this, the smallest that . . .

There are many proposed applications for fullerenes. For example, they might be used as the basis for producing unique lubricants. As was mentioned above, the C_{60} molecule is very strong both chemically and mechanically. Its mechanical strength was tested as follows: a flux of buckyballs was accelerated to a velocity of 30,000 km/h (about orbital velocity) and then sent crashing into a steel wall. The fullerenes bounced off, and were none the worse for the wear! Such strength is just what one wants in a lubricant. So not only is the fullerene C_{60} the world's smallest soccer ball, it's the smallest and strongest ball bearing as well.

The chemical stability and hollow structure of fullerenes suggest ways in which they might be used in chemistry, microbiology, and medicine. For example, fullerenes seem to have no match as a packing material for individual atoms. Scientists have learned how to pack fullerenes even with such heavy atoms as lanthanum and uranium. Fullerenes filled with such atoms open unexpected possibilities for chemists. For instance, fullerenes can be used to pack and transport not only atoms but entire molecules to the required destination. Not a bad idea for pharmacists and microbiologists! So the world's smallest soccer ball is also the world's smallest packing box—or should I say, pillbox.

Nowadays molecular biologists engaged in genetic engineering use viruses (many of which, by the way, are shaped like buckyballs). If scientists manage to use fullerenes to transport the necessary organic molecule to a particular site in a protein, it would mean the creation of an artificial, specialized virus, and again—the smallest one (for the benefit of life on Earth, we all hope).

Now let's talk about microelectronics. It's well known that the process of miniaturization of electronic chips has recently reached its natural limits—that is, molecular and atomic dimensions. As a matter of fact, another term is used more and more for the new technology: not micro-, but nanoelectronics. The characteristic lengths of the elements are nanometers. In nanoelectronics the most interesting objects from the viewpoint of possible applications are quantum dots—microcrystals or other formations incorporated into a nanoelectronic circuit—that can retain (localize) electrons. Such dots have a number of unique optical properties that make it possible to use them either as control elements in fiber-optic communications or as the basic processor elements in the optical supercomputer currently being designed. Fullerenes are in many respects ideal quantum dots. Adding to our list of records, we can say that fullerenes have a good chance of becoming the smallest microchip in a computer nanoprocessor.

And, last but not least, high-temperature superconductivity.[4] Following the discovery of high-temperature oxide superconductors in 1986 by Bednorz and Müller, new substances have continually been tested for possible superconductivity. A pure fullerite, of course, was an unlikely candidate for superconductivity, since it's a dielectric (as was mentioned above). But everyone knows how to turn a dielectric into a conductor: you dope it. Atoms of a suitable impurity can, for example, be donors of the electrons needed to conduct electric current. It was doping that produced the first high-temperature superconductor $La_{2-x}Sr_xCuO_4$ (here x = 0.1–0.2 is

[4]See "Meeting No Resistance" in the September/October 1991 issue of *Quantum*.—*Ed.*

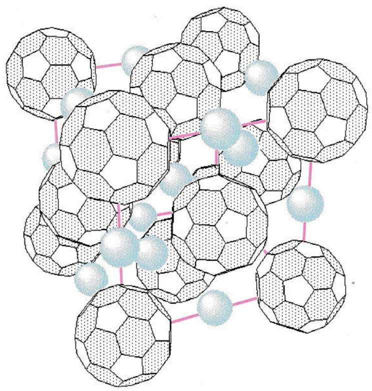


Figure 4

Crystal structure of superconducting fullerite K_3C_{60}. The little balls correspond to atoms of the doping impurity (potassium).

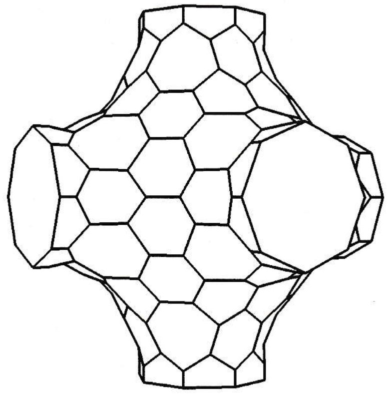


Figure 5

Elementary cell of a schwartzite—an infinite surface made of carbon atoms. (Six of its 24 heptagons are completely visible here.)

the concentration of the impurity—strontium, in this case).

In the beginning of 1991 a new discovery grabbed the scientific headlines: A. Hebard and his colleagues discovered that a fullerite doped with potassium—K_3C_{60}—became a superconductor at 18 K (–255°C). This temperature wasn't a record, but when rubidium was substituted for potassium, the superconductivity transition temperature jumped 28–29 K. Before the race began in 1986 to find high-temperature superconductors, no one had found a material that was superconducting above 24 K. Now materials that are superconducting at 126 K have been found, and there are reports of even higher temperatures.

Superconductors based on C_{60} molecules appear to enjoy superior stability due to the strength of these molecules. This is what makes them stand out from the oxide high-temperature superconductors. The crystal structure of a superconducting fullerite is shown in figure 4. The doping impurity occupies positions in the crystal between the fullerenes.

Still another form of carbon: schwartzite

So, we see that cellular structures made of carbon pentagons and hexagons have been discovered and are now the subject of intense research. But what about heptagons? Such structures have not yet been obtained experimentally, but theoreticians are already modeling their properties on computers. In fact, heptagons offer even more possibilities than fullerenes. For example, a carbon "sponge" has been found whose complex surface consists of hexagons and heptagons that separate three-dimensional space into two subspaces.

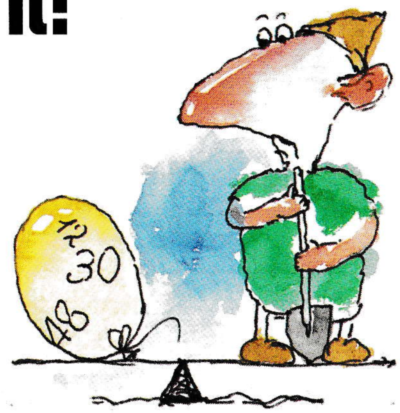
These structures were named *schwartzites* after the German mathematician who was the first to study such surfaces at the end of the last century. Figure 5 shows but one elementary schwartzite cell. The entire crystal is obtained by an infinite repetition of such cells. Schwartzite has the same type of crystal lattice as the cubic face-centered fullerite in figure 2. The elementary cell has 216 corners, 24 heptagons, 80 hexagons, and 3 handles. Note that in this periodic structure the handles connect adjacent crystal cells; in figure 5 each handle is cut in two and only half of it is shown.

Again, such structures have not yet been observed experimentally. But if researchers manage to synthesize this new allotrope of carbon, they would obtain a substance with unique mechanical, physical, and chemical properties.

The author is grateful to Grigory Kopelevich, who prepared the computer-generated illustrations for this article.

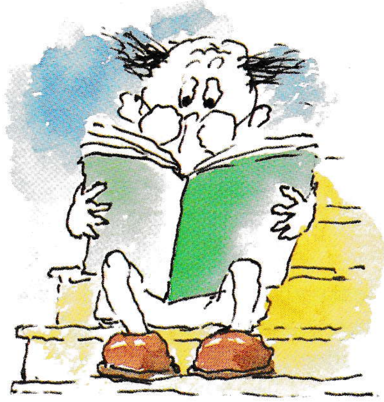
BRAINTEASERS

Just for the fun of it!



B111

Breaking even. According to a contract, a worker is to be paid 48 francs for each day worked and is to give up 12 francs for each day not worked. After 30 days the worker is owed nothing. How many days did the worker work during these 30 days? (Etienne Bezout [1730–1793])



B112

What's wrong? Once I found a strange notebook. A hundred statements were written in it. They said:

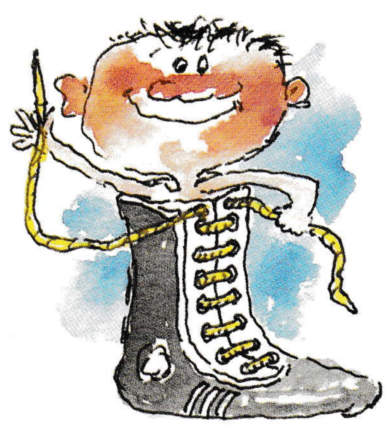
"There is exactly one wrong statement in this notebook."
"There are exactly two wrong statements in this notebook."

⋮

"There are exactly one hundred wrong statements in this notebook."
Which of these statements is true? (A. Savin)

B113

"Fire!" Which is more effective in extinguishing a fire—cold water or boiling water? (S. Krotov)



B114

Arithmetic of lacing. There are many ways to lace wrestling shoes, as is shown in the figure, although we can't see how the shoelace is arranged inside the shoe. Can you tell exactly how many? (N. Zilberberg).

B115

In half. Cut the figure at right into two congruent parts.

Art by Pavel Chernusky

ANSWERS, HINTS & SOLUTIONS ON PAGE 58

A strange emperor and a strange general

A case study with two leaps into the past

by Igor Akulich

WHEN TERENTIUS, A BRAVE Roman military leader, decided to retire, he came to the emperor and asked for a payment of 5 million brasses (a "brass" was a copper coin with a mass of 5 g). The emperor, however, was tight with his money, so he decided to cheat the general. He said, "I wouldn't want you to be content with such a pitiful reward. Go to the treasury and carry out one brass the first day, a two-brass piece the second day, then a four-brass, eight-brass, sixteen-brass piece, and so on, doubling the value of the coin each day. I'll have a coin minted every day of the appropriate size. As long as you're able to carry the coin out on your own, with no help, it's yours. But as soon as a coin is beyond your power to carry, you'll have to stop, and our agreement will be null and void." Terentius was very happy. He imagined an enormous pile of coins, each one bigger than the next, that he'd carry out of the treasury.

What actually happened? Terentius's enrichment lasted only 18 days, because the coin on the 18th day weighed about 655 kilos (he managed to roll it out of the treasury with the greatest difficulty, using his spear as a lever). The next coin was absolutely unmovable. So the total sum he received amounted to 262,143 brasses—that is, slightly more than 1/20 of what he originally asked for. The emperor exulted, while Terentius suffered miserably.

This story, with a lot of interesting details and the brilliance typical of its author, can be found in Yakov Perelman's book *Mathematics Comes Alive.*[1]

Into the distant past

More than a century ago, during a lecture in Baltimore, Lord Kelvin asked the following rhetorical question: "Of all the two hundred billion men, women, and children that have walked across wet sand from the beginning of time down to the meeting of the British Association in Aberdeen in 1885, how many would answer anything but 'yes' to the question: 'Did the sand become compressed under your foot?' " Why Aberdeen in 1885? That was where O. Reynolds showed that the sand actually expands rather than contracts under our feet, contrary to common sense.

But let's not digress too far from our subject. It would be better to ask a similar question about the emperor's award: "Of all the millions of readers of Perelman's book and the thousands of *Quantum* readers, how many noticed that the behavior of both the emperor and the general described in the story was at least *strange* and absolutely *illogical*?"

What was so strange and illogical? We'll soon see.

First, let's try to estimate some of the values we'll need. It's clear from the story that a coin with a mass of 655 kg was just about at the limit of Terentius's physical resources: a little more and he would be unable even to budge it. We'll estimate this "little bit" as 45 kg—that is, assume that the biggest coin that would yield to Terentius's efforts has a mass of 700 kg (which corresponds to a denomination of 140,000 brasses). In addition,

[1] *Quantum* readers may already be acquainted with this outstanding Russian popularizer of math and physics. See, for instance, the Kaleidoscope of the November/December 1992 issue.—*Ed.*

assume that Terentius's state of health will allow for daily visits to the treasury and removal of new coins for ten thousand days (about 25 years).

So, the emperor decided to lure his combative general into a trap that is often called the *avalanche*. (Indeed, it's hard to think of a better name: the coins grow like an avalanche, and this is what the miserly and cunning emperor counted on.) In this case, the multiplication factor is $k = 2$—that is, each coin is twice as massive as the previous one.

And it is this choice of multiplication factor that leads one to suspect that the emperor was a strange person, because of all positive integers k, he had chosen the one that brought the greatest profit to Terentius!

Consider, for instance, the case in which each new coin is three (and not two) times as massive as the preceding one. How many coins would Terentius be able to lift? The value of the $(n + 1)$st coin would then be 3^n brasses. The general can lift a coin that is equivalent to no more than 140,000 brasses. What is the largest n such that $3^n \le 140{,}000$? This n satisfies the inequalities $3^n \le 140{,}000 < 3^{n+1}$, or

$$\log_3 140{,}000 - 1 \le n \le \log_3 140{,}000.$$

Since $\log_3 140{,}000 = 10.7\ldots$, we get $n = 10$. So the last coin Terentius would be able to lift is the eleventh: on the first day he'd receive 1 brass, on the second day 3 brasses, on the third $3^2 = 9$ brasses, and so on. The total reward would come to $1 + 3 + 3^2 + \ldots + 3^{10} = 88{,}573$ brasses. Remember, with $k = 2$ he received 262,143 brasses—almost three times as many!

A similar situation would occur for larger values of k. In general, the sum S that Terentius could receive in n days given a factor k equals

$$S = 1 + k + k^2 + \ldots + k^n,$$

where $n = [\log_k 140{,}000]$ (and $[a]$ denotes the greatest integer not exceeding a). By the formula for the sum of a geometric sequence,

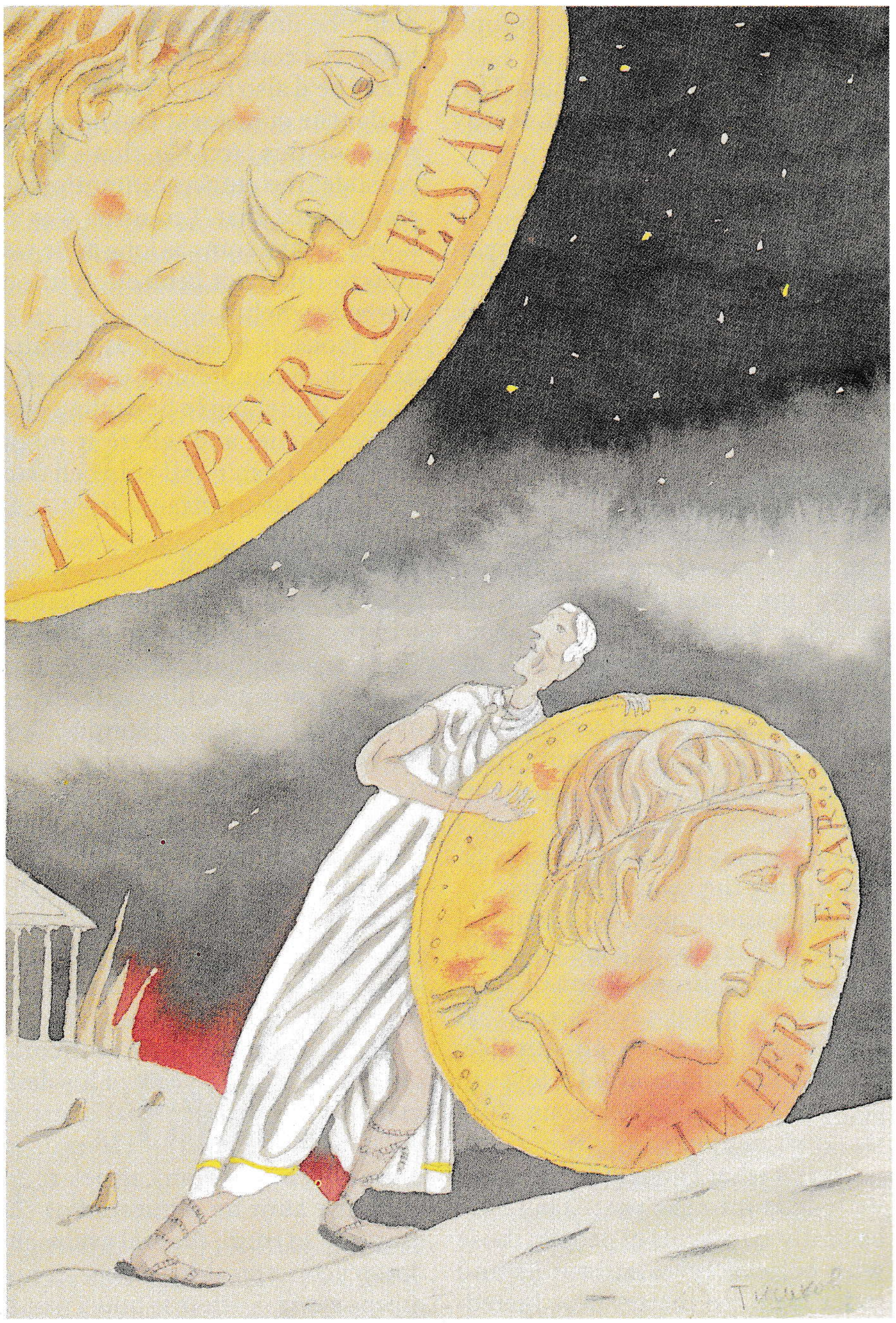
$$S = \frac{k^{n+1} - 1}{k - 1}.$$

If n is large enough, we can assume $n = [\log_k 140{,}000] \approx \log_k 140{,}000$; then

$$S \approx \frac{140{,}000k - 1}{k - 1} = 140{,}000 + \frac{139{,}999}{k - 1}.$$

This means that S actually decreases as k increases. In our case, however, this is not exactly true: the logarithm isn't very large, so in fact $S(k)$ decreases "irregularly." Here are a few values of S: $S(4) = 87{,}381$; $S(5) = 97{,}658$; $S(6) = 55{,}987$; $S(7) = 137{,}257$; $S(8) = 37{,}499$; $S(10) = 111{,}111$; $S(20) = 8{,}421$; $S(50) = 127{,}551$; $S(100) = 10{,}101$ (in the last case, Terentius would come for his reward only three times!).

And what happens for $k = 1$? Perhaps in this case the sum S turns out to be greater than for $k = 2$? Alas, that's not the case. Another factor comes into play here—the somber fact of human mortality. We've already estimated the time allotted to Terentius for receiving his reward as



Art by Leonid Tishkov

10,000 days. Consequently, in the case $k = 1$, he would get simply 10,000 brasses.

Of course, we could adopt other limitations instead of 700 kg and 10,000 days. Then our conclusions would have been somewhat different—for instance, with 1,000 kg as the greatest possible mass of the coin we'd have $S(2) < S(3)$—but basically they would remain the same.

Thus, the (supposedly) cunning emperor, having decided to cheat the general using the avalanche effect, chose the worst multiplication factor (or at least one of the worst). And this gives us grounds to consider him a strange person—to put it mildly.

And what about Terentius? This is a little more complicated. It may help to take a leap . . .

Into the recent past

Way back when I was in elementary school (and, I should add, after I read Perelman's book), I used to ask my friends—the ones who didn't particularly like math—to estimate how many grains of wheat you would need to put one grain on the first square of a chessboard, two grains on the second square, four on the third, and so on, doubling the number of grains each time. "Half a sack," my friend would naively reply. Then I would happily set about convincing my victim that this answer wasn't just wrong, it was very wrong—that in actual fact the number of grains keeps growing from square to square, like an avalanche, and becomes unimaginably large. I would present the result of calculations using the geometric series . . . and I'd be interrupted by a skeptical snort: "What the heck are you talking about, 'trillions of tons'? I say half a sack, and it's half a sack!" A total fiasco! And it's no surprise, because the human mind refuses to comprehend such enormous, "unworldly" numbers.

Here's what I conclude from this little leap into the past. There are basically two kinds of people: those who believe in calculations and strict logic, no matter how incredible the result is; and those who rely on common sense and don't trust anything that contradicts it. Both attitudes are absolutely normal and natural.

Which of the two categories does our character Terentius belong to? On the one hand, he was very satisfied with the emperor's suggestion, because he understood at once how large the coins will grow even if you start with one brass. (And this is just what Perelman's story says: "He imagined an enormous pile of coins, each one bigger than the next.") So we can definitely include Terentius in the first category. But then, why didn't he understand that since the value of a coin is proportional to its mass, he'd also have to cope with an avalanche of masses? Did he simply overlook this fact? No, that's hardly possible: the emperor intentionally emphasized that Terentius was allowed to take coins as long as he was able to lift them himself, without anybody's help.

It looks as if Terentius simultaneously did and didn't understand that he was going to be dealing with an avalanche, and where it might lead him. I can't call this behavior anything but *strange*.

Under the false bottom

By now, I imagine you've figured out that this story has a kind of "false bottom," like a jewelry box. But there's something interesting even under the *second* bottom. Let's allow for non-integer values of k. Then what value of k will make Terentius's reward the greatest? It's clear enough that for this k the mass of the coin he takes on the 10,000th day must be exactly 700 kg—that is, it must have a denomination of 140,000 brasses, which means that $k = 140{,}000^{1/9999} \approx 1.0012$.

Then Terentius's total income over more than 25 years of his daily visits to the treasury will come to $S = (k^{10000} - 1)/(k - 1) \approx 120$ million brasses! This is many times more than the sum he requested of the emperor. In truth, the real avalanche doesn't come crashing down—it just creeps along. So here is how I would advise Terentius to respond to the emperor's seemingly attractive offer:

"Sire! Such a reward is too generous for me. Not only that, it will lessen the treasury so rapidly that severe damage will he inflicted on you and on the entire state. So I can't agree to such a sharp growth in the coins' value. But it would be impudent of me to turn your offer down completely. Might I ask only one thing of you: let the value of the coins grow, but not so rapidly. I'd be completely satisfied if each coin would be more massive than the previous by twelve hundredths of a percent." (Note: most probably they didn't know percentages at that time. I imagine, though, that Terentius could have expressed his wish in some other way.)

Nothing ventured, nothing gained. Maybe the emperor would have swallowed the bait without noticing the hook—which would eventually lead to the bankruptcy of the empire.

Actually, in this case a certain difficulty arises: the values of the coins won't be expressed as integers, which probably wasn't allowed at that time. No matter—Terentius could propose a magnanimous correction: rounding *down* to the nearest integer! This wouldn't cost him too much, because the damage will definitely be less than ten thousand brasses, which is nothing compared to his income.

Of course, it's easy for us to solve the financial problems of the brave general. But how would Terentius himself respond to my advice? It's not unlikely that he would find the proposal *strange*, to say the least. After all, he would have to wait 20 long years for the bulk of his reward. In the first five years Terentius would receive less than 6,600 brasses, and during the first year and a half he'd have to come every day for a one-brass coin! So who of the three is the strangest: the emperor, the general, or I? It's up to you to decide. At any rate, I can't help wondering what Yakov Perelman would have thought of this interpretation of his story. I'd like to think he would have been amused.

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M111

Factorials and powers. (a) Prove the identity

$$\frac{1\cdot 2!}{2}+\frac{2\cdot 3!}{2^2}+\cdots+\frac{n(n+1)!}{2^n}$$
$$=\frac{(n+2)!}{2^n}-2$$

($n! = 1 \cdot 2 \cdot \ldots \cdot n$).

(b) Find the sum

$$\frac{1\cdot 3!}{3}+\frac{2\cdot 4!}{3^2}+\cdots+\frac{n(n+2)!}{3^n}.$$

(V. Zhokha)

M112

Meeting on the diagonal. A line drawn through a point K in a square $ABCD$ intersects two opposite sides AB and CD at points P and Q (fig. 1). Two circles are drawn: through points K, B, P and through points K, D, Q. Prove that their second point of intersection (the point other than K) lies on the diagonal BD. (V. Dubrovsky)

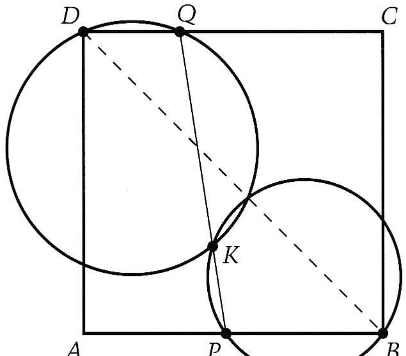


Figure 1

M113

Playing with quadratics. The coefficients in a quadratic equation are replaced with asterisks: $*x^2 + *x + * = 0$. The first player names three numbers. The second one writes them instead of asterisks at will. Can the first player ensure that the resulting equation has distinct rational roots no matter how the second one arranges the coefficients? (A. Berzins)

M114

Rolling to almost everywhere. (a) A regular octagon is rolled over the plane by repeatedly turning it over (reflecting about) any of its sides. Prove that the sequence of rolls can always be chosen in such a way that the octagon's center ends up inside or on a given (arbitrarily small) circle. (b) Solve a similar problem for a regular pentagon. (c) For what regular q-gons is a similar statement true? (G. Galperin)

M115

Composite sum of squares. Prove that $235^2 + 972^2$ is a composite number. (D. Fomin)

Physics

P111

Spring in water. A long homogeneous spring of length L in the relaxed state consists of a large number of identical turns. When the spring is placed vertically inside a tall cylinder with a smooth wall, the spring is half as long as it originally was. Water is then poured into the cylinder up to the level $L/2$. How long is the spring after the water is added? The density of the spring is ρ and the density of water is ρ_0. (S. Krotov)

P112

A charge isn't alone. A point particle of mass m and charge Q is placed at a distance L from an infinite conducting plane and then released. How long does it take the particle to reach the plane? Neglect the effects of gravity. (Hint: use the method of images and compare it to previous problems that you have done with the same force law.) (A. Bytsko)

P113

Sublime self-rescue. According to a science fiction story, an astronaut of mass M = 100 kg was at a distance L = 100 m from her spaceship with a glass of frozen water in her hand. Using the sublimation of the ice, the astronaut returned to her ship. Is such a mode of rescue possible? Determine the time needed to return to the ship. Assume that the sublimation of the ice occurred at a constant temperature T = 272 K. The pressure of saturated vapor at this temperature is P = 550 Pa. The gas constant R = 8.3 J/mol · K). The size of the glass and the mass of the ice can be any values you wish. (A. Stasenko)

CONTINUED ON PAGE 45

The bounding main

The forces controlling the sea swells

by Ivan Vorobyov

THE HURRICANE IS HUNdreds of miles away, the air is calm hereabouts, but the walls of water roll one after another as far as the eye can see. It's the sea swell—the steady surging of the Earth's great oceans. The chain of parallel curves stretches for tens of kilometers, and the waves go on like that for hours on end.

Near the Cape of Good Hope the waves can reach 9–11 m with a wavelength of 100–300 m. Only the ocean's great depth (2 km) reassures us that a particularly high wave won't expose the very floor of the ocean. The speed of these colossal waves is quite impressive: 40–70 km/h.

What are the forces that produce this regular movement of so much water? What does the velocity of the waves depend on? What is their characteristic shape (profile)? What's going on beneath the roiling surface? I'll try to answer these questions. But first, it will be worthwhile to take a close look at the wave itself.

Layered flow

With waves, it's more convenient to study them when they aren't moving. Imagine we're flying in a helicopter with the velocity of the wave motion c. Relative to us, the curves of the water's surface don't change, and along their unchanging profile the water flows steadily. Both the level and inclination of the surface, as well as the velocity V of the water flow along the stationary profile, repeat themselves over a distance of the wavelength λ. Next to the surface layer is another one just below it, and next to that is another layer below, and so on. (Physicists call this *laminar flow.*) As the water moves smoothly along, there are no gaps or ruptures, and the curves of the deepest layers remain stationary with respect to the profile of the surface layer, repeating with the same distance λ (fig. 1).

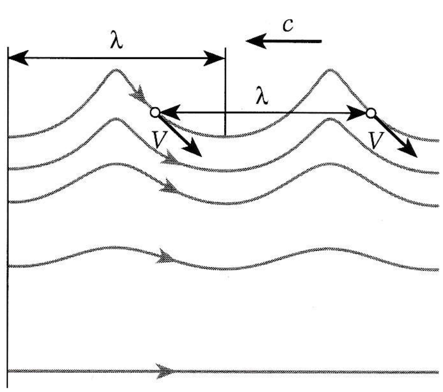


Figure 1

So, our "stopped wave" reference frame turns the movement of the water into a steady-state flow along the curved layers. The stationary boundaries are formed by water particles moving along the same trajectory. Water doesn't leave a layer, which means that the same mass of water passes through any cross section of a given layer per unit time.

The profiles of the different layers are not identical. Their amplitudes decrease gradually with depth. This becomes clearer if we examine the flow between the boundaries of one layer. Because the flow is steady, the layer is thicker where the velocity is less, and vice versa. Flowing downhill, the water particles gain velocity; climbing upward, they slow down. Therefore, the distance between boundaries is larger at the crests and smaller at the troughs (fig. 2). Because

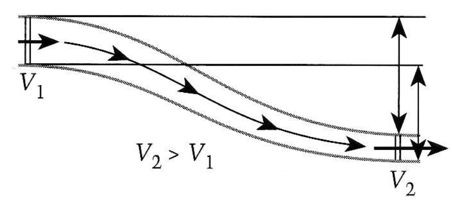
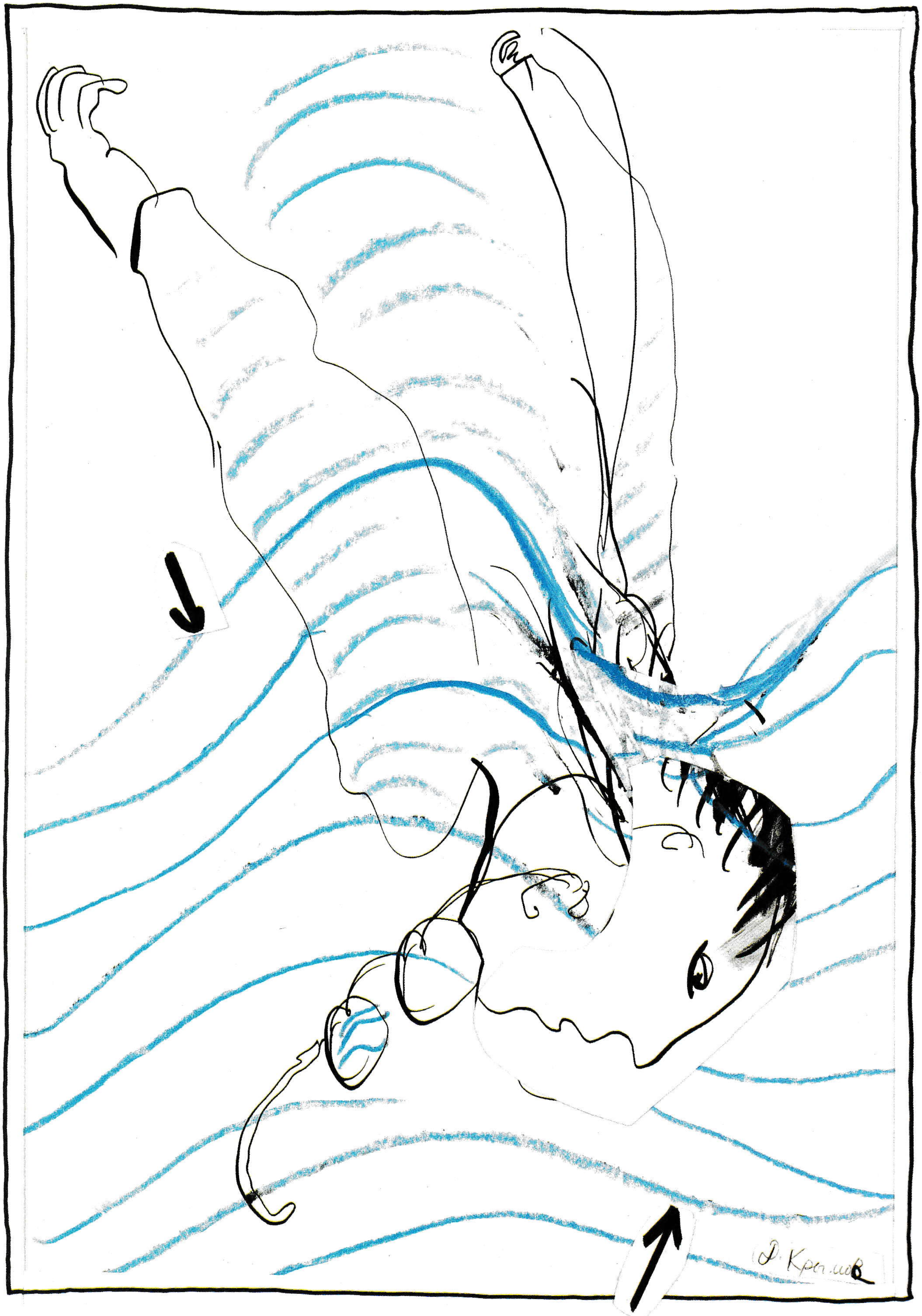


Figure 2

of this, the lower boundary of each layer is less inclined than the upper boundary. The difference in their heights becomes smaller, the changes in the velocity during ascent and descent are less pronounced, and the layers become more homogeneous in thickness as a result of the damping of the curves.

At the lower limit we come upon horizontal layers of still water. But

Art by Dmitry Krymov



actually the water is stationary only relative to the ocean floor—in our moving (stopped-wave) reference frame, it travels with a velocity c directed opposite to the wave. The ocean floor moves with the same velocity and in the same direction.

If the water is deep enough to damp the wave appreciably, the ocean floor will be in still water and will not affect the movement in the upper layers. A quantitative treatment of this condition will be found at the end of this article. But for now it's clear why the uneven relief of the ocean floor doesn't disturb the waves at the surface.

To find the velocity $\mathbf{v}$ of a water particle in the "moving wave" reference frame (in which the ocean floor and the shore are stationary), we need to sum the velocity of a particle moving along the wave profile $\mathbf{V}$ and the velocity of this profile $\mathbf{c}$ (fig. 3):

$$\mathbf{v} = \mathbf{V} + \mathbf{c}.$$

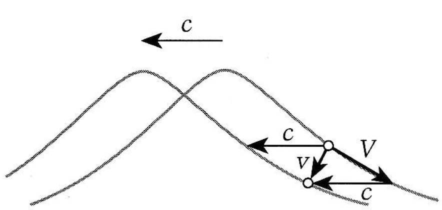


Figure 3

This simple equation will play an important role later on.

Gravity and pressure

The term "surge" is used when waves are anywhere from one meter to hundreds of meters long. For waves that long we can neglect surface tension. The fact that they travel hundreds of kilometers with no appreciable damping is evidence of the small role friction plays here. So the ocean's surge is basically determined by the interplay of just two forces: gravity and pressure.

The pressure along the water's surface is identical everywhere and equal to the atmospheric pressure. At a very great depth the layers are almost horizontal and the water in them is almost stationary. The pressure at a given depth in still water is the same throughout. In deep water it differs from the atmospheric pressure, but there are no pressure variations along the wave profile in either the surface wave or in the deep waves. In the intermediate layers the variations in pressure cannot be caused by air pressure, since it is uniform everywhere on the surface. But there can't be any heterogeneities coming from below either. This is a compelling argument in favor of the pressure being identical at the boundary of any layer.

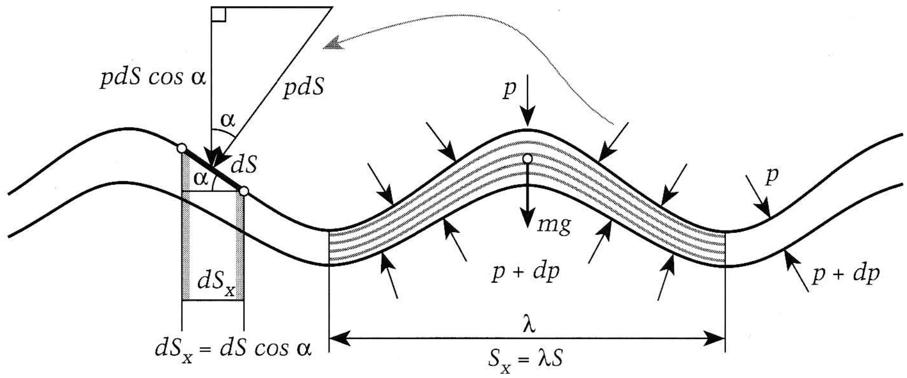


Figure 4

Although the pressure is the same at any point in a profile, it changes at the transition from one boundary to another. The pressure difference and the force of gravity accelerate the water particles. We can determine this pressure difference by the following reasoning. A layer as a whole doesn't move up or down. The forces are counterbalanced for every fragment of the layer with a length equal to the wavelength λ (fig. 4).

To calculate the force of the pressure acting on the upper curved boundary, where the pressure at any point is p, we begin with a small inclined fragment. The force is equal to the pressure times the area of the fragment and is directed perpendicular to it. The vertical component of the force is equal to the pressure times the area of the horizontal projection of the fragment (fig. 4). The common factor (pressure) is taken out of the brackets when we sum the vertical components, and the sum of the areas of the projections gives the area λL of the horizontal cross section of the wave fragment (where L is its width). Thus, the total force of pressure acting on the upper boundary is $p\lambda L$ and is directed downward.

For a lower boundary with a pressure $p + dp$, the corresponding force is directed upward and is equal to

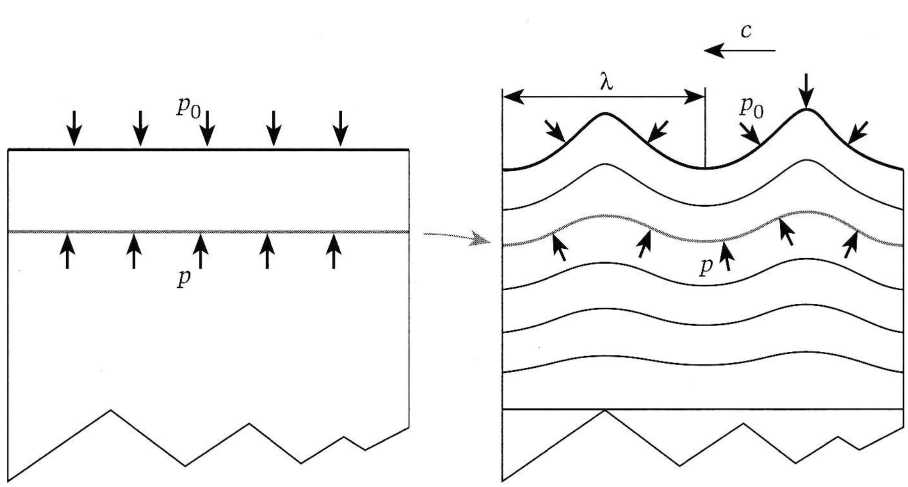


Figure 5

$(p + dp)\lambda L$. The difference between these forces is counterbalanced by the force of gravity mg, where m is the mass of the fragment. Therefore,

$$dp = \frac{mg}{\lambda L}.$$

Let's take a closer look at this equation for the increase in pressure. A curve doesn't change a layer's mass or the mass of a fragment. It's the same as it was between the horizontal boundaries of this fragment in still water. So in still water there is the same pressure difference, and the same pressure. (In every case it begins at the surface with the atmospheric pressure.) The pressure at the curved boundary is equal to the initial hydrostatic pressure acting on its particles in still water (fig. 5).

Acceleration of a water particle

Consider a small fragment of a thin layer (fig. 6). Its butt-ends are

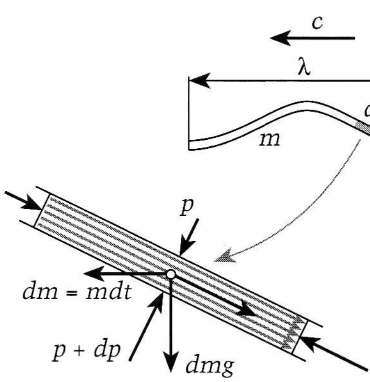


Figure 6

perpendicular to the velocity V of the flow. During the time dt of passage through this fragment, the length of the inclined boundaries is $V\,dt$, and their area is $LV\,dt$. In order to apply Newton's second law and find the acceleration, we need to know the mass of the fragment dm and the sum of the forces acting on it.

In the period $T = \lambda/c$ the entire mass m of the wave fragment is replaced, so the mass that passes through the fragment per second is $m/T = (m/\lambda)c$. In the period dt the mass entering the fragment is equal to

$$dm = \left(\frac{m}{\lambda}\right)c\,dt.$$

From this it follows that the force of gravity acting on the fragment is

$$dm\,g = \left(\frac{mg}{\lambda}\right)c\,dt.$$

This force is directed downward at a right angle to the wave velocity c.

The pressure doesn't change along the flow in a layer, and the sum of the forces directed at one another at the butt-ends is zero. The pressure difference at the inclined boundaries dp results in the force

$$dp\,LV\,dt = \left(\frac{mg}{\lambda}\right)V\,dt.$$

This force is directed perpendicular to the boundary, at a right angle to the velocity of the flow $\mathbf{V}$.

Both forces can be obtained from the vectors $\mathbf{V}$ and $\mathbf{c}$ in the same way: by rotating them 90° and multiplying by the same factor $(mg/\lambda)\,dt$. Thus, the sum of the forces can be obtained from the vector sum $\mathbf{V} + \mathbf{c}$ by the same two operations: rotation by 90° and multiplication by the aforementioned factor (fig. 7). Since $\mathbf{V} + \mathbf{c} = \mathbf{v}$,

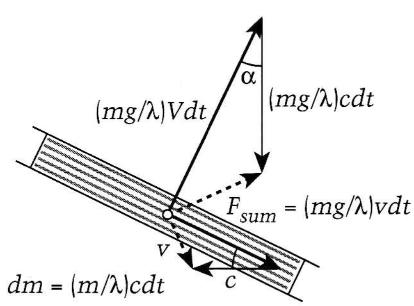


Figure 7

the total force can be expressed by the velocity of a water particle relative to the ocean floor (that is, in the moving-wave reference frame). The force is perpendicular to this velocity and is equal to

$$F_{\text{sum}} = \left(\frac{mg}{\lambda}\right)v\,dt.$$

Dividing the force by the mass dm gives us the acceleration of the fragment:

$$a = \frac{g}{c}v.$$

The acceleration is directed at a right angle to the velocity v.

This is a turning point. We've determined the acceleration that results from the forces of gravity and pressure, and now we're ready to unravel the details one by one.

A picture of motion

An acceleration that is perpendicular to the velocity does not change its value. In this case the acceleration itself $a = (g/c)v$ has a constant absolute value. Constant acceleration at a right angle to the velocity clearly indicates uniform circular rotation (fig. 8). For a circle

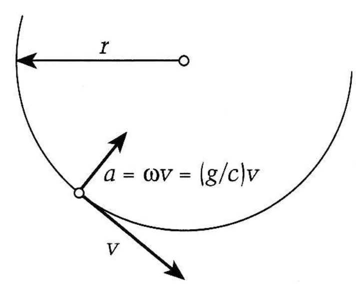


Figure 8

of radius r, the centripetal acceleration is $a = v^2/r$; and since the angular velocity is $\omega = v/r$, then $a = \omega v$. Comparing the last equation and the formula for acceleration found previously, we obtain the angular velocity:

$$\omega = \frac{g}{c}.$$

Now we essentially have the whole picture of the motion: the water particles move with the same angular velocity in a circle of constant radius. For particles forming the wave's profile, the radii are the same, but their centers lie on a horizontal line. The particles rotate simultaneously; the angular displacement between particles doesn't change, but the wave profile as a whole moves

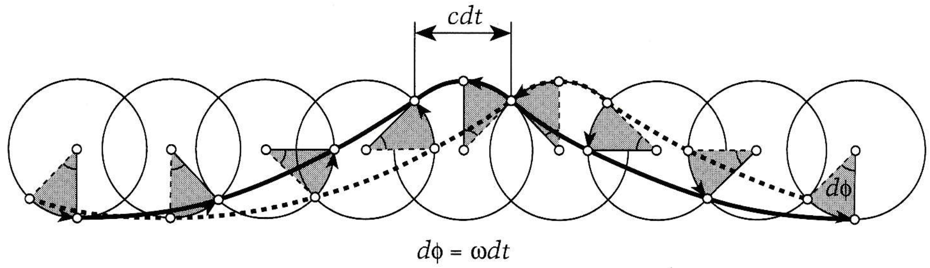


Figure 9

with a velocity c (fig. 9). This occurs at every depth in the water—only the radius of the circular motion changes from layer to layer.

After a time $T = 2\pi/\omega$ a particle will return to its initial position and find itself in the same fragment, but in the next wave (fig. 10)—the one that traveled the distance $\lambda = cT = 2\pi c/\omega$ during this time. Since $\omega = g/c$, then $\lambda =$

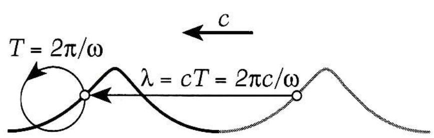


Figure 10

$2\pi c^2/g$. This gives us the following equation for the velocity of the wave:

$$c^2 = \frac{g\lambda}{2\pi}.$$

It's easier to analyze the wave profile in the stopped-wave reference frame. The profile then is drawn by the particle itself, whose movement is a combination of rotation and translation with a velocity c. This allows us to construct the profile and find the dependence of the coordinates on time:

$$x = ct - r \sin \omega t,$$
$$y = r \cos \omega t.$$

The coordinate origin (fig. 11) is

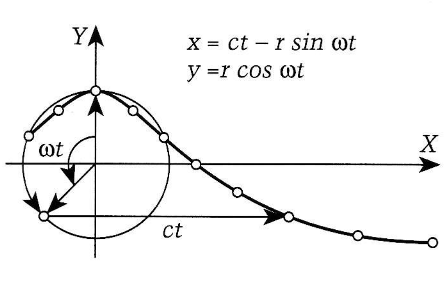


Figure 11

chosen at the center of the circle, and zero time corresponds to the moment when the particle is at the crest.

For a weak wave (when $v = \omega r$ is small compared to c), the horizontal velocity $V_x = c - \omega r \cos \omega t$ can be considered constant and equal to c. Then $x \cong ct$. In the formula for y we replace t with x/c and obtain an approximate equation for the profile:

$$y = r \cos\left(\frac{\omega x}{c}\right) = r \cos\left(\frac{2\pi x}{\lambda}\right).$$

So it turns out that a weak wave is sinusoidal. The "weakness" condition mentioned in the opening section is equivalent to the inequality

$$r << \frac{\lambda}{2\pi}.$$

For moderate waves we see a sharpening at the crests and a flattening in the troughs (fig. 12). The

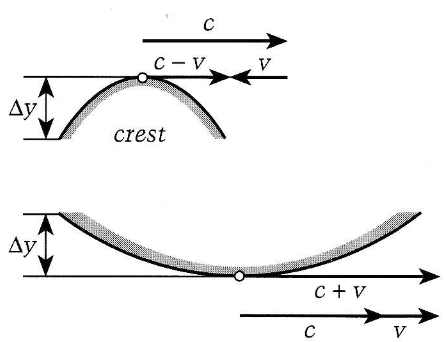


Figure 12

vertical deflections from the crests and troughs are equal for small time periods, but the horizontal deflections differ: the velocity at a crest is $c - v$, while the velocity in a through is $c + v$. The closer the value of v is to c, the stronger the horizontal "compression" of the crests.

Our physics intuition tells us that we should expect something unusual when the natural limit of the wave velocity is passed. Formally, at $v = c$ an infinitely sharp vertical spike appears, and at $v > c$ the profile crosses over itself and a loop emerges (fig. 13). For a wave in the ocean this scenario would look pretty strange. According to our

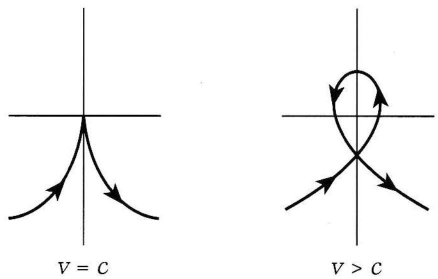


Figure 13

theory, when v is smaller than but close to c, the layers near the sharp crests curve steeply, and the velocity of flow changes sharply from layer to layer. Doubts arise not only about the stability of such motion, but even whether such a motion is close to being stable. So the case of large waves seems to demand a more complicated approach.

Damping of waves at depth

The crests of laminar waves are located precisely one beneath the other, and likewise the troughs. At the points of the profile that are symmetrical relative to the crests, the absolute value of the flow velocity is equal, and also at these points the thickness of a layer is the same because the flow is constant, which precludes any skewing.

The boundaries of a layer in a stopped wave are formed by the trajectories of its particles. For each of them the motion consists of a combination of rotation and translation. The difference is that the centers of the circles are located at different depths and their radii aren't equal.

To obtain the dependence of the radius on depth we use the fact that

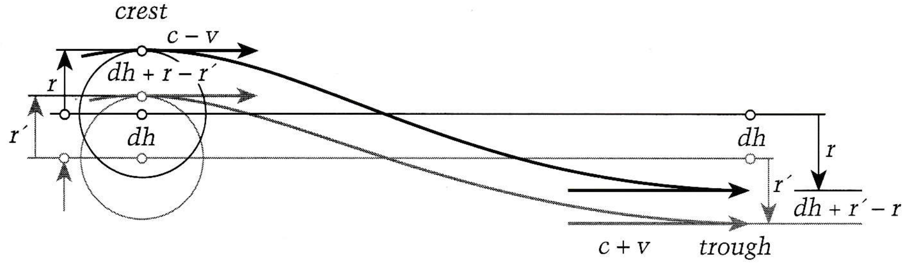


Figure 14

the flow is constant throughout a layer. Let's consider two cross sections (fig. 14). The first is in a crest, where both radii "look" up; the flow velocity is $c - v$ and is directed horizontally. The second is in a trough, where the radii "look" down; the flow velocity is $c + v$. When the distance dh between the centers is small, the radii r and r' of the upper and lower circles differ only slightly. The thickness of the layer in the crest is $dh + r - r'$, and in the trough it's $dh + r' - r$. The equality of the flows in these cross sections gives us

$$(dh + r - r')(c - v) = (dh + r' - r)(c + v).$$

Therefore, we can find the increase in the radius $dr = r' - r$, which is negative—

$$dr = -\frac{v}{c}dh$$

—and this is correct because the radius decreases with depth. Since $v = \omega r$, the decrease in the radius is proportional to the radius itself. When the center of the circle is lowered by dh, the radius decreases by the same proportion:

$$\frac{dr}{r} = -\frac{\omega}{c}dh.$$

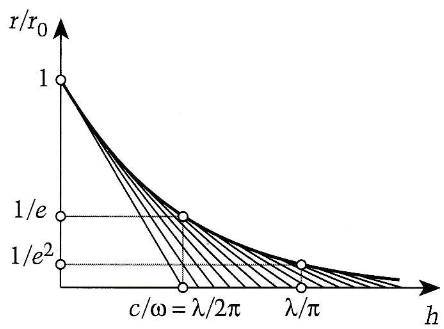


Figure 15

Starting from the circle of radius r_0 at the surface and descending by small steps from its center, we can find either graphically or numerically the radius at any depth h (fig. 15). Readers who can integrate will be able to obtain the analytical expression

$$r = r_0 e^{-\frac{\omega h}{c}} = r_0 e^{-\frac{2\pi h}{\lambda}}$$

(They will also have obtained the expression $\omega/c = 2\pi/\lambda$ with no great difficulty.)

The amplitude of the waves and the velocity of the water decrease with depth geometrically. The number e is approximately equal to 2.72 $\cong 10^{0.43}$. At a depth of $\lambda/2\pi$ the wave is damped roughly by a factor of 3. But at a depth equal to the wavelength, it's damped by a factor of $e^{2\pi}$, which is close to 535. When the depth is of the order of the wavelength, agitation from the ocean floor is far weaker than the surface agitation, which allows us to neglect the influence of the ocean floor on the waves.

And now, a few final remarks. Our basic assumption at the outset was the absence of a pressure drop along the curved layers. In essence this meant that a deep wave is similar to a surface wave, only it has a different "atmospheric" pressure. The discovery of this kind of "self-similarity" has helped lead to the solution of complex problems in fields ranging from fluid mechanics to elementary-particle physics. In our problem this approach made it possible to describe the motion of the entire mass of water by applying Newton's second law to a single droplet.

Six challenging dissection tasks

And a visit from a close relative of ϕ

by Martin Gardner

KARL SCHERER, A COMPUTER scientist in Auckland, New Zealand, recently posed the following six tasks:

1. Cut a square into three congruent parts.
2. Cut a square into three similar parts, just two of which are congruent.
3. Cut a square into three similar parts, no two congruent.
4. Cut an equilateral triangle into three congruent parts.
5. Cut an equilateral triangle into three similar parts, just two of which are congruent.
6. Cut an equilateral triangle into three similar parts, no two congruent.

The solution to the first task is obvious (see figure 1). It is surely unique, though I know of no proof. Ian Stewart and A. Womstein have shown that no rectangle can be divided into three congruent polyominoes unless the pieces are rectangles.[1]

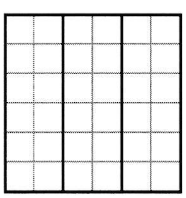
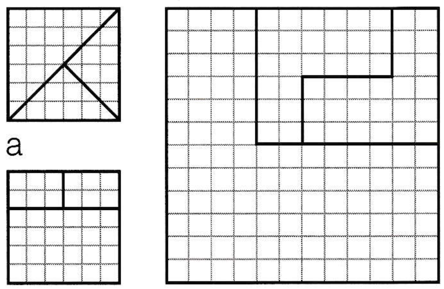


Figure 1

Figure 2 shows three solutions to the second dissection task.

Task 3 is more difficult. Scherer found the pattern shown in figure 3. The solution is not unique, because the slanting line can assume an infinity of positions. The one shown may be the one in which line segments have the smallest possible integer lengths.

[1]*Journal of Combinatorial Theory*, Series A, Vol. 61, September 1992, pp. 130–36.



a b c

Figure 2

As mathematician Robert Wainwright of Plainview, New Jersey, has observed, figure 2b results when the slanting line is orthogonal.

We turn now to the three equilateral triangle tasks.

The fourth task obviously has an infinity of solutions, obtained by rotating the three trisecting lines about the central point (fig. 4). The trisecting lines need not be straight. They can be as wiggly as you like, provided that they are identical and do not intersect (fig. 4b).

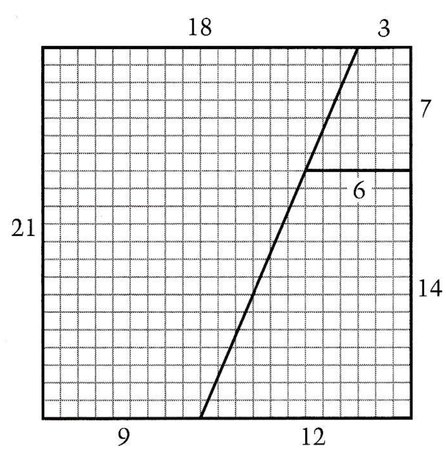


Figure 3

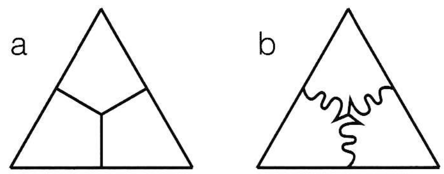


Figure 4

Scherer found an elegant solution to the fifth task (fig. 5). It's believed to be unique. Note its similarity to figure 2c.

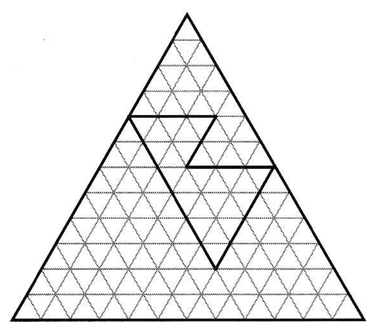


Figure 5

The sixth task is easily solved (fig. 6). It's probably unique, though no proof is known.

My only contribution to the six tasks was the rediscovery of a second solution to the third task

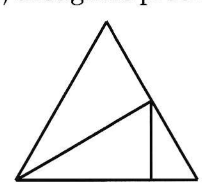


Figure 6

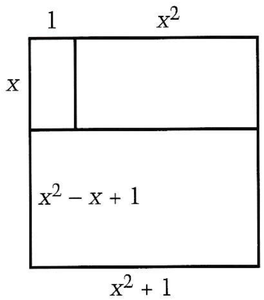


Figure 7

(fig. 7). I later learned from Scherer that he had found it years earlier. What is the value of x, assuming the smaller side of the smallest rectangle is 1? I thought this would be a simple question to answer. If x isn't rational, surely it's a recognizable irrational, such as 1.732... (the square root of 3), or 1.618... (the golden ratio, often called phi), or some other well-known irrational.

To my amazement, x turned out to be an irrational number I had never encountered before.

The cubic equation relating the ratio of the sides of the smallest rectangle to the ratio of the sides of the similar largest rectangle is

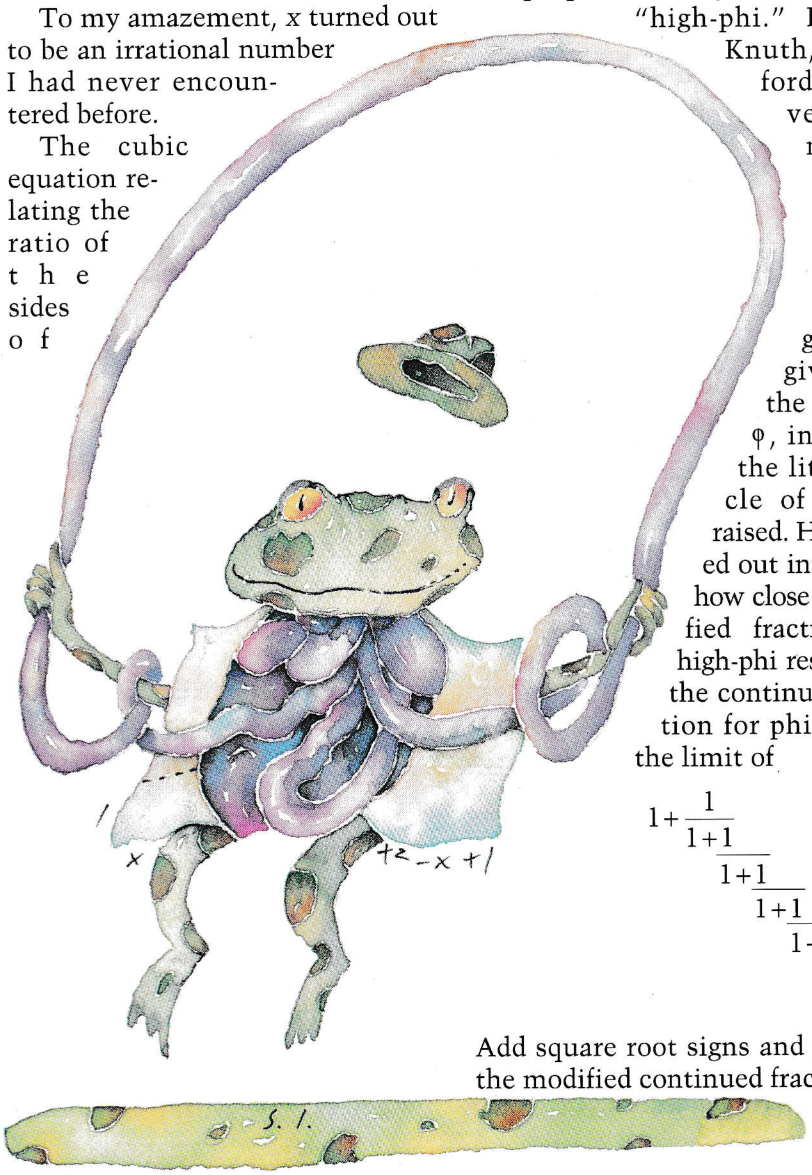
$$\frac{1}{x} = \frac{x^2 - x + 1}{x^2 + 1},$$

$$x^3 - 2x^2 + x - 1 = 0,$$

$$(x^2 - x)(x - 1) = 1.$$

The decimal expansion of x is 1.75487766624669276... . As Wainwright pointed out, the number is closely related to phi, the golden ratio. The reciprocal of phi equals phi minus one. The reciprocal of x equals $(x - 1)^2$. Other equalities are

$$\frac{1}{x^2} = \sqrt{x} - 1; \quad \sqrt{x} = \frac{1}{x - 1}.$$



I propose calling this number "high-phi." Donald Knuth, Stanford University's noted computer scientist, suggested giving it the symbol φ, in which the little circle of phi is raised. He pointed out in a letter how close a modified fraction for high-phi resembles the continued fraction for phi. Phi is the limit of

$$1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \ddots}}}}$$

Add square root signs and you get the modified continued fraction for high-phi:

$$1 + \cfrac{1}{\sqrt{1 + \cfrac{1}{\sqrt{1 + \cfrac{1}{\sqrt{1 + \cfrac{1}{\sqrt{1 + \ddots}}}}}}}}$$

As Knuth write, the series converges more rapidly than the series for phi, giving values that are alternately over and under the true value: 1, 2, 1.71, 1.765, 1.753, 1.7554, ...

Knuth also called attention to the following equality for high-phi:

$$1 + \frac{1}{\varphi - 1} = \varphi + \frac{1}{\varphi}.$$

Karl Scherer points out that the three rectangles in my figure have areas of x, x^3, and x^4. And if the original square has a side length of 1, the rectangles have areas of $1/x$, $1/x^2$, and $1/x^4$. This shows that $1 = 1/x + 1/x^2 + 1/x^4$, and the ratio of the largest rectangle to the rest of the square is $\sqrt{\varphi}$.

Scherer suggests the terms phi-two, phi-three, and so on, for the first terms of the series of solutions for the equation

$$\frac{1}{x} = (x - 1)^n.$$

He conjectures that 1 is the sum of the infinite series of the reciprocals of phi-two, phi-three, phi-four, and so on. In brief,

$$1 = \sum_{0}^{n} \varphi_n^{-2^n}.$$

Can any reader prove or refute this conjecture?

Is it not surprising that such a simple geometrical construction would generate such a curious number? Note that 666, the number of the beast in the Book of Revelations, follows its first six decimal digits. Perhaps *Quantum* readers know of other properties, serious or numerological.

I would welcome hearing from anyone who can find other solutions to any of the six tasks.

Art by Sergey Ivanov

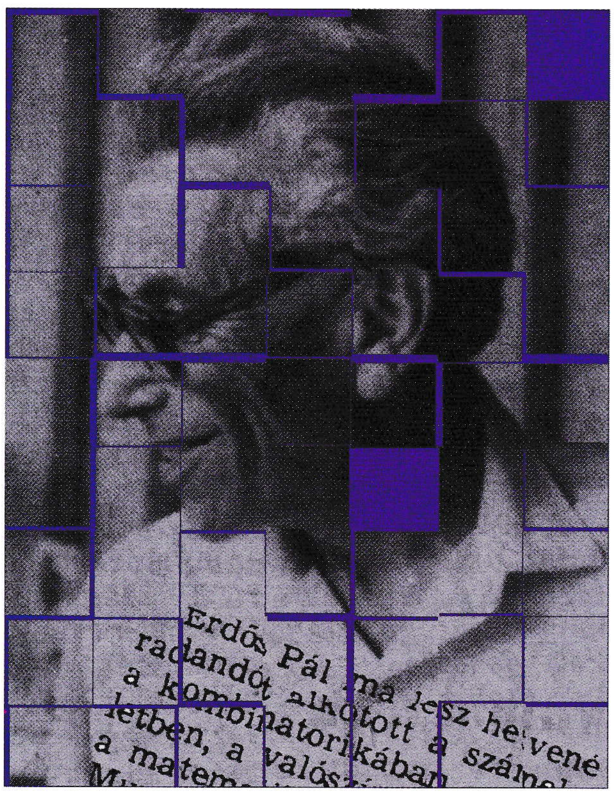
Happy birthday, Uncle Paul!

Erdős is "squared," and we hope he will be squared yet once more

by George Berzsenyi

PAUL ERDŐS, MATHEMATICIAN extraordinaire, recently celebrated his 81st birthday and referred to the occasion as being "squared," since $81 = 9^2$. We hope he will celebrate many more birthdays, write yet another 1,500+ papers, inspire many more mathematicians throughout the world to decrease their "Erdős number" to 1, and pose many more wonderful problems.[1] In this column I'll share some of his problems with my readers in the hope that they will assist Uncle Paul, as he is affectionately called by his many friends and admirers, in his constant quest for the mathematical unknown.

Paul Erdős is a native of Hungary. His unique mathematical talents were recognized very early and were encouraged by his parents, both of whom were excellent mathematics teachers. His development as a mathematician was much enhanced by the *Középiskolai Matematikai Lapok*, Hungary's 100-year-old high school mathematics journal, which is greatly responsible for the high level of mathematical life in that small country. Following the completion of his formal studies for a doctorate in mathematics, Erdős became the only truly universal university professor of mathematics. In the words of one of his admirers, J. W. S. Cassels of Trinity College, Cambridge University: "He has executed an almost Brownian motion amongst the mathematical centers of the world, being the focus of mathematical activities wherever he goes. Just as a bumblebee goes from flower to flower carrying its load of pollen, he goes from mathematical center to mathematical center with his problems and information, thereby being an agent of mathematical cross fertilization." The problems below are but a few samples of his many interesting queries.



Paul Erdős in 1983.

Problem 1: Let P be an arbitrary point interior to a triangle, and denote by a_1, a_2, a_3 the distances from P to the triangle's vertices, and by x, y, z be the distances from P to the three sides of the triangle. Determine the minimum of $(a_1 + a_2 + a_3)/(x + y + z)$.

Problem 2: Let n points be given in a plane, with no three of them on a line. Maximally how many pairs of points can be unit distance apart?

Problem 3: Prove that $\sum \frac{1}{n!-1}$ is irrational.

Problem 4: In a convex n-gon, let $s_1, s_2, \ldots$ denote the multiplicity of the occurrence of the distances between the vertices of the n-gon. (Note that $\sum s_i = \binom{n}{2}$.) Prove that

[1]If you coauthor a paper with Erdős, your Erdős number is 1; if you coauthor one with someone who coauthored one with Erdős, your Erdős number is 2; and so on.

there exists a constant c such that $\sum s_i^2 < cn^3$.

Problem 5: Prove that a convex n-gon always has a vertex that is not equidistant from any four of the other vertices.

The problems above were posed in the centennial issue of *Középiskolai Matematikai Lapok*; they are all open. Problem 6 below was his first serious problem, posed in 1931; it, too, is still unresolved. Problems 7 through 10 are reproduced here from recent letters of Pali Bácsi (Hungarian for "Uncle Paul") to the author. The prizes offered by him for the solution of his problems are also typical, except for the fact that most of them are in the thousands. It has been estimated that he couldn't possibly cover his promises if all of his problems were solved at the same time. However, since most of them are very deep and difficult, he seems to be safe from ever going bankrupt.

Problem 6: Let $a_1 < a_2 < a_3 < \ldots < a_k$ be distinct positive integers such that the 2^k sums, $\sum_1^k \epsilon_i a_i$, where $\epsilon_i =$ 0 or 1, are all distinct. Estimate or determine the value of min a_k.

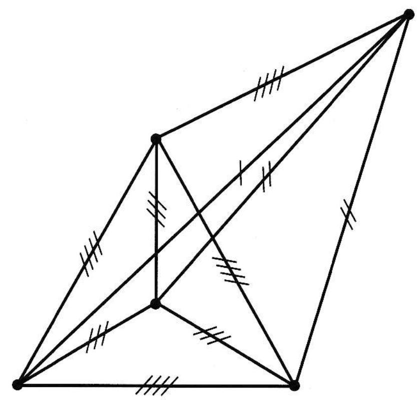
Problem 7: Let $a_1 < a_2 < \ldots$ be the set of integers of the form $2^\alpha 3^\beta$. Prove that every n can be written as the sum of a_i's, no one of which divides any other. Is it in fact true that if n is sufficiently large, then there exist such a_i's with

$$a_{i_1} + a_{i_2} + \cdots + a_{i_k} = n$$

and

$$a_{i_1} < a_{i_2} < \cdots < a_{i_k} < 2a_{i_1}\ ?$$

Problem 8: Let $x_1, x_2, \ldots, x_n$ be n points in the plane in general position—that is, no three on a line and no four on a circle. Erdős believes that for large n the points determine at least n distinct distances. For small n this is certainly false. In fact, for $n < 9$ it can happen that one distance occurs $n - 1$ times, one $n - 2$ times, and so on. For $n = 4$ an isosceles triangle and its center constitute an example. For $n = 5$ Carl Pomerance constructed such an example, while for $n = 6$, 7, and 8 Ilona Palasti did so. (To whet your appetite, Pomerance's example is illustrated below, with equal distances bearing the same markings.) Erdős offers \$10 for an example for $n = 9$, \$25 for a proof that for large n such an example does not exist, and \$50 for a proof or disproof that for large n the points determine at least n distinct distances.



Problem 9: Let $f(n)$ be the largest integer for which there are integers a, b for which $n!/(a!b!)$ is an integer, and $a + b = n + f(n)$. Prove that there exist c_1 and c_2 such that $f(n) < c_1 \log n$ for all n, and try to prove that $f(n) > c_2 \log n$ for infinitely many n. Moreover, refer to n as a *champion* if $f(n) > f(m)$ for all $m < n$. For example, 10 is a champion, since $10!/6!7! = 1$, $6 + 7 = 10 + 3$, and $f(m) < 3 = f(10)$ for all $m < 10$. Try to determine all champions. Let $g_n(a) = b$ be the largest b for which $n!/a!b!$ is an integer, and let $f_n(a) = a + g_n(a) - n$. Determine or estimate the value of $\frac{1}{n}\sum_{a=1}^{n/2} f_n(a)$.

Problem 10: Let $x_1, x_2, \ldots, x_n$ be the vertices of a convex n-gon in the plane. Construct all of its diagonals; there will be $\binom{n}{2} - n$ of them. Consider the interior points of intersection thereof. If we assume that no three of them go through a point, then it is trivially true that there are $\binom{n}{4}$ of them, since every choice of 4 points yields a point of intersection. Hence let's not assume anything about the number of diagonals meeting at a point. Then there are two questions: How many distinct points of intersection can there be? And what is the minimum of the number of intersection points? Moreover, for $2n$ points, can the number of intersection points be smaller than that for regular $2n$-gons?

As the above problems illustrate, the questions posed by Paul Erdős constantly probe the frontiers of the known mathematical universe. His discoveries cover most branches of mathematics from number theory to combinatorics, from foundations to analysis, from geometry to probability, and many new areas which were initiated by his own investigations. It is not unusual that he is simultaneously working on several papers with different mathematicians.

"Pali Bácsi" loves to work with young people, many of whom have been inspired by him to great accomplishments. He was also one of the founders of the famous Budapest Semesters in Mathematics program, which is briefly described in the Bulletin Board in this issue (page 52).

Paul Erdős is a member of the Hungarian Academy of Sciences and was also elected to membership in the Academies of the Netherlands, Australia, India, and England. He has also been the recipient of numerous honorary degrees. Most recently, the World Federation of National Mathematics Competitions honored him by creating the Erdős Prize, to be awarded to mathematicians whose efforts in the popularization of mathematical competitions have resulted in an increased awareness of the important role of mathematics.

The purpose of this column is to direct the attention of *Quantum*'s readers to interesting problems in the literature that deserve to be generalized and could lead to independent research and/or science projects in mathematics. Students who succeed in unraveling the phenomena presented are encouraged to communicate their results to the author either directly or through *Quantum*, which will distribute among them valuable book prizes and/ or free subscriptions.

Geometric summation

Infinite algebraic tilings

by M. Apresyan

HIGH SCHOOL STUDENTS know that for $|q| < 1$, the sum of an infinite geometric sequence $1 + q + q^2 + q^3 + \ldots$ equals $1/(1 - q)$, and study an analytical proof of this formula. For $q = 1/n$, where $n \in \mathbf{N}$, this sum can be computed geometrically. You might have seen similar geometric proofs of simple algebraic identities—for example, $(a + b)^2 = a^2 + 2ab + b^2$—that make use of a device called *algebraic tiling*.

We'll begin with the case $n = 2$. Let's find the sum

$$1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots.$$

Take an arbitrary rectangle of area 2. Cut it into two rectangles of unit area (fig. 1). Cut one of these rectangles again into two equal halves of area 1/2. Do the same with one of them to get two small rectangles of area 1/4, and so on. This process creates a sequence of rectangles whose areas are equal to 1, 1/2, 1/4, 1/8, ..., $1/2n$, ... The union of these rectangles coincides with the initial rectangle (without one corner point). So the sum of the areas of all these rectangles is equal to the area of the initial rectangle—that is, to two. Thus, $1 + 1/2 + 1/4 + 1/8 + \ldots = 2$.

Now let's find the sum $1 + 1/n + 1/n^2 + 1/n^3 \ldots$. Start with a rectangle of area n and cut it into n rectangles of area 1 (fig. 2). Next, cut one of these rectangles into n rectangles of equal area $(1/n)$, do the same with one of these smaller rectangles, and so on. This process yields rectangles of areas $1, 1/n, 1/n^2, \ldots, 1/n^k, \ldots$ ($n - 1$ rectangles of each of these areas). The union of all these rectangles is again the initial rectangle with one corner point deleted. So the area of the union is n; on the other hand, it's equal to

$$\begin{aligned}(n-1)+\frac{n-1}{n}+\frac{n-1}{n^2}+\cdots\\ =(n-1)\left(1+\frac{1}{n}+\frac{1}{n^2}+\cdots\right).\end{aligned}$$

Thus,

$$1+\frac{1}{n}+\frac{1}{n^2}+\cdots=\frac{n}{n-1}=\frac{1}{1-1/n}.$$

The same idea works for the sum of an infinite geometric sequence with an arbitrary rational ratio q, $0 < q < 1$. Indeed, let's take the last construction in a rectangle of area n, but each time let's subdivide m $(m < n)$ rectangles of the n obtained in the previous step (fig. 3), leaving intact the remaining $n - m$ rectangles. Do you think you can take it from here? Give it a try, and derive the formula for $1 + m/n + (m/n)^2 + \ldots$.

This method can be applied to some other infinite sums as well.

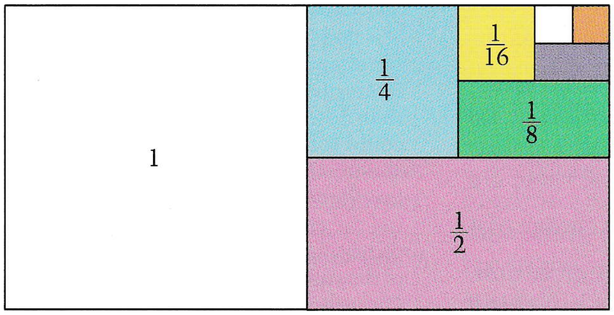


Figure 1

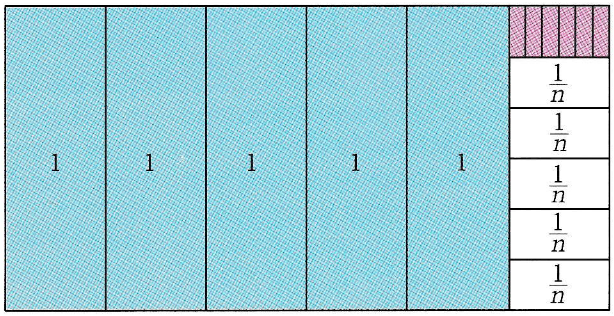


Figure 2

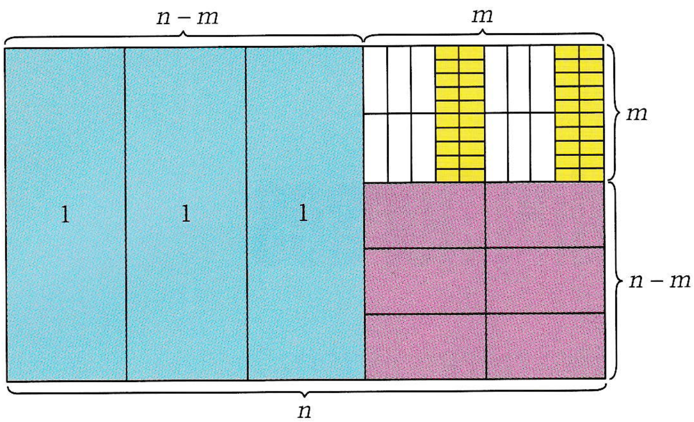


Figure 3

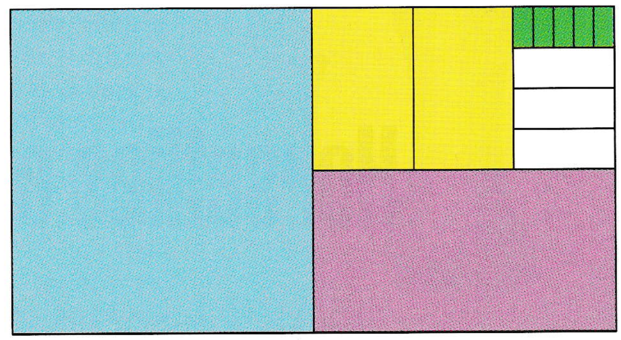


Figure 4

For instance, let's prove that

$$1+\frac{1}{2!}+\frac{2}{3!}+\frac{3}{4!}+\cdots+\frac{n}{(n+1)!}=2.$$

Again we'll take a rectangle of area 2. The first two steps are the same as in the first example: we cut the rectangle into two unit rectangles, and then one of them into two rectangles of area 1/2 (fig. 4). But now we cut one of the last two rectangles into *three* parts equal in area (the area of each of these parts is $1/2 \div 3 = 1/3! = 1/6$). One of the rectangles of area 1/6 is now cut into *four* equal parts, one of these parts is cut into *five* parts, and so on. This yields one rectangle of area 1, one of area 1/2, two rectangles of area 1/3!, three of area 1/4!, ..., n rectangles of area $1/(n + 1)!$, and so on. So the sum arising in this case is $1 + 1/2! + 2/3! + 3/4! + \ldots$, which is equal to 2—the area of the whole rectangle. ◘

drop us a line at *Quantum*,
1840 Wilson Blvd.,
Arlington VA 22201-3000
or
zip us an electron or two at
72030.3162@compuserve.com.

Horseflies and flying horses

Questions of scale in the animal kingdom

by A. Zherdev

SIZE IS ONE OF THE MOST important characteristics of a living thing, but the difference in size is so obvious that we often pay little attention to it. Everybody knows that an elephant is bigger than a mouse, but rarely do we think *how much* bigger: by a factor of 100,000. The smallest adult shrew is a tenth the size of a mouse, which makes it 1/1,000,000 the mass of an elephant. The difference is even more striking when we compare animals of different phyla—say, a protozoan and a whale.

Does a size difference lead to a qualitative change? For a long time both scientists and lay persons believed that it doesn't and thought that all the characteristics of living creatures change in proportion to their size. The classic example of such an assumption is the world created by Jonathan Swift in his classic satire *Gulliver's Travels*. Readers then as now were delighted by Swift's "wondrous realism" as expressed in the precision of his arithmetical calculations: Gulliver was 12 times larger than Lilliputians and was smaller than the Brobdingnagians by exactly the same factor, and everything in both countries was scaled by the appropriate amount.

Real life is not so simple, however, and the first person who realized it was none other than Galileo Galilei. He wrote: "It is not possible to decrease in equal measure the surface and the weight of a body and preserve similarity of form. It is absolutely clear that the decrease in weight is proportional to the decrease in volume, and therefore every time the volume decreases more than the surface (while preserving similarity of form), the

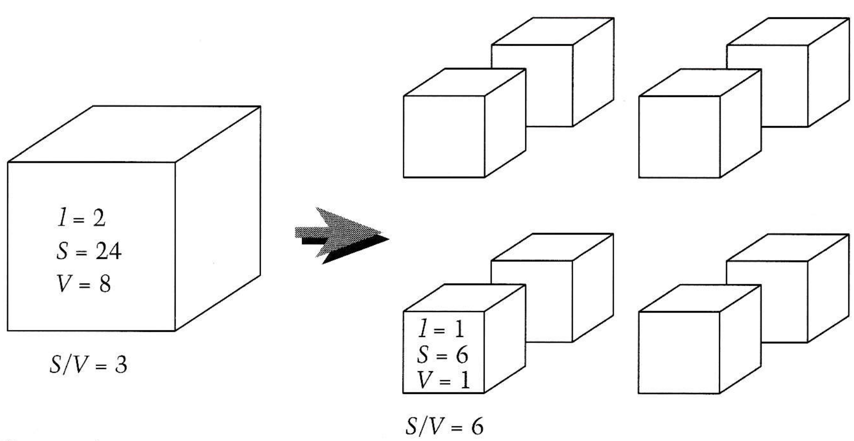
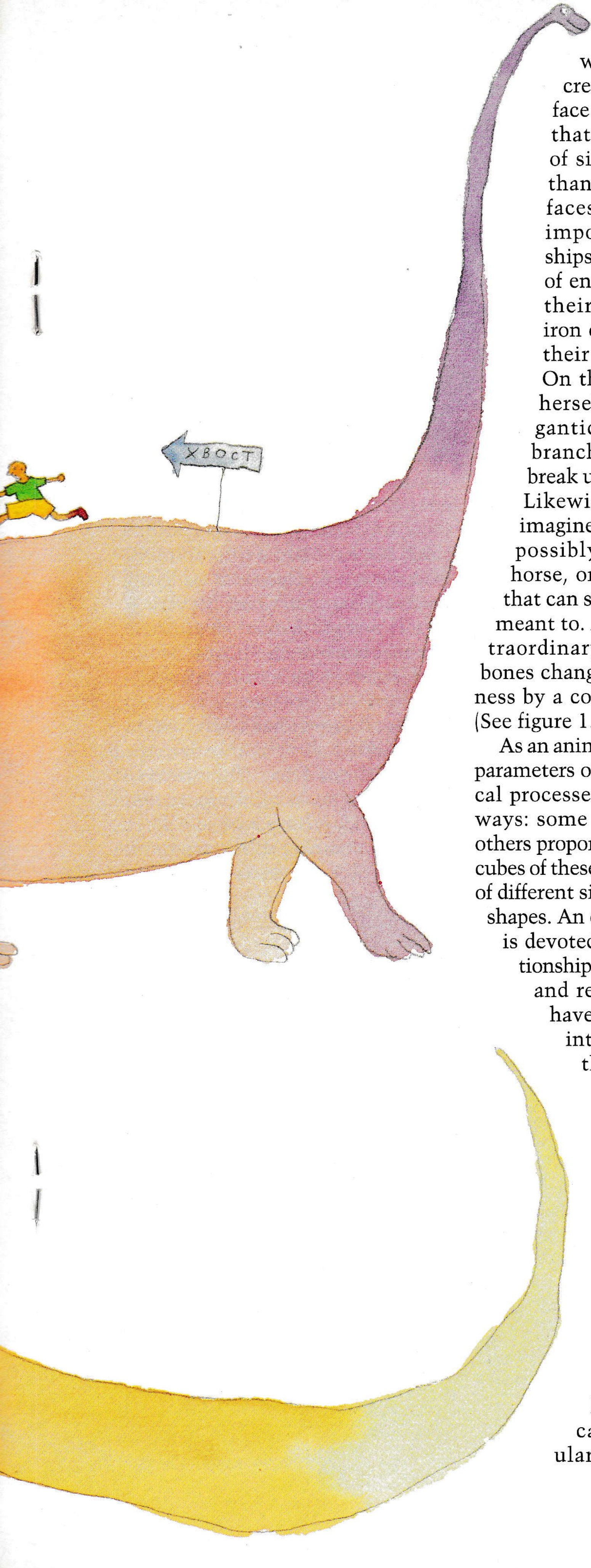


Figure 1

Art by Leonid Tishkov



weight will also decrease more than the surface. But geometry teaches that the ratio of volumes of similar bodies is larger than the ratio of their surfaces. . . . Therefore, it is impossible to construct ships, palaces, and churches of enormous size such that their oars, masts, beams, iron clamps—in a word, all their parts—hold together. On the other hand, Nature herself cannot produce gigantic trees because their branches would ultimately break under their own weight. Likewise, it is impossible to imagine the skeleton of an impossibly huge human being, horse, or other living creature that can support the body as it is meant to. Animals can attain extraordinary sizes only if their bones change, increasing in thickness by a corresponding amount." (See figure 1.)

As an animal increases in size, the parameters of its various physiological processes increase in different ways: some linearly proportional, others proportional to the squares or cubes of these values. And so animals of different sizes must have different shapes. An entire branch of biology is devoted to analyzing the relationship between size and shape, and researchers in this area have obtained a number of interesting results. Work through the questions presented below and you'll become acquainted with some of them.

Questions

1. On the microscopic level the muscles of the most various kinds of animals do not differ all that much in structure. Muscular contraction is caused by intermolecular chemical complexes

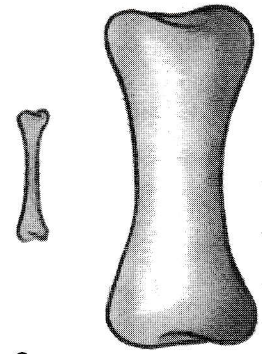


Figure 2

whose structure and arrangement are basically the same. Still, a distinct differentiation exists: the smaller the animal, the greater the mass (relative to its own) it can lift. How can this be explained?

2. Why do animals with approximately the same shape (a grasshopper and a locust or a kangaroo rat and a kangaroo) jump to the same height, regardless of their size?

3. In the treatise cited above, Galileo introduces a drawing that shows that a large bone is disproportionately thicker than a small bone (fig. 2). He made a small arithmetical error, though. What was it?

4. A scientific expedition discovered a new creature: a one-footed mammal (fig. 3). Its dimensions are given in retems—the unit of length

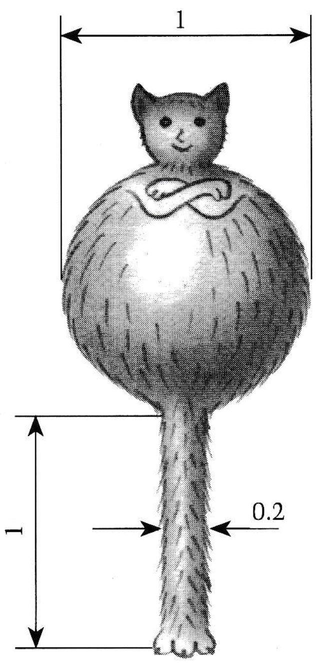


Figure 3

used by the natives. Estimate the mass and height of the "monoped" using human data (average measurements or your own dimensions). Determine the length of 1 retem in meters.

5. A baby was born weighing 8 lbs. It tripled its weight in one year, weighing in at 25 lbs. Continuing to grow at the same rate, by the end of its second year it would weigh $25 \times 3 = 75$ lbs; after three years, $75 \times 3 = 225$ lbs; after four, $225 \times 3 = 675$ lbs; and after five years, the baby would weigh $675 \times 3 = 2{,}025$ lbs. The daily food intake of this one-ton "toddler," at 1/6 to 1/7 of its body mass (as is normal for growing children), comes to 300 lbs. From this estimate the famous Polish teacher Janusz Korczak drew the conclusion that one should not force little children to eat against their will! What do *you* think the daily food intake of such a three-year-old baby should be? How much do you weigh, and how much food do you eat each day? Do your figures jibe with Korczak's calculations?

6. (a) Insects don't try to buzz or drone—it just happens. The sound comes from the flapping of their wings. The force of the flapping (pushing against the air) must compensate for the insect's tendency to fall due to gravity. So why do gnats buzz while bees drone? What is the relationship between the insect's tone and its size? (b) Our friend Gulliver complained about the droning of Brobdingnagian flies: at dinnertime these insects didn't give him a moment's rest. He may have been bothered by them, but did they really drone?

7. The higher the body temperature of birds and mammals, the smaller the animal is. Why?

8. Parents who are in a hurry walk quickly, and if they happen to be holding their child by the hand, the child must run to keep up. Why do adults and children achieve the same speed in such different ways?

9. Which desert animals are able to live without water for a longer time—small ones or big ones?

Take a good look around . . .

Compare the height and thickness of nearby trees and stalks of tall grass. What formula describes the relationship between these two parameters? What factors are at play here?

It's interesting that . . .

. . . there are so many large animals around. If it's so disadvantageous to be large, why didn't evolution produce a prevalence of small animals? The reason is that large animals are stronger than smaller species in absolute terms, although the smaller animals are *relatively* stronger. This gives them the advantage in head-to-head competition and also allows them to occupy new ecological niches. But bigness is unfavorable in other respects. In order to find enough food, a horse must cover more terrain than a mouse does. Since the food supply is usually limited, a mouse may have the upper hand over the horse because it needs less food. A hectare of meadowland can support a huge population of mice, but no more than one or two horses. Appreciable populations of large animals can gather only on great expanses of land. Thus, both smallness and bigness have their own ecological advantages.

. . . a few years ago the newspapers trumpeted the achievement of a certain seven-year-old who set a record for doing push-ups (about 5,000). The boy grows up and probably takes up gymnastics, but he has no chance of beating his own record. We hope by now you understand why.

. . . the biggest land animal was the brachiosaurus, which was up to 20 m long and weighed 80 tons. Speculation about how such a huge animal could exist led paleontologists to the hypothesis that this kind of dinosaur lived in tidal areas, so that its entire body (except the head) was immersed in water (fig. 4). Otherwise, in the opinion of these authors, the bones of the brachiosaurus could not have borne the enormous load. However, this mode of living creates certain problems. In particular, it would be impossible to breathe, because the lung muscles would not be able to continually overcome the pressure of the water. And so biologists had to go back to the blackboard and recalculate the strength of the bones. They came to the conclusion that the bones *could* support such a huge load.

. . . comparing different mammals, biologists came up with a

CONTINUED ON PAGE 37

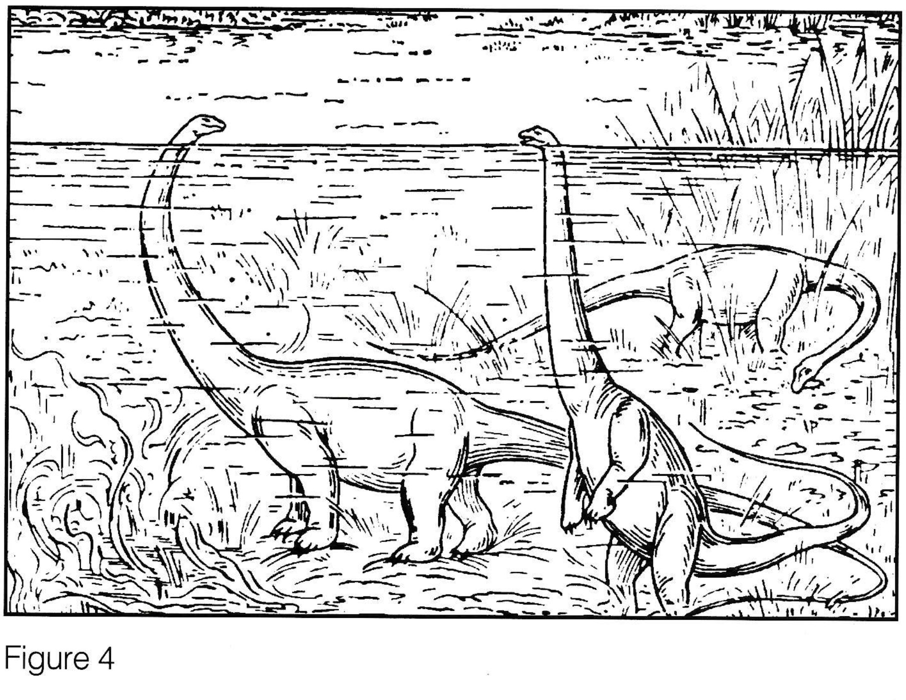


Figure 4

AT THE BLACKBOARD I

A little lens talk

Then let's go to the movies

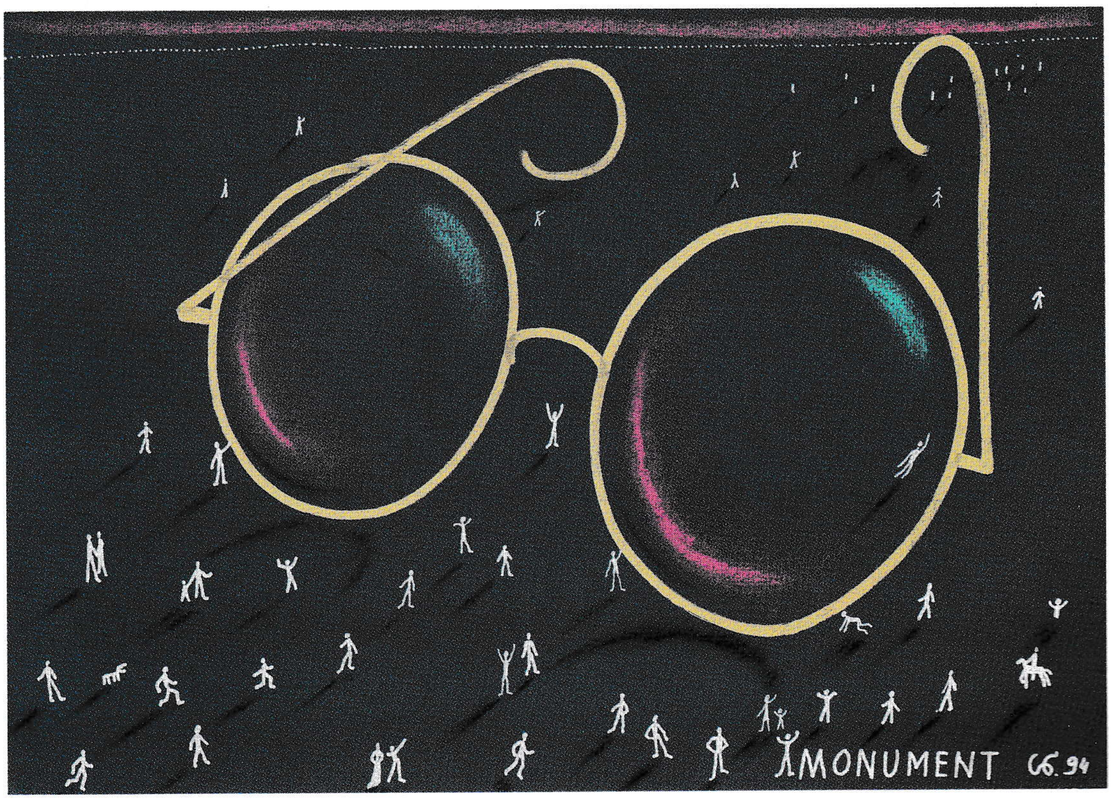
by Alexander Zilberman

THE PHENOMENON OF REfraction of light at the boundary of transparent media (say, air and glass) can be used for a number of purposes in various optical devices, including parallel plates, prisms, and lenses.

Most often lenses are used to form images of luminous (or illuminated) objects. A lens makes it possible to produce an image at the right place (on the film in a camera, on the screen in a movie theater) or at a distance that is comfortable for viewing (eyeglasses, a magnifying glass, contact lenses). We can obtain an image of the object that is either greatly increased (in a microscope or film projector) or decreased (in a telescope or binoculars—yes, *decreased* by a factor of thousands, but brought nearer by a factor of hundreds of thousands, which makes it possible to examine the object in all its details). We can also obtain an image of a luminous point at infinity—in this case the light beam is almost parallel (for example, the beam from a searchlight).

In many cases of practical importance, the optical system consists of several lenses. For example, the objective of a camera can consist of more than ten different lenses—concave and convex, thin and thick, made of different and special kinds of glass. One can even have lenses with a more complicated geometry than the usual spherical surfaces. Admittedly, the simplest calculations of optical systems that can be done within the framework of high school physics will not meet the practical demands of actual devices: even the lenses for high-quality eyeglasses (not to mention contact lenses!) are often calculated by computer, and not because of a surplus of computers.

Nonetheless, even a simple theory can come in handy, because approximate calculations will often



Art by Sergey Barkhin

be good enough. Let's look at the paths of beams falling on a very simple, plano-convex lens with a spherical convex surface of radius R. I've chosen this lens because the paths of the beams will be simplest in this case. Let the lens be made of glass and placed in air.

Let's imagine that a parallel beam of light falls on the plane surface of our lens. We know that after refraction in a convergent lens it must converge at a point lying in the focal plane. Let's show how. First we need to formulate the problem more accurately. We'll assume that we're using a thin lens (in due course I'll explain more rigorously what the term "thin lens" means and what we should neglect), and we'll take the angle of incidence to be small (this keeps the error small when we replace the functions sin α and tan α with the angle α itself, which greatly simplifies the calculations). These are reasonable conditions that correspond in general to the actual situation in simple optical experiments.

Let's draw the lens and the path of one of the incident beams (fig. 1—here the lens is thick and the angles are large for the sake of legibility).

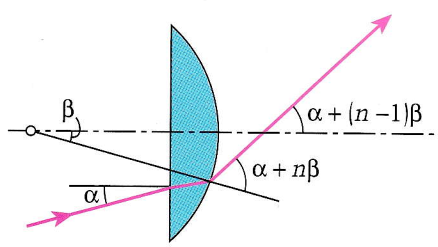


Figure 1

The chosen beam strikes the lens at an angle α to the principal optical axis, and after refraction at the plane surface the angle decreases by a factor of n, where n is the refractive index of glass. We can see this by using Snell's law

$$\sin\alpha = n\sin\phi$$

and approximating these sines by the angles to get $\phi \cong \alpha/n$. The beam falls on the spherical boundary between the glass and air at an angle of $\alpha/n + \beta$, where β is the angle between the principal optical axis of the lens and radius drawn to the point of incidence of the beam. The beam then exits the lens at an angle of

$$\left(\frac{\alpha}{n}+\beta\right)n-\beta=\alpha+(n-1)\beta.$$

From now on we'll need the thin-lens condition: we'll consider that the points of entry of the beam into the lens and exit from it are at the same distance $R\beta$ from the principal optical axis. At a distance L to the right of the lens the beam is shifted vertically relative to the exit point by $L[(n-1)\beta+\alpha]$, and the distance to the principal optical axis is

$$H = L[(n-1)\beta+\alpha]-R\beta.$$

Two different rays from the original incident beam intersect to the right of the lens (fig. 2). Equating the

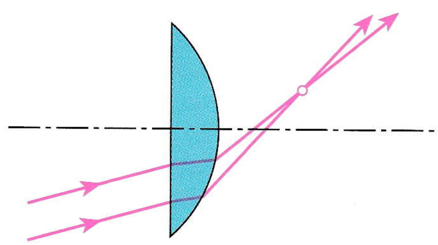


Figure 2

distances H for both rays, we find the distance L to the point of intersection:

$$\begin{aligned}L[(n-1)\beta_1+\alpha]-R\beta_1\\ = L[(n-1)\beta_2+\alpha]-R\beta_2,\end{aligned}$$

from which we get

$$L=\frac{R}{n-1}.$$

We see that the distance obtained doesn't depend on the angle β—it's the same for all the rays from our beam. Thus, we have proved that the refracted beams converge at one point, and we've found the distance from this point to the lens. Notice that all the points of intersection (corresponding to various angles of incidence) lie in the plane perpendicular to the principal optical axis of the lens and are located at a distance $R/(n-1)$ from the lens. In other words, we have calculated the focal length of our lens:

$$f=\frac{R}{n-1}.$$

In the same way, after more lengthy calculations we can obtain an analogous formula for the focal length of a lens formed by two spherical surfaces with radii R_1 and R_2. It's more convenient to present this formula in a slightly different form:

$$\frac{1}{f}=(n-1)\left(\frac{1}{R_1}+\frac{1}{R_2}\right).$$

The radii of the lens can be both positive (biconvex lens), both negative (biconcave), or they can have opposite signs. (In our case of a plano-convex lens, one of the radii is infinitely large.) If the resulting focal length turns out to be positive (the focus of the lens is real)—that is, the refracted beams indeed converge at a point—the lens is called converging (or positive). On the other hand, if the focal length is negative (the focus is imaginary)—that is, the parallel rays diverge after refraction—the lens is called diverging (or negative).

There is simple relationship known as the lens formula (it can be deduced geometrically) that links the distance s between the source and the lens, the distance s' between the lens and the image, and the focal length f of the lens:

$$\frac{1}{s}+\frac{1}{s'}=\frac{1}{f}.$$

Thus, after refraction by the converging lens, the beam of parallel rays converges at a point in the focal plane. This makes it easy to determine the path of any beam after refraction. We need only draw the auxiliary ray parallel to it that passes through the center of the lens (this ray is not refracted) and find the point of intersection of this ray with the focal plane—the original ray must pass through this same point

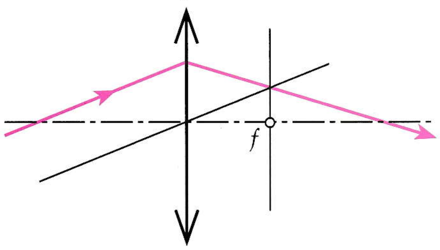


Figure 3

after refraction (fig. 3). The possibility of sketching the path of an arbitrary ray helps greatly when we need to form an image in a complicated optical system consisting of a dozen lenses. There is no need to obtain intermediary images after each lens.

I should point out that with this method of drawing we can use any rays—even those at large angles to the principal axis. If necessary we can increase the diameter of the lens so that the rays land on it. The point is, this method of drawing corresponds to a "paper lens"—that is, not to a lens but to the lens formula. For a real lens such a result corresponds to "correct" rays only—that is, to rays at small angles to the principal axis—and this is exactly what we need.

Let's use the method described to solve a simple problem. Figure 4 shows a converging lens, its principal axis, and the path of one ray before and after refraction by the lens.

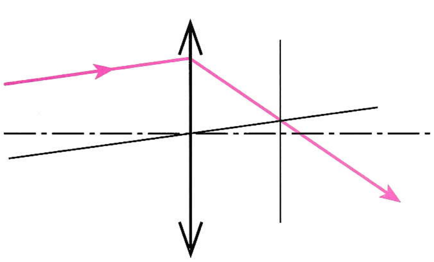


Figure 4

We need to find the position of the focal plane. Let's draw the auxiliary ray parallel to the incident ray such that it passes through the center of the lens—this ray doesn't refract, but its point of intersection with the refracted ray lies in the focal plane. By the way, we can "reverse" the rays—taking the incident ray to be the refracted one and the refracted ray as the incident ray—and, as in the first case, draw the focal plane on the other side of the lens. However, both focal planes must be at the same distance from the lens.

Now let's return to "real" lenses. Usually it's assumed that the point source emits light evenly in all directions. Wherever the eye is located, it receives the diverging beam of light and we perceive the luminous point. It's quite another matter if we want to see the image of this point in a lens. Let figure 5 show a "real" lens of a certain size. In this

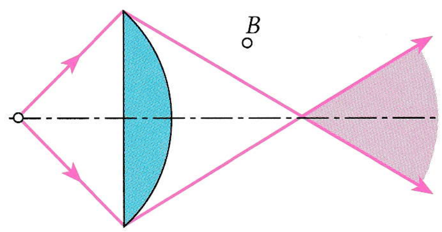


Figure 5

case the image of a point source is formed by a beam of rays whose marginal rays are limited by the diameter of the lens. These marginal rays also restrict the beam of rays emerging from the lens. We can see the image only if we are within the solid angle formed by the rays that passed through the point of intersection after passing through the lens.

It's interesting that there are points from which neither the source nor its image can be observed. For example, say we were at the point B: the source is covered by the lens, and we're outside the solid angle from which the image can be seen. In order to expand the region where the real image can be observed, we can use a screen. If we place a screen where the refracted rays emerging from the lens intersect, the rays reflected from the screen image will travel in every direction.

This is how a movie is shown in a theater. If we try to watch a movie by looking into the objective of the projector, or by using a mirror instead of the screen, we'll see at best only a small fragment of the overall picture. Different parts of it will be visible from different places—not exactly what the director had in mind!

"KALEIDOSCOPE" CONTINUED FROM PAGE 34

curious relation (Bergmann's rule): as we move from the poles to the equator, the size of warm-blooded animals of the same or similar species decreases. Why? One of the important factors determining the energy balance in an organism is heat loss, so animals in polar regions "try" to decrease it by any means possible. Because mammals are warm-blooded and their temperatures don't change, they can't decrease heat losses by lowering their body temperatures. However, increasing the body's dimensions lets them decrease the relative (that is, per unit mass) surface that transfers heat. Since heat transfer is higher in a cold climate, a larger size is more important for animals in polar regions, accounting for Bergmann's rule.

. . . hummingbirds with a mass of 3–5 g are the smallest birds, and their heat losses are particularly high. To keep their body temperature constant at night, the hummingbird must store up fat or glycogen during the day. However, this would be extremely inconvenient for the bird, since it would increase its body mass in the daytime and lead to higher energy expenditures when flying. In addition, the conversion of the original carbohydrates to stored, energy-rich substances also requires energy. Also, the hummingbird's method of feeding—hovering over a flower—consumes quite a bit of energy. In the course of evolution, the hummingbird was faced with two options: get bigger, or reduce nocturnal heat losses. As a result, their record (among birds) high body temperature of 43–45°C is maintained only during the day. At night it drops all the way down to 10–20°C. This explains why hummingbirds live in the tropics—otherwise they would experience large heat losses around the clock and couldn't have the normal metabolism of warm-blooded animals.

ANSWERS, HINTS & SOLUTIONS ON PAGE 59

PHYSICS CONTEST

Laser levitation

"Suspend here and everywhere, eternal float of solution!"
—Walt Whitman, Leaves of Grass

by Arthur Eisenkraft and Larry D. Kirkpatrick

HOW CAN SOMEONE LEVItate an object? Magicians do it all the time. Can physicists do it as well? The easiest technique is to attach a string to the object and secure the string to the ceiling. The weight of the object is balanced by the tension in the string. If the suspended object is a magnet, then a second magnet can keep it in place. A third technique is to shoot pellets at the object so that the force of the pellets balances the weight of the object.

Let's assume that the object we wish to suspend is a rectangular box oriented so that its bottom is horizontal. If we shoot pellets vertically upward at the box, the pellets just provide an average force on the box that is equal to its weight. If the pellets rebound from the box downward with the same speed, then the momentum change of each pellet is given by

$$\Delta p_{\text{pellet}} = 2mv_0,$$

where v_0 is the initial speed of the pellets and m is the mass of each pellet. The impulse-momentum theorem and Newton's third law tell us that the beam of pellets exerts a force on the box equal to

$$F_{\text{box}} = R\Delta p_{\text{pellet}},$$

where R is the number of pellets hitting the box each second.

We can get a feeling for the problem by solving it with some appropriate values. If the pellet gun shoots 5 pellets per second, and each of these 2-g pellets hits the box with a speed of 50 m/s and rebounds with the same speed, what is the heaviest box that can remain suspended? Let's work it through:

$$p_{\text{pellet}} = mv = (2 \cdot 10^{-3}\ \text{kg})(50\ \text{m/s}) = 0.1\ \text{kg m/s},$$

$$\Delta p_{\text{pellet}} = 0.2\ \text{kg m/s},$$

$$F = (5\ \text{pellets/s})(0.2\ \text{kg m/s}) = 1\ \text{N}.$$

Therefore, a 0.1-kg box can be suspended with these high-speed pellets.

A. The problem becomes more challenging to solve if the pellets hit the box at an angle. Assume that the pellets are identical to those in the example, but that they hit the box at an angle of 53° from the vertical. Once again, the pellets rebound at the same speed (50 m/s) and at the same angle. (Assume that the pellets hit at random orientations about the vertical so that there is no horizontal component of the net force.) What is the heaviest box that can remain suspended?

What do we do if the object to be suspended is so small that its weight is of the order of 10^{-10} newtons? If the object is transparent, it can be levitated by a laser beam! How to do this was one of three theoretical problems that were given to students who participated in the XXIV International Physics Olympiad, which was hosted in the United States in July 1993. This theoretical problem was created by Charles Holbrow of Colgate University. We have adapted it for *Quantum* readers.[1]

By means of refraction a strong laser beam can exert appreciable forces on small transparent objects. To see that this is so, consider a small glass triangular prism with an apex angle $A = \pi - 2\alpha$, a base of length $2h$, and a width w. The prism has an index of refraction n and a mass density ρ.

Assume that the prism is placed in a laser beam aimed horizontally in the x-direction. (Throughout this problem assume that the prism does not rotate—that is, its apex always points opposite to the direction of the laser beam, its triangular faces

[1]The entire XXIV International Physics Olympiad Examination has been published in *Physics Today* (November 1993) in an article by Anthony P. French, chair of the examination committee.

Art by Tomas Bunk


ELEVATOR
Tomas Bunk

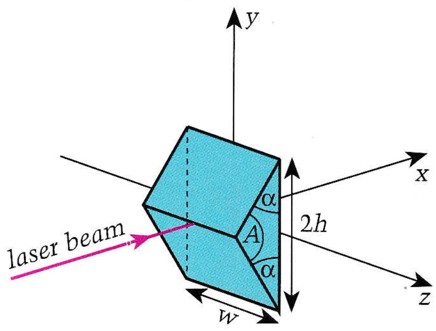


Figure 1

are parallel to the xy-plane, and its base is parallel to the yz-plane, as shown in figure 1.) Take the index of refraction of the surrounding air to be $n_{air} = 1$. Assume that the faces of the prism are coated with an antireflective coating so that no reflection occurs. The momentum of a photon is given by $p = E/c$.

The laser beam has an intensity that is uniform across its width in the z-direction but falls off linearly with the vertical distance y from the x-axis such that it has a maximum value I_0 at $y = 0$ and falls to zero at $y = \pm 4h$ (fig. 2).

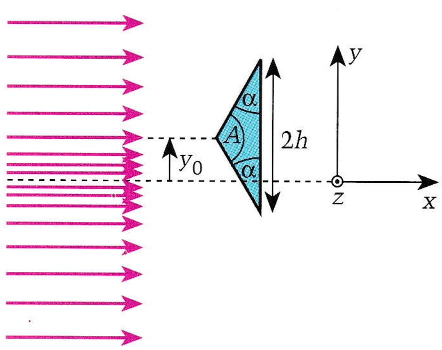


Figure 2

B. Write equations from which the angle θ (see figure 3) may be determined in terms of α and n for the case when the laser beam strikes the upper face of the prism.

C. Express, in terms of I_0, θ, h, w, and y_0 the x- and y-components of the net force exerted on the prism by the laser light when the apex of the prism is displaced a distance y_0 from the x-axis, where $h \le y_0 \le 3h$. If we want the prism to be suspended, should the prism be placed above or below the axis of the laser beam?

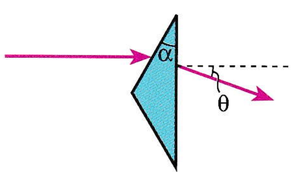


Figure 3

D. Plot graphs of the values of the horizontal and vertical components of force as functions of the vertical displacement y_0.

E. Suppose that the laser beam is 1 mm wide in the z-direction and 80 μm thick (in the y-direction). The prism has the following characteristics: $\alpha = 30°$, $h = 10$ μm, $n = 1.5$, $w = 1$ mm, and $\rho = 2.5$ g/cm^3. How many watts of power would be required to balance this prism against the pull of gravity when the apex of the prism is at a distance $y_0 = 2h = 20$ μm?

This problem is certainly difficult enough. Olympiad students from 42 countries took the problem one step further and solved parts C, D, and E for prism positions where $y_0 < h$! And some of them correctly completed this analysis within the allocated time of 100 minutes!

Please send your solutions to *Quantum*, 1840 Wilson Boulevard, Arlington VA 22201 within a month after receipt of this issue. The best solutions will receive special certificates from *Quantum*.

Electricity in the air

In the November/December issue of *Quantum* we asked our readers to use Gauss's law to examine the electric field near the Earth's surface. We will follow the solution provided at the International Physics Olympiad held in the United States in July 1993.

Part A of our problem was to find the total charge and charge density on the Earth's surface given the electric field near the surface. We begin by assuming that we have a spherical gaussian surface that is only slightly above the Earth's surface. Therefore, the radius of this surface is R. Because the electric field points radially, the total electric flux through this surface is just the product of the surface area of the sphere A and the electric field E_0. Gauss's law tells us that

$$-E_0 A = \frac{Q_0}{\varepsilon_0},$$

where Q_0 is the total charge enclosed by the surface and the minus sign is included because the electric field is directed into the sphere. Because

$$Q_0 = \sigma_0 A,$$

where σ_0 is the Earth's surface charge density, we can solve for either the charge density or the total charge. Let's find the charge density:

$$\begin{aligned}\sigma_0 &= -\varepsilon_0 E_0 \\ &= \left(-8.85 \cdot 10^{-12} \frac{\mathrm{C}^2}{\mathrm{N \cdot m^2}}\right)\left(150 \frac{\mathrm{N}}{\mathrm{C}}\right) \\ &= -1.33 \cdot 10^{-9} \frac{\mathrm{C}}{\mathrm{m}^2},\end{aligned}$$

where we have replaced the units V/m by N/C. The minus sign tells us that the charge on the Earth is negative, which we also know from the direction of the electric field. We can now find the total charge on the Earth:

$$Q_0 = \sigma_0 A = \sigma_0 4\pi R^2 = -6.85 \cdot 10^5 \text{ C}.$$

Part B required our readers to calculate the average net charge per cubic meter of the atmosphere given the electric field at a height of 100 m. Many students at the International Physics Olympiad solved this part of the problem by considering the gaussian surface to consist of two concentric spheres, one with a radius R and the second with a radius $R + h$ with $h = 100$ m. However, since $R << h$, the Earth's surface is relatively flat on the scale of the problem. Therefore, it's simpler to

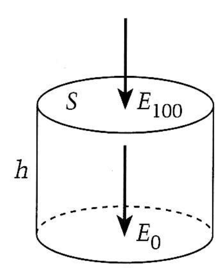


Figure 4

consider a cylinder with a cross-sectional area S and a height h sitting just above the Earth's surface, as shown in figure 4.

The walls of the cylinder do not contribute to the electric flux, because the electric field is parallel to the walls. Therefore, Gauss's law tells us

$$S(E_0 - E_{100}) = \frac{Q_{\text{enc}}}{\varepsilon_0} = \frac{\rho S h}{\varepsilon_0},$$

where ρ is the average charge density inside the cylinder and the contribution of E_{100} to the flux is negative. Using the data in the problem, this yields

$$\rho = \frac{\varepsilon_0}{h}(E_0 - E_{100}) = 4.42 \cdot 10^{-12} \frac{\text{C}}{\text{m}^3}.$$

Notice that the charge density is positive.

To solve part C, we first note that under the influence of the electric field, the positive ions move downward and the negative ions move upward. Therefore, only the positive ions can contribute to the cancellation of the surface charge density. The current per unit area j is given by

$$j = n_+ q v = (1.44 \cdot 10^{-14})E,$$

where we have used the values and relationship given in the statement of the problem. Note that the constant must have units of A/V · m.

Now, j is the rate of change of the surface charge density $(\Delta\sigma/\Delta t)$, and $E = -\sigma/\varepsilon_0$ from part A. Therefore,

$$\frac{\Delta\sigma}{\Delta t} = -1.63 \cdot 10^{-3}\sigma = -\frac{1}{613}\sigma.$$

This is the same relationship that we encounter in radioactive decay. Therefore, its solution is an exponential decrease of σ with time:

$$\sigma(t) = \sigma_0 e^{-t/\tau},$$

with $\tau = 613$ s. Putting $\sigma(t) = \sigma_0/2$ gives

$$t = \tau \ln 2 = 425 \text{ s} \cong 7 \text{ min}.$$

QUANTUM SMILES

A mathematical handbook with no figures

A little book that gives endless "dolgics" of reading pleasure to children of all ages

by Yuly Danilov

LEGEND HAS IT THAT WHEN the famous American puzzlemeister Sam Loyd invented Taquin (or "15"), it was greeted with almost disastrous enthusiasm. Farmers stopped farming, shopkeepers forgot to open their shops in the morning, government bureaucrats stood all night long under street lamps, trying in vain to solve a puzzle that looked simple but was really unsolvable: to make two small squares numbered 14 and 15 change places without taking them out of the frame.

Something like that happened, and continues to happen, with those lucky enough to get their hands on a copy of a little book by the poet Grigory Oster with the seemingly innocent, but actually subtly ironic, sarcastic, and even insidious title *Problem Book*. Without paying attention to the subtitle—"A Mathematical Handbook with No Figures," which to some extent should warn of impending danger,[1] you credulously open the book and promptly lose yourself in a world of curious characters struggling with unmathematical problems: captain Flint, boatswain Fedya, sad uncle Borya, little baby Kuzya, the she-goat Lucy, the scientist Innokenty of world renown, a criminal, the criminal's grandma, a cannibal, kittens, ducklings, an elephant, persons from other planets, an octopus, Bryaka, Mryaka, Slyunik, Hryamzik, and many others. It's an enchanting blend of fairy tale, thriller, adventure novel, and problem book! Most people can't tear themselves away from the *Problem Book* until they've read it cover to cover.

The book has led to untold dramas, and even tragedies. People are late for work, a son argues with his father about who will read the book first, but while they struggle on, the boy's mother picks up the *Problem Book* and won't let go until she's finished. Hoary scholars greet each other like the courtiers in Hans Christian Andersen's fairy tale "The Emperor and the Nightingale"[2] and, interrupting one another, recite their favorite problems from Oster's little book. And it's easy to understand why. How would *you* react to "problems" like these (the first ones in the *Problem Book*)?

1. In the tiny hold of a pirate ship Captain Flint and boatswain Fedya divided one and the same dividend by different divisors: Captain Flint, with a dark smirk, divided by 153, and boatswain Fedya, with a pleasant smile, divided by 8. Boatswain Fedya got a quotient of 612. What quotient did Flint get?

2. The personal parrot of Captain Flint learned 1,567 swear words in different languages: 271 swear words in English, 352 in French, and 127 in Spanish. The rest of the swear words the parrot got from the great and powerful Russian language. How many swear words did Captain Flint's personal parrot get from the Russian language?

3. Ten pirates divided among themselves in equal shares 129 captured maidens, and the rest were put in a boat and sent home to their parents. How many captured maidens were sent back to their parents?

[1]The subtitle—Ненаглядное пособие по математике—contains a rather outrageous pun. Наглядный means "visual" (something you can easily understand by looking at it); while ненаглядный—with the negative prefix не added—usually means "beloved" (someone you can't see enough of)!

[2]For those who haven't reread Andersen for a while, I'll explain: when three persons would meet, instead of "Good day" the first would say "Nigh-," the second "-tin-," and the third "-gale"—so great was their admiration for the mechanical nightingale presented to the emperor.

Art by Yury Vashchenko

Bryaka with marphoophochka on the first thing that came in hands...

How many times was poor Marphoophochka GRABBED by the tail and fryacked on the first thing that came in hands?????????????????

...and bryaka 9 times fryacked mryaka with the same marphoophochka on the first thing thatcame in hands

4. To find a hidden treasure, you need to start from the old oak and go 12 steps to the north, 5 steps to the south, 4 more steps to the north, and 11 more steps to the south. Find where the hidden treasure is.

5. The 14 best friends of Captain Flint, after various pirating exploits, were left with one leg each, and Captain Flint was left with two legs. How many legs, not including wooden ones, could be counted under the table when all 15 men sat down to eat breakfast?

So, do you like them? So do I. Many people have tried to imitate Oster, but all have fallen short. The *Problem Book* even got an official seal of approval: the Ministry of Education of Russia (no joke!) recommended the "mathematical handbook with no figures" as a textbook for schools! Apparently bureaucrats in the Ministry of Education actually have a sense of humor and appreciate a good joke.

And when, after they had read the *Problem Book* over and over, from cover to cover, and had learned their favorite problems by heart, charmed readers turned at last to the foreword (because a real reader reads a foreword only after diving deep into the book itself), they found that the author had no intention of misleading them. He had honestly warned everyone, children and adults alike, about his intentions. He even went the extra mile and wrote two (or possible three, depending on how you count) forewords. Here they are.

Forewords

"Would you like to hear a sadistic joke? One day a children's author comes to his readers and says: 'Dear children! I have written a new book for you—a mathematical problem book!' That's like getting a bowl of oatmeal instead of a pretty cake on your birthday. But to tell you the truth, the book you're looking at isn't exactly a problem book."

Here the text splits: one foreword for kids, one foreword for grown-ups.

For grown-ups: "Don't worry, don't worry, these are real problems. For second, third, and fourth grades, as a matter of fact. All of them are solvable and help cement the material studied in the classroom. But the main aim of the *Problem Book* isn't to cement any material, and the problems have nothing in common with what is called 'recreational mathematics.' I don't think these problems will elicit any professional interest from math olympiad champions. These problems are for those who don't like mathematics, who find it tedious and enervating to solve problems. Let them have their doubts about it!"

For kids: "Dear kids, this book is called *Problem Book* on purpose. It's so you can read it on your lap behind your desk. And if your teacher gets upset, just say: 'I don't understand. This book has been approved by the Ministry of Education.' "

Needless to say, the spoofing, high spirits, and inventiveness displayed by Oster are enjoyed greatly not only by those who don't like math but also by olympiad champions. And their teachers, and the teachers of their teachers—professors of mathematics—were filled with the warmest feelings (and the darkest envy) toward the author, because they didn't come up with the idea of creating such wonderful problems.

Now I think it's about time I offered those who haven't seen Grigory Oster's *Problem Book* a few more problems, taken almost at random.

Problems

1. Mryaka drooses poosics. To droose one poosic it takes Mryaka a half-dolgic. How many dolgics will Mryaka spend droosing 8 poosics?[3]

2. Mryaka and Bryaka droosed a poosic. Mryaka took for herself 2 farics, and Bryaka took 1. How many hroonechkas does Mryaka have, and how many does Bryaka have?

3. Bryaka and Mryaka quarreled. Mryaka kryacked Bryaka 7 times with a marfoofochka on his whatever, and Bryaka kryacked Mryaka 9 times with the same marfoofochka on her whatever. The question is, how many times was the poor marfoofochka grabbed by the tail and kryacked on somebody's whatever?

4. Bryaka hid 3 poosics under a coolyuk, shoved 5 poosics in a mlijechka, and buried 12 poosics in a gryazinuce. Mryaka went out to look for Bryaka's poosics, found 17 of them, and droosed them into hroonichkas. Where did Bryaka most likely find his undroosed poosics?

5. Mryaka and Bryaka came to a meadow and started to jump. Mryaka jumped on 7 lygs and Bryaka jumped on 8. How many lygs remained uncrushed if 39 lygs had been sitting in the grass, softly singing their pensive song?

6. Mryaka and Bryaka found a chalochka that was 9 tyatoosics long. Mryaka nibbled 4 tyatoosics and gave the rest to Bryaka. How long in dlinnics was the piece of chalochka that Bryaka got? (Keep in mind that there are 7 dlinnics in a tyatoosic.)

7. Mryaka, Bryaka, Slyunik, and Hryamzik walked and walked and walked, covering 200 dlinnics in 5 dolgics. How many dolgics will it take for them to cover 360 dlinnics if they walk and walk and walk with the same velocity?

8. If Hryamzik is called a slyunik, he starts to butt and doesn't stop until he has butted the offender 5 times with each horn. One day Bryaka called Hryamzik that very thing, and Hryamzik butted Bryaka 35 times. How many horns does Hryamzik have?

9. Every time they go out for a walk, Mryaka puts on 3 foofiras, while Bryaka puts on only 2. Both of them always return home buck naked. How many foofiras did Mryaka and Bryaka lose in one summer if it's known that Mryaka went for a walk 150 times and Bryaka 180 times this summer?

10. If Slyunik is teased, she begins to kick and doesn't settle down until she kicks the teaser 3 times with each leg. One day Mryaka called Slyunik a hryamzik, and Slyunik

[3]There are 3 farics in a poosic, 5 blyakas in a faric, and 2 hroonichkas in a blyaka. (You'll need this for the next problem.)

kicked Mryaka 27 times. How many legs does Slyunik have?

11. One day two numbers—5 and 3—came to a place where a lot of different differences were scattered about, and they started looking for their own. Find the difference of these numbers.

12. Once upon a time there lived two numbers—5 and 3. They had a sack of average size that they took with them wherever they went. When they came across something dangerous, they would quickly jump into the sack, close it from inside, and press against each other so tightly that sometimes they became one number. And then the sack would contain their sum. Find the sum of the numbers 5 and 3 in the sack.

Editor's note: We have retained the Russian flavor of the nonsense words in these problems. A true English translation might turn "dlinniki" into "longies," for instance, since длинный ("dlinny") means long, as in "long distance." But what is the poor translator to do with "dolgiki"? Долгий ("dolgy") also means long—but "time" this time! English readers can perhaps imagine a similar work written by Lewis Carroll or Ogden Nash, but we hope they experience some of the giddy silliness that Russians feel when they read Oster's exuberant coinages.

"HOW DO YOU FIGURE?" CONTINUED FROM PAGE 19

P114

Rings on the move. Two identical wire rings of radius r, each of mass m, are located in a homogeneous magnetic field B directed perpendicular to the plane of the rings and into the page (fig. 2). The rings make electrical contact at the points of intersection A and C. What is the velocity that each ring gains when the magnetic field is switched off? The electrical resistance of each ring is R and the angle $\alpha = \pi/3$. Neglect the self-inductance and mutual inductance of the rings, the displacements of the rings while field is turned off, and any frictional effects. (V. Mozhayev)

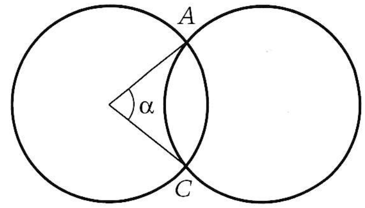


Figure 2

P115

Image vs. reality. A point source of light moves parallel to the principal axis of a converging lens with focal length F. Determine the distance of the source from the lens when the absolute value of the velocity of its image is equal to that of the source. The distance from the source to the principal axis is $H = F/4$. (A. Zilberman)

ANSWERS, HINTS & SOLUTIONS ON PAGE 54

Nine solutions to one problem

And integer angles galore

by Constantine Knop

THE PROBLEM THAT I'M GOing to discuss has appeared repeatedly in geometry problem books. Although its statement seems simple, it's very difficult to solve. This might be why most of the books give more or less the same solution, and not even the best one, in my view. But let's start with the statement.

Problem. *In an isosceles triangle ABC, AB = AC and the angle BAC measures 20°. Points D and E are taken on the sides AC and AB, respectively, such that angle ECB is 50° and angle DBC is 60°. Find the angle EDB.*

Before you go on, try to solve the problem on your own. Give yourself a few hours (or maybe minutes) to think it over. You'll find real pleasure . . . if you manage to find the answer.

The first solution. Draw segment DF parallel to BC with F on AB (fig. 1), draw CF, label as G the intersection of BD and CF, and draw GE. Clearly the triangle BGC is isosceles (with $BG = GC$) and, therefore, equilateral (since $\angle DBC = 60°$). Then the triangle GDF is equilateral as well. Further, we notice that $\angle BEC = 180° - \angle BCE - \angle CBE = 180° - 50° - 80° = 50°$ ($\angle CBE$ can be found from the given isosceles triangle ABC), so $\angle BEC = \angle BCE$, which means that triangle BGE is also isosceles ($BE = BC = BG$) with $EBG = 80° - 60° = 20°$ and $\angle EGB = 80°$. Now we find $\angle FGE = 180° - 60° - 80° = 40°$, and $\angle EFG = 40°$ (say, from triangle BGF). This reveals another isosceles triangle, FEG ($FE = EG$). Finally, by the SSS property, triangles EFD and EGD are congruent, and so

$$\angle EDB = \angle EDF = \frac{60°}{2} = 30°.$$

I want to point out one fact mentioned in this proof that will be repeatedly used in what follows: the triangle BCE is isosceles—that is, $BE = BC$.

Well, this solution, borrowed from a problem book, uses two additionally constructed points, and five triangles are under consideration. It couldn't be called too complex or too long, really, and yet it didn't seem very elegant or beautiful to me. So when, after thinking long and hard, and unsuccessfully, I finally found another solution, I was happy. The only thing that distressed me was that this solution was analytical (trigonometric) rather than geometric.

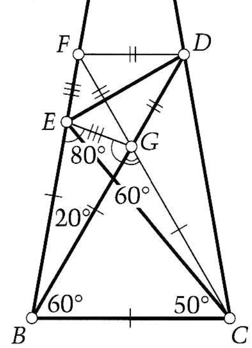


Figure 1

The second solution. Let x be the measure of the unknown angle EDB. Then $\angle BED = 160° - x$. By the Sine Law, from triangle BED we find $BD : BE = \sin(160° - x) : \sin x$, and from triangle BCD, $BD : BC = \sin 80° : \sin 40° = 2\cos 40°$ (since $\angle BDC = 180° - 60° - 80°$).

Using the aforementioned equality $BE = BC$, we get the equation

$$\frac{\sin(160°-x)}{\sin x} = 2\cos 40°.$$

Let's rework and solve it:

$$\sin(20° + x) = 2\cos 40° \cdot \sin x$$
$$= 2\cos(60° - 20°)\sin x;$$
$$\sin 20° \cos x + \cos 20° \sin x$$
$$= (\cos 20° + \sqrt{3}\sin 20°)\sin x;$$
$$\sin 20° \cos x = \sqrt{3}\sin 20° \sin x;$$
$$\tan x = \frac{1}{\sqrt{3}};$$
$$x = 30°.$$

Art by Yury Vashchenko

Trigonometry is a powerful and universal tool. But does our problem really have no other geometric solutions?

Fortunately, such a solution does exist—in fact, there are a number of them. I created the next two in several hours of leisure time.

The third solution. As with the first solution, I'll again try to prove that DE is the bisector of $\angle BDF$ (fig. 1). To this end, I'll create a triangle, one of whose angles will be BDF, with its incenter at E. Draw DH and BH parallel to CB and CD, respectively, to obtain a parallelogram $BCDH$ (fig. 2). Draw CG as we did in figure 1 (to make an equilateral triangle BCG). Now we have

(1) $BH = CD$ (by a property of the parallelogram);

(2) $BE = BC = CG$;

(3) $\angle HBE = \angle HBA = \angle BAC = 20°$, and $\angle GCD = 80° - 60° = 20°$, so $\angle HBE = \angle GCD$.

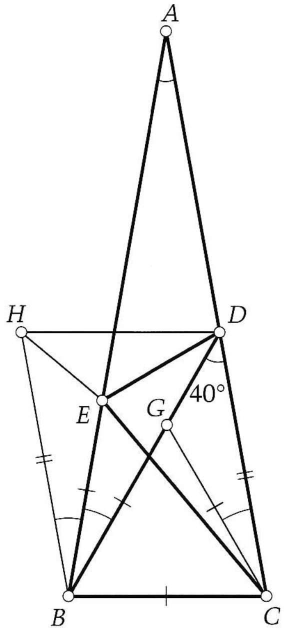


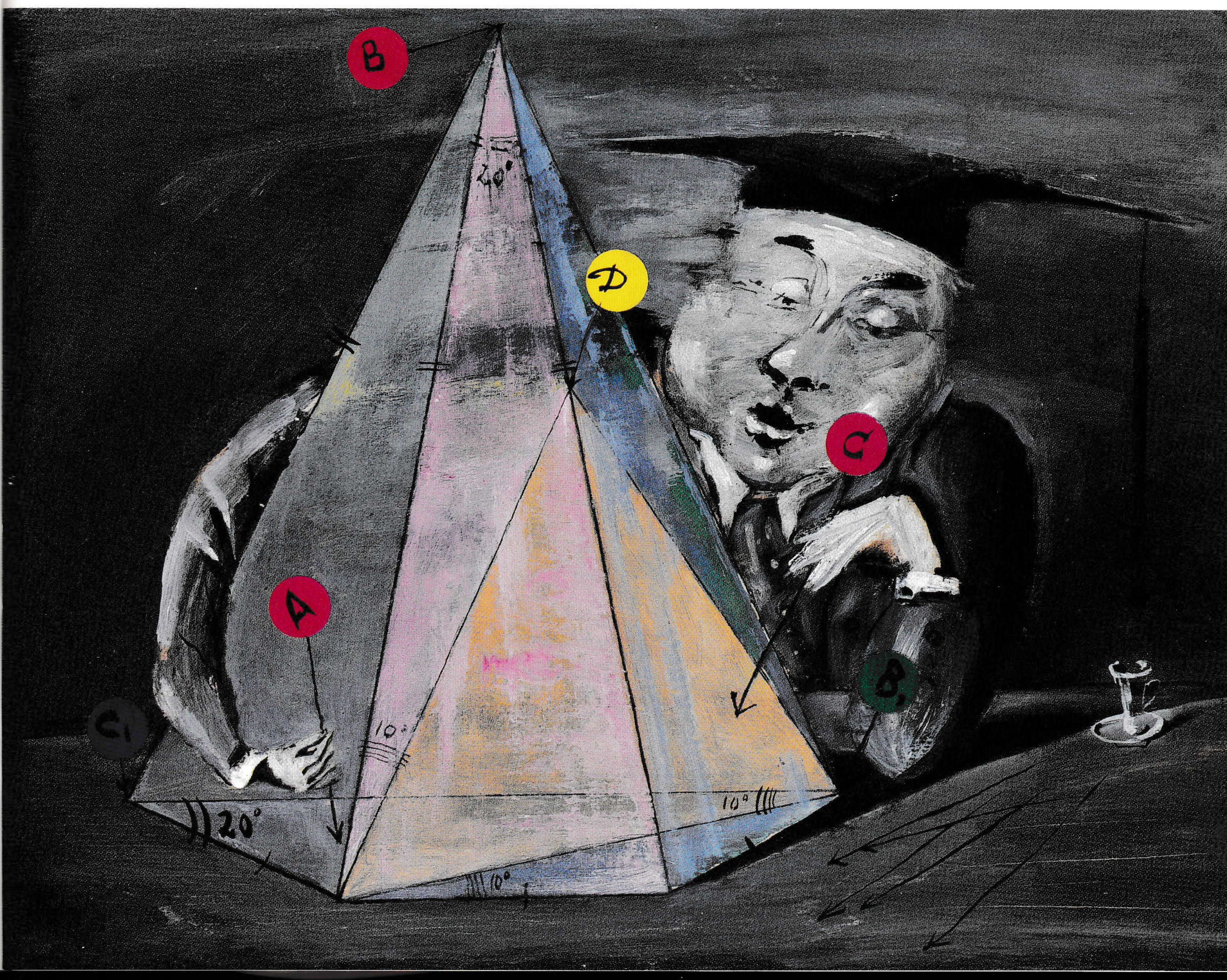
Figure 2

Therefore, the triangles BEH and CGD are congruent by the SAS property; consequently, $\angle BHE = \angle CDG = 40° = \frac{1}{2}\angle BCD = \frac{1}{2}\angle BHD$. It follows that HE bisects $\angle BHD$, and at the same time, BE bisects $\angle HBD$ (since $\angle HBE = \angle DBE = 20°$). Therefore, E is the center of the incircle of triangle BDH, and DE is the bisector of $\angle BDH$.

This was a different solution, but it's hardly simpler than the first one. The next solution seems more attractive to me.

The fourth solution. Mark point K on AC such that $\angle KBC = 20°$, and join it to B and E (see figure 3 on the next page).

Then . . . On second thought, why don't you try to finish this proof yourself?



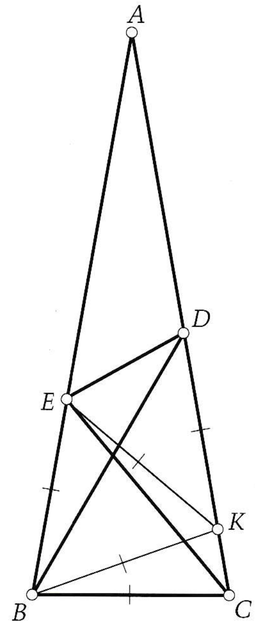


Figure 3

Exercise 1. Prove that the marked segments in figure 3 are congruent and use them to find the unknown angle.

After I found these solutions, I was obsessed with the idea of offering this problem to students at some serious math competition. These clever kids might discover something new! Unexpectedly, about a year ago my dream came true: the problem was proposed to the candidates for Ukraine's International Mathematics Olympiad team. And my collection of solutions grew by four new items. It's interesting that three of these solutions involve an auxiliary construction based on drawing the bisector of the angle B of the given triangle ABC. All three solutions use the fact that this line is also the perpendicular bisector of the segment CE (since $BC = BE$), but other than that they are surprisingly different.

The fifth solution (Maria Gelband). Let M be the reflection of E across AC (fig. 4). Then $CE = CM$ and $\angle ECM = 2\angle ECD = 60°$, so the triangle CEM is equilateral, $CM = EM$, and therefore M lies on the bisector mentioned above. Now we notice that the point D is the intersection of the perpendicular bisector of EM and BD, the bisector of the angle EBM of the triangle BEM ($\angle EBD = 20° = \frac{1}{2}$ BM). It follows that D lies on the circumcircle of BEM (if D_1 is the intersection of the circumcircle with the perpendicular bisector of EM, then $ED_1 = D_1M$, and the angles EBD_1 and MBD_1, inscribed in this circle, are subtended by congruent chords ED_1 and DM_1; so these angles are equal, which means that $D_1 = D$). Thus, the angles EDB and EMB are inscribed in this circumcircle and subtended by the same chord BE. So $\angle EDB = \angle EMB = \frac{1}{2}\angle EMC = 30°$, and we're done.

The sixth solution (Sergey Saprikin). Let the bisector of $\angle ABC$ intersect AC at T (fig. 5). Then $\angle ETB = \angle BTC$ (why?). But $\angle BTC = 180° - 40° - 80° = 60°$, so $\angle ETD = 60°$, and TD is the bisector of the exterior angle of triangle BET at T. On the other hand, as we've already seen, BD bisects angle EBT, so D is equidistant from the lines BA, BT, and ET, and, therefore, ED is the exterior bisector of triangle BET at E.

Note, by the way, that D is the excenter (center of the escribed circle) of this triangle. Now we find $\angle BED = \angle BET + \angle TED = \angle BET + (\frac{1}{2}180° - \angle BET) = 90° + \frac{1}{2}\angle BET = 130°$, because $\angle BET = \angle BCT = 80°$. Finally, from triangle BED we get what we want: $\angle EDB = 180° - 130° - 20° = 30°$.

In the last part of this proof we've actually proved the following property of the excenter D of an arbitrary triangle EBT that lies inside its angle EBT: $\angle EDB = \frac{1}{2}\angle ETB$ (which is independent of specific values of the angles).

The seventh solution (Alexey Borodin). Consider the circumcenter O of the triangle EDC. Since $EO = OC$, the line BO is just the bisector of the angle EBC used in the two previous proofs.

Exercise 2. Finish this proof using figure 6. (Hint: prove the congruence of triangles BED and BOD.)

Perhaps one of the most natural ways to tackle our problem is to notice that the measure of $\angle A$ of the given triangle is 20°—that is, one third of 60°—and try to make use of this observation. This idea is implemented in the next solution.

The eighth solution (Alexander Kornienko) (fig. 7). Reflect the given triangle about AB (to get ABC_1) and AC (to get ACB_1). Then $\angle AC_1E =$

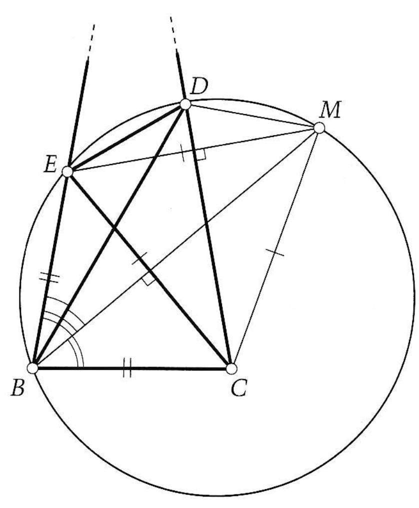


Figure 4

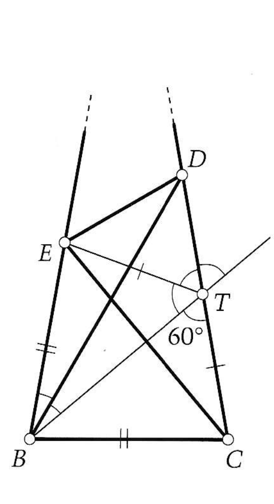


Figure 5

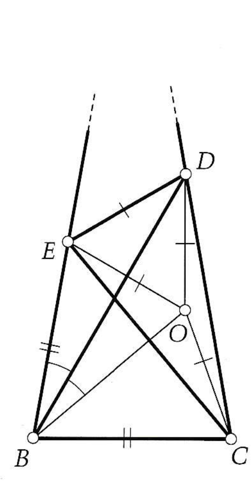


Figure 6

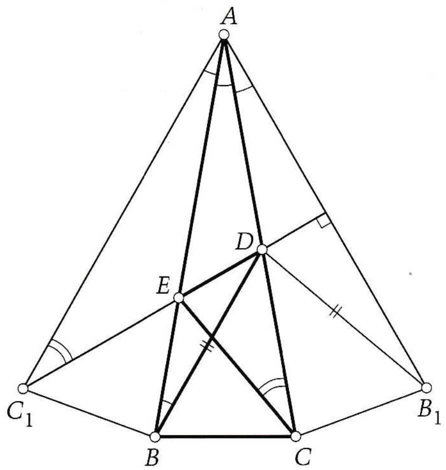


Figure 7

$\angle ACE = 30°$, and $\angle AC_1B_1 = 60°$ (since triangle AB_1C_1 is equilateral—$AC_1 = AB_1$, $\angle C_1AB_1 = 60°$), so C_1E bisects $\angle AC_1B_1$, which means that C_1E is the perpendicular bisector of AB_1. On the other hand, $AD = BD$ (because $\angle ABD = \angle BAD = 20°$), and $BD = B_1D$ (by construction). So D is equidistant from A and B_1 and, therefore, lies on line CE. Now the angle EDB can be found from triangle C_1BD, in which $\angle BC_1D = 80° - 30° = 50°$ and $\angle C_1BD = 80° + 20° = 100°$. We find that $\angle EDB = 100° - 50° - 80° = 30°$.

(In fact, the idea underlying this proof is to consider figure 7 a part of the regular 18-gon centered at A and segments C_1B, BC, and CB_1 as three consecutive sides. Segments C_1D and BD turn out to be parts of its two diagonals—which has to be proved, of course, but this enables us to quickly find the unknown angle.)

I hope you liked these eight solutions, full of many clever constructions and useful properties of triangles. "But where's the ninth one?" you ask. Why, I've left it for you to find!

I'll leave you with another problem about the same triangle (proposed by a ninth-grader, Sergey Yurin).

Exercise 3. In an isosceles triangle ABC, $AB = AC$, and $\angle A = 20°$. Point P is taken on the side AC such that $AP = BC$. Find the angle PBC.

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

A tale of one city

It was the best of times . . .
Problems from the International Mathematics Tournament of Towns in Beloretsk, Russia

by Andy Liu

THE ROAD NETWORK OF A certain city consists of a continuous chain of circles. At the point of tangency of two adjacent circles, the roads cross over as shown in figure 1, which illustrates the case with four circles.

A ring road is constructed around the city, and is integrated with the inner chain at various points. At each "integration point," the ring road is crossed over with the inner chain as shown in figure 2, which illustrates the case with two integration points.

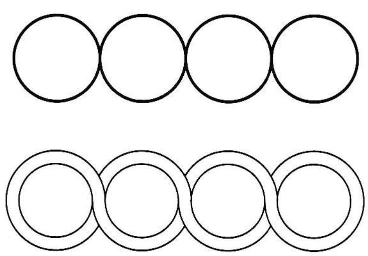


Figure 1

Note that in figure 2 the integrated network consists of two mutually inaccessible components. We call an integrated network regular if it consists of only one component, and irregular otherwise. We wish to find a necessary and sufficient condition for an integrated network to be regular.

This is the main question in one of five problems posed in a Problem-solving Workshop conducted in Beloretsk, in the Bashkirian Republic of Russia, from August 1 to August 9, 1993. The participants were mostly from the former Soviet bloc, but included representatives from England, Austria, Canada, Germany, and Colombia. They were high school students invited on the strength of their performances in the International Mathematics Tournament of Towns.[1]

Each of the five problems is very carefully constructed. It is divided into many questions, leading the solvers step-by-step to the main one. It also raises other related questions along the way. The problem we feature here is proposed by S. Loktev and M. Vialyj, based on a problem of Prof. V. Arnold.

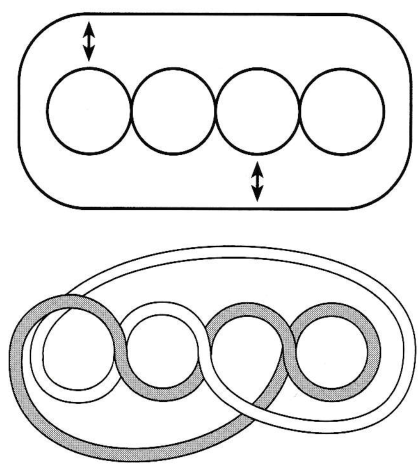


Figure 2

[1]For more on the Tournament of Towns, see the articles in the Happenings department in the January 1990 and November/December 1990 issues of *Quantum*.

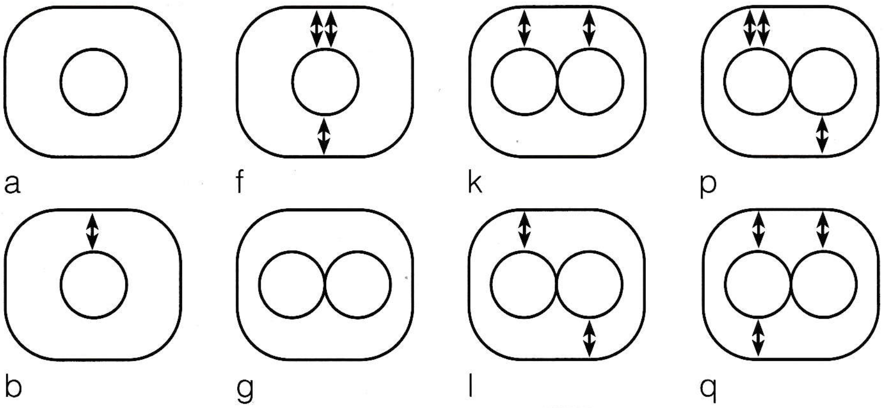
We begin our investigation by studying some small integrated networks.

Question 1. Which of the integrated networks in figure 3 are regular?

It may be observed that each of the irregular networks in figure 3 consists of two components. Is this true in general? In other words, how bad can an irregular network be?

Question 2. Is there an integrated network that consists of three or more components?

Some of the integrated networks in figure 3 have, at the end of the inner chains, circles without integration points. Clearly, the deletion of such circles does not affect regularity. From now on, we shall assume that they have been deleted. If the inner chain consists only of such circles, it will be reduced to the empty set. However, the integrated network is still considered to be irregular, although the ring road now constitutes the only component.



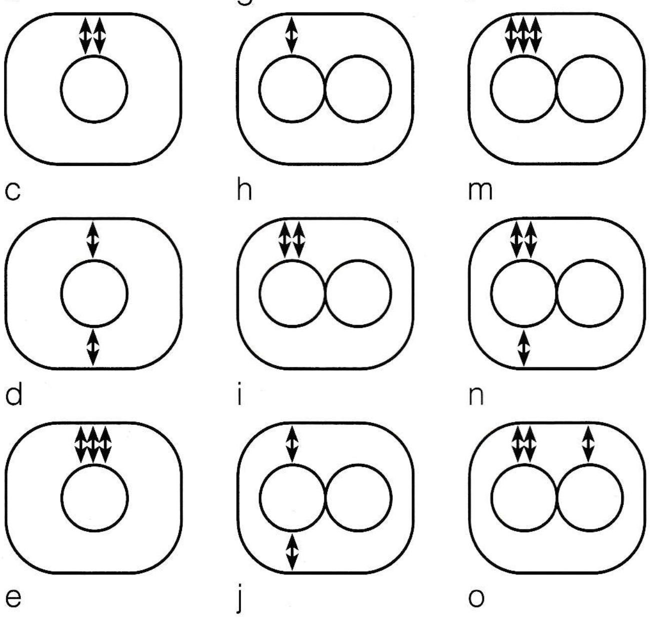


Figure 3

As we are interested only in the regularity of integrated networks, we can simplify them in many ways. The following is also motivated by observations of the simple cases in figure 3.

Question 3. Prove that neither of the following operations affect the regularity of an integrated network: (a) adding two integration points to a circle in the inner chain; (b) removing two integration points from a circle in the inner chain.

A simple but useful corollary of question 3 is the following. It makes figures 3d, 3f, 3j, 3l, 3n, 3p, and 3q irrelevant.

Question 4. Prove that in an integrated network, we can move all integration points in each circle of the inner chain to the same side of the ring road.

A more important corollary of question 3 is that we may assume that each circle of the inner chain has either one integration point or no integration points. We say that such an integrated network is normalized, and represent it by a sequence of 0's and 1's.

Question 5. Which of the following normalized networks are regular? (a) 101 (b) 111 (c) 1001 (d) 1011 (e) 1111 (f) 10001 (g) 10011 (h) 10101 (i) 10111 (j) 11011 (k) 11111.

The next result allows us to reduce every normalized network to one of three in figure 3—namely, 3a, 3b, and 3k.

Question 6. From a binary sequence, we delete all subsequences of the forms 00 and 111. We also replace all subsequences of the form 101 by 0. Prove that the reduced sequence represents a regular network if and only if the original one does.

It turns out that it is possible to determine whether a binary sequence represents a regular network or not without actually carrying out the reduction process in question 6. Consider the sequence as blocks of consecutive 1's separated by single 0's. For example, 111001101 consists of four blocks of consecutive 1's, with 3, 0, 2 and 1 of them in the respective blocks. The alternate sum of these numbers is $3 - 0 + 2 - 1 = 4$, and the alternate sum of any binary sequence can be defined in the same way.

Question 7. Prove that a binary sequence represents a regular network if and only if its alternate sum is not divisible by 3.

Amazingly, what we have gone through so far constitutes only part of the problem of Loktev–Vialyj. A more general setting replaces the city with a metropolis, where the road network is not a chain but may look like figure 4. There are also other questions one can ask about integrated networks.

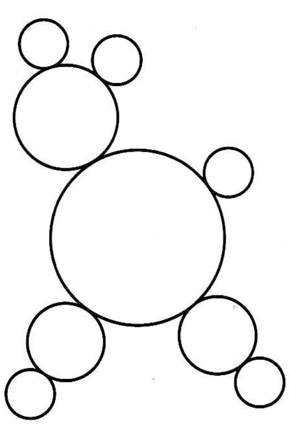


Figure 4

The top prize winner for this problem at the workshop was Clemens Heuberger, a graduating high school student from Graz, Austria. Other winners were M. Alekhnovich and M. Ostrovsky (joint effort), and I. Nykonov. All three were high school students from Moscow, Russia.

The principal organizer of the workshop was Prof. Nikolay Konstantinov of the Independent University of Moscow, a recent winner of the prestigious Erdős Award from the World Federation of National Mathematics Competitions. The assembly was honored by the presence of Prof. Nikolay Vasilyev, chairman of the problem committee of the Tournament of Towns, and Prof. A. A. Yegorov, an editor of *Kvant*, the sister journal of

Quantum. Both posed problems for the workshop.[2]

Most of the group assembled at Moscow on July 30, 1993, and took a 36-hour train ride across European Russia to Beloretsk. Five problems were distributed on board to whet the participants' appetites. We conclude with one of them.

Training Problem. (a) Let a and b be integers and p a prime number. Prove that

$$a \equiv b \pmod p \Rightarrow a^p \equiv b^p \pmod{p^2}.$$

(b) $ABCD$ is a parallelogram (fig. 5). The circumcircle of triangle BAD, with center O, cuts the extensions of BC, AC, DC, and AO at K, L, M, and N, respectively. (1) Prove that N is the circumcenter of triangle KCM. (2) Express the length of LC in terms of $a = KL$ and $b = LM$.

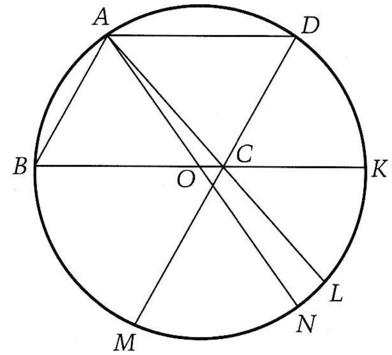


Figure 5

[2]See "A Mathematical Journey" by A. Liu in *Crux Mathematicorum*, January 1994 (pp. 1–5).

ANSWERS, HINTS & SOLUTIONS IN THE NEXT ISSUE

Bulletin Board

Budapest Semesters in Mathematics

Initiated by Paul Erdős, László Lovász, and Vera T. Sós in 1984, the Budapest Semesters in Mathematics program offers a unique opportunity to North American undergraduates for a semester or year of study abroad, in one of the most advanced mathematical centers of the world. Through this program, mathematics and computer science majors in their junior/senior years can take a variety of courses in all areas of mathematics under the tutelage of eminent Hungarian scholar–teachers, most of whom have had years of teaching experience in North America. The classes are small, all of the courses are conducted in English, and the credits are transferable to the students' home institutions.

The classes are held on the International College Campus of the Technical University of Budapest, which is near the historic city center. The accommodations are excellent, the living costs are modest, and the tuition is most reasonable. The fall semester usually begins during the first week of September and ends before Christmas, while the spring semester starts in early February and ends in late May. There is a brief orientation program prior to the start of semesters, and one can also take part in an optional two-week language program prior to the beginning of the regular program. Arrangements can be made for taking the Putnam Examination in Budapest and/or for taking the Graduate Record Examination in Belgrade or Vienna, both about four hours by train from Budapest.

During the past ten years, hundreds of North American students, representing more than 120 universities, took advantage of Hungary's long tradition of excellence in mathematical education and in creative problem solving through this unique program. Many of them have stayed in close contact with one another and the faculty/organizers of the program.

To learn more about the Budapest Semesters in Mathematics, please contact its North American Director, Professor Paul D. Humke, at Saint Olaf College (telephone: 800 277-0434 or 507 646-3113; e-mail: humke@stolaf.edu). You can also obtain a copy of the application materials and a brochure describing the program via anonymous ftp (ftp.stolaf.edu). The registration deadline for the fall semester is April 30; early applications are encouraged, but late ones are sometimes accepted. The size of each class is usually around 30.

—George Berzsenyi

Quantum vs. Kvant

Student readers of *Quantum* and its sister magazine *Kvant* will go head to head in a friendly mathematics competition in Moscow in the summer of 1995. The event will be sponsored by American University in Moscow and administered by the two magazines. Participants will be selected on the basis of answers to several rounds of questions published here and in *Kvant*. According to Edward Lozansky, *Quantum*'s international consultant, a team of five US high school students will be chosen for the one-week, all-expense-paid trip to Moscow.

Watch this space in the months ahead for further details.

Duracell Scholarship winners

Seventeen-year-old Tracy Phillips of Long Beach, New York, was the first-place winner in the Duracell/NSTA Scholarship Competition. She invented Money Talks, an electronic device neatly built into a wallet that helps blind people distinguish between different values of paper money. After a bill is placed in the wallet, the device "talks," giving the denomination. How does it know what to say? It uses an infrared light emitter/detector that lets varying amounts of light pass through the printed patterns of a bill, identifying key points that distinguish the bill's denomination. The amount of light passing through each point is

CONTINUED ON PAGE 61

Criss×cross science

by David R. Martin

1	2	3	4	■	5	6	7	8	9	■	10	11	12	13
14				■	15					■	16			
17				■	18					■	19			
20				21			■	■	22	23				
■	■	■	24			■	25	26				■	■	■
27	28	29			■	30			■	31		32	33	34
35			■	36	37				38	■	39			
40			■	41			■	42		43	■	44		
45			46	■	47		48				■	49		
50				51	■	52			■	53	54			
■	■	■	55		56			■	57			■	■	■
58	59	60				■	■	61				62	63	64
65				■	66	67	68			■	69			
70				■	71					■	72			
73				■	74					■	75			

Across

1 Group of legislators
5 Common electric circuit element
10 Wild duck
14 Of aircraft
15 Decrease
16 Stare flirtatiously
17 Birds
18 Wooden pin
19 Hardy cabbage
20 Semiconductor
22 Rotating vector
24 Negative prefix
25 Type of salt
27 Hawthorne's home
30 1051
31 1958 Physiology Nobelist Edward L. ___
35 Aunt in Spain
36 Tooth covering
39 "Citizen ___"
40 Intuitive letters
41 Crag
42 4.19 joules: abbr.
44 One who transmits: abbr.
45 Landed
47 Eye's outer coat
49 ___ douloureux
50 Elementary particle
52 Dawn goddess
53 Scandinavian
55 Sways
57 Understand
58 Am. inventor
61 Fluid acceleration machine
65 Fashion designer ___ Gernreich
66 Disulphuric(vi) acid
69 First garden
70 European capital
71 Scoop
72 Peruse
73 British gun
74 Deforming force
75 Plant fluids

Down

1 Sheep cries
2 ___-Civita symbol
3 City on the Oka
4 Trigonometric function
5 Radioactive gas
6 Black
7 Fundamental physical truth
8 Devoured
9 Dog's cries
10 Plasma confining device
11 1949 Physiology Nobelist Moniz, Antonio ___
12 Chemical prefix
13 Sly smile
21 Halley's ___
23 Highly radioactive
25 Graceful tree
26 Parts
27 Vaporized water
28 Corridor
29 ___ lazuli (blue rock)
30 French anthropologist ___ Mauss
32 Flavor
33 Untied
34 U.S. dancer ___ Cunningham
37 1 and 101, e.g.: abbr.
38 Roman spirit
43 Monochromatic radiation source
46 Twisting deformation
48 ___ Angeles
51 Hydrocarbon prefix
54 Magnetic flux units
56 Compounds containing –CH=C(OH)–
57 Ancient Mesopotamian region
58 God of love
59 Dirt
60 Unloaded engine speed
61 River near Mexico City
62 Concept
63 Lowest tide
64 Tails
67 ___-di-dah
68 Dutch town

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE MARCH/APRIL PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M111

(a) The equality can be proved by induction. It's obviously true for $n = 1$: $(1 \cdot 2!)/2 = 3!/2 - 2 = 1$. Assuming that it's already been proved for $n = k$, let's make sure it's true for $n = k + 1$. Using the inductive assumption, we get

$$\frac{1\cdot 2!}{2}+\frac{2\cdot 3!}{2^2}+\cdots+\frac{k(k+1)!}{2^k}+\frac{(k+1)(k+2)!}{2^{k+1}}$$
$$=\frac{(k+2)!}{2^k}+\frac{(k+1)(k+2)!}{2^{k+1}}-2$$
$$=\frac{(2+k+1)(k+2)!}{2^{k+1}}-2$$
$$=\frac{(k+3)!}{2^{k+1}}-2.$$

(b) The sum equals $(n + 3)!/3^n - 6$. This can also be proved by induction, as in part (a), but we'll give a somewhat modified argument with "telescoping" sums. Note that any term of the given sum can be written as the difference

$$\frac{(k+3)!}{3^k}-\frac{(k+2)!}{3^{k-1}}=\frac{(k+2)!(k+3-3)}{3^k}=\frac{k(k+2)!}{3^k}.$$

Writing out all these differences for $k = 1, 2, \ldots, n$ and adding them up we get, after canceling out terms of opposite sign,

$$\frac{(n+3)!}{3^n}-3!=\frac{1\cdot 3!}{3}+\frac{2\cdot 4!}{3^2}+\cdots+\frac{n(n+2)!}{3^n}.$$

Similarly, for any positive integer d, we have

$$\frac{(k+d)!}{d^k}-\frac{(k+d-1)!}{d^{k-1}}=\frac{k(k+d-1)!}{d^k}$$

from which the following generalization of statements (a) and (b) can be derived:

$$\frac{1\cdot d!}{d}+\frac{2(1+d)!}{d^2}+\cdots+\frac{n(n+d-1)!}{d^n}=\frac{(n+d)!}{d^n}-d!.$$

(N. Vasilyev, V. Zhokha)

M112

Let L be the point of intersection of the diagonal BD and the circle KDQ (fig. 1). We have to prove that L lies on the second circle KBP. For the sake of definiteness, we'll assume that points K and Q are on different sides of BD; then K and P are on the same side of BD, because the segment PQ intersects BD. Then the angles DQK and DLK are inscribed in the same circle, subtend the same chord DK, and have their vertices on the same side of the chord, so $\angle DQK = \angle DLK$. On the other hand, since $DC \parallel AB$, $\angle DQK = \angle KPB$. Therefore,

$$\angle KPB + \angle KLB = \angle DQK + (180^\circ - \angle DLK) = 180^\circ,$$

which means that points B, P, K, L are concyclic—that is, L lies on the circle KBP.

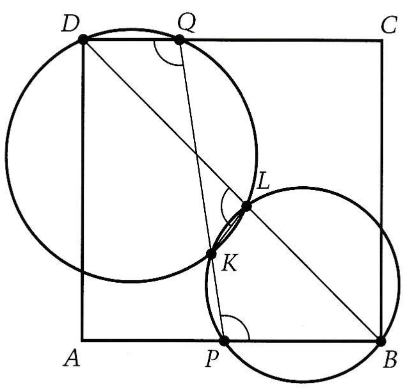


Figure 1

M113

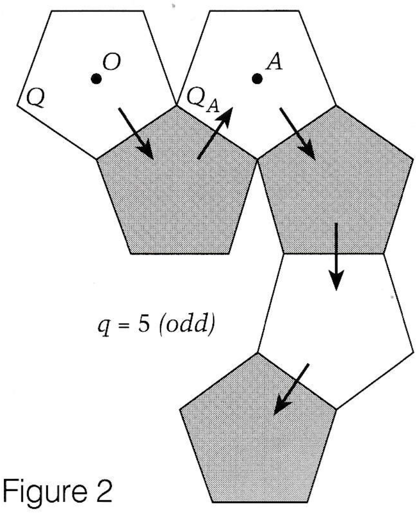
The answer is yes. If the first player names any three different integers whose sum is zero (say, 1, –3, 2), then regardless of the order chosen by the second player, the resulting equation $ax^2 + bx + c = 0$ will have a root $x_1 = 1$ (because $a \cdot 1^2 + b \cdot 1 + c = 0$) and a different second root $x_2 = c/a \neq 1$ (since the product x_1x_2 is always equal to c/a).

If you liked this problem, here is a much more elaborate extension.

Two players create an equation of the form $x^3 + *x^2 + *x + * = 0$. The first one names a number, the second writes it in place of any of the asterisks; then the first player names a second number and the other player inserts it in place of any of the two remaining asterisks; finally, the first player replaces the last asterisk with some number. Can the first player ensure that the resulting equation has three distinct integer roots?

M114

The specific problems (a) and (b) have specific solutions—based, for instance, on the Fractional Parts



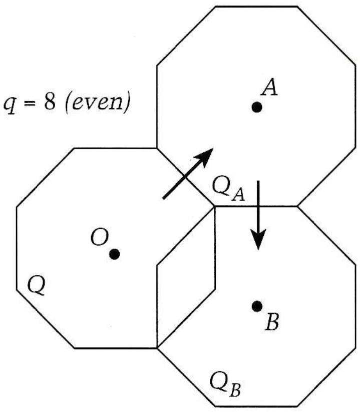


Figure 2

Theorem from "Ones Up Front in Powers of Two" in the November/ December 1993 issue of *Quantum*. But we'll consider the general problem (c) right away. The solution below involves some ideas from the solution to M100 in that same issue.

First let's introduce a number of convenient notations and terms. Let Q be the fixed initial position of the given regular q-gon and O its center. Denote by Q_A the q-gon (with center A) obtained from Q under translation by vector $\overrightarrow{OA}$. Two points A and B are said to be *connected* if the polygon Q_A can be rolled into the polygon Q_B. The sequence of its successive positions in this series of rolls will be called the *track AB*. (Notice that if q is odd, then any track AB necessarily consists of an odd number of polygons (including Q_A and Q_B), because in this case a single rolling yields a polygon turned 180° with respect to the initial one, so only an even number of rollings restores the initial orientation of the polygon—see figure 2.) The points connected to O will be called *attainable*. Clearly any two attainable points A and B are connected to each other (we can construct a track from A to O to B); conversely, any point B connected to an attainable point A is attainable itself (there is a track from O to A to B).

Next we prove two important properties of attainable points.

(1) *If points A and B are attainable and C is obtained from B under a rotation r through 360°/q about A, then C is attainable.*

Indeed, the rotation r takes Q_A into itself and any track AB (which exists because A and B are connected) into a track AC (fig. 3). So C is connected to an attainable point A.

(2) *If A, B, and C are attainable, then the translation t by vector* $\overrightarrow{AB}$ *takes C into an attainable point D.*

To prove this, it suffices to construct a track AC and notice that our translation takes Q_A and Q_C into Q_B and Q_D, and the constructed track into a track BD. So D is connected to an attainable point B.

Of course, this statement is true for the translation by $\overrightarrow{BA}$ as well.

Now let's prove that for $q \geq 7$ any circle contains an attainable point inside it. Let ε be the radius of the circle. Take any two attainable points A and B and construct C as specified in statement (1) (fig. 4). Then $BC/AB = k = 2 \sin (180°/q) < 2 \sin 30° = 1$. Applying the same construction to B and C instead of A and B, we'll get a pair of attainable points C and D such that $CD = kBC = k^2AB$ (fig. 4). Then we repeat the construction with C and D, and so on, until we get attainable points X and Y such that $XY = k^nAB < \varepsilon$ and also the next point Z in the sequence $A, B, C, D, \ldots, X, \ldots$ (such that $XZ = XY$ and $\angle ZXY = 360°/q$). Repeatedly applying statement (2) to points X, Y, O and X, Z, O, we see that any number of translations by the vectors $\overrightarrow{XY}$, $-\overrightarrow{XY} = \overrightarrow{YX}$, or $\pm\overrightarrow{XZ}$—that is, a translation by vector $n\overrightarrow{XY} + m\overrightarrow{XZ}$ with any integer m and n—takes O into an attainable point. These points make up a grid of rhombi with side length less than ε (fig. 5).

The center of the given circle falls into one of the rhombi, and it's easy to see that its distance from one of the vertices of this rhombus is smaller than ε. So this vertex lies inside the circle, which proves the statement of the problem for $q \geq 7$, because all the nodes of the grid are attainable.

For $q = 5$ we can take two attainable points A and B, rotate B about A through $3 \cdot 360°/q = 216°$ to get an attainable point P, and translate A by $\overrightarrow{PA}$. This yields an attainable

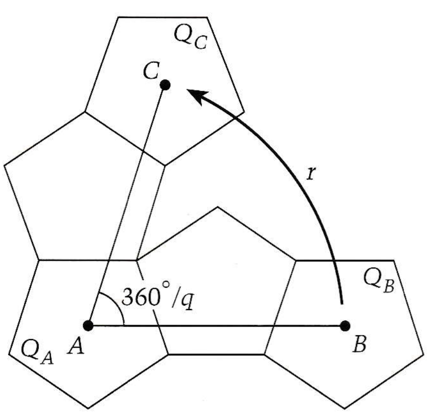


Figure 3

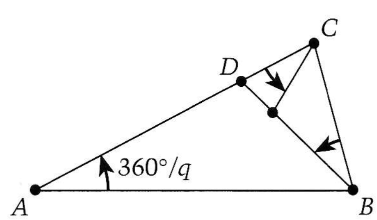


Figure 4

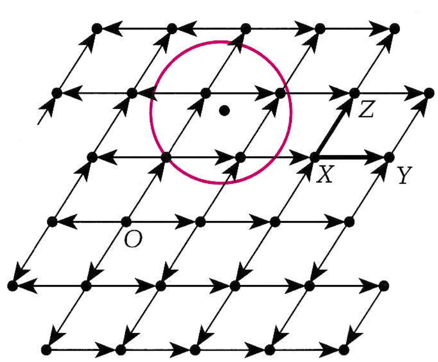


Figure 5

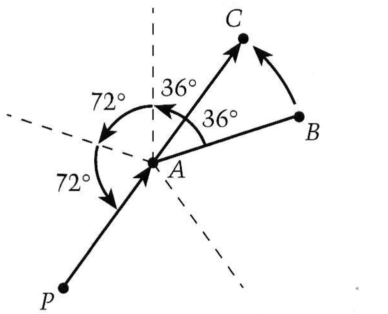


Figure 6

point C (fig. 6), which can alternatively be constructed by rotating B about A through $216° - 180° = 36° = 360°/10$. So in this case we can apply the above argument for $q = 10$.

In the cases $q = 3$, $q = 4$, and $q = 6$ the statement is obviously wrong (the rolling produces a hexagonal, square, or triangular grid of centers, respectively). Thus, the answer is $q = 5$ or $q \geq 7$. (V. Dubrovsky)

M115

The given number is equal to $1{,}000{,}009 = 3^2 + 1{,}000^2$. Let's prove the following general fact: if an integer m of the form $4k + 1$ (with a positive integer k) can be represented as the sum of two squares in two or more different ways, then m is a composite number.

Suppose that

$$m = x^2 + y^2 = u^2 + v^2. \qquad (1)$$

Since m is odd, we can assume that x and u are odd, $x > u$, and y and v are even, $y < v$, so $x \pm u$ and $v \pm y$ are positive even numbers. Then m can be written as

$$m = \left(\frac{x+u}{2} + \frac{x-u}{2}\right)^2 + \left(\frac{y+v}{2} - \frac{v-y}{2}\right)^2$$
$$= \left(\frac{x+u}{2}\right)^2 + \left(\frac{x-u}{2}\right)^2 + \left(\frac{y+v}{2}\right)^2 + \left(\frac{v-y}{2}\right)^2$$
$$= p^2 + q^2 + r^2 + s^2,$$

where $p = (x + u)/2$, $q = (x - u)/2$, and so on. In this calculation we've used the equality

$$pq = \frac{x+u}{2} \cdot \frac{x-u}{2} = \frac{v+y}{2} \cdot \frac{v-y}{2} = rs, \qquad (2)$$

which is equivalent to $x^2 - u^2 = v^2 - y^2$ and so to equation (1).

Let a be the greatest common divisor of p and r. Then $p = ab$, $r = ac$, where b and c are relatively prime. Substituting into equation (2) yields $abq = acs$, or $bq = cs$. It follows that q is divisible by c—that is, $q = cd$, and so $bcd = cs$, or $s = bd$. Now we have

$$m = a^2b^2 + c^2d^2 + a^2c^2 + b^2d^2$$
$$= (a^2 + d^2)(b^2 + c^2).$$

In our particular case this reasoning results in the factorization

$$1{,}000{,}009 = (17^2 + 2^2)(7^2 + 58^2)$$
$$= 293 \cdot 3{,}413.$$

A stronger statement can be proved: a number $m = 4k + 1$ is prime if and only if it's uniquely representable as the sum of two squares. For details, see the article by V. Tikhomirov in this issue. (N. Vasilyev, D. Fomin)

Physics

P111

Let the spring consist of N turns (by the statement of the problem, $N >> 1$). Consider first the compression of the spring due to its own weight in the absence of water. Turn number j (counting from above) must support the upper $(j - 1)$ turns. The change in its length is determined by the weight of these turns $Mg(j - 1)/N$, where M is the spring's mass, and by the spring constant for a single turn, which is N times that for the entire spring k. Therefore,

$$\Delta L_j = \frac{Mg(j-1)}{N}\frac{1}{Nk}.$$

The total shortening of the spring is equal to the sum of the changes in each turn's length:

$$\Delta L = \sum_{j=1}^{N} \Delta L_j = \frac{Mg}{2k} = \frac{L}{2},$$

where

$$\sum_{j=1}^{N} (j-1) = N(N-1) \cong N^2.$$

Therefore, we get $Mg/k = L$.

Let's now consider the case with the water at a height $L/2$. If there are n turns underwater, the other $(N - n)$ will be above the surface. Let's find the value of n/N. We begin by finding the load for each immersed winding and the change in its length. We then sum the changes and equate the result to $Ln/N - L/2$ (this is the difference between the lengths of n turns in the relaxed and compressed states).

The weight of the immersed windings is reduced due to the buoyancy of the water. The effective weight of each turn is

$$\frac{Mg(\rho - \rho_0)}{\rho N} = \frac{\alpha Mg}{N},$$

where $\alpha = (\rho - \rho_0)/\rho$. So the force compressing the $(i + 1)$th turn (counting down from the surface) is

$$F_i = \frac{Mg(N-n)}{N} + \frac{\alpha Mgi}{N},$$

and the sum of the changes in the lengths of the immersed windings is

$$\sum_{i=0}^{n} \frac{F_i}{kN} = \frac{Mgn(N-n)}{kN^2} + \frac{\alpha Mgn^2}{2kN^2}$$
$$= L\left(\frac{n}{N} - \frac{1}{2}\right).$$

Setting $n/N = x$ and using $Mg/k = L$, we obtain the equation

$$x(1-x) + \frac{\alpha x^2}{2} = x - \frac{1}{2},$$

from which we get $x = 1/\sqrt{2-\alpha}$.

Now the entire length of the spring can easily be obtained by summing the length of the immersed part $L/2$ and that of the deformed $(N - n)$ windings above the surface:

$$L^* = \frac{L}{2} + L(1-x) - \frac{L(1-x)^2}{2}$$
$$= L\left(1 - \frac{x^2}{2}\right),$$

or

$$L^* = \frac{L(3-2\alpha)}{2(2-\alpha)}.$$

P112

When a point charge is placed near a conducting plane, charges are induced in the plane that attract the particle. Their effect is equivalent to the action of an image charge $-Q$ located at the same distance on the other side of the plane (see figure 7). The resulting force is obtained from Coulomb's law:

$$F = \frac{kQ^2}{(2x)^2} = \frac{kQ^2}{4x^2}.$$

Because the force of gravity also depends on the inverse square of the separation, let's replace the electrical force with an equivalent gravitational force. We can do this by assuming that we have a mass M at O with

$$M = \frac{Fx^2}{Gm} = \frac{kQ^2}{4mG}.$$

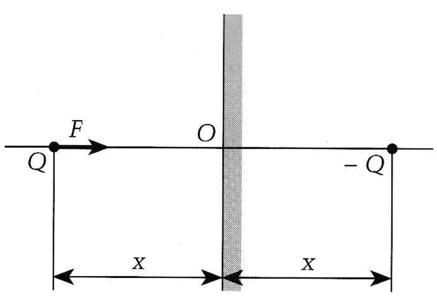


Figure 7

Now we can describe the trajectory of the particle using Kepler's laws. It can be considered a very elongated ellipse with semimajor axis $a = L/2$ and semiminor axis $b << a$ (the foci are at O and the initial position of our point particle).

Because the period of an elliptical orbit with semimajor axis r is the same as that for a circular orbit of radius r, let's find the period of rotation T of the particle about the mass M:

$$m\frac{4\pi^2}{T_0^2}r = \frac{GmM}{r^2},$$

$$T_0 = 2\pi\sqrt{\frac{r^3}{GM}}.$$

It's clear that the time necessary for the particle to reach the plane is equal to the half the rotation period:

$$t = \frac{1}{2}T = \frac{\pi}{2}\sqrt{\frac{L^3}{2GM}}$$
$$= \frac{\pi L}{Q}\sqrt{\frac{Lm}{2k}}.$$

P113

In a closed vessel the number of molecules striking the surface of the ice per unit time is equal to the number of molecules sublimating from the surface (dynamic equilibrium). These are the conditions under which the saturated vapor pressure P_s is measured. Both fluxes of mass are equal to $\rho_s \bar{v} S/6$, where $\rho_s = P_s\mu/RT$ is the saturated vapor density, S is the surface area of the ice, μ is the molecular mass of water, and $\bar{v} \cong \sqrt{3RT/\mu}$ is the mean molecular velocity. The factor 1/6 (or strictly speaking, 1/4) accounts for the choice of a particular direction among six possible directions. If the vessel is open, the flow of evaporating molecules remains the same, but there is no return traffic; now the pressure is $P_s/2$.

Let's estimate the time necessary for complete evaporation by assuming an initial mass of ice $m \sim 0.2$ kg and a cross-sectional area for the glass $S \sim 30\ \text{cm}^2$ and using $\mu = 18$ g/mol:

$$\tau \sim \frac{m}{\frac{1}{6}\rho_s \bar{v} S} \cong \frac{6m}{P_s S}\sqrt{\frac{RT}{3\mu}}$$
$$\cong 150\ \text{s}.$$

In reality, however, the evaporation requires more time; thus, we have obtained a lower bound for the evaporation time.

As the ice evaporates the acceleration of the astronaut is $a = P_s S/2M$, and she will cover a distance $d = a\tau^2/2 \cong 90$ m. Since this is comparable to the distance given in the problem, the rescue will work!

Recalling the approximate nature of our computations, we can say that the astronaut will return to her spaceship in a time $t \gtrsim 100$ s.

P114

When the external magnetic field is switched off, the value of the magnetic field drops from the initial value B to zero. The changing magnetic field induces an electric current in each ring. Let's determine this current at time t after the magnetic field is turned off.

Consider the closed loop $AfCbA$, which coincides with the left ring (fig. 8). According to Lenz's law, the current flows clockwise. Let the current in the AfC section be $I_1(t)$, and let the current in the CbA section be $I_2(t)$. The electromotive force (emf) induced in this loop is

$$\mathscr{E}_{\text{ind}} = -\pi r^2 \frac{\Delta B(t)}{\Delta t}.$$

According to Ohm's law, for a closed circuit we have

$$\mathscr{E}_{\text{ind}} = I_1(t)\frac{R}{2\pi r}l_{AfC} + I_2(t)\frac{R}{2\pi r}l_{CbA},$$

or, taking into account that the

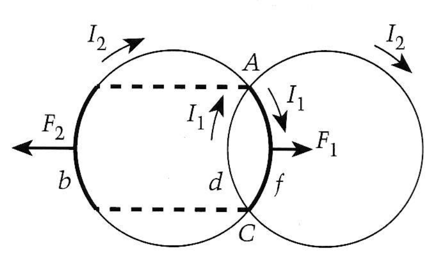


Figure 8

lengths of the arcs l_{AfC} and l_{CbA} are $\pi r/3$ and $5\pi r/3$, respectively, we obtain

$$I_1(t) + 5I_2(t) = -\frac{6\pi r^2}{R}\frac{\Delta B(t)}{\Delta t}.$$

In the same way we write down Ohm's law for the loop $AfCdA$:

$$I_1(t) = -\frac{\left(2\pi - 3\sqrt{3}\right)r^2}{2R}\frac{\Delta B(t)}{\Delta t}.$$

Inserting this value of $I_1(t)$ into the equation above gives us

$$I_2(t) = -\frac{\left(10\pi + 3\sqrt{3}\right)r^2}{10R}\frac{\Delta B(t)}{\Delta t}.$$

Each element of a ring Δl carrying electric current $I(t)$ experiences the Amperean force $\Delta F = I(t)\Delta l \cdot B(t)$, which is directed along the radius of the ring. Due to the bilateral symmetry of these forces relative to the horizontal axis connecting the centers of the rings, the resulting force acting vertically on each of the rings is zero. The absence of symmetry of the forces relative to the vertical axis passing through the center of the left ring ($I_1(t) \neq I_2(t)$) results in a horizontal force. This force is equal to the difference between the forces acting on the arc AfC and the symmetrical arc on the opposite side of the ring:

$$F = F_2 - F_1 = I_2(t)l_{AC}B(t) - I_1(t)l_{AC}B(t),$$

where l_{AC} is the chord subtending the arc AfC ($l_{AC} = 2r \sin \alpha/2 = r$). The final expression for the force F is

$$F = -\frac{9\sqrt{3}r^3}{5R}B(t)\frac{\Delta B(t)}{\Delta t}.$$

The action of this force during a small time period t results in a change in the ring's momentum:

$$m\Delta v = F\Delta t = -\frac{9\sqrt{3}r^3}{5R}B(t)\Delta B(t)$$

$$= -\frac{9\sqrt{3}r^3}{10R}\Delta\left[B^2(t)\right],$$

and so the ring will gain a velocity

$$v = \frac{9\sqrt{3}r^3}{10mR}B_0^2.$$

P115

Denoting the distance between the source and the lens as d and the distance between the image and the lens as f, we write the lens formula as follows:

$$\frac{1}{d} + \frac{1}{f} = \frac{1}{F}.$$

In a small time period Δt the distance between the source and the lens decreases by $\Delta d = v_0\Delta t$, and the distance between the lens and the image increases by $\Delta f = u \cos \alpha\Delta t$. Then (see figure 9)

$$\frac{1}{d - v_0\Delta t} + \frac{1}{f + u\cos\alpha \cdot \Delta t} = \frac{1}{d} + \frac{1}{f},$$

or

$$\frac{v_0\Delta t}{d^2} = \frac{u\cos\alpha \cdot \Delta t}{f^2}.$$

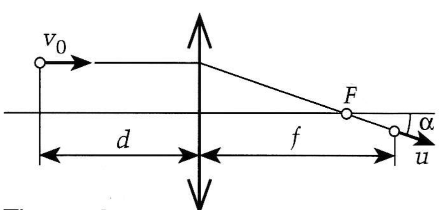


Figure 9

The image's velocity u equals that of the source v_0 when $f_1 = d_1\sqrt{\cos\alpha}$. Taking into account that $\cos \alpha = F/\sqrt{F^2 + H^2}$, we obtain

$$\frac{1}{d_1} + \frac{1}{d_1\sqrt{\cos\alpha}} = \frac{1}{F},$$

$$d_1 = F\left(1 + \frac{1}{\sqrt{\cos\alpha}}\right) = F\left(1 + \sqrt[4]{1 + \frac{H^2}{F^2}}\right)$$

$$= F\left(1 + \frac{\sqrt[4]{17}}{2}\right).$$

Brainteasers

B111

Six days. The number of days off must be $48 \div 12 = 4$ times the number of days worked. So the number of days worked is 1/5 of 30, the total number of days.

B112

Since every two statements contradict each other, only one of them can be true. All of them can't be wrong, because in this case the hundredth one would be true. So there are exactly one true and 99 wrong statements. This means that the only true statement is the ninety-ninth.

B113

Hot water puts out a fire quicker than cold water, because it evaporates more rapidly and the vapor impedes the access of air feeding the flames.

B114

The answer is 384. Suppose we start to lace the shoe by passing the lace through the top right hole. Then it may go out through any other hole except the top left one—that is, in eight ways. Next we have to pass the lace through the "parallel" hole on the other side and let it out through any of the six holes. This makes $8 \cdot 6$ ways. Proceeding in the same manner, we'll get four choices on the next step, then two choices, and, finally, we'll pass the lace through the top left hole. All in all we have $8 \cdot 6 \cdot 4 \cdot 2 = 384$ ways of lacing the shoe.

B115

See figure 10: point A is the midpoint of the arc; the two pieces can be made to coincide by rotation about O through 45°.

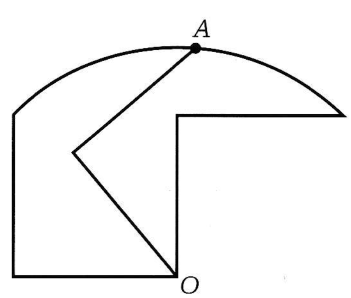


Figure 10

Kaleidoscope

1. The molecular force is proportional not to the muscle's mass but to its cross-sectional area. So if all the linear dimensions of an animal are decreased by a factor of n, its mass will be reduced by a factor of n^3, and the force will decrease only by a factor of n^2. Thus, the relative force (that is, the force per unit mass) increases as the animal's size decreases. Of course, the geometry is not precisely the same in different animals, but the influence of the dimensional factors far outweighs the role of specific features, which allows us to establish a clear relationship between the relative force and the animal's size.

2. The force produced by a muscle is proportional to its cross-sectional area (see the previous problem), and the distance the muscle contracts is proportional to its initial length. Because the mechanical energy output of a single contraction is the product of force and distance, it is proportional to the cube of the organism's linear size (and correspondingly to its mass). The same amount of muscular energy per unit mass corresponds to the same potential energy at the top of the jump. Thus, geometrically similar animals should be able to jump to the same height.

3. If the linear dimensions of an animal are increased by a factor of n, its body mass increases by a factor of n^3. Suppose that, when this happens, the thickness of a bone is increased by a factor of m. To the extent that we assume (in accordance with actual conditions) that the composition of the bone doesn't change, the pressure on the bony tissue (per unit cross section) must be preserved—that is, $n^3/m^2 = 1$. In other words, m is proportional to $n^{1.5}$. As one can see from the figure, Galileo increased the linear dimensions by a factor of 3. So the animal's mass increased by a factor of $3^3 = 27$. It seems that for the sake of clarity Galileo increased the bone's thickness in the figure by a factor of $27/3 = 9$ (rather than $3^{1.5} \cong 5.2$). In Galileo's defense, it should be pointed out that all the calculations in the text of his book are correct.

4. As in the previous problem, we assume that the pressure exerted on the foot's cross section in both a human and in a "monoped" is the same. The body density for all mammals is about 1 g/cm^3. The mass of the foot itself can be neglected in such approximations. In humans this pressure is about

$$\frac{700 \text{ N}}{2 \cdot 3.14 \cdot (15 \text{ cm})^2/4} \cong 1 \text{ N/cm}^2$$

(two feet!). Let 1 retem be z meters. Then the weight of the monoped will be

$$3.14z^3 \cdot 1{,}000 \cdot \frac{9.8}{8} \cong 4{,}000z^3 \text{ N}$$

(estimate the contribution of the foot's mass on your own), and the cross-sectional area of the foot is

$$3.14 \cdot \frac{(0.2z \cdot 100)^2}{4} = 80z^2 \text{ cm}^2.$$

This gives us

$$50z = 1,$$
$$z = 0.02.$$

It should be noted, however, that comparing the monoped to an antelope instead of a human would drastically change the result.

5. Of course, 30–40 lbs of food per day is an absolutely fantastic figure both for a three-year-old child and for our readers. (Don't forget that water requirements weren't included in our estimates.) Little children eat greater amounts of food because of their faster metabolisms and perhaps even more because of the higher heat losses typical of smaller mammals. Another instance of this factor can be seen in problem 9.

6. The weight of an insect (that is, the force of gravity acting upon it) is proportional to n^3, and the pressure on the air created by each stroke of its wing is proportional to the wing's area (n^2) and to the muscular force (another factor of n^2). Although a decrease in size makes an animal relatively stronger (see problem 1), it doesn't help with "rowing" types of movement. In order to hang in the air, a smaller animal has to increase the number of strokes. The pitch of the sound emitted by the wings increases correspondingly. As for the flies that bothered Gulliver, their flight should be more or less silent (to the human ear), just as the flight of birds is relatively quiet. The sounds that we hear when birds fly are produced by other types of wing movement.

There is an even more energy-consuming mode of flight based upon the rotation of a propeller. In your leisure time try to estimate how much jam Karlsson would have to eat so as not to lose weight in flight. (Karlsson is a character in the stories of A. Lindgren.)

7. Warm-blooded animals expend a significant amount of energy keeping their temperature constant. This is a particularly challenging problem for a small animal with a relatively large surface area. Its surface area is inversely proportional to the square of its linear dimensions, while its mass is inversely proportional to the cube of its linear dimensions. In order to achieve equilibrium between heat production and heat loss, small animals maintain higher body temperatures.

8. When we walk, our center of mass—which is located in the lower part of the body, just below the navel—moves along the arc of a circle whose radius is approximately equal to the length of our legs. It's known that a body moving with velocity v along a circle of radius l has an acceleration v^2/l directed toward the center of the circle. Two forces act on a person when walking: the force of gravity and the supporting force. The resultant of these forces—the centripetal force—clearly cannot be more than the force of gravity (mv^2/l does not exceed mg, where m is the mass of the body). So the maximum walking velocity equals $v = \sqrt{gl}$, which for a human comes to about 3 m/s (a

reasonable value). Children have shorter legs, so they have to run to keep up with their parents, but their running is actually a succession of jumps and not rotations of the body about the axis of the leg.

9. As the problem deals with desert animals, we can assume that water loss is caused by evaporation from the body's surface. Since the area of this surface is proportional to the square of the animal's linear dimensions (n^2) and the amount of stored water is proportional to the body's volume (n^3), larger animals can survive longer after drinking their fill of water. And yet, desert animals vary widely in size. How do we explain this? It turns out that our reasoning is applicable to animals that are closely related taxonomically (for instance, jerboa and camel), for which a decrease in size is not accompanied by a reduction in the water permeability their tissues. It makes no sense to compare beetles, lizards, and mice according to the similarity principle because their tissues are fundamentally different.

Toy Store

1. If r, l, u, d are the numbers of the R, L, U, D moves, respectively, required to get from (0, 0) to (m, n), then $r - l = m$, $u - d = n$, and the total number of moves equals $r + l + u + d = m + n + 2(l + d)$.

2. Required short sequences to reach a colored square (m, n) from (0, 0) marked face up are, for instance: UR^mD for $m \geq 1$, $n = 0$; $UR^{m-2}URDR$ for $m \geq 2$, $n = 1$; similar sequences for $(0, n)$ and $(1, n)$; and $RULURD$ for $(m, n) = (1, 1)$.

3. The three possible half-turns are given by RUL^2DR, UR^2DL^2, U^2RD^2L.

4. $H(a^{-1}, b)$ generates the half-turn h_v of the central cube, and $H(a, b^{-1})$ generates h_s; both operations can be reduced to 38 moves.

5. Consider the chessboard coloring of the squares. If the vacancy comes back to its initial place, then the number of cubes that moved

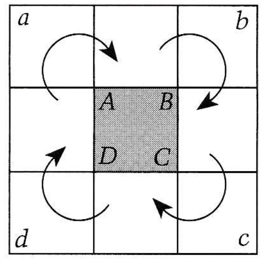


Figure 11

from the black squares to the white squares equals the number of cubes that changed from white to black. In particular, the total number of cubes that changed their underlying color is always even. So the answer to (a) is no (because a 90° turn is possible only with a change of color). The answer to (b) is no, too, because all corner squares in a 3 × 3 board are the same color, and so we have to make only two like color changes without any opposite changes.

We'll use the shortened notations a, b, c, d for the four-roll cyclic moves about points A, B, C, D in figure 11 (as we did in the article). When verifying the number of single rolls, take cancellations into account.

6. $(dcba)^3dc$.

7. $a^2b^2c^2dcd^2c^{-1}d^{-1}$.

8. $ab^{-1}ad^{-1}cd^{-1}ab^{-1}ad^{-1}c^2ab^{-1}c^{-1}b^{-1}dcbada^2d$.

9. $a^{-1}dcb^2d^{-1}a^{-2}d^{-1}c^2bd^{-1}c^{-1}d^{-1}a^{-1}c^2$.

10. $b^2a^{-1}(b^{-1}a)^3ba^{-2}D$ (the vacancy is shifted; the direction of rotation here is the opposite of that in the previous solution).

11. $a^{-1}ba^2$.

Mt. Fermat-Euler

1. If $a^2 + b^2$ is an odd prime, then one of the squares—say, a^2—is even. Therefore, $a = 2n$, and the other one is odd, so $b = 2k + 1$. Then $a^2 + b^2 = 4n^2 + 4k^2 + 4k + 1 = 4m + 1$, where $m = n^2 + k^2 + k$.

2. For any integer a, $1 \leq a \leq p - 1$, consider the products $1 \cdot a, 2 \cdot a, \ldots, (p - 1)a$. The remainders of all these products when divided by p are all different (because if $ka \equiv la \pmod p$, $k > l$, then $(k - l)a$ is divisible by p, which is impossible since both factors $k - l > 0$ and a are less than p, and p is prime). None of these products is divisible by p, so the remainders take each of the values $1, 2, \ldots, p - 1$ once. So for the given value of a there is a unique b, $1 \leq b \leq p - 1$, such that $ab \equiv 1 \pmod p$. For $a = 1$ and $a = p - 1$ the corresponding number b is equal to a. For any other a ($2 \leq a \leq p - 2$), $b \neq a$, because $a^2 \equiv 1 \pmod p$ implies that $a^2 - 1 = (a - 1)(a + 1)$ is divisible by b only for $a - 1 = 0$ or $a + 1 = p$. This proves the first statement. The proof of Wilson's lemma is now completed just as was done for $p = 13$ in the article:

$$
\begin{aligned}
(p-2)! &= 2 \cdot 3 \cdot \cdots \cdot (p-2) \\
&= \left(2 \cdot \frac{p+1}{2}\right) \cdot \cdots \cdot \left[\frac{p-1}{2} \cdot (p-2)\right] \\
&\equiv 1 \pmod p,
\end{aligned}
$$

and $(p - 1)! \equiv p - 1 \equiv -1 \pmod p$.

Nine solutions

1. In figure 3 in the article, the angles of triangle CBK are $\angle C = 80°$ (by the condition), $\angle B = 20°$ (by construction), $\angle K = 180° - 80° - 20° = 80°$, so $BC = BK$. We saw in the first solution that $BC = BE$, so $BE = BK$; and, since $\angle KBE = 80° - 20° = 60°$, the triangle BEK is equilateral. Further, $\angle KBD = 60° - 20° = 40° = \angle BDK$ (the last equality was proved in the third solution), so $KD = KB = KE$. To finish the solution, we can notice, for instance, that K is the center of the circle that passes through B, E, and D; it follows that $\angle EDB = ½\angle EKB = 60° \div 2$.

2. In figure 6 in the article, $\angle EOD = 2\angle ECD = 60°$, because $\angle EOD$ is a central angle in the circumcircle of triangle CED, and $\angle ECD$ is an inscribed angle subtended by the same chord ED. We deduce that DEO is an equilateral triangle and $ED = DO$. Thus, D lies on the perpendicular bisector p of the segment EO; also, we know that BD bisects the angle

EBO. If p and BD were different lines, we could apply the argument from the fifth solution to show that D lies on the arc EO of the circumcircle of triangle BEO. But in this case we'd have $\angle EDO + \angle EBO = 180°$, whereas in fact this sum equals $60° + 40° = 100°$. So BD is the perpendicular bisector of EO, which means that triangles BDE and BDO are congruent, and $\angle EDB = \angle BDO = 60° \div 2 = 30°$.

3. Apply the construction with two reflections from the eighth solution (fig. 12) (compare figure 7 in the article). Then $AP = CB_1$, $AB = $

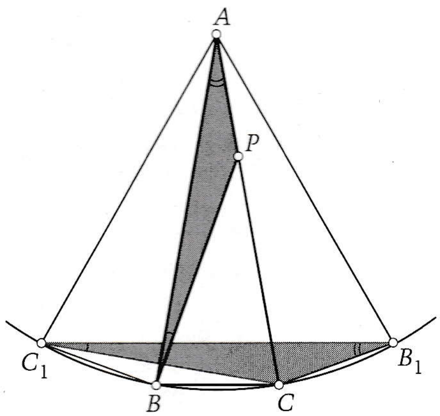


Figure 12

C_1B_1. Using the fact that points C_1, B, C, and B_1 all lie on a circle with center A and applying the Inscribed Angle Theorem, we get $\angle C_1B_1C = \frac{1}{2}\angle C_1AC = \angle BAC$ (and, therefore, $\triangle C_1B_1C$ is congruent to $\triangle BAP$), and $\angle B_1C_1C = \frac{1}{2}\angle B_1AC = 10°$. By the congruence of the triangles mentioned above, $\angle ABP = \angle B_1C_1C = 10°$, so $\angle PBC = 80° - 10° = 70°$. (V. Dubrovsky)

Yo! That's not how it goes!

The author of "The Good Old Pythagorean Theorem" in the January/February issue is V. N. Berezin (not "Beryozin," as printed). Likewise, the author of "Flexible in the Face of Adversity" in the September/October 1990 issue is A. P. Veselov (not "Vesyolov"). Our apologies for mangling these names.

"BULLETIN BOARD" CONTINUED FROM PAGE 52

measured by converting light energy to electrical energy. The signals are sent to a digital logic circuit, which matches codes and identifies the bill. The logic circuit activates a voice chip. The device is powered by four AAA batteries.

Phillips said she wanted to combine her interest in electronics with her desire to do something for blind people. (Tracy's brother is blind.) For her scientific skill, Phillips will receive a \$20,000 savings bond from Duracell.

Second-place winners were Chris Hyun Cho, 16, of East Setauket, New York; Seth Frankel, 17, of Demarest, New Jersey; Eric Magnuson, 18, of Uniontown, Ohio; David Monson, 15, of Boise, Idaho; and Robbie Lynn Slaughterbeck, 17, of Oklahoma City, Oklahoma. Each of these student inventors received a \$10,000 savings bond.

Cho was inspired to create the Automated Page-Replacing Contrivance because, as a member of his school's chamber orchestra, he found quick page turning to be a problem. Frankel created Safe-T-Eyes, a device that protects a power tool operator from injury by requiring a face shield to be in place before the tool can be turned on. Magnuson developed the Safe Distance Brake System, which shows, in different colors, the amount of pressure a driver is applying to the brakes. Monson invented the RF Interconnectable Smoke Alarm, a wireless unit that causes every smoke alarm in a house to sound when only one is activated by smoke or fumes. Slaughterbeck devised the Rx-Locker, a timed, internally locked pill dispenser designed to prevent overdoses. (He was inspired by the accidental overdose of a family friend.)

The first- and second-place winners, their parents, and their science teachers were guests of Duracell at an awards ceremony in Anaheim, California, on March 30. The young inventors demonstrated their devices for a luncheon audience and exhibited them for thousands of science teachers at the 42nd annual convention of the National Science Teachers Association (NSTA).

Ten students were also awarded \$1,000 third-place savings bonds; 25 students won \$200 fourth-place bonds; and 58 finalists were selected for \$100 bonds.

To enter the Duracell/NSTA Scholarship Competition, students in grades 9 through 12 design and build a device that is educational, useful, or entertaining and is powered by one or more Duracell® batteries. Entries are judged on energy efficiency, practicality, and inventiveness. Every student who enters receives a gift from Duracell and a certificate of participation from NSTA. Proposals for entries are due at NSTA each January. For more information, write to Eric Crossley, NSTA, 1840 Wilson Blvd., Arlington VA 22201-3000, or phone 703 243-7100.

Index of Advertisers

The rolling cubes

Can you roll your way into the record books?

by Vladimir Dubrovsky

THIS ARTICLE COMPLETES (for the time being) the discussion of different kinds of rolling-block puzzles in the September/October and November/December issues. In most of the puzzles we considered earlier you had to roll pyramid-shaped blocks. This time we'll be rolling cubes around.

Tumbleweed revisited

We'll begin with one of the simplest problems posed in connection with the "Tumbleweed" game in the November/December issue. Slightly generalized, it reads as follows.

Consider a unit-square grid in the first quadrant (fig. 1) and a unit cube sitting on the corner square of this grid. Suppose five faces of the cube are white, and one face—the top one—is colored (we'll call this the *marked* face). We have to roll the cube over to a given square of the grid in such a way that the marked face appears on the top again, and figure out the possible number of moves (rolls) this procedure will take.

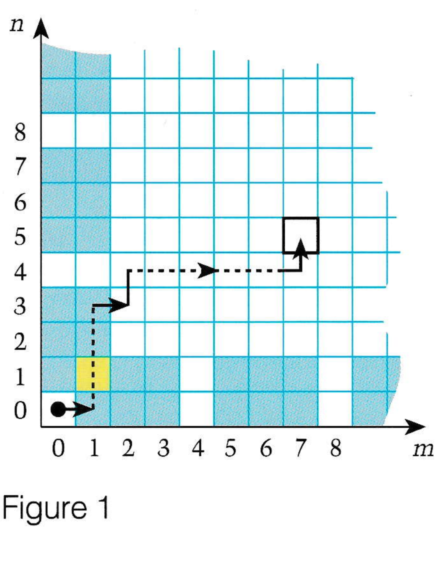


Figure 1

Well, this problem isn't all that difficult. The answer is given in figure 1: if we number the horizontal and vertical lines of the grid 0, 1, 2, ..., thus linking a pair (m, n) with each grid square (m and n are the numbers of the column and the row, respectively, that contain this square), then the *minimal* number N of rolls sufficient to reach the square (m, n) marked face up is equal to $m + n$ for the white squares, $m + n + 2$ for the blue ones, and $m + n + 4 = 6$ for the single yellow square $(1, 1)$.

To see that this is indeed true, we note first that after four rolls in the same direction the cube restores its initial orientation (it makes a full turn). Therefore, for $(m, n) = (4l, 0)$ or $(0, 4l)$ we have $N = m + n = 4l$. Obviously it's impossible to get to these squares in a smaller number of moves, and it's just as obvious that we can't get to any square (m, n) in less than $m + n$ moves.

Now, imagine you roll the cube from the initial position once to the right. Then the marked face appears on the right side and will stay on this side after any number of subsequent rolls in the vertical direction. In particular, if you make only one "up" roll and then continue rolling to the right, the cube will assume the initial orientation after three additional right moves, as well as after $3 + 4k$ moves for any k. Thus, we can roll the cube to the square $(m, n) = (4l, 1)$ in $m + n = 4l + 1$ moves. Denoting the right, left, up, and down moves as R, L, U, and D, we can write this sequence of rolls as RUR^{4l-1} (R^k denotes k successive R moves).

Similarly (and symmetrically) we can get to the square $(m, n) = (1, 4l)$ in $m + n = 4l + 1$ moves: URU^{4l-1}. The simple trick used here—rolling the cube with its marked face on a side along this face—enables us to solve the problem in $m + n$ moves for any square (m, n) with $m \geq 2$, $n \geq 2$. Formally, the sequence of moves that carries the cube over to this square marked face up is $RU^{n-2}RUR^{m-2}U$. (See the example for $(m, n) = (7, 5)$ in figure 1—rolling along the dotted lines in this figure doesn't change the relative position of the marked face.)

Next, you can easily make sure that $m + n$ rolls aren't enough to get to the colored (blue and yellow) squares in figure 1, and that to get to the yellow square $(1, 1)$ you have to make even more than $4 = m + n + 2$ moves (there are only a few conceivable routes of length $m + n$ from $(0, 0)$ to (m, n) with m or n no greater than 1, and you can check all of them).

Problem 1. Prove that if you can reach the square (m, n) in k moves, then $k - (m + n)$ is a nonnegative even number.

According to this problem, the blue squares require no less than $m + n + 2$ moves, while the yellow one—$(m, n) = (1, 1)$—requires at least $m + n + 4 = 6$ moves.

Problem 2. Find the solutions for the yellow and all the blue squares in 6 and $m + n = 2$ moves, respectively.

If you compare these results with the numbers given in the November/December issue for the Tumbleweed game, you'll see at once that these numbers are the minimum possible (that is, $m + n$ for the white squares, $m + n + 2$ for the blue ones, and 6 for the yellow one).

Rotations and rolling tours

More interesting questions come up when we have a cube whose faces are all colored differently. Can we roll a cube from the square (0, 0) to (m, n) in such a way that it ends up in exactly the same position as it started, or, more generally, in a certain given position?

It turns out that possible final positions depend on the coordinates (m, n) of the final square—more exactly, on the parity of $m + n$. This becomes clear when you look at figure 2. We draw a line on each face of the cube as shown in this figure (the lines on the faces we can't see are parallel to the lines on their opposite faces), and draw lines on all grid

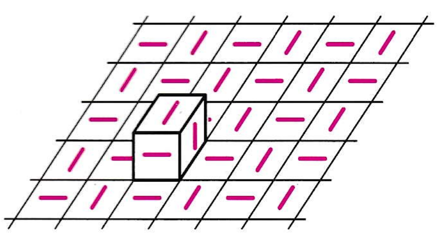


Figure 2

squares alternating their directions in chessboard order. Put the cube on a grid square so that the line on its bottom coincides with the line on the square. Now roll the cube. The lines on the cube and on the plane will fit each other again, so they will coincide after a second roll and, in general, after any number of rolls in any direction. This means that only a half of all possible positions of the cube on a given square are accessible by rolling it from another given square. Indeed, a cube can rest on any of its six faces and we can turn it four different ways on this face. So there are $6 \cdot 4 = 24$ positions on a given square. But if we

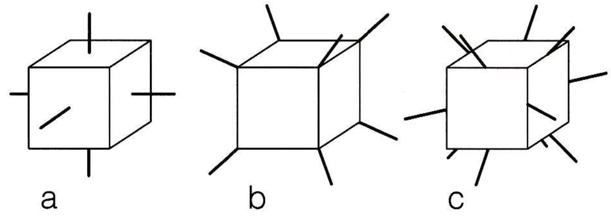


Figure 3

want the line on the cube's bottom to fit the line on the square, we'll be left with only two ways to turn the cube on the bottom, which amounts to $6 \cdot 2 = 12$ positions.

Now imagine that we've rolled the cube somewhere and then slid it back without turning it (that is, by a parallel translation). Then the final position of the cube can be obtained from the initial one by a certain rotation. It's a very good exercise to find all the 24 rotations of the cube, and the 12 that preserve our pattern of lines on the faces. Try to do it yourself.

And here's the answer to verify your investigation: there are three rotations (by 90°, 180°, and 270°) about each of the three axes perpendicular to the cube's faces (fig. 3a); two rotations (120° and 240°) about each of the four diagonals (fig. 3b); and the (least obvious) rotations, or half-turns, about each of the six axes through the midpoints of the cube's opposite edges (fig. 3c). This makes $3 \cdot 3 + 2 \cdot 4 + 6 = 23$ different rotations; the one missing is the identity transformation. Our pattern of lines is preserved by the three half-turns about the "face axes," eight diagonal turns, and, of course, the identity transformation.

Since every single roll changes the direction of the line on the cube's bottom, the rotations that preserve the pattern of lines emerge after any even number of rolls, and we can call them *even* rotations. Twelve other rotations will be called *odd*. By problem 1, the number of rolls required to get from the square (0, 0) to (m, n) has the same parity as $m + n$, so the rotations that can emerge as the result of rolling the cube from (0, 0) to (m, n) are even or odd depending on whether $m + n$ is even or odd.

In particular, when a cube traces any closed path, it ends up on the initial square rotated "evenly." But can we actually obtain all 12 even rotations in such a way? The simplest closed path consists of four moves around a 2×2 square (fig. 4). We can see immediately that this "rolling tour" results in a 120° rotation about the diagonal passing through the center of the square (point A in figure 4a), as shown in figure 4b. The sense of

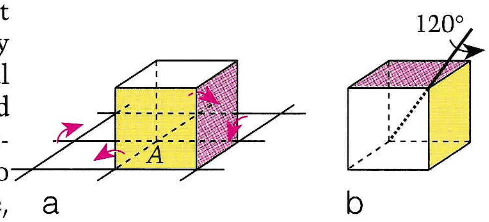


Figure 4

the rotation depends on the sense of the rolling tour. In our notation the tour in figure 4a is written as $LURD$. Since the moves R and D are the inverses of L and U, we can rewrite it in the form $LUL^{-1}U^{-1}$. Visitors to our Toy Store may remember that operations of this form (in general, $XYX^{-1}Y^{-1}$, rendered in shorthand as $[X, Y]$) are called *commutators* (of X and Y) and are often useful with transformational puzzles. (See, for instance, the January/February issue.) The commutator $[L, U]$ together with the seven other commutators of L, U, and their inverses—$[L, U^{-1}] = LU^{-1}L^{-1}U = LDRU$, $[U, L] = ULDR$, and so on—yield all eight diagonal rotations of the cube.

Problem 3. Find the three 6-move rolling tours that turn the cube 180° in its place about each of the three face axes (shown in figure 4a).

By taking the cube on rolling tours, we can obtain all 12 of its even rotations and return it to its initial location.

Rolling in a crowd

Now we're ready to investigate puzzles with many rolling cubes and only one empty space. One version of such a puzzle is shown in figure 5 (on the next page). It consists of eight

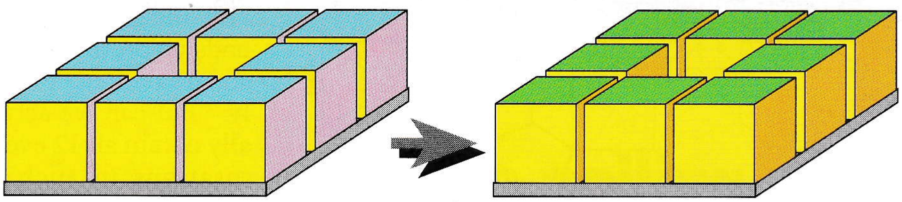


Figure 5

identical 6-color cubes in a square box with an empty space in the middle. Initially all the cubes are oriented in the same way so that each face of the square "ring" is the same color. You have to roll the cubes in such a way that they all again become oriented alike but not as they were initially. It's not hard to see that the final coloring in this problem can be chosen in 23 different ways, which is the number of non-identical rotations of the cube.

This particular problem was proposed by puzzle designers A. Dryomov and G. Shevtsova, whom we've already mentioned in connection with other rolling-block puzzles. It's interesting how they managed to prevent the cubes from slipping: the cubes they use have grooves along their edges (fig. 6), and the bottom of the box has a 3×3 square grid of narrow laths glued to it; the laths fit into the grooves, so when you tilt a cube to roll it to an adjacent square, it "stumbles" on a lath and lands exactly where you want. Of course, other antislipping devices exist, and other problems, too.

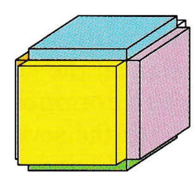


Figure 6

For instance, the set of rolling cubes manufactured in Poland some time ago comprised eight identical cubes that had only one colored face each. It was supplied with instructions saying that the puzzle could be used for evaluating the IQ of the player. The maximum score was given for the following "Royal Problem." Initially the four corner cubes are set in the box, their colored faces up, and the remaining four cubes at the edges are turned so that their colored faces are down. You have to roll the cubes so that the colored faces appear on the bottom in the corners and on the top at the edges. According to the instructions, "if you find a solution in no more than 36 moves, you're a genius; if not, you fall just a little short."

A natural way to tackle the problems we described, and others like them, would be to try to find sequences of rolls that result in rotating a single cube in its place and leave all the other cubes unaltered. This won't be economical; but the advantage of this method is universality: if we learn how to turn a single cube in every possible way, we'll be able to obtain the required arrangement by rotating the cubes one by one. Notice that this approach didn't work with the rolling pyramids (see the September/October Toy Store), because a pyramid always returns to its initial location in the same orientation that it had at the start.

I'll describe two types of operations that rotate a single cube. Both can be performed in a "small" box measuring 3×2; for definiteness, we'll assume that the empty square is at the middle of the longer side of the box, as shown in figure 7. Any sequence of rolls in this box that brings the empty space back to its initial location can be represented as a combination of the four 4-roll cyclic moves about the centers A and B in figure 7: $DRUL$ (the clockwise cycle about A, which will be denoted by a), $LDRU$ (the clockwise cycle b about B), and their inverses $a^{-1} = RDLU$ and $b^{-1} = DLUR$. This is easy to verify, so I'll give only one illustration: $RDLLUR = RDLUDLUR = a^{-1}b^{-1}$—the successive moves U and D in the intermediate expression cancel out. Below we'll use this shorthand notation—that is, the cycles a and b—rather than the notation for single rolls R, L, U, D.

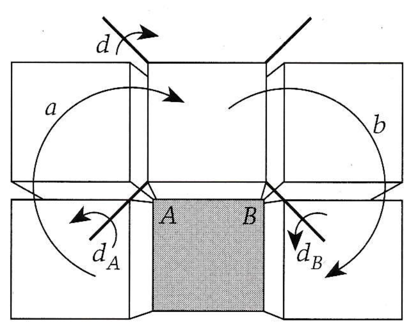


Figure 7

So, in cyclic notation, operations of the first type are commutators (déjà vu!) of the *triple* cyclic moves a^3, b^3, and their inverses. More exactly, under $[a^3, b^3] = a^3b^3a^{-3}b^{-3}$ the four corner cubes stay in place, while the central cube makes a half-turn h_l about the axis parallel to the long sides of the box; then, $[a^{-3}, b^{-3}] = a^{-3}b^3a^3b^{-3}$ performs a half-turn h_s of the central cube about the axis parallel to the short sides of the box; and $[a^3, b^{-3}]$ rotates the central cube by 180° about the vertical axis (perpendicular to the box). Notice that the last half-turn h_v is produced by $4 \cdot 3 \cdot 4 = 48$ single rolls, while the actual number of rolls in h_l and h_s is 46 (we've seen that in the combination $a^{-1}b^{-1}$ two rolls—U and D—cancel out; the same happens with ba).

It's interesting to look into why these operations behave as they do. The 4-roll move a cycles the three cubes about A clockwise by one position. Repeated three times, it brings the cubes to where they started after rolling one full circuit about A—that is, a^3 twists these cubes 120° about their diagonals drawn from their vertices at A (compare figure 4). The operation b^3 behaves similarly about point B instead of A. What happens under $a^3b^3a^{-3}b^{-3}$? The two left corner cubes are twisted by a^3, left intact by b^3, untwisted by a^{-3}, and left intact again by b^{-3}. So eventually these cubes, as well as the two right cubes, stay put. At the same time, the central cube is successively subjected to 120° rotations about its diagonals from A and B and their inverses, which results in the half-turn specified above.

Operations of the second type were presented by John Harris in one of the first publications on rolling cubes. They can be written in the following general form:

$$H(x, y) = xyxyxy^{-1}x^{-1}y^{-1}x^{-1}y^{-1},$$

where x and y stand for a, b, or their inverses, and they also result in rotations of a single (central) cube. In particular, taking $(x, y) = (a, b)$, we get an operation $H(a, b) = a(ba)^2(b^{-1}a^{-1})^2b^{-1}$ that performs the diagonal rotation d_B shown in figure 7; $H(b, a)$ gives the inverse rotation d_B^{-1}; $H(a^{-1}, b^{-1})$ and $H(b^{-1}, a^{-1})$ produce rotations d_A and d_A^{-1}, respectively. Although these operations are defined as sequences of ten 4-roll cyclic moves, they contain a number of mutually annihilating successive rolls (in the combinations ba and $a^{-1}b^{-1}$,), so after cancellations they can be reduced to 32 rather than 40 single rolls. The four other operations of this form—$H(a^{-1},b)$, $H(a,b^{-1})$, and their respective inverses $H(b,a^{-1})$ and $H(b^{-1},a)$—don't add much to what we already have: they only reproduce some of the rotations generated by the commutators above, though in fewer moves.

Problem 4. Exactly what rotations are generated by the operations $H(a^{-1}, b)$ and $H(a, b^{-1})$, and how many rolls do they comprise after possible cancellations?

So, it turns out we have comparatively short operations for seven of the eleven even non-identity rotations of the central cube on a 3×2 board: the three half-turns h_l, h_s, h_v, and four 120° rotations about two of the diagonals. As to the remaining four rotations about the other two diagonals, I don't know a nice, elegant way to obtain them, but they can be represented as combinations of d_A and d_B. For instance, the rotation d in figure 7 results when d_A is followed by d_B^{-1}—that is, $d = d_Ad_B^{-1}$. A shorter way to perform d is (1) turn the central cube (and maybe some other cubes) so that the axis of d fits onto the axis of, say, d_B; (2) perform d_B; (3) "undo" operation 1—that is, $d = a^3d_Ba^{-3}$, or $a^{-3}d_A^{-1}a^3$. Here d is represented as a *conjugate* of d_B (or d_A^{-1})—see the January/February Toy Store.

Conjugation can also be used to rotate separately any of the corner cubes in its place. For instance, the top left cube is rotated by an operation of the form ara^{-1}, where r is any of the rotations of the central cube considered above: a rolls the top left cube to the central location, r rotates it, a^{-1} rolls it back.

Now we can do any rolling-cube problem in which the "target" position differs from the initial one by an even rotation on each square turning cubes in their places one by one. As for problems with odd rotations, I'll illustrate them with one example: *turn all the cubes in a box through 90° about their vertical axes.*

Imagine that the squares of the box are colored black and white in chessboard order so that the empty square is white. To be "oddly rotated," a cube must make an odd number of rolls; therefore, it must move to a square of the other color (because one roll changes the color of the cube's underlying square). But initially our cubes occupied three black and two white squares. One white square for the "black cubes" is lacking, and the problem seems to be unsolvable! However, a solution does exist. Here it is: $a^2b^2a^2b^2a^2D$. It wasn't stipulated that the vacant place remain the same, and in the last move it's shifted to a black square, while an extra "black cube" changes its underlying color.

Tricks aside, this solution illustrates a general rule for solving rolling-cube puzzles. First, using the chessboard coloring, we determine which cubes must change their underlying colors. Then we roll the cubes so as to change colors as required, without paying much attention to the cubes' orientations. After that, the required final orientations become attainable by even rotations of the cubes in their new places, which can be done by using the operations we described above.

Problem 5. In a 3×3 box with an empty space in the center, all the cubes are colored the same and have the same initial orientation. Is it possible to obtain a position in which (a) one cube is rotated 90°, (b) two corner cubes are rotated 90°, while all the other cubes retain the initial orientation?

Beat these records

Although the explanations above allow us to solve any (solvable) rolling-cube problem, they don't help much if you have to find a solution in a small enough (or the smallest) number of moves. I count such problems among the most intriguing and challenging transformational puzzles. Here are some of them, together with the lengths of the best solutions I know of. The empty square must remain in its place in all the problems below except one (figure out which).

Problems

The first three problems are borrowed from John Harris's article in the *Journal of Recreational Mathematics*, Vol. 7, No. 3.

6. In a 3×3 box with an empty space in the middle each of the eight cubes has only one colored face. Initially it's on the left side of each cube. You must move it onto the right side. (30 moves).

7. Starting with the initial position of the previous problem, turn the colored faces onto the front sides of the cubes. (44 moves)

(In the next problem the coloring is different.)

8. Initially all the external faces of the cubes are red, and all the hidden faces are white. You must hide all the red faces. (84 moves[1])

The remaining problems are for 6-color cubes. The record solutions for the first two are held by A. Panteleyev, a Moscow mathematician.

9. Turn all the cubes in the 3×3 box through 90° about the axis parallel to the lower edge of the board. (60 moves)

10. Do the same as in the previous problem in the 3×2 box (45 moves.)

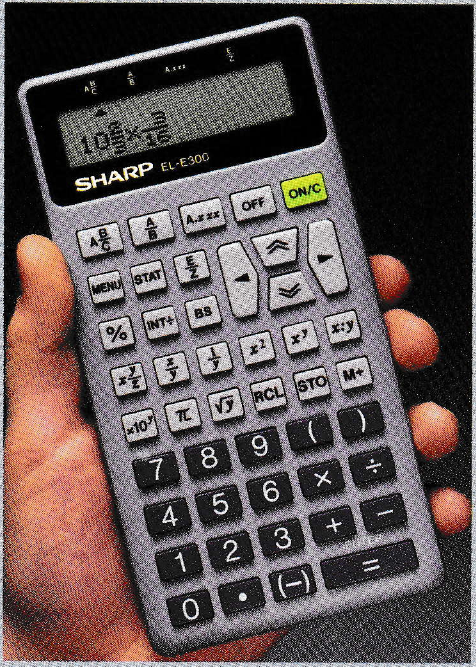
11. You're given five cubes in the 3×2 box having the same orientation. Roll them so as to obtain five different colors other than the initial one on their top faces. (The author, V. Rybinsky, can do this in 14 moves.)

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

[1]According to Martin Gardner, Harris later found a 74-move solution.

THE NEW WORKPLACE™

[Sharp EL-E300 *Teaching* **Calculator**]



"Hey, now I get it!"

It's a new kind of calculator. For a new way of teaching. Doing what no other calculator has ever done before. Using Sharp's *Equation Editor* to display entire equations on screen-just as you'd write them. Explaining in words on the screen-how to solve fraction problems. And making statistics easy-since there are no other intimidating functions. So it's *easier than ever* for 4th-8th grade students to learn math–even when you're not there. And, there's a *10-unit Classroom Pack* that comes with a versatile calculator application guide. The Sharp EL-E300 Teaching Calculator. So convenient, it makes anyplace, your workplace. For a brochure and the name of your nearest educational distributor, call

1-800-BE-SHARP

OVERHEAD MODEL THAT DISPLAYS SCREEN & KEYS WILL BE AVAILABLE FOR CLASSROOM USE.

SHARP
FROM SHARP MINDS COME SHARP PRODUCTS™

LCD

COPIERS • FACSIMILE • LASER PRINTERS • NOTEBOOK COMPUTERS • COLOR SCANNERS • PERSONAL DIGITAL ASSISTANTS • CALCULATORS
CORDLESS TELEPHONES • TYPEWRITERS • WORD PROCESSORS • WIZARD® ORGANIZERS

Durable plastic case included.

Circle No. 3 on Reader Service Card