MAY/JUNE 1993

$5.00

QUANTUM

THE STUDENT MAGAZINE OF MATH AND SCIENCE

SPRINGER INTERNATIONAL

Kosti Ruohomaa, *Black Star*

GALLERY Q

"YOU'RE NEVER TOO OLD TO swing"—that's what this classic photograph seems to be saying. It appeared in the photographic exhibition "The Family of Man" at the Museum of Modern Art in 1955. The work of 273 photographers from around the world—503 photos culled from over two million—showed humanity in the breadth and depth of its common experience: birth and death, work and play, learning, loving, building, arguing, dreaming—the gamut of life. "The Family of Man" was seen by over nine million people, and a book of the same name is still in print.

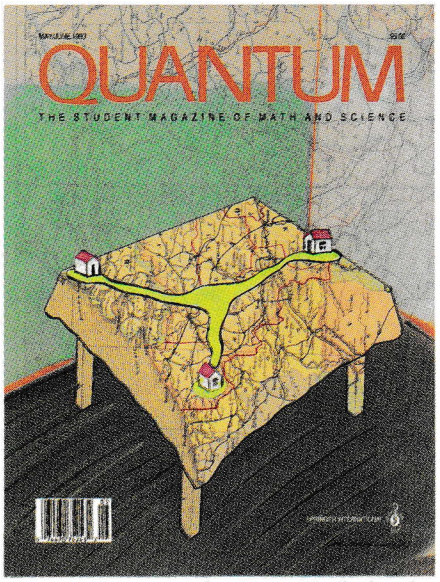
The exhibition was the brainchild of Edward Steichen (1879–1973), a legendary figure in American photography. His first photographs reflected his early training in lithography and painting. In 1902 the great Alfred Stieglitz invited Steichen and eleven other photographers to join him in founding the Photo-Secession, dedicated to promoting photography as fine art. The group was open to all styles and approaches. Steichen himself soon moved away from soft-edged, sentimental pictures toward cleaner, more direct images. In a search for a perfect rendering of gradations of black, gray, and white, he photographed a white cup and saucer against a black velvet background more than a thousand times. He also photographed such celebrities as Greta Garbo and Charlie Chaplin for *Vogue*.

Steichen was 70 years old when "The Family of Man" opened. He married for the third time at eighty. Steichen felt that "boredom and disinterest and lack of awareness are what characterize a superannuated mind, no matter what the chronological age may be. Happily, growth is not something for children only."

Nor is play, as our swinging couple demonstrates. Some prefer to swing sitting down, some standing up. In this photo, taken by Kosti Ruohomaa somewhere in the US, both styles are represented simultaneously. Here the woman apparently is keeping the swing going by pumping her legs. What if she weren't there—how would the man keep swinging? The Toy Store at the other end of this issue explores the physics of swinging.

QUANTUM

MAY/JUNE 1993 VOLUME 3, NUMBER 5



Cover art by Yury Vashchenko

When we consult a map, usually it's to find the best route from where we are to where we want to go. And "best" usually means "shortest." Unfortunately (or perhaps not), in real life the shortest distance from A to B is generally *not* a straight line. Obstacles present themselves: rivers and mountains if we're driving; bad weather or protected air space if we're flying; trees, fences, and all manner of things if we're walking.

Mathematicians suffer no such impediments when they deal with the distance between points. But they usually end up wanting to go to n places (not just one) or trying to connect them. Say you wanted to create the shortest possible network connecting n points—how would you do it? The article beginning on page 4 tackles this very problem.

Speaking of maps, we seem to recall that you need only four colors when you make a map . . . the details escape us. Maybe "The Mapmaker's Tale" on page 46 will refresh our recollection.

FEATURES

DEPARTMENTS

Of microscopes, e-mail, and word of mouth

It's not enough to have a good product—you have to sell it

IN THE JANUARY/FEBRUARY issue I spoke of the difficulties facing ex-Soviet science and described some of the steps Russian scientists are taking to ensure the survival of a prodigious scientific legacy. I'm pleased to announce that the National Science Teachers Association, through an opportunity created by *Quantum* magazine and our colleagues in Moscow, is offering high-quality Russian scientific equipment for sale in the United States. The Russian Academy of Sciences has created Russian–American Science, Inc. (RAS) to serve as its commercial representative in the US. We are working with RAS in marketing microscopes and other items that Russian teachers can no longer afford to buy.

I'm very excited about this project. It will bring a much-needed infusion of capital into the strapped Russian economy and establish a foothold for Russian technology in the world market. It will help *Quantum* magazine balance its books (we do not as yet take in as much as we spend on the magazine). And it will offer the US consumer a taste of the kinds of products the former Soviet Union is capable of producing, and at a very reasonable price.

If you're interested in the product line we will initially be offering, please write to me. I'll send you a brochure.

Feed us!

I've said it more than once, but it remains true as ever: we want and need your feedback. When we print and mail each issue of *Quantum*, it's as if we "shoot an arrow into the air." Like the arrow, *Quantum* "comes to Earth we know not where." We need to know when we hit the target and when we miss. Bull's-eyes are nice to hear about, but we also need to hear about the shots that land in the next county. Not only do we want to provide the best magazine for our current readers, we need to expand our circulation if we are to survive.

You may have noticed an ad in the last few issues, calling on all "modem maniacs" to contact us by e-mail. I don't know about you, but nowadays I write and read more electronic mail than the paper variety. We've heard from some of our readers, but what about the rest of you? Write! Make it a habit to tell us which article was your favorite in each issue, which one fell short. Do you like the science crossword puzzle (a feature we added in the past year)? Let us know a topic you really want to see covered. We can't promise immediate results, but your input will definitely affect future issues of *Quantum*.

So, tell our managing editor, Tim Weber, what you think. His e-mail address is

72030.3162@compuserve.com

Take us with you!

Summer vacation is almost upon us, and I'm sure many of you have travel plans. Why don't you take *Quantum* with you? Work through the problem that resisted your efforts during the school year. Catch up on an article that you skipped. And, please, show *Quantum* to lots of people. We try hard to put the magazine in the hands of people we think will enjoy it, but word of mouth is a powerful—perhaps *the* most powerful—way of publicizing a product that "isn't for everyone." *Quantum* will probably never be a mass-market magazine, but its audience could certainly be larger. *You* know someone who would like it—*tell* them about it!

Thanks for your continuing support. I wish all our readers a pleasant summer, exciting and relaxing by turns, and a safe one. Look for the July/August issue of *Quantum* to break up any midsummer doldrums. It'll have an article on the "omnipresent and omnipotent neutrino" and a tongue-in-cheek investigation of what the Moon is *really* made of. Enjoy!

—Bill G. Aldridge

QUANTUM

THE STUDENT 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)

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York Inc. Volume 3 (6 issues) will be published in 1992–1993. *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, 3140 North Washington Boulevard, 2nd Floor, Arlington, VA 22201. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010. Production Editor: Madeline R. Kraner.

Advertising:

Advertising Representatives: (Washington) Paul Kuntzler (202) 328-5800; (New York) Bob Vrooman (212) 460-1700; and G. Probst, Springer-Verlag, Heidelberger Platz 3, 1000 Berlin 33, FRG, telephone (0) 30-82 07-1, fax 30-82 07-1.

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, 44 Hartz Way, Secaucus, NJ 07096-2491.

Subscription Information: North America

Student rate: \$14; Personal rate (nonstudent): \$18; Institutional rate: \$30; 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; 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, Heidelberger Platz 3, 1000 Berlin 33, FRG.

Shortest networks

The answer to the perennial question: what's the shortest distance between three (or more) points?

by E. Abakumov, O. Izhboldin, L. Kurlyandchik, and N. Netsvetayev

THIS ARTICLE IS DEVOTED to the famous Shortest-Network Problem, proposed by Jacob Steiner, an outstanding Swiss geometer of the last century:[1]

The residents of a number of villages are going to build a system of roads that would connect every village to every other village and would have the least possible total length. How should they do this?

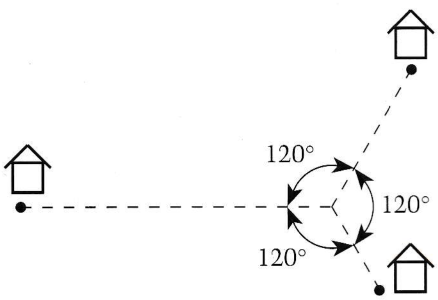


Figure 1

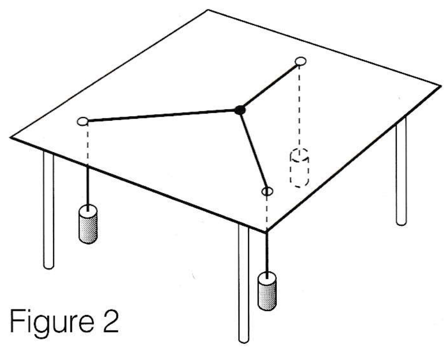


Figure 2

Three villages

It's rather simple to connect three villages in the shortest way: if all the angles of the triangle formed by these villages are less than 120°, its vertices should be connected to one point, defined by the condition that the angles subtended by the sides of the triangle at this point are each equal to 120° (fig. 1).

This can be proven using the device shown in figure 2.[2] The three holes in the table are drilled where the villages are located on the imaginary map spread on the table. Three strings are tied together and passed through the holes and are pulled by equal masses attached to their ends. When let free, the weights must end up at the position with the lowest total potential energy—that is, with the greatest possible total length of string below the table. Therefore, the string on the table's surface will indicate the required shortest connection. But why will the angles between them measure 120°? Note that there are three forces of *equal magnitude* acting on the knot (fig. 3), and they can balance each other only if they make equal angles.

Exercise

1. Prove this.

At this point the proof would be finished, if we were sure that the knot wouldn't fall through a hole. But when could this happen? When the sum of two of our tension forces directed along the triangle's sides (fig. 4) is not greater (in magnitude, of course) than the third force. Since

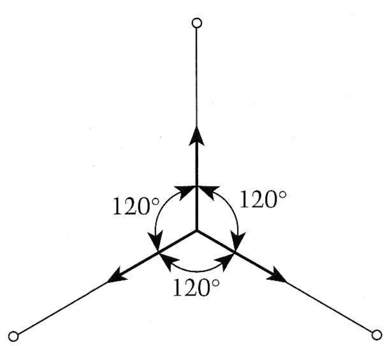


Figure 3

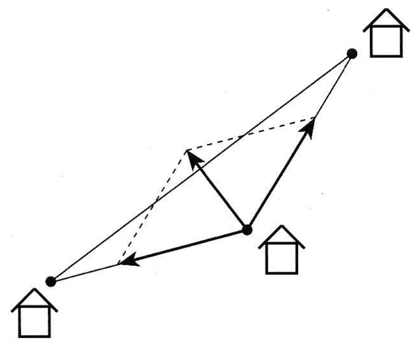
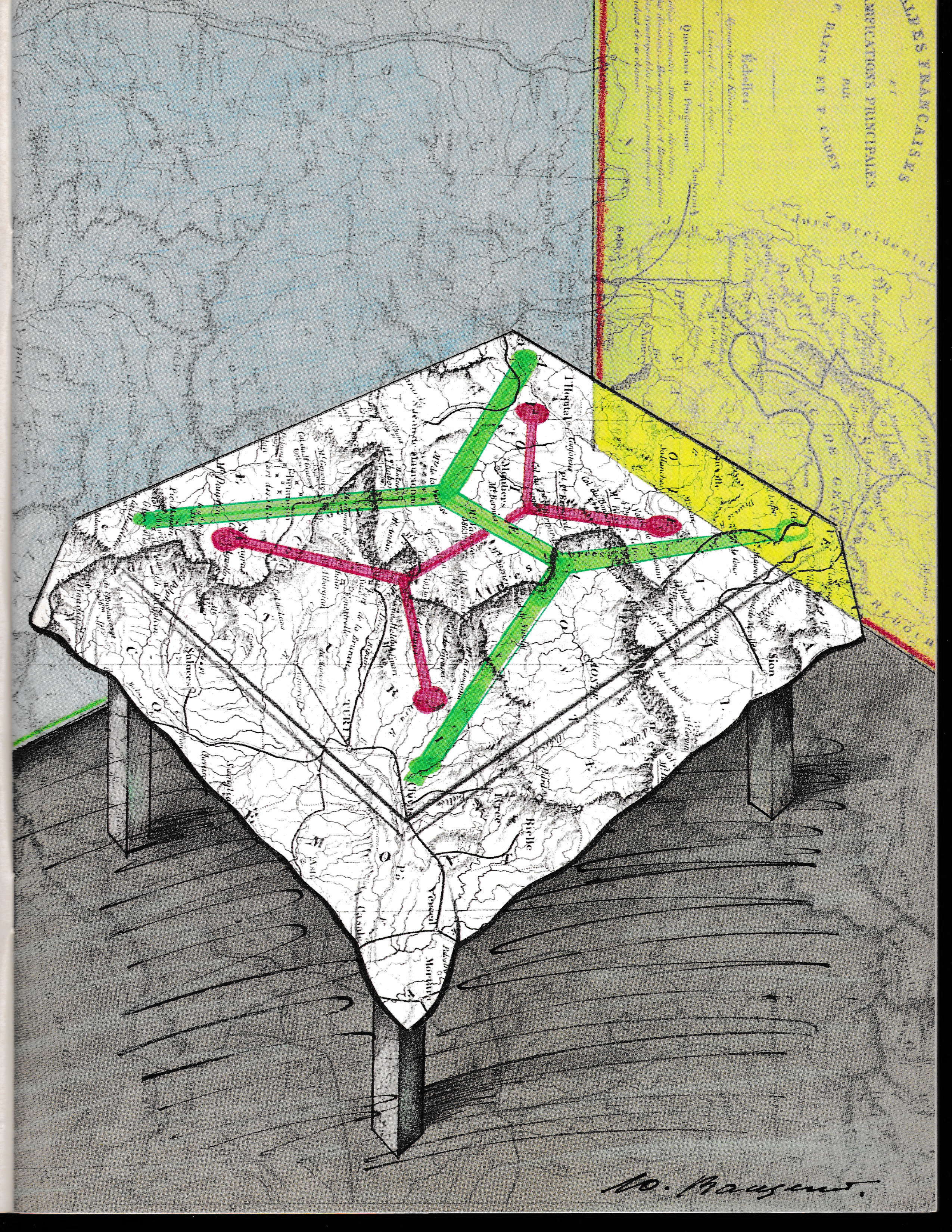


Figure 4

Art by Yuri Vashchenko

[1]The story of his life and work will soon be published in *Quantum*.—*Ed.*

[2]We've already described it in "Botanical Geometry," the Kaleidoscope in the September/October 1990 issue of *Quantum*. There you can also find more on the subject—for example, the case of four points—as well as the solutions to some of the exercises below.—*Ed.*


ALPES FRANÇAISES
PAR
F. BAZIN ET F. CADET
Échelles:
Jura Occidental

all the forces are of the same magnitude, this is possible only when the angle between the first two forces is not less than 120°.

So our discussion of the case of three villages can be summarized as follows:

If the angles of the triangle formed by the given points are all less than 120°, its vertices should be connected to the point at which each side of the triangle subtends an angle of 120°.

If one of the triangle's angles is not less than 120°, the vertex of this angle should be connected to the other two vertices.

This completes the solution, but we want to add three notes.

NOTE 1. In actual fact, we've replaced the initial question about the shortest network with the problem of finding the point with the smallest sum of distances to three given points. (This point is called the *Fermat point* of the corresponding triangle).

Before going further, think about whether this replacement is legitimate. Is a similar reformulation of the problem for four given points also legitimate?

NOTE 2. There is another mechanical device that "constructs" the shortest network (for three points). Instead of drilling holes, let's drive three nails into the table at the same places and stretch a rubber band around them, passing it through a small ring, as shown in figure 5. In the absence of friction, the rubber band will tend to become as short as possible, which is just what we need.

NOTE 3. We've learned how to find the shortest network in a triangle, yes, but at considerable cost: we had to damage the table. It would be better to find a purely geometric construction of the point we need. One way is shown in figure 6: construct equilateral triangles on the sides of a given triangle, outside it, and join their "remote" vertices to the opposite vertices of the given triangle. The three lines thus drawn meet at the Fermat point. Of course, here we assume that the triangle's angles are less than 120°.

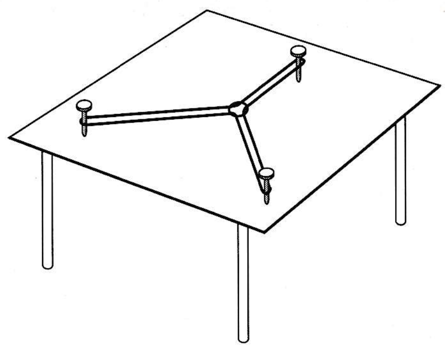


Figure 5

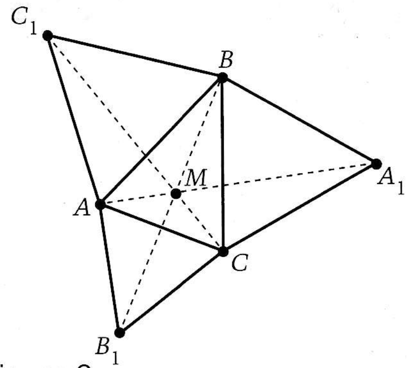


Figure 6

Exercises

2. Prove that the lines described above (AA_1, BB_1, CC_1 in figure 6) have a common point M.

3. Prove that $\angle AMB = \angle BMC = \angle CMA = 120°$.

4. Prove that $AM + BM + CM = AA_1 = BB_1 = CC_1$.

5. Prove that $AX + BX + CX > AM + BM + CM$ for any point $X \neq M$ (if the angles of triangle ABC are less than 120°).

6. Show that if $\angle BAC \geq 120°$, then $AB + AC < XA + XB + XC$ for any $X \neq A$.

Clearly, exercises 2, 5, and 6, together with note 1 above, constitute a complete geometric solution of the shortest network problem for three points.

The properties of a shortest network

Now let's proceed to the general Shortest-Network Problem:

Connect n points given in the plane with a network of line segments of the smallest total length (the "shortest network").

For the discussion that follows, let's agree on some terms. We define a *network connecting n given points* as a finite set of line segments such that any two of the given points are the endpoints of some polygonal path made up of segments of this set. The given points will be called villages; all the other endpoints of the network's segments will be called *forks* (even if such a fork has only one or two offshoots). Figures 7a and 7b give examples of networks. The diagrams shown in figures 7c and 7d aren't "networks" in our sense.

In figure 7a the fork α is in fact superfluous (to say nothing of its not being a "fork" in the usual sense of the word). It's also clear that we can do without the "fork" β (which is rather a dead-end). These observations suggest that we can restrict ourselves to networks whose forks have not less than three offshoots. More rigorously, if there are only two segments that meet at some fork, we can replace them with one segment joining their endpoints. In so doing, the total length of the network doesn't grow, and the number of two-seg-

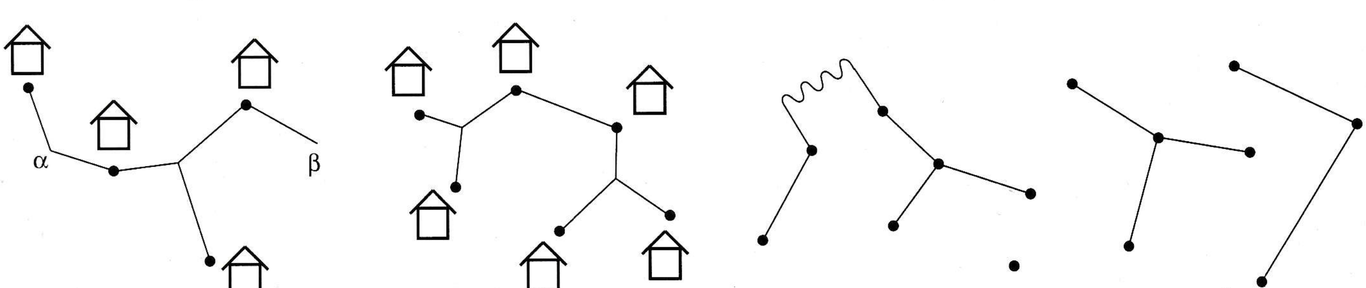


Figure 7 a b c d

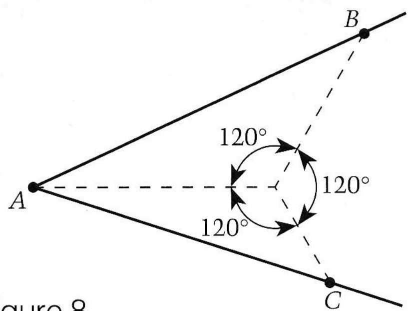


Figure 8

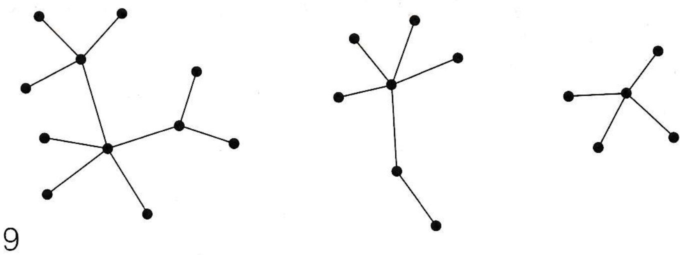


Figure 9

ment forks becomes smaller by one, even though the new segment may cross some of the old segments, creating new "regular" forks. If a fork is an endpoint of only one segment, we can simply remove it, making the total length smaller.

We can also show that the angle BAC between two segments AB and AC of a shortest network cannot be less than 120°. Otherwise, we could replace these segments with the paths through the Fermat point of triangle ABC. The three points are connected by this network, and also by the segments AB, AC. Since the network through the Fermat point is the shortest, it must be shorter than the sum of the two segments (see figure 8). Hence our replacement will shorten the entire network. It follows, in particular, that the number of segments issuing from a village or a fork is never greater than three (because four or more segments form at least two angles less than 120°).

So here are the two main properties of a shortest network:

Property 1. *Any fork is the common endpoint of exactly three segments that make angles of 120° with each other.*

Property 2. *There are one, two, or three segments issuing from every village; if the number of segments is two, the angle between them is not less than 120°, and if the number is three, the angles between them are each 120°.*

One more property follows from a simple observation that a network containing a closed circuit of segments can be shortened by erasing any of these segments without breaking the connection between villages. So we have another property:

Property 3. *The segments of a shortest network never form a closed circuit.*

How many forks can a network have?

We've proven three fundamental properties of a shortest network. If you know anything about *graphs*, it's already clear to you that a shortest network is a *tree*. For the rest of our readers, we'll give a definition of this notion adjusted to our purposes: a tree is a set of line segments on the plane that has no closed circuits and is "connected"—that is, any two of these segments are connected by a chain of segments in which every two neighbors have a common endpoint. The segments making a tree are called its *edges*, and their endpoints are the *vertices* of the tree.

Three different trees are illustrated in figure 9. As you can see, for each of them *the number of vertices V is one greater than the number of edges E*. This is a general property of trees and can be proven by induction over the number E. The crux of the proof is that any tree contains an edge one of whose vertices does not belong to any other edge (otherwise, a closed circuit of edges must appear); removing this edge together with its "hanging" vertex, we obtain a new tree with a smaller number of edges but the *same* difference $V - E$, which must be equal to one by the induction assumption. The details of this proof are left to the reader.

For our networks, this means that the number of villages plus the number of forks is one greater than the number of segments.

But we can count the segments another way: since three segments emerge from each fork, and at least one segment emerges from each village, the number of segments is not less than $(3f + v)/2$ where f and v are the numbers of forks and villages, respectively, and we divide by two because $3f + v$ accounts for every segment twice—a segment has two endpoints. It follows that $f + v - 1 \geq (3f + v)/2$, or $f \leq v - 2$.

Property 4. *The number of forks is at least two less than the number of villages.*

Exercise

7. Prove the number of forks of a shortest network is exactly $v_1 - v_3 - 2$, where v_1 is the number of villages with one "road" starting at them and v_3 is the number of villages with three "roads."

Shortest network in a square

Now we know enough to be able to effectively apply our knowledge to a concrete situation. Consider four villages located at the vertices of a square. What is the shortest network connecting them?

It's not hard to see that the network will not extend beyond the square. It follows that there is only one road issuing from each village. Otherwise, the two roads issuing from the same village would form an angle less than 90°, hence less than 120°. This contradicts our second property of shortest networks.

In the notations of exercise 7, $v_1 = 4$, $v_3 = 0$, so the number of forks

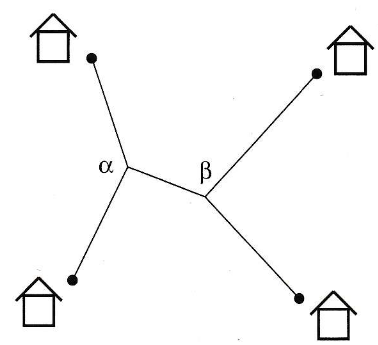


Figure 10

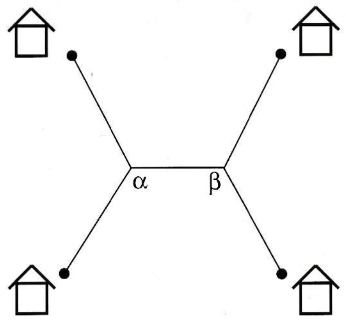


Figure 11

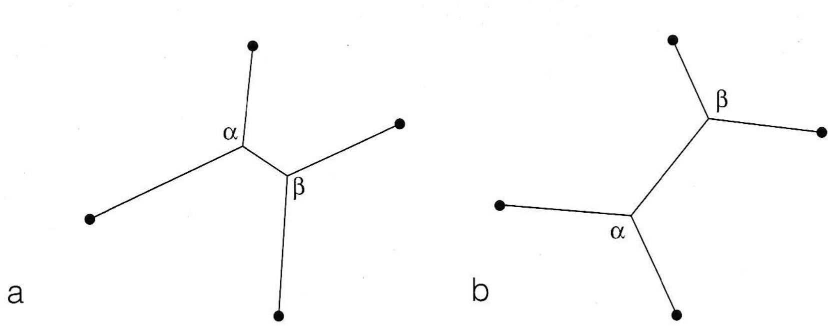


Figure 12

is $v_1 - v_3 - 2 = 2$. The only way to draw segments in accordance with these requirements is to join each fork to two villages and to the other fork, as shown in figure 10. It remains for us to locate exactly the forks α and β in this figure.

Using Property 2 again, we conclude that the six angles in figure 10 are each 120°. Now it's easy to draw the entire network (fig. 11).

Four villages

Now let's solve a more difficult problem. Four villages are located at the vertices of a convex quadrilateral whose angles are less than 120°. What is the shortest network connecting them?

All the arguments for the case of a square are applicable again, and they yield the same structure for the network. But now there are two (generally, distinct) conceivable ways to draw such a network—see figures 12a and 12b. For a square they could be considered the same, because one of them could be obtained from the other by a 90° rotation of the square.

So which of the two ways is the one we need? To answer this question we must simply measure the total lengths of the two networks and choose the shorter one. But to do this, we need to know how to *construct* the shortest network of a given shape (say, as in figure 12a)—that is, to correctly locate forks α and β.

A construction is shown in figure 13, where points A, B, C, and D are the villages. We note that segments αA, αB and $\alpha\beta$ form the shortest network for triangle $AB\beta$, so α is the Fermat point of this triangle. According to exercise 3, angle $A\alpha B$ = 120°, so point α lies on the segment $O_1\beta$, where O_1 is the vertex of the properly constructed equilateral triangle ABO_1 (see the construction in figure 6). Similarly, β lies on the line αO_2, where O_2 is found from the equilateral triangle CDO_2. So points α and β both lie on the line O_1O_2. On the other hand, $\angle AO_1B + \angle A\alpha B = 60° + 120° = 180°$, so point α lies on the circumcircle of triangle ABO_1, and, for a similar reason, β lies on the circumcircle of triangle CDO_2. So what we have to do is construct equilateral triangles ABO_1 and CDO_2 outside the quadrilateral $ABCD$; draw the circumcircles of ABO_1 and CDO_2; and, finally, take the intersections of the circles with the line O_1O_2 to be the forks of the desired network.

The other possible shortest network (for figure 12b) is constructed in the same way. Of these two, the shortest one yields the answer to the problem.

It may amuse the reader to devise a "mechanical" solution to this problem using a rubber band, two rings, and four nails.

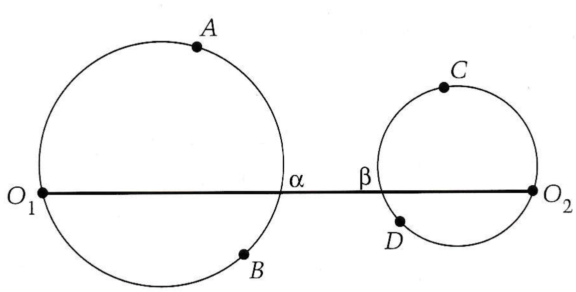


Figure 13

Exercises

8. We required that the angles of $ABCD$ be less than 120°. Why?

9. Prove that the total length of the shortest network constructed according to figure 13 is simply equal to O_1O_2.

10. Find the length of the shortest network in a rectangle measuring 3×4.

In a regular pentagon

Now let's try to find the shortest network for five villages forming a regular pentagon. The angles of the pentagon are each equal to 108° < 120°, so there's only one segment from each village again, and the number of forks is $v_1 - v_3 - 2 = 5 - 0 - 2 = 3$. Essentially there's only one connection scheme—the one shown in figure 14; the other four are obtained from it by rotations of the pentagon. It remains for us to find the actual positions of the forks α, β, γ (fig. 15). We'll apply the idea we've used already: construct equilateral triangles ABO_1 and CDO_2 outside the pentagon. An argument similar to one already given will show that the forks α and γ lie on segments $O_1\beta$ and $O_2\beta$, respectively. Hence the three segments βO_1, βO_2, and βE form three angles each measuring 120°, so β is the Fermat point of triangle O_1O_2E, and we can construct it by the method described above.

Now points α and γ can be found

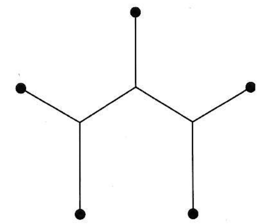


Figure 14

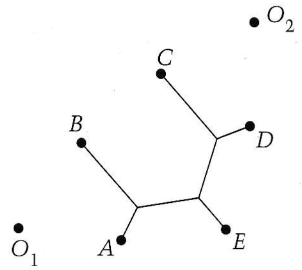


Figure 15

as the Fermat points of triangle $AB\beta$ and $CD\beta$, and all that remains is to join them as in figure 15.

Shortest networks in the general case

How should the Shortest-Network Problem be handled in the general case of n arbitrarily located villages? A thoughtful reader must have noticed that our solutions for 4 and 5 villages consisted of two steps. First we found the connection scheme—that is, determined the possible number of forks and which forks and villages can be connected by segments. In the second step, we used this information to determine the actual location of the forks. In the general case we'll proceed along the same lines.

What do we have to do in the first step? As we know, the number of forks is not greater than $n-2$, so there is a finite number of possible connection schemes (to be more exact, "graphs") for the given n points. Therefore, we can make an exhaustive list of all possible connection graphs. Figure 16 illustrates a number of connection graphs for five villages.

In the second step, searching through the list of graphs, we need to construct the shortest network for each of them, measure its length, and choose the absolutely shortest one. How is the shortest network for a particular graph constructed? It's sufficient to show how the problem for n villages can be reduced to the construction of the network for a smaller number of villages.

So, suppose we're given some connection graph for $n \geq 4$ villages. If it contains two villages connected by a segment, then, removing it, we'll obtain two graphs with fewer than n villages; the problem can be solved for each of them separately. For, if some network with the given graph is the shortest for this graph, then each of the two subnetworks emerging after our "surgery" must be the shortest one among all those that have the same connection graph. Otherwise, we could shorten one of the subnetworks—and thereby the enitre network—without spoinling the entire connection graph.

If every edge that starts at some village leads to a fork, there are two villages connected to the same fork (because the number of forks is less than the number of villages). Label these villages and the fork A, B, and α, respectively. Consider the third segment issuing from α. If it leads to some third vertex C, we place α at the Fermat point of triangle ABC and then take away edges αA, αB, αC, thus dividing the graph into three mutually disconnected graphs with a smaller number of vertices and completing the (first) reduction. Finally, if the third edge issuing from α leads to another fork β, we again apply the method applied repeatedly above. We construct the equilateral triangle ABO_1 and note that the fork α must lie on $O_1\beta$ and that $\alpha A + \alpha B + \alpha\beta = O_1\beta$ (see figure 13 and exercise 4). Now, replacing two villages A and B by one village O_1 and the segments αA, αB, and $\alpha\beta$ by $O_1\beta$, we reduce the problem to constructing the shortest network for a graph with $n-1$ vertices. Having done this, we'll locate all the forks except α; after that, we find α at the intersection of $O_1\beta$ and the circumcircle of triangle ABO_1.

An alert reader will certainly have discovered a number of inaccuracies in this argument. The following exercises will help you fill (and sometimes, perhaps, become aware of) the gaps in the construction.

Exercises

11. What is to be done if there are two or three roads issuing from the village A (or B) above?

12. There are two equilateral triangles with a side AB. Which of them must we choose?

13. Does our method of constructing a network, given a connection graph, always yield a "network" in actuality?

14. What should be done if point O_1 coincides with some village? Or if one of the segments $A\alpha$ or $B\alpha$ crosses one of the previously constructed segments?

What do existence theorems exist for?

After all this, can we now say that the Shortest-Network Problem for n points is completely solved? No, not yet!

Our solution lacks one essential detail—a proof that the shortest network exists. But is that really an important question? Yes, it certainly is: the very algorithm of construction is based on the existence of the network we seek. However, one might come away with the impression that the fact of existence in itself is obvious, and that a proof is needed simply because mathematicians commonly want to prove everything. But that impression would be deceptive. To understand why, look at the following problem:

Connect two villages by the shortest network of roads such that they remain connected even when the traffic along an arbitrary road in the network is blocked.

The setting is quite like the original one, and the problem is undoubtedly much simpler—there are only two villages. But the similarity is

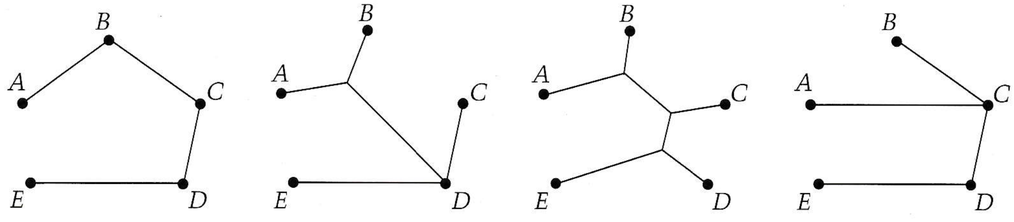


Figure 16

CONTINUED ON PAGE 31


Куда ты улетаешь,
гортензия моя?
Тишков 1993

A burst of green

How fast does a leaf grow?

by Alexander Vedenov and Oleg Ivanov

NO DOUBT MANY OF OUR readers have noticed how quickly a tilled and sown field becomes covered with fresh verdure. It remains black or brown for some time. But then little green sprouts appear, and before your eyes—in just a week or so—the field looks like a green carpet. (Of course, this is contingent on favorable conditions: that most of the seeds germinate, the soil is moist enough and well fertilized, and the days are sunny.)

Experimental studies have shown that under constant external conditions (constant lighting, temperature, and humidity, regular watering, and well-tilled soil), the weight and size of a plant and its parts increase exponentially with time during the initial stages of growth. Figure 1 shows the results of experiments on growing wheat in a solution of nutrients. The scale of the abscissa is linear and that of the ordinate is logarithmic—that is, the evenly spaced distances on the horizontal axis show the days since the start of growth, and the vertical axis shows the logarithms of the measured values. Notice that the scale for ln *y* is linear and that the graphs are straight lines that can be represented by

$$\ln y = \ln A + Cx,$$

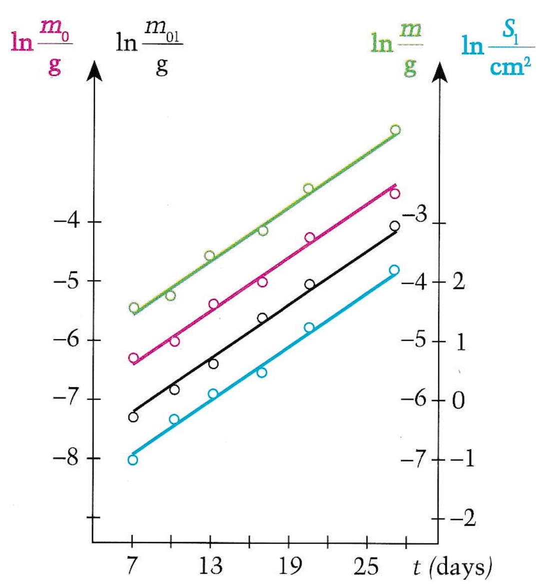


Figure 1

where C is the slope of the line and $\ln A$ is the y-intercept. It's easy to determine the slope C from any of the lines by calculating the rise over the run—that is, the change in ln y divided by the corresponding time interval. What value do you get?

We can see in figure 1 that, beginning with the seventh day after germination for a period of three weeks, the total weight of the plant m, the dry weight of the plant m_0 (that is, the weight of the organic material remaining after drying), the dry weight of the leaves m_{01}, and the area of the leaves S_1 increase exponentially with time. In nature, exponential growth is a widespread phenomenon. The growth of the number of bacteria or single-cell organisms in a culture medium is a good example.[1]

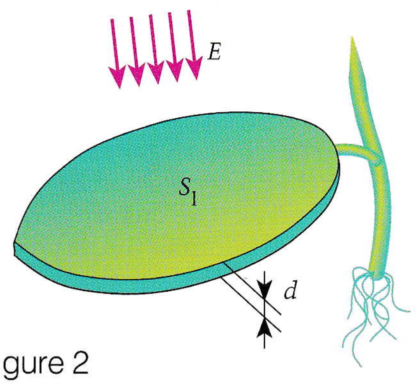


Figure 2

Let's find the physical explanation for the exponential growth of plants and determine the constant C of the growth rate from the laws of conservation of mass and energy. First we need to construct a model that reflects the reality of plant life and is convenient for our investigations. Suppose we consider only one leaf (whose area is equal to the area of all leaves of a given plant), a stem, and roots (fig. 2). Let's suppose further that the ratio of the total mass to the dry mass of the entire plant (as well as that of its separate parts) is a constant:

$$\frac{m}{m_0} = \frac{m_1}{m_{01}} = \alpha,$$

Art by Leonid Tishkov

At left: "Where are you off to, my dear Hydrangea?"

[1]See "[Getting to Know] The Natural Logarithm" in the November/December 1990 issue of *Quantum*.—*Ed.*

where m_1 and m_{01} are the total mass and dry mass of the leaf. Finally, we'll consider the thickness of the leaf and the ratio of the leaf's dry mass to the plant's dry mass to be constants:

$$\frac{m_{01}}{m_0} = \varepsilon.$$

(Such suppositions are based on quantitative observations during the period of exponential growth of grassy plants. Of course, one can't use such a model for describing the growth of other plants. For instance, the ratio of the mass of tree leaves to the mass of the entire tree is a variable.)

From biology we know that a plant increases its mass due to the process of photosynthesis. Organic substances are formed out of carbon dioxide in the air and water in the plant when light energy is absorbed by the plant's leaves. The quantity of dry organic matter formed Δm_0 is proportional to the light energy absorbed ΔE:

$$\Delta m_0 = \gamma \cdot \Delta E. \qquad (1)$$

If the intensity of the light flux I (the light energy falling on a unit area per unit time) is constant, then

$$\Delta E = IS_1 \cdot \Delta t. \qquad (2)$$

Using our plant model, let's express the leaf's area S_1 in terms of the plant's dry mass m_0. The leaf's area is equal to the leaf's volume V_1 divided by its thickness d: $S_1 = V_1/d$. We can find the volume by using the leaf's raw (predried) mass and its density: $V_1 = m_1/\rho$ ($\rho \cong 1$ g/cm^3). And we can express the leaf's raw mass in terms of m_0, using the formulas for the coefficients α and ε: $m_1 = \alpha\varepsilon m_0$. As a result, we get

$$S_1 = \frac{m_0 \varepsilon\alpha}{\rho d}. \qquad (3)$$

Combining equations (1) through (3), we obtain the equation for the rate of change in the amount of dry organic matter:

$$\frac{\Delta m_0}{\Delta t} = \frac{\gamma\varepsilon\alpha I}{\rho d} m_0.$$

We know from the graph in figure 1 that the slope C is given by

$$C = \frac{\Delta(\ln m_0)}{\Delta t}.$$

If you know calculus, you know that $\Delta(\ln m_0) = \Delta m_0/m_0$. However, you can obtain this expression without calculus:

$$\begin{aligned}\Delta(\ln m_0) &= \ln(m_0 + \Delta m_0) - \ln m_0 \\ &= \ln\left(\frac{m_0 + \Delta m_0}{m_0}\right) \\ &= \ln(1 + \Delta m_0/m_0) \\ &\cong \Delta m_0/m_0,\end{aligned}$$

where we have used the approximation $\ln(1 + x) \cong x$ for $x << 1$. Therefore,

$$C = \frac{\Delta m_0}{m_0 \Delta t} = \frac{\gamma\varepsilon\alpha I}{\rho d}. \qquad (4)$$

Thus, the constant for the growth rate is proportional to the intensity of the light falling on the plant and depends on the parameters α, ε, d, and γ. For wheat, it's possible to determine α, ε, and d from the graph in figure 1 by using the values m, m_0, m_{01}, and S_1 for a given day. For example, on the thirteenth day, $m = 0.025$ g, $m_0 = 0.0043$ g, $m_{01} = 0.0021$ g, and $S_1 = 1$ cm^2 (we obtained these values with a calculator). So

$$\alpha = \frac{m}{m_0} = 5.8, \quad \varepsilon = \frac{m_{01}}{m_0} = 0.49,$$

and we can determine the thickness of the leaf by using the same values in the ratio of the leaf's volume $V_1 = m_{01}\alpha/\rho$ to its area:

$$d = \frac{m_{01}\alpha}{\rho S_1} = 0.012 \text{ cm}.$$

The value of the *photosynthetic equivalent* of light (γ) determines how much the plant's dry mass increases when it absorbs a certain amount of light energy. To gain a better understanding of this value, let's derive it by examining the energy processes at play in a plant cell.

As a whole the process of photosynthesis can be given by the equation

$$H_2O + CO_2 \xrightarrow{\text{light}} (CH_2O) + O_2.$$

On the left side of the equation you see the molecules of water and carbon dioxide (certainly familiar to you). On the right side we have an oxygen molecule and a minimal structural group of carbohydrates CH_2O. A molecule of plant sugar (glucose) $C_6H_{12}O_6$ consists of six of these groups—$C_6H_{12}O_6 = 6(CH_2O)$—and a molecule of ordinary sugar $C_{12}H_{24}O_{12}$ consists of 12 such groups. The word "light" above the arrow means that light energy must be absorbed for the transformation to occur.

Only the initial and final products are given in the equation. The entire process of photosynthesis includes many intermediate elements and reactions. In the cells of higher plants—more exactly, in their *chloroplasts*—molecular "factories" churn out goods. Protein molecules (*enzymes*) in the chloroplasts are the "machines" that, powered by light energy, generate electrical energy, carry out electrolysis, and rearrange atoms and atomic groups into molecules.

Chlorophyll molecules in the chloroplasts absorb light. When one photon is absorbed by a chlorophyll molecule, one of the electrons in the molecule is knocked to a higher energy level. Ultimately the energy from these excited electrons is used to transfer a charge along an electric circuit consisting primarily of protein molecules. (Chlorophyll molecules and the proteins associated with

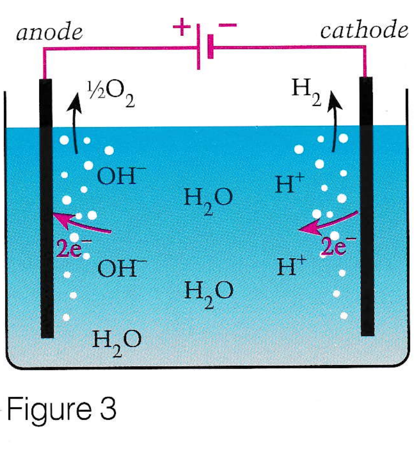


Figure 3

them play the role of photoelectric cells, or photoelements, within the biological cell.) The current flowing through the circuit electrolyzes water to form ions. In addition, the current electrolyzes the substance that carries hydrogen atoms from the water molecules to a carbon dioxide molecule. "A" and "A^+" will stand for a molecule of the carrier and its ion, respectively.

This electrolysis can be compared with the electrolysis of pure water, where hydrogen is formed at the cathode and oxygen at the anode. True, in the usual electrolysis of pure water and other electrolytes, the source of the electromotive force (emf) and the lead wires of the circuit are external—outside of the electrolyte—and only the electrodes are inserted in the electrolyte (fig. 3). But in the case of a biological cell, the entire electrolytic circuit is immersed in the electrolyte—that is, in the ion solution that fills the cell (fig. 4).

A cell can be imagined as a small tube. Inside the tube there are emf sources isolated from the electrolyte, proteins, and other organic molecules. The emf sources are the chlorophyll molecules combined with proteins (cellular photoelements) and the lead wires are proteins and other organic molecules. There are also electrodes that come into contact with the electrolyte at the ends of the tube. These electrodes are also organic molecules. In order to transfer one electron along the entire electrolytic circuit, the energy of two excited molecules of chlorophyll is needed. Therefore, the voltage at the electrodes that is necessary for electrolysis is generated by two cellular photoelements connected in series. The following processes occur at the electrodes:

$$\text{cathode: } 2H^+ + 2A^+ + 4e^- = 2AH;$$
$$\text{anode: } 4OH^- = O_2 + 2H_2O + 4e^-.$$

Now we can calculate how many quanta of light are necessary to obtain one carbohydrate group. According to the equation for photosynthesis, two atoms of hydrogen are to be joined to one molecule of carbon dioxide. Consequently, two A^+ ions must be subjected to electrolysis to carry out the transformation. To do this, four electrons must be carried along the circuit, and eight photons are needed to carry four electrons (you'll recall that two chlorophyll molecules must be excited for one electron to be transferred). The energy of each photon must exceed 1.8 eV, which is the minimum energy required to excite chlorophyll molecules. The current from the electrolysis produced by light in green leaves is significant. Given solar light with an intensity equal to 400 W/m^2, the total electrolytic current in all chloroplasts of all the cells in one square centimeter is about 0.005 A. This means that a current equal to 0.15 A flows in a leaf area of 30 cm^2—as much as in a penlight, although there is as yet no technology that can tap into this current directly.

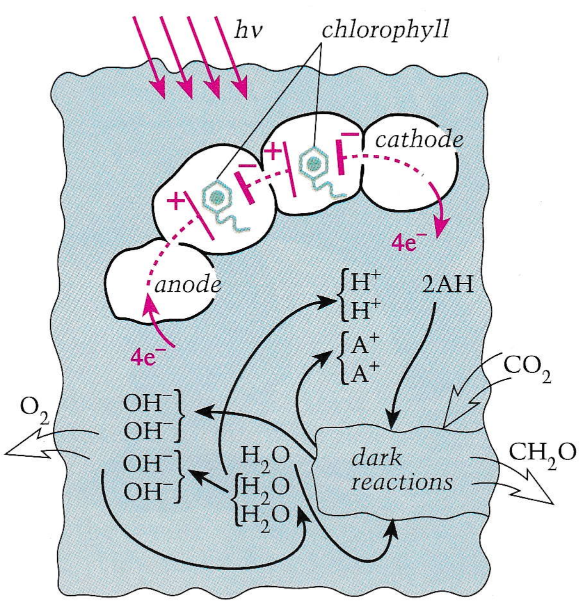


Figure 4

In the next phase of photosynthetic reactions (called "dark reactions" because light energy isn't needed), hydrogen atoms are transferred from AH to the molecule of carbon dioxide, and thus carbohydrates are formed. The carrier molecule, in the form of the A^+ ion, is again ready to take part in electrolysis.

The light illuminating a plant is usually not monochromatic. The portion of the light radiation represented by quanta capable of exciting chlorophyll is called *photosynthetically active radiation* (PAR). The excitation efficiency of chlorophyll molecules depends on the energy of the quanta. Figure 5 shows the distribution of photons over energies in the flux of solar radiation—the red curve $n(h\nu)$—and a curve giving the dependence of chlorophyll excitation efficiency on the energy of the quanta—the green curve $p(h\nu)$. With these curves we can calculate the average energy of a quantum of PAR in solar light—it's ~ 2.1 eV. But the total energy of PAR is approximately equal to one half of all the energy from solar radiation falling on the Earth. (We should note that the proportion of PAR in the overall radiated energy is different for different sources of light.)

We can now use equation (1) to determine the photosynthetic equivalent for PAR. Let Δm_0 be the mass of one mole of carbohydrates, $m_0 = \mu_{CH_2O} = 30$ g. The energy ΔE required to do this is equal to $8 \cdot N_A$ photons with an average energy of 2.1 eV, where N_A is Avogadro's number. This gives an energy of 10^{25} eV = $1.6 \cdot 10^6$ J. Therefore,

$$\gamma = \frac{\Delta m_0}{\Delta E} = 18 \text{ g/MJ}.$$

This calculated value of γ can be observed only under special experimental conditions—when there is no oxygen in the air surrounding the plant. Usually a portion of the intermediate products in the chain of

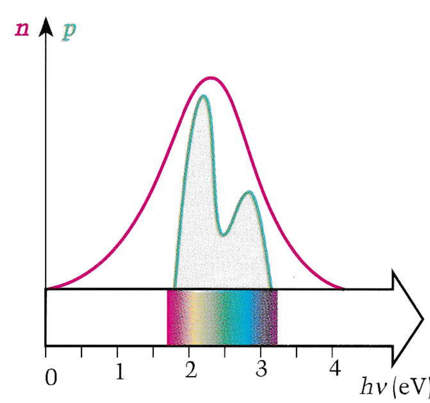


Figure 5

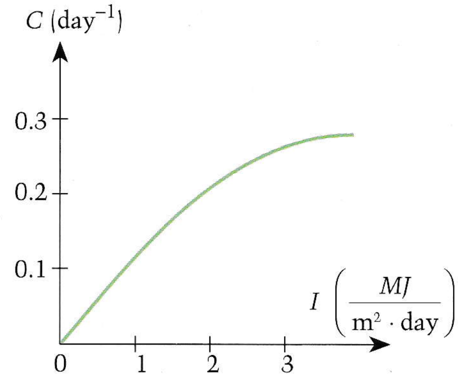


Figure 6

photosynthetic reactions is oxidized by the oxygen in the air and cuts the production of carbohydrates in half. This process is called *photorespiration*. About 30% of the synthesized carbohydrates is used later by the plant as a source of energy for synthesizing substances it needs to grow and for other processes. That's why γ is usually smaller than the maximum value by at least a factor of 3. We'll take γ to be equal to 6 g/MJ.

Now we know everything we need to know in order to calculate the growth rate constant C. For the experiments shown in figure 1, the PAR intensity in artificial light was $G = 30$ W/m^2. Every day the plants were illuminated for 16 hours, and then for 8 hours they were in the dark. The daily average intensity was $\bar{I} = 1.7$ MJ/(m$^2 \cdot$ day). Substituting numerical values for all the magnitudes into equation (4), we obtain $C = 0.24$ day^{-1}. Figure 6 shows the dependence of C on the PAR intensity obtained experimentally. As you can see, the value of C that we calculated coincides with the experimental results.

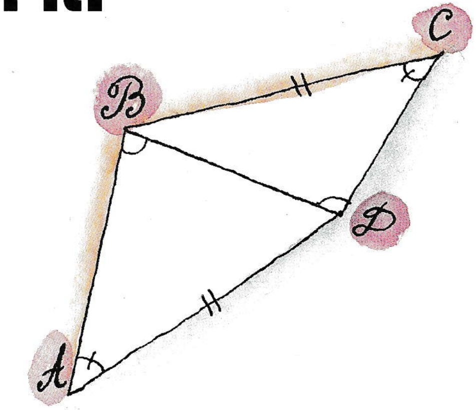
On a bright spring day the PAR intensity reaches 200 W/m^2 at noon, and the average daily intensity $\bar{I} \cong$ 3.5 MJ/(m$^2 \cdot$ day). According to equation (4), the growth rate constant must be twice as large under such conditions. But when the light intensities are greater, ε becomes smaller, because the relative weight of the roots increases—it's necessary to satisfy the increased demands of the plants for mineral nutrients. Therefore, C increases with I nonlinearly but more slowly. When I = 3.5 MJ/(m$^2 \cdot$ day), C is 0.3 day^{-1}.

The calculations given above show that if the conditions are ideal, a plant can grow by a factor of $e^{0.3} = 1\frac{1}{3}$ in a day—that is, it will be a third larger, and such a change is quite noticeable. If the conditions are not very favorable for the plant to grow—say, C = 0.1 day^{-1}—then a noticeable change will occur in two days ($e^{0.1 \cdot 2} = 1.24$). This means that perfectly normal growth rates cause a perceptible change in the size of plants practically every day.

Afterword

The problem we solved in this article—determining a plant's growth rate—is a simple example of the kinds of problems solved by the discipline called *agrophysics*. It's a rather young science where atmospheric physics, soil physics, biophysics, plant physiology, and applied mathematics come together. Agrophysicists study the growth and development of individual plants and of farming regions. They also seek to learn how vegetation is affected by external conditions: light intensity, temperature, soil moisture, humidity, and wind speed.

Just for the fun of it!

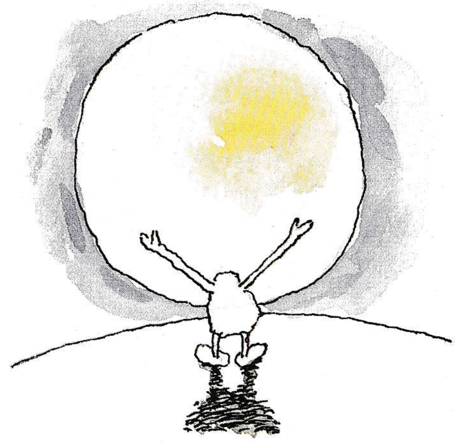


B81

Equal sides, equal angles. In a quadrilateral *ABCD* the sum of the angles *ABD* and *BCD* is 180°, and the sides *AD* and *BC* are congruent (see the figure at right). Prove that angles *A* and *C* of the quadrilateral are congruent. (V. Proizvolov)

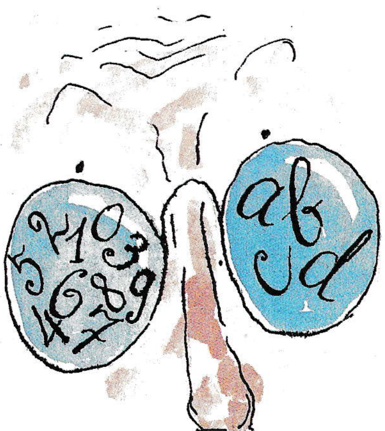
B82

Retrochess. A chess knight traversed a 6×6 chessboard and returned to the starting square after visiting all the other squares once. Some of the squares still bear a trace of the knight's visit—the number of the square in the sequence of the knight's route (see the figure at left). Restore the numbers of all the squares. (A. Savin)



B83

View of the Moon. In which case is the angular diameter of the Moon greater: when it's near its zenith or at the horizon? (V. Surdin)



B84

Calendar puzzle. Denoting different digits by different letters and the same digits by the same letters, I discovered that in the second half of a certain century *bd* there was a year *abcd*. What was that year? (I. Akulich)

B85

Worming out the truth. There are 12 persons in a room. Some of them always tell the truth, the others always lie. One of them said, "None of us is honest"; another said, "There is not more than one honest person here"; a third said, "There are not more than two honest persons here"; and so on, until the twelfth said, "There are not more than eleven honest persons here." How many honest persons are there in the room? (D. Fomin)

ANSWERS, HINTS & SOLUTIONS ON PAGE 59

Art by Pavel Chernusky

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M81

The way to weigh. You have 1993 coins, 20 of which are counterfeit. A counterfeit coin differs from a genuine one only in its weight: it can be heavier or lighter, but always by exactly 1 gram. You're given a pan balance with a pointer showing the difference of the masses on its two pans. Can you tell whether any single coin chosen from the given 1993 is genuine or not by weighing only once? (S. Fomin)

M82

Angled aright. On the sides AC and AB of an equilateral triangle ABC, points D and E are given such that $AD : DC = BE : EA = 1 : 2$. The lines BD and CE meet at point P. Prove that angle APC is a right angle. (A. Krasnodemskaya)

M83

Composite powers and divisibility. Prove that (a) there is an odd number n such that, for any even number k, none of the terms of the infinite sequence

$$k^k+1,\ k^{k^k}+1,\ k^{k^{k^k}}+1, \ldots$$

is divisible by n (here $k^{k^k} = k^{(k^k)}$, and so on); (b) for any natural number (that is, positive integer) n there exists a natural number k such that all the numbers

$$k+1,\ k^k+1,\ k^{k^k}, \ldots$$

are divisible by n. (S. Lavrenchenko)

M84

Skewb lines. Let's define a "skewb" as a hexahedron all of whose six faces are (arbitrary) quadrilaterals joined like the faces of a cube (fig. 1).[1] Prove that if three "big" diagonals of a skewb (that is, lines through the pairs of vertices that don't lie in one face) meet at one point, then the fourth big diagonal also passes through this point. (V. Dubrovsky)

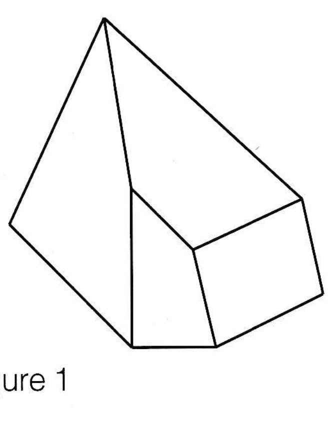


Figure 1

M85

Roots distinct and real. (a) A quadric equation without the square term, $x^4 + ax^3 + bx + c = 0$, has four distinct real roots. Prove that $ab < 0$. (b) An equation of the nth degree, $n \geq 2$, without the term of degree k, $a_nx^n + a_{n-1}x^{n-1} + \ldots + a_{k+1}x^{k+1} + a_{k-1}x^{k-1} + \ldots + a_0 = 0$, has n distinct real roots. Prove that $a_{k-1}a_{k+1} < 0$. (V. Vavilov)

[1]The term was coined by the Englishman Tony Durham as a name for his mechanical puzzle, akin to Rubik's cube.

Physics

P81

Strange velocity dependence. A body is moved slightly from a position of unstable equilibrium. Its velocity increases according to the formula $v(x) = A\sqrt{x}$, where x is the distance from the starting point and A is a constant. How long will it take for the body to travel a distance L? (Z. Rafailov)

P82

Pressure cooker. A small amount of water is poured into a pressure cooker, which is then closed tightly and placed on a burner. The initial temperature is 20°C. At the moment when all the water has evaporated, the temperature T of the pressure cooker is 115°C, and the internal pressure is 3 atm. What portion of the pressure cooker was initially occupied by water? (A. Sheronov)

P83

Fill to capacity. Two capacitors are connected in series. The first has a capacitance C_1 and a maximum voltage V_1. The corresponding values for the other capacitor are C_2 and V_2. What is the maximum voltage that this combination of capacitors can be fed? (I. Slobodetsky)

P84

Ring in a magnetic field. A ring of diameter d, mass m, and resistance R falls in a vertical magnetic field from a great height. The plane of the ring is always horizontal. Find the terminal velocity of the ring if the magnetic field strength B changes with height H according to $B = B_0(1 + \alpha H)$. (I. Slobodetsky)

P85

Drops in a fog. Calculate the number of water drops in 1 m^3 of fog if the visibility is 10 m and the fog settles in 2 hours. The height of the layer of fog is 200 m. The air resistance acting on a drop of water with a radius R meters, falling with a speed v meters per second, is $4.3 \cdot Rv$ newtons.

ANSWERS, HINTS & SOLUTIONS ON PAGE 55

AT THE BLACKBOARD I

Keeping cool and staying put

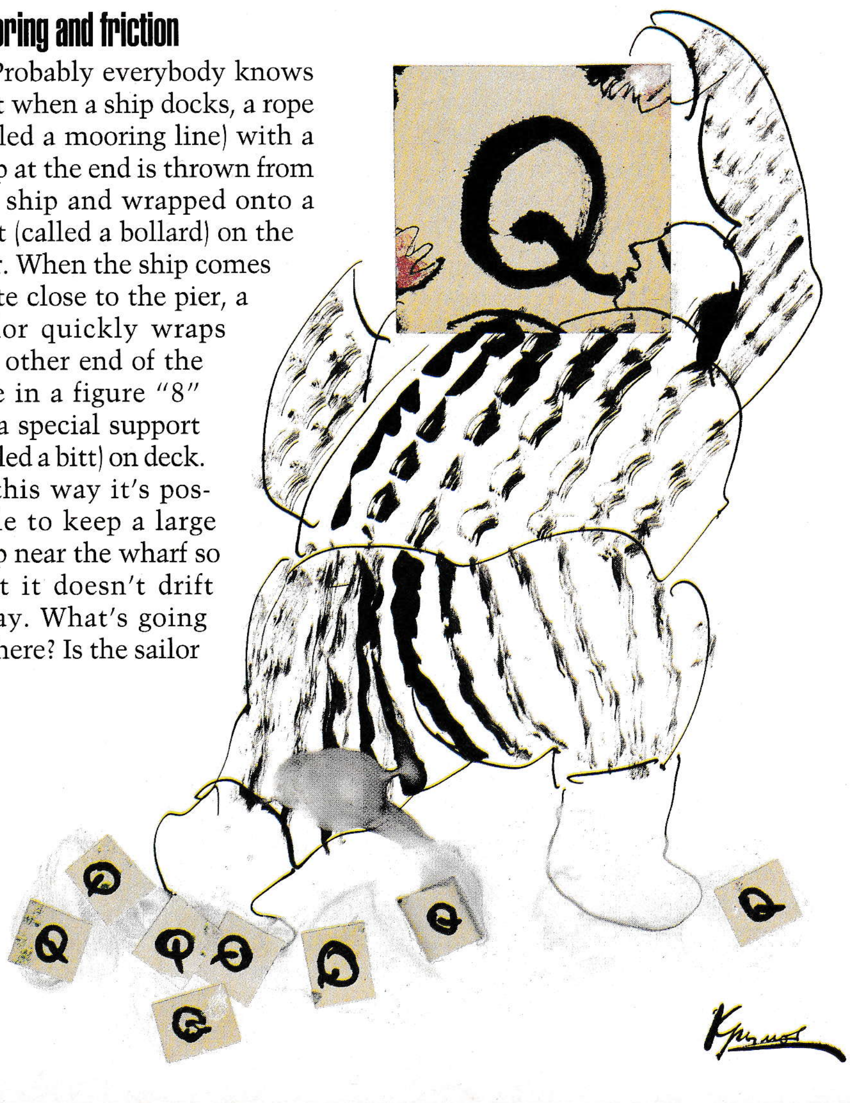
The physics of two workaday phenomena

by Alexander Buzdin

IF YOU'RE LIKE ME, HARDLY A day goes by that you don't say to yourself "I wonder how that works?" or "What's going on here?" or "Is *this* really better than *that*?" Well, I came across something in a novel, and something in real life, that got these questions started in my mind . . .

Mooring and friction

Probably everybody knows that when a ship docks, a rope (called a mooring line) with a loop at the end is thrown from the ship and wrapped onto a post (called a bollard) on the pier. When the ship comes quite close to the pier, a sailor quickly wraps the other end of the line in a figure "8" on a special support (called a bitt) on deck. In this way it's possible to keep a large ship near the wharf so that it doesn't drift away. What's going on here? Is the sailor possessed of superhuman strength?



Art by Dmitry Krymov

Many people, including the well-known French writer Jules Verne, have been inclined to conclude just that. Yakov Perelman, whom you've met in these pages, retold an episode from Verne's book *Matthias Chandor* in which the strong man and athlete Matiphu pulled off an amazing feat. As the Trabokolo, a new ship, was being launched, our hero prevented it from colliding with a small pleasure yacht, which would have been destroyed in the encounter.

"The Trabokolo was sliding quickly downward. A puff of white smoke from the friction curled from the front of the bow, while the stern was already entering the water of the bay.[1]

"Out of the blue a man appears, grabs the mooring line hanging from the front of the Trabokolo, and tries to stop the ship. In a moment he wraps the mooring line around a metal pipe driven into the ground and—although he risked being crushed himself—holds the rope in his hands with superhuman strength for 10 seconds."

Jules Verne was right to emphasize the role that friction played as the ship slid down the ramp: the heating of its bow and the resulting smoke. But he underestimated the role of friction (and thus overestimated Matiphu's role) in describing the athlete's heroics.

Let's try to understand what force is needed to keep the rope wrapped on the support (the pipe or bitt).

First we'll neglect friction and consider a stationary section of the rope, bent by the support into a small angle $\Delta\alpha$ (see figure 1). Let's assume that the rope has a tension T and that the sup-

[1]The ship was being launched stern first.

port exerts a normal force N. Since the section of the rope is in equilibrium, the net force must be equal to zero. For small angles, the vector diagram in figure 1 shows that

$$N \cong T\Delta\alpha$$

(here we've taken account of the fact that for small angles $\sin \Delta\alpha = \Delta\alpha$).

With friction, a rope can remain stationary when the forces of tension to the left and right of the section differ slightly. A rope begins to slide when the difference between these forces reaches the maximum value for the static friction:

$$\Delta T = F_{fr} = \mu N \cong \mu T \Delta\alpha,$$

where μ is the coefficient of friction between the rope and the support. It follows from the last equation that the rate of change in the tension with respect to the wrap angle is proportional to the tension:

$$\frac{\Delta T}{\Delta\alpha} \sim T,$$

or

$$\frac{\Delta T}{\Delta\alpha} \cong -\mu T.$$

Here the minus sign indicates that the tension decreases as the wrap angle increases.

There are many situations in physics when the rate of change of some quantity is proportional to the quantity itself. For example, think of radioactivity: the decrease in the number of radioactive nuclei per unit time is proportional to their number. Another example is the discharge of a charged capacitor through a resistor:

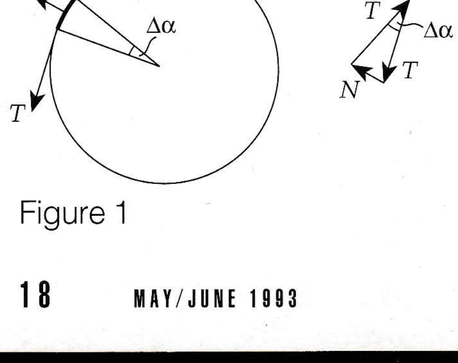


Figure 1

the decrease in the capacitor's charge is proportional to the current through the resistor, which in turn is proportional to the charge on the capacitor. In all these cases there is a very rapid change in the corresponding quantity. If, for instance, the rate of change of the velocity (acceleration) of a body in motion is constant, then the velocity increases linearly with time. If the acceleration is proportional to the velocity, the velocity increases much more rapidly (exponentially).

The same dependence holds true in our case for the change in the rope's tension. (Again, we're dealing here with the minimum possible difference between the forces of tension in the rope—when the rope just barely starts sliding over the support.) The great mathematician, engineer, physicist, and astronomer Leonhard Euler (1707–1783) was the first to consider this problem. He showed that the tension T changes according to the following law:

$$T = T_0 e^{-\mu\alpha},$$

where $e = 2.72\ldots$ is the base of the natural logarithm and T_0 is the initial tension of the rope (which hasn't been wrapped around the support yet).

The angle α (measured in radians) is linked to the number of turns n of the rope around the bitt by a simple relationship: $\alpha = 2\pi n$. So, if the rope's tension decreases by a factor of k after one turn—that is,

$$\frac{T_1}{T_0} = e^{-2\pi\mu} = \frac{1}{k}$$

—after n turns it decreases by a factor of k^n:

$$\frac{T_n}{T_0} = \frac{T_1}{T_0}\frac{T_2}{T_1}\cdots\frac{T_n}{T_{n-1}} = e^{-2n\pi\mu} = \frac{1}{k^n}.$$

For a coefficient of friction $\mu = 0.3$, for example, one turn of the rope around the bitt decreases the tension by a factor of 6.6. And if two more turns are made, the tension decreases by a factor of 43. As the number of turns increases, the rope's tension (thanks to friction) gets smaller and smaller and gradually tends toward zero.

Returning to Verne's hero Matiphu, we can now say that when he wrapped the rope on the iron pipe, he made his job a lot easier. Yakov Perelman took the data on the Trabokolo provided by the novel and made a few calculations. He discovered that if Matiphu managed to wrap the rope around the pipe three times, a child could have done what he did. The same is true of sailors. They don't need fantastic strength. They just need to pay attention and be quick about wrapping the rope around the bitt.

I need hardly point out that each of you encounters this phenomenon practically every day, whenever you tie something—shoelaces, a scarf, any old string. A knot is nothing but a rope wrapped around a "support" (the rope itself)!

The heat pump

We're all used to electric space heaters. All you have to do is plug in the heater and turn it on to get welcome warmth in a chilly room. The heater turns the energy of the electric current into heat. The electric heater's design is very simple—the working portion is simply a heating coil (that is, a resistor). All of the electric energy is transformed into heat except for the portion that turns into luminous radiation, if the coil is heated enough and emits light. By the way, from this point of view a light bulb isn't as good a heater as an electric space heater, since several percent of the lamp's electric power is expended on luminous radiation. Nevertheless, no matter how strange it might be, it's more of a heater than a source of light—that is, it heats more that it illuminates.

The electric heater would seem to have coped ideally with the problem of transforming practically all the electric energy into the required heat. Would it be possible for us to use a certain amount of energy and get, say, twice as much heat, thus cutting down our expenses for heating? At first glance this seems out of the question—it contradicts the law of conversation of energy. However, let's not be in too much of a hurry—we should consider this problem in more detail. Let's start with a refrigerator.

A refrigerator takes heat from an internal reservoir where a low temperature is maintained and releases the heat into the room. Such a process can't take place all by itself. Heat cannot flow from cold to hot (this is one of the formulations of the second law of thermodynamics).

A "reverse" transfer of heat requires a constant supply of energy. This is the work supplied by the refrigerator's compressor. We don't need to know all the details of refrigerator design, but we should note that energy is always needed for the device to function.

Let's assume that as a result of work W the refrigerator removed an amount of heat Q_2 from the freezing compartment. According to the law of conservation of energy, an amount of heat $Q_1 = Q_2 + W$ is released into the room.

Let's determine the refrigerator's efficiency, using the idea that a refrigerator is simply a heat engine working in reverse.

As you know, a heat engine (fig. 2) obtains an amount of heat Q_1 from a hot reservoir, performs work W', and releases an amount of heat $Q_2 < Q_1$ to a cold reservoir. The maximum efficiency of an ideal heat engine is equal to

$$\eta_{max} = \frac{W'}{Q_1} = \frac{Q_1 - Q_2}{Q_1} = \frac{T_1 - T_2}{T_1},$$

where T_1 is the temperature of the hot reservoir and T_2 is the temperature of the cold reservoir.

The efficiency of the refrigerator η_r is defined by the ratio of the amount of heat removed from the freezing compartment (the cold reservoir) to the work required to accomplish that. For an ideal refrigerator (fig.3), this is equal to

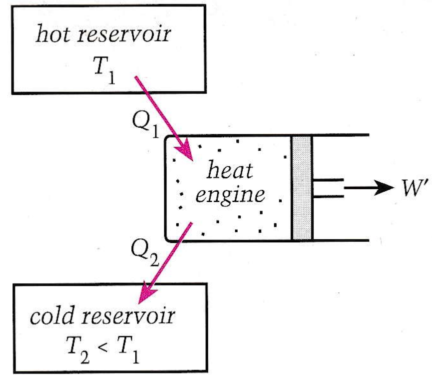


Figure 2

$$\eta_{r\,max} = \frac{Q_2}{W} = \frac{Q_2}{Q_1 - Q_2} = \frac{T_2}{T_1 - T_2} = \frac{1 - \eta_{max}}{\eta_{max}}.$$

Notice that this formula implies that the refrigerator's efficiency can be greater than unity.

Now it's easy to guess how a refrigerator can be used to heat a room during the fall and winter: you put the refrigerator's motor and compressor outside (but keep all the rest of the device indoors!). When we perform an amount of work W (taking energy from the electric circuit) and bring an amount of heat Q_2 in from outside, we transfer an amount of heat $Q_1 = W + Q_2 > W$ into the room. It's clear that there isn't any contradiction with the law of conservation of energy: additional energy in the form of heat is removed from the cold, outside air.

A refrigerator that works like this is called a "heat pump," since heat is pumped into the house from outside. As a result of the work per-

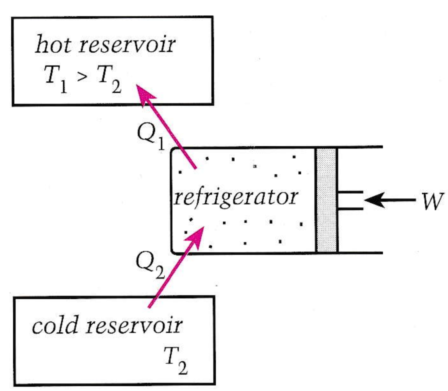


Figure 3

formed by the heat pump, it gets warmer indoors, but outside it gets even colder (but not much colder, of course—the atmospheric effect from one heat pump is negligible). The heat pump's efficiency η_{hp} is defined as the ratio of the amount of heat brought into the room to the external work needed to do that. In the ideal case this is equal to

$$\eta_{hp\ max} = \frac{Q_1}{W} = \frac{Q_1}{Q_1 - Q_2} = \frac{T_1}{T_1 - T_2}$$

$$= \frac{1}{\eta_{max}}$$

and always greater than 1.

For example, consider the case when the temperature of the outside air is –20°C (T_2 = 253 K), while inside the house we need to maintain a temperature of 20°C (T_2 = 293 K). Then η_{hp} = 293/40 ≅ 7.3—that is, by using electric energy to run the heat pump, we get seven times more heat than we do using an electric space heater. The actual efficiency, of course, is always lower. Also, the heat pump doesn't transform all the energy consumed into work. Nevertheless, the heat pump is considerably more economical than the electric heater.

By the way, did you know that your air conditioner is actually a heat pump? It pumps heat from the room, expelling it outside. If its "air intake" and "exhaust outlet" were reversed, it would make an economical heater during the colder months.

So why, despite their lower energy requirements, haven't heat pumps replaced electric heaters? The thing is, electric heaters are extremely simple and cheap, while a heat pump is a rather complicated, bulky, expensive piece of machinery. But mark my words: in the future heat pumps will be widely used and will replace our wasteful electric heaters. ◘

The Worm Problem of Leo Moser

Part III: From warm blanket to cold steel

by George Berzsenyi

As promised in part II of this account, in this article I'll describe some of the conjectures known to me, as well as two more of the "special worms" you'll need to keep in mind throughout your investigations. The conjectured minimal regions shown below are based on the article cited in part II. Before discussing them, I'd like to express my appreciation to George Poole for introducing me to this problem many years ago, for sharing his thoughts with me, and for sending me a copy of his most recent article.

Figure 1 is a sector of a circle with central angle 30° and radius 1. Its area is approximately 0.26180.

Figure 2 is a 30°–60°–90° right triangle with hypotenuse of length $\left(3+4\sqrt{3}\right)/9$. Its area is approximately 0.26350.

The Worm Problem: Find the area of the smallest convex blanket that will cover every worm of unit length.

Figure 3 is a right triangle with legs of length 1 and 1/2. Its area is exactly 0.25.

Figure 4 is a 30°–60°–90° right triangle with hypotenuse $\left(6+2\sqrt{3}\right)/9$. Its area is approximately 0.23450.

Note that the first and third of these triangular regions will automatically accommodate the ⊔-worm, as shown in these figures. In addition, they will also cover the ≀-worm of Besicovitch as shown in figure 5, as well as the V-worm (known in the literature as the "broad worm" of Schaer), shown in figure 6. These are the two other special worms promised in the first two parts of this account. Every viable candidate for the blanket must also cover them!

In the ≀-worm (of Besicovitch), the total length of the straight line segments XX_1, X_1Y_1, and Y_1Y is 1; the worm XX_1Y_1Y is symmetric about Z, the midpoint of X_1Y_1; and $\angle Y_1XZ = \arctan\left(1/\left(9\sqrt{3}\right)\right)$.

In figure 6, which shows the V-worm, $PP'Q'Q$ is a rectangle in which X and Y are the midpoints of PP' and QQ', respectively, and $Q'P'/PP' = \frac{1}{6} + \frac{4}{3}\sin\left(\frac{1}{3}\arcsin\frac{17}{64}\right)$; UV is a circular arc with center P' and radius $Q'P'$;

CONTINUED ON PAGE 31

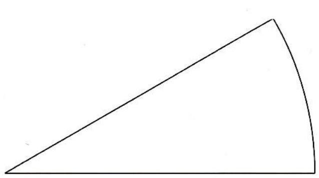


Figure 1

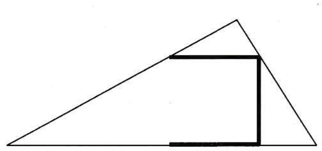


Figure 2

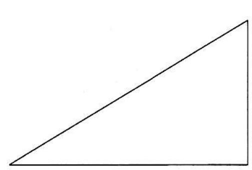


Figure 3

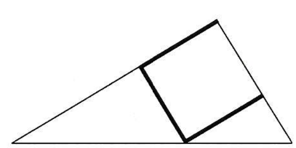


Figure 4

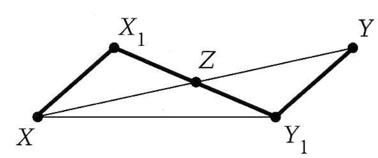


Figure 5

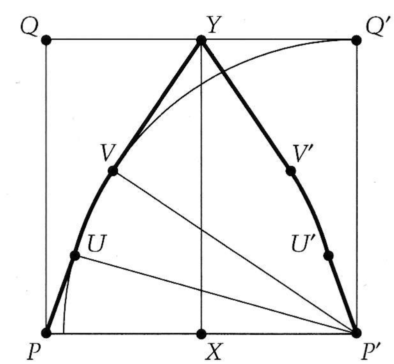


Figure 6

The symmetry of chance

An introduction to geometric probability

by Nikolay Vasilyev

IN "COMBINATORICS–POLYnomials–probability" (in the last issue of *Quantum*), we touched on the simplest kind of probability problems—those dealing with random experiments that can have a finite number of equally likely outcomes. The hypothesis of equal likelihood in such experiments is usually based on their intrinsic symmetry. In this article we'll get acquainted with another kind of problem, in which probabilities can also be found by using considerations of symmetry. The number of elementary outcomes will be infinite, though, and they'll be representable as points in a coordinate plane.

Here are two examples of such problems.

Date problem. Two friends made an appointment to meet at Red Square sometime between 5:00 and 6:00 P.M. Each of them arrives at some random moment, waits 20 minutes, and, if the other one doesn't arrive, leaves. What is the probability they'll meet?

Acute triangle problem. Three points are chosen at random on the circumference of a circle. What is the probability that the triangle with vertices at these points is acute?

But let's begin with simpler, "finite" examples that will help introduce (or dust off) some fundamental concepts in probability theory.

Rolling a die

The best way to get acquainted with probabilities is to play with dice. Since the cube is geometrically symmetrical, it's only natural to assume that each of its six faces has the same chance of turning up (we assume, of course, that our die is fair). So the probability of rolling, say, a 6 is 1/6; the probability of rolling a number not less than 3 is 4/6 = 2/3; the probability of rolling an odd number is 1/2. (The corresponding *events*—sets of "favorable outcomes"—are shown in red in rows a, b, and c, respectively, of figure 1.)

If there are no dice at hand, one can roll a six-sided pencil with the numbers 1 to 6 written on its sides. And it's easy to imagine a pencil with any number n of sides.

Suppose we have a device that, after being put into action, can end up in n equally likely *outcomes*. Then, by definition, the probability of each of these outcomes is assumed to be $1/n$, and the probability of any event A comprising k outcomes is equal to k/n.

Of course, for $n = 6$ it's easy to simply draw a chart and count all the necessary outcomes. But when n is large, one has to use some rules for calculating probabilities.

Let's denote by $p(A)$ the probability of event A. In general, an event is just a subset of the set S of all possible outcomes ($p(S) = 1$, of course). For a single toss of the die, S consists of the six elements 1, 2, ..., 6. The simplest relations between probabilities emerge from relations between sets—or, to be more exact, between the numbers of their elements.

The probability of the event $\overline{A}$—the *complement* of A (which consists of all the elements of S that don't belong to A)—is equal to

$$p(\overline{A}) = 1 - p(A). \qquad (1)$$

For instance, if A is the event "the roll is a number divisible by three"—that is, consists of 3 and 6—then $\overline{A}$ consists of numbers not divisible by three; so $p(A) = 1/3$, $p(\overline{A}) = 4/6 = 1 - 1/3 = 2/3$ (in figure 1d the outcomes constituting A are red, and blue denotes $\overline{A}$).

The *union* $A \cup B$ of A and B is the event consisting of the occurrence of one of two events: event A *or* event B. If the events are *incompatible*—that is, if the sets A and B are disjoint—then the following *addition rule* clearly holds:

$$p(A \cup B) = p(A) + p(B). \qquad (2)$$

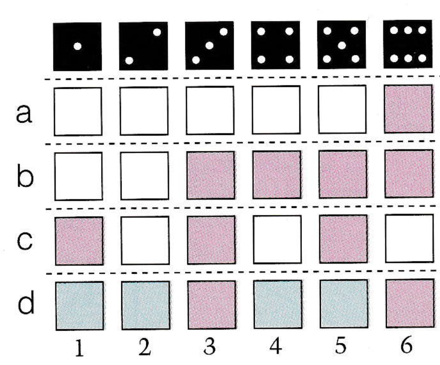


Figure 1

Art by Dmitry Krymov


y
x
0

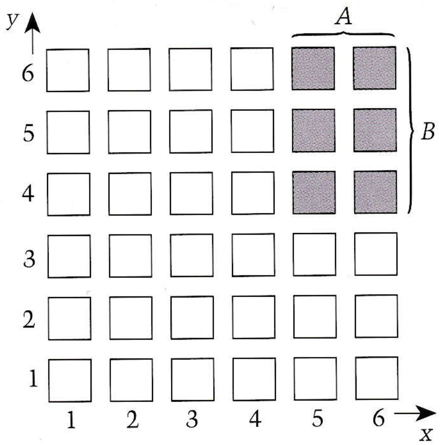


Figure 2

If the events A and B are *compatible*—that is, they have a nonempty intersection AB—then

$$p(A \cup B) = p(A) + p(B) - p(AB). \quad (3)$$

(The intersection of A and B—the event consisting the occurrence of both event A *and* event B—is denoted in probability theory simply by AB.)

Repeated trials

Suppose a fair die was rolled twice, or that two dice were rolled at the same time. In each case, two independent trials were conducted.

Problem 1. What is the probability that the first roll is a number not less than 5 and the second roll a number not less than 4?

The set S here consists of pairs (x, y), where x and y are any numbers from 1 to 6; x is the number that shows on the first die, y on the second. All pairs (x, y) are equally likely, and there are $6 \cdot 6 = 36$ such pairs. It's convenient to represent them as a 6×6 square array: the unit cell with coordinates x and y represents the pair (x, y) (fig. 2). We must choose the cells satisfying the problem's condition $x \geq 5$, $y \geq 4$. They fill up a rectangle of 2×3 cells. So 6 cells out of 36 are "favorable," and the probability is

$$6/36 = 1/6.$$

In general, if an event A is determined by the result of a first trial and B by a second trial (that is, A is a certain set of columns and B a certain set of rows in our array), then the probability of the simultaneous occurrence of A and B is found by the *multiplication rule*

$$p(AB) = p(A)p(B). \quad (4)$$

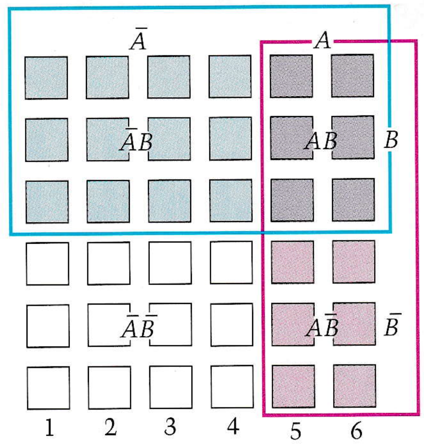


Figure 3

As figure 3 illustrates, this equality means that the ratio of the probabilities of B and $\bar{B}$ remains the same regardless of whether event A occurs. This is exactly what is meant by *independence* in probability: the multiplication rule is considered the *definition* of the independence of any events A and B not necessarily connected with repeated trials. This rule is also applied in the case of more than two independent events. For instance, the probability of rolling a 5 or 6 in each of three tosses of a die is $(1/3)^3 = 1/27$.

Of course, the multiplication rule works only if the event whose probability we want to find can be represented as the *intersection* of two independent events. Let's consider two rather more involved examples based on the same set of outcomes (the set of pairs (x, y), $1 \leq x, y \leq 6$).

Problem 2. A die is rolled twice. What is the probability of getting a number not less than 5 at least once?

The corresponding pairs are colored green in figure 4, so the answer is $20/36 = 5/9$. We can also represent the event in question as the union $A_1 \cup A_2$, where A_i $(i = 1, 2)$ is "the ith roll is not less than 5"; then, according to the addition rule (3) and by the independence of A_1 and A_2,

$$p(A_1 \cup A_2) = p(A_1) + p(A_2) - p(A_1 A_2)$$
$$= \frac{1}{3} + \frac{1}{3} - \frac{1}{3} \cdot \frac{1}{3}$$
$$= \frac{5}{9}.$$

Alternatively, one might like to find the complementary probability $\bar{p}$ of rolling not more than 4 in the first *and* second toss. This way we have the intersection of two independent events, and so we can apply the multiplication rule for $\bar{p}$; thus, the probability we need equals

$$1 - \bar{p} = 1 - \left(\frac{2}{3}\right)^2 = 1 - \frac{4}{9} = \frac{5}{9}.$$

Problem 3. Find the probability that the numbers rolled in two tosses of a die differ by not more than 1.

The outcomes in question are colored in figure 5—they are 6 cells on the diagonal $x = y$ and two parallel lines next to it consisting of 5 cells each. So the answer is $16/36 = 4/9$.

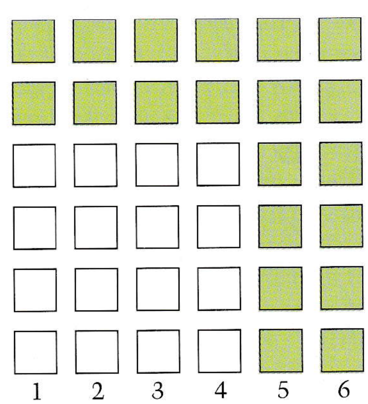


Figure 4

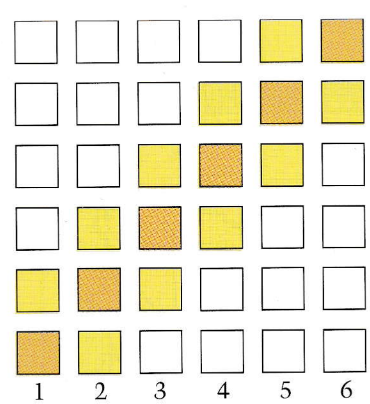


Figure 5

Random numbers and points: uniform distribution

Now we'll turn to random points on a line segment, a circle, a square ... How are the probabilities defined in this case? What "events" can one consider?

Let's roll a round pencil (a cylinder) along a table. Imagine its surface is yellow but a strip (or several strips) of total width α are colored red (fig. 6). What is the probability that the pencil ends up on a red and not a yellow line? (Here α is an arc, or angle, measured in, say, degrees.)

Suppose the pencil is not round but rather has a large number n of sides, k of which are colored red. Then the unknown probability would equal k/n. For a round pencil the set S of "elementary outcomes" can be identified with a circle, and the probability of each particular outcome—the probability of stopping on a particular line—is zero. But the probability of stopping on one of the red lines should naturally be the ratio $\alpha/360°$.

In the same way, when we say that we are choosing a *random point* on a line segment (or on a circle) of length L, we mean that the probability of our choosing the point on an arbitrary segment (or arc) of length d is d/L. Another way to put this is to say that the points are *uniformly distributed* on the segment or arc. According to the addition rule, the probability of hitting any of a number of disjoint segments of total length d is also equal to d/L. For example, the probability that the first digit of a random number in the segment [0, 1] after the decimal point is a prime number—that is, 2, 3, 5, or 7—is equal to 4/10 = 2/5 (in figure 7 the points satisfying this condition are colored red: the points in the interval [0.5, 0.6) have the digit 5 right after the decimal point, and so on).

Choosing a random point in a square or some other figure of area A is defined similarly: the probability that such a point will be in a region of area a is assumed to be equal to a/A. Note that both the length and the area satisfy addition formulas (1), (2), and (3) (of course, the 1 in formula (1) should be replaced by L or A—the measure of the entire segment or plane figure). The branch of probability theory that studies problems about choosing random points is called *geometric probability.*

The independence of events receives a nice graphic interpretation in terms of random points. Two random numbers x and y chosen independently in the interval [0, 1] can be viewed as one random point (x, y) in the unit square $\{0 \le x \le\ ,\ 0 \le y \le 1\}$: the probability that x hits an interval of length a and y hits an interval of length b, by the multiplication rule, is equal to the product ab, which is just the area of the intersection of two perpendicular strips representing these events in the square (a rectangle $a \times b$ with sides parallel to the coordinate axes), or the probability that a point (x, y) hits this rectangle.

For instance, the probability that a random number in [0, 1] is not more than 0.1 units from the midpoint of the interval is equal to 0.2 (the points in question fill up the segment from 0.4 to 0.6). If x and y are two random numbers in [0, 1], the probability that *both* of them are not more than 0.1 units from the point 0.5 equals $0.2^2 = 0.04$, while the probability that at least one of them meets this condition is 0.36 (fig. 8). The latter can be found by adding the area of the rectangles that make up the "cross" in figure 8; by formula (3) (or formula (4)): $0.2 + 0.2 - 0.2^2 = 0.4 - 0.04 = 0.36$; or by switching to complements—by the expression $1 - (1 - 0.2)^2$ (compare with problem 2 above).

Date problem

Now we can solve the "date problem" formulated at the beginning of the article. Let's specify it as follows.

Assume that each of the friends arrives at Red Square at a random moment in time chosen in the interval [0, 60] (within 60 minutes of the agreed-upon hour) and waits 20 minutes for the other (if the friend hasn't arrived yet). They will meet if the difference between the moments x and y of their arrival (in absolute value) is not greater than 20.

The set of all possible outcomes of this "experiment" can be represented as the square $0 \le x \le 60$, $0 \le y \le 60$. Then the set of favorable outcomes $|x - y| \le 20$ is the part of the square bounded by the lines $y - x = 20$ and $y - x = -20$ parallel to the diagonal $x = y$ (fig. 9; compare with problem 3). The area of this set

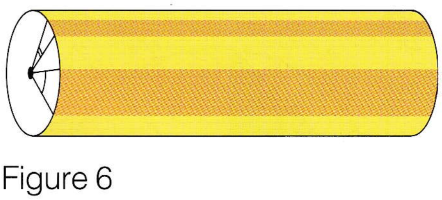


Figure 6

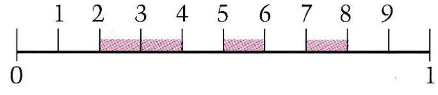


Figure 7

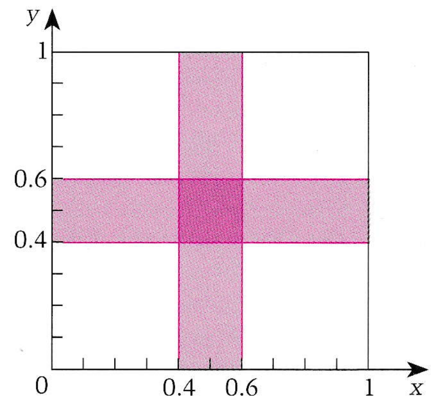


Figure 8

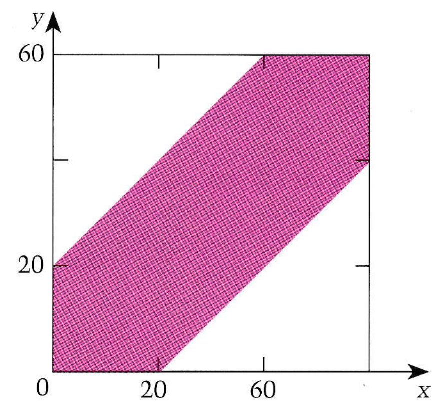


Figure 9

equals $60^2 - 40^2$ (the two white triangles put together make a square with side length 40), so the unknown probability, equal to the ratio of this area to the total area of the square, is

$$1-\left(\frac{40}{60}\right)^2=1-\left(\frac{2}{3}\right)^2=\frac{5}{9}.$$

Let's solve one more problem concerning a random point (x, y).

Problem 4. Find the probability $p = p(a)$ that the sum $x + y$ of two independent random numbers x and y on the interval [0, 1] is greater than the given number a.

The equation $x + y = a$ defines a line parallel to the diagonal $x + y = 1$ of our square. The unknown probability equals the part of the area of the square that lies above this line. For $a \geq 1$ this will be a triangle; for $a < 1$ it's a pentagon, and it's easier to calculate the area of the complement. Draw the corresponding picture and verify that

$$p(a)=\begin{cases}\dfrac{(2-a)^2}{2} & \text{for } a\geq 1,\\ 1-\dfrac{a^2}{2} & \text{for } a\leq 1.\end{cases}$$

Considerations of symmetry: points on a circle

Note that the answer for $a = 1$ in the last problem is $p(1) = 1/2$ (the corresponding points of the square lie above the diagonal). We could guess this without drawing anything: if x and y are random numbers in [0, 1], the condition $x + y \leq 1$ can be rewritten as $x \leq 1 - y$ and read as "x is closer to 0 than y is to 1." The complementary condition is obtained by simply switching the points x and y and the endpoints 0 and 1 of the segment, or by reflecting the segment with two random points chosen in it about its midpoint. None of these operations changes the probability of the event in question. Hence the probabilities of the event and its complement are equal.

Here's one more example of this sort.

Problem 5. Three numbers are chosen at random in the interval [0, 1]. What is the probability that (a) the number chosen last is the largest of the three; (b) the numbers were chosen in ascending order?

We could represent random triples (x, y, z) of chosen numbers as coordinates of a point in the unit cube and compute the volumes of the cube's pieces defined by the appropriate inequalities: $x \leq z$, $y \leq z$ in part (a), $x \leq y \leq z$ in part (b) of the problem. But that's not necessary—it's quite clear that all six possible orders of our numbers—$x<y<z$, $x<z<y$, $y<x<z$, $y<z<x$, $z<x<y$, $z<y<x$—are equally likely, and so each has a probability of 1/6. Thus, the answer to part (b) is 1/6, and the answer to part (a) is 1/3.

Acute triangle problem

Let's turn to the second problem stated at the beginning of the article. Clearly, any rotation of the circle preserves the probability of any event (the set of triples of the circle's points) and the condition of acuteness of the triangle. So we can fix one of the three chosen vertices A, B, C—say, C—and choose the other two at random. Their locations can be given by the angular measures of arcs $CA = \alpha$ and $CB = \beta$ reckoned counterclockwise (fig. 10—note that β is the *major* arc CB). If the arcs are measured in radians, a pair (α, β) is a point in the square $0 < \alpha < 2\pi$, $0 < \beta < 2\pi$. By the Inscribed Angle Theorem, the angles of triangle ABC are equal to $\pi - \beta/2$, $\alpha/2$, $(\beta - \alpha)/2$ (for the case $\beta > \alpha$; the case $\alpha > \beta$ is similar—we simply swap the letters α and β, A and B). We can now think of the pair of numbers (α, β) as specifying a point inside the square $0 < x < 2\pi$, $0 < y < 2\pi$. If $\alpha < \beta$, we can restrict this point to the red region in figure 11. If the three angles A, B, and C are each less than $\pi/2$, we have $\beta > \pi$, $\alpha < \pi$, and $\beta - \alpha < \pi$. These inequalities describe the region colored blue in figure 12. The case $\alpha > \beta$ gives a region symmetric to this region with respect to the diagonal $\alpha = \beta$ of the square. Hence the required probability is 1/4.

This problem also has another, strikingly elegant solution that allows one to solve an analogous problem for n points (see exercise 9 below). I learned about it from the physicist V. Fok and the mathematician Y. Chekanov.

Let's try to figure out the complementary probability that three points on a circle are the vertices of an *obtuse* triangle.

For each point M of circle O consider the diametrically opposite point M' and the semicircle whose arc is bisected at M'. Note that for any point D on this semicircle, angle DOM is obtuse. A triple A, B, C is "obtuse" if

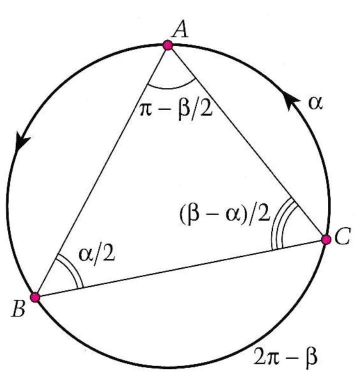


Figure 10

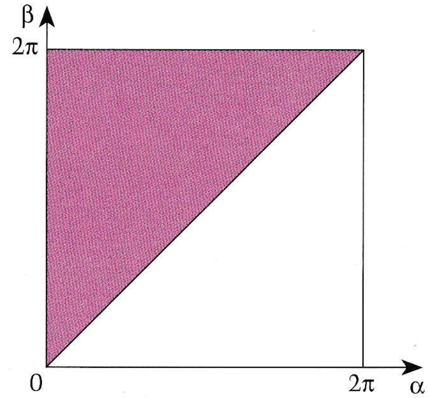


Figure 11

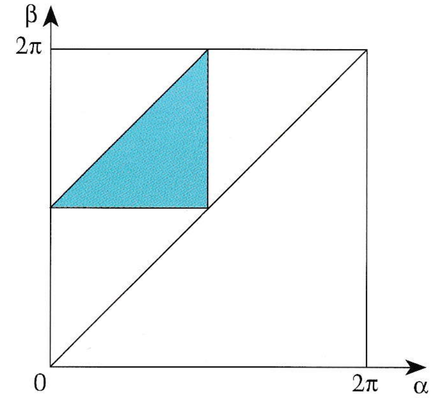


Figure 12

and only if the semicircles corresponding to points A, B, C have a nonempty intersection (the blue sector in figure 13). Indeed, if D belongs to the intersection, then the three angles DOA, DOB, and DOC are all obtuse; such a point exists if and only if points A, B, C lie on the same semicircle (the one corresponding to D) or, equivalently, if they make an "obtuse" triple.

Let's choose our random points in two steps. First, we take at random three diameters; then, for each diameter independently, we "toss a coin" and choose one of its endpoints with a probability of 1/2. We claim that exactly 6 of 8 possible choices in the second step will yield an "obtuse" triple A, B, C.

To prove this, draw three diameters perpendicular to the diameters chosen in the first step. They divide the circle into 6 sectors each of which is the intersection of some three semicircles corresponding to a certain choice of points A, B, C. So there are 6 ways to choose three semicircles with a nonempty intersection, and 6 ways to choose an "obtuse" triple.

Finally, we get a probability of 6/8 = 3/4 for an obtuse triangle and the complementary probability 1/4 for an acute triangle.

The number π in geometric probabilities

In problems with a finite number of outcomes the probabilities are fractions, usually with a small-integer numerator and denominator. Such is the case in many problems in geometric probability as well. But to close out this article, let's look at two problems whose answers are expressed in terms of the number π. The first is easy.

Problem 6. A point is tossed at random onto a large sheet of paper ruled by a grid of unit squares. What is the probability that it will end up less than 1/2 from the center of some square?

It will suffice to consider one square. The points at a distance not greater than 1/2 from its center fill up a circle of area $\pi/4$. And that's our answer: the required probability (the ratio of the area of the circle to that of the square) is $\pi/4$.

Problem 7 (Buffon's needle). A plane is ruled by strips of width 1. A needle (line segment) of length 1 is tossed onto the plane at random. What is the probability that the needle will intersect one of the lines?

This problem has a remarkable answer: $2/\pi$. This result even caused a spate of experiments to see if the theory agrees with reality. Try to prove this answer yourself, or look into "Delusion or Fraud?" in the September/October 1990 issue of *Quantum*, where you can find the solution and an interesting and useful discussion of the statistical results obtained in the aforementioned experiments.

I leave you with some exercises similar to the problems we've investigated in this article.

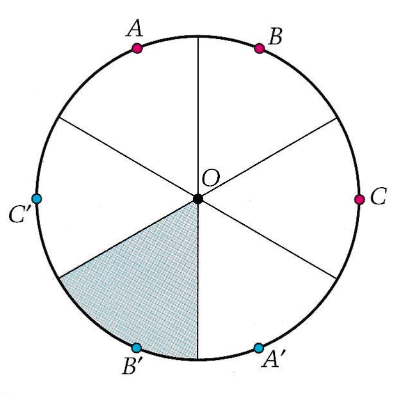


Figure 13

Exercises

1. Find the probability of rolling a die twice and getting (a) two numbers whose sum is not less than 10; (b) two numbers the first of which is divisible by the second.

2. A passenger comes to a bus stop at a random moment in time and waits for a bus from either of two bus lines. The interval between buses is 10 minutes for one line and 15 minutes for the other. Find the probability $p = p(t)$ that she will have to wait at least t minutes. (Assume that the schedule for either line is made independently of the other.)

3. A line segment is divided into three equal parts. What is the probability that three points tossed at random onto the segment will hit three different parts?

4. Four points A, B, C, D are chosen at random on a circle. What is the probability that the segments AC and BD intersect each other?

5. *Bertrand's paradox.* (a) On a given diameter drawn in a given circle a point is chosen at random. What is the probability that the chord through this point perpendicular to the diameter is longer than the radius? (b) Two points are chosen at random on a circle. What is the probability that the chord joining them is longer than the radius? (c) Through a random point in a circle the chord bisected by this point is drawn. What is the probability that this chord is longer than the radius? NOTE: These three questions are variants of the following: what is the probability that a chord intercepted by a given circle on a random straight line is longer than the radius of the circle? Three different ways of choosing the line lead to three different answers! (d) Answer the same question for chords of length greater than $r\sqrt{3}$, where r is the radius of the circle.

6. The vertices of a triangle are chosen at random on a given circle. Find the probability that (a) one of the angles of the triangle is greater than 30°; (b) all its angles are greater than 30°; (c) all the angles are less than 120°.

7. What is the probability that two random points chosen on a segment divide it into three parts out of which a triangle can be constructed?

8. A grid of lines divide the plane into (a) squares, (b) equilateral triangles with side length 1. What is the probability that a coin of diameter 1 tossed onto the plane covers a node of the grid?

9. (a) Find the probability that a convex n-gon with vertices at n random points chosen on a given circle contains its center. (b) Show that n random points on a sphere lie in one and the same hemisphere (on one side of some plane through the sphere's center) with a probability of $(n^2 - n + 2)/2^n$.

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

PHYSICS CONTEST

Animal magnetism

"Ask the female Palme how shee
First did woo her husbands love;
And the Magnet, ask how he
Doth th'obsequious iron move . . ."
—Thomas Stanley (1625–1678)

by Arthur Eisenkraft and Larry D. Kirkpatrick

THE MAGNETIC FORCE IS A strange beast indeed. It doesn't exist at all for neutral particles. And it only exists for charged particles if those particles are moving. Finally, the direction of the force isn't toward the magnetic field or away from the field but "sideways." This magnetic force protects us from cosmic rays by creating Van Allen belts of charged particles around the Earth; entertains us by creating pictures on our television sets; and provides a vital component of research in all areas of physics.

Let's begin to investigate this magnetic force by placing a charged particle (an electron) near a permanent magnet. If the particle is stationary, there is no force. If the electron is moving to the right and the magnet's north pole is behind this page (signified by dots in figure 1), then the force is toward the top of the page. But not for long! As the electron changes its velocity as a result of this force, the direction of the force changes as well. In fact, we observe that the electron moves in a circle. We conclude that the force must be a centripetal force. The magnetic force is always perpendicular to the velocity and to the magnetic field.

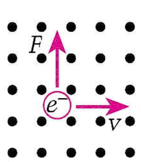


Figure 1

We have a way of describing this mathematically: the vector cross product. The magnetic force is

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B},$$

where **F** is the magnetic force, q and **v** are the charge and velocity of the particle, and **B** is the external magnetic field. In fact, by measuring the force, charge, and velocity, this equation provides a definition for the magnetic field strength. Physicists have invented a number of different rules to help remember the direction of a cross product. One such rule states that if you rotate the velocity vector **v** into the **B** field vector through the smaller angle between them in the same way that you would turn a screwdriver, the force is in the direction that the screw would move—for a positively charged particle. It would be opposite for a negatively charged particle. But remember, this is true only for velocity components that are perpendicular to the magnetic field. Charged particles moving parallel to the magnetic field experience no force whatsoever.

One way in which magnetic fields are used in research is in a mass spectrometer. A schematic of a mass spectrometer is shown in figure 2. Assume that we accelerate a particle with a positive charge q through a potential difference V. The particle gains kinetic energy

$$\tfrac{1}{2}mv^2 = qV.$$

Once the particle enters the magnetic field, the magnetic force provides the centripetal acceleration. Therefore,

$$qvB = \frac{mv^2}{R},$$

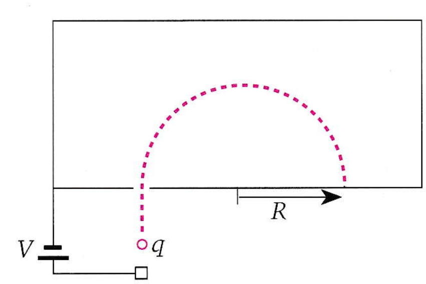


Figure 2

Art by Tomas Bunk


WE CHARGE
VISA
WE MAKE YOU MOVE!
MOVE IT!!
MAGNETIC FIELD ELECTRONS ONLY
ONE WAY
NO U TURN

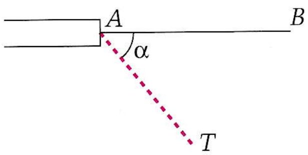


Figure 3

where R is the radius of the particle's circular path.

The mass spectrometer allows us to measure the radius of the particle and to determine its mass.

Parts of our contest problem were first given in the International Physics Olympiad (IPhO) in Czechoslovakia in 1977. This problem also appeared in the semifinal examination used to select the 1992 US team for the IPhO.

A. Show that the mass of the particle is given by

$$m = \frac{qB^2R^2}{2V}.$$

B. An electron gun accelerates electrons through a potential difference V and emits them along the direction AB, as shown in figure 3. We want the particles to hit the target T located a distance d from the gun and at an angle α relative to AB. Find the strength of the uniform magnetic field B required for each of the following situations: (a) the field is perpendicular to the plane defined by AB and AT; (b) the field is parallel to AT.

C. What are the numerical values of the magnetic field if V = 1,000 V, d = 5 cm, and $\alpha = 60°$?

Please send your solutions to *Quantum*, 3140 North Washington Boulevard, Arlington, VA 22201 within a month after receipt of this issue. The best solutions will be noted in this space and their authors will receive special certificates from *Quantum*.

A topless roller coaster

The contest problem in the November/December issue of *Quantum* was used on the semifinal exam to select the 1992 US Physics Team that competed in the XXIII International Physics Olympiad held in Helsinki, Finland, during July 1992. Our solution is modeled after the one submitted by Ben Davenport of the North Carolina School of Science and Mathematics in Durham.

When the point-mass train leaves the track, it becomes a projectile subject only to the force of gravity. Because of the symmetry of the problem, we know that the train will meet the track on the other side with the correct speed and angle if the parabola describing the motion is also symmetric. We can, therefore, begin by remembering (or deriving) the formula for the range L of a projectile over flat ground in the absence of air resistance:

$$L = \frac{2v^2 \sin\alpha \cos\alpha}{g}, \qquad (1)$$

where v is the speed of the train and g is the usual acceleration due to gravity. Note that the angle of the projectile with respect to the ground is α.

The range of the train must be equal to the horizontal distance across the opening in the track, which we can get from trigonometry:

$$L = 2R \sin\alpha. \qquad (2)$$

Equating equations (1) and (2) leads to the following condition on the speed of the train at the time it leaves the track:

$$v^2 = \frac{gR}{\cos\alpha}. \qquad (3)$$

Since there is no friction, conservation of mechanical energy requires that the sum of the kinetic energy and the gravitational potential energy be the same at height H and at the height of the end of the track. Therefore,

$$mgH = mgR(1 + \cos\alpha) + \tfrac{1}{2}mv^2.$$

After canceling the common factor m, we substitute in the condition on v^2 from equation (3):

$$gH = gR(1+\cos\alpha) + \frac{gR}{2\cos\alpha}.$$

Defining $k = H/R$, we can simplify the equation to

$$k = 1 + \cos\alpha + \frac{1}{2\cos\alpha}. \qquad (4)$$

Substituting $\alpha = 0$, we find $k = 5/2$, which is the answer we expect to get from a complete loop by setting the gravitational force equal to the centripetal force required for the train to execute the loop.

The graph of H/R versus α is given in figure 4. We have not plotted the graph beyond 75° so that we can see the details of the graph in the range of interest.

To find the value of α for the limiting case of $k = 3$, we solve the quadratic equation in $\cos\alpha$ to obtain

$$\cos\alpha = \frac{(k-1) \pm \sqrt{(k-1)^2 - 2}}{2}. \qquad (5)$$

Plugging in the value of $k = 3$, we get

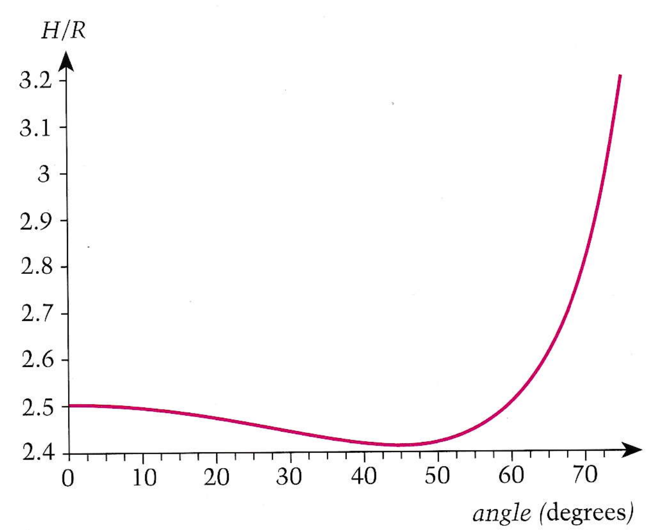


Figure 4

$$\cos\alpha = 1 \pm \frac{1}{\sqrt{2}}.$$

Throwing out the positive root because $|\cos\alpha| \le 1$, we obtain a maximum angle of 73°, which can be verified on the graph in figure 4. The graph also tells us that the angle α can be as small as we want.

The case $k = 2.5$ is an interesting one, since this corresponds to the minimum height for which the car can complete the closed loop without falling off the top. We can also see from the graph that there are two possible values of α for this case, since the curve is multivalued in this range. We could have anticipated this because there are two angles, $45° \pm \theta$, that produce the same range in projectile motion. Plugging $k = 5/2$ into equation (5) yields

$$\cos\alpha = \frac{3}{4} \pm \frac{1}{4}$$

or

$$\alpha = 0°, 60°.$$

Notice that this is also the maximum height for which there are two possible angles.

The minimum value of k is obtained by setting the argument of the radical in equation (5) equal to zero and solving for k:

$$k = 1 + \sqrt{2} \cong 2.414,$$

which corresponds to an angle of 45°. Alternatively, we could set the derivative of equation (4) with respect to $\cos\alpha$ (or α) equal to zero.

Ben states that "this makes sense physically, since 45° is the optimum angle for maximum range. That is, we get the most out of our velocity if the projectile leaves the track at this angle." Even in this case, the maximum height of the train exceeds $2R$. ◙

"SHORTEST NETWORKS" CONTINUED FROM PAGE 9

superficial: in this case the shortest network simply doesn't exist! (See figure 17.)

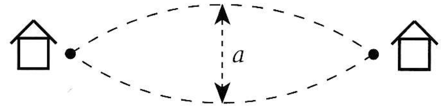


Figure 17
The smaller the distance a, the shorter the network!

Now we hope you're convinced not only that it's necessary to prove the existence of the shortest network but that the very fact of existence is far from self-evident.

Unfortunately, we can't present the proof here—it's far beyond the scope of this article (and the high school curriculum as well). We only wanted to demonstrate here the fundamental significance of existence theorems.

In conclusion, we propose three projects for your own investigation (a computer might be a great help in this research).

1. Solve the Shortest-Network Problem for a regular n-gon.
2. Try to solve the problem of the shortest connection for a number of villages when there are obstacles (say, a round lake).
3. Consider the Shortest-Network Problem for points in space.

And in the end—a few words about the efficiency of algorithms. It often happens in mathematics that there is an algorithm for solving a problem, but its actual realization takes too much time. Then it becomes a substantial problem to find quicker algorithms (the development of computers has made such problems especially important). A typical example is the "algorithm" of long multiplication, which all of you know very well. Nowadays, much faster multiplication algorithms based on profound mathematical ideas (like the Fourier transform, for instance) have been discovered. Another example is the problem of factoring an integer (which underlies some methods of coding). Various ways of making the algorithms for constructing the shortest network more efficient can be found in "The Shortest-Network Problem" by M. W. Bern and R. L. Graham in *Scientific American* (January 1989).

It should be pointed out, though, that even the best algorithms found so far can cope with at most 20–30 points and are unable to construct the shortest network for a set of about 100 points in a reasonable time. And it seems likely that a substantial improvement is impossible. ◙

"MATH INVESTIGATIONS" CONTINUED FROM PAGE 21

PU and YV are tangent to this circle; and U' and V' are the reflections of U and V, respectively, in XY. The base of the rectangle is chosen so that the total length of the V-worm $PUVYV'U'P'$ is 1.

My first challenge to you is this: **Prove that the length of the V-worm is indeed 1 and that the ≀-worm cannot be covered by an equilateral triangle of sides 1 unit long.** After completing these tasks, you might carry on: **Address the conjectures above, and put forth your own conjectures.** I hope your efforts will be rewarded by success. Otherwise, you may wish to replace the warm blanket with the cold steel of a hammerhead, in search of the shape and size of that minimal flat surface that, when it hits the worm, will smash it from stem to stern, regardless of what planar shape it wiggles into. ◙

KALEIDOSCOPE

Always a new face to show

The multifaceted polyhedron

THE STRANGE FRUIT YOU see in figure 1 was bred not so long ago, in 1971, in the "geometric nursery" of the French mathematician Adrien Douady. He named his creation the "shaddock[1] with six spines." It's not just an attractive-looking species of polyhedron. It has a remarkable property that apparently contradicts our intuition (and has found an application in algebraic geometry).

A two-dimensional illustration will make things clearer. We say that a figure (like the shaded polygon in figure 2) is *starlike* relative to its interior point *O*, the "center of the star," if it contains the entire segment joining *O* to any of its points. A starlike polygon may not be convex, but one can always move its vertices along the rays from the center so as to make it convex—one can "puff it out," so to speak. (For instance, we can make the polygon inscribed in some circle—see figure 2). Our common sense tells us this must be true for polyhedrons, too. But it isn't, and the shaddock provides a counterexample. This "fruit" is a starlike polyhedron, relative to its center of symmetry—the common midpoint of *AB*, *CD*, and four other "big diagonals" (see figure 1), but it can't be "puffed out"! (Try to figure out why.)

The problems below will afford you an opportunity to test and exercise your imagination in creating three-dimensional shapes of your own and on your own. Often you'll have to decide for yourself whether the things you're asked about exist at all! Perhaps they won't be as whimsical as the shaddock, but every one of them will certainly be some sort of surprise to you. And they'll help us acquaint you with some important theorems about polyhedrons.

[1]Named for Captain Shaddock, a seventeenth-century English ship commander, the shaddock is a large, thick-rinded, usually pear-shaped citrus fruit.—*Ed.*

1. A triangle with congruent sides is a regular triangle. Is any tetrahedron with congruent faces a regular tetrahedron? (That is, are its faces equilateral triangles?)

2. Can two opposite faces of a quadrilateral pyramid be perpendicular to its base? Can three faces of a hexagonal pyramid be perpendicular to its base?

3. Can a cube be defined as a polyhedron all of whose faces are squares?

4. The Oxford English Dictionary defines a prism as "a solid figure of which the two ends are similar, equal, and parallel figures, and the sides are parallelograms." Do you agree with that definition?

The weak point of the definitions of polyhedrons in terms of their faces in the last two problems is that the information is incomplete. For instance, in problem 3 it would have sufficed to fix the number of faces (six). In general, according to the classical Cauchy theorem, a *convex* polyhedron can be uniquely restored given its faces (their shapes and sizes) and the order in which they should be joined to one another. This statement is not valid for nonconvex polyhedrons. And not only that, it turned out that a polyhedron with rigid faces can be flexible, which was one of the greatest surprises in the modern theory of polyhedrons. (You can read more on this subject in "Out of Flexland" in the July/August 1992 issue of *Quantum*.) Here's another, less sophisticated variation on this theme.

5. Bill cut a convex cardboard polyhedron along its edges and sent the set of faces thus obtained to Carol. She glued the faces back together into a convex polyhedron. Could the two polyhedrons be different?

What Bill had to do to be sure that Carol would exactly reconstruct his polyhedron was to mark the edges and vertices on different pieces that

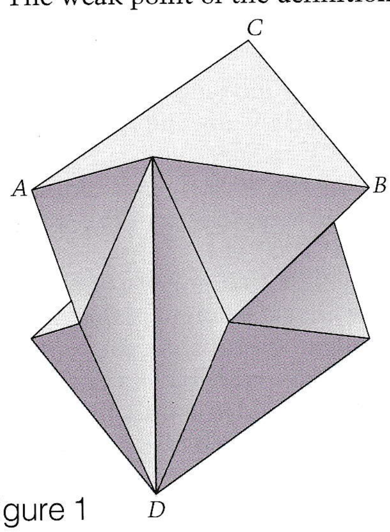


Figure 1

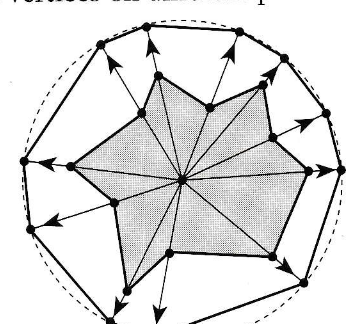
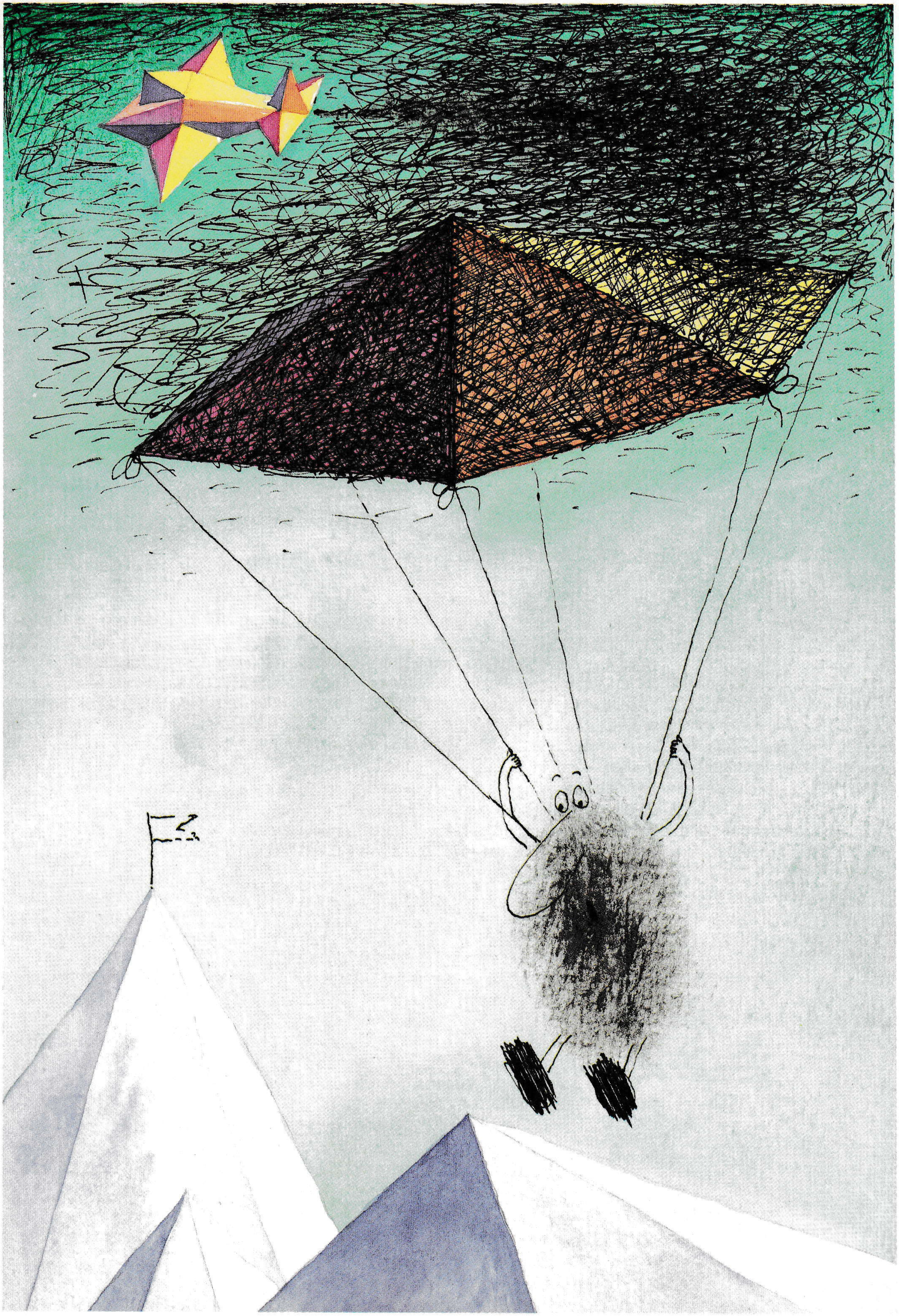


Figure 2

Art by Pavel Chernusky



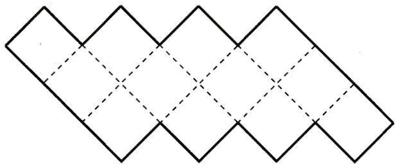


Figure 3

were to be glued together. When the surface of a polyhedron, after a number of straight cuts, is spread out on the plane, and it's specified what points on the boundary of the polygon (or polygons) obtained should be considered the same (or should be glued together), we say we're given a *development* of the polyhedron. The cuts may not coincide with the edges, and, of course, one polyhedron has many different developments.

6. Can the surface of a cube be developed into the saw-toothed polygon in figure 3?

In this problem you had to find the "gluing rule" that turns the given polygon into the given polyhedron. Now try to solve a problem that is, in a sense, converse to that.

7. Find a polyhedron whose development is (a) a rectangle measuring $1 \times \sqrt{3}$ (fig. 4a), (b) an isosceles triangle with an angle of 120° at the vertex (fig. 4b), if in both cases each side of the given figure must be folded at its midpoint and glued to itself (so that its points symmetric about the midpoint stick together)?

When is a polygon (or a set of polygons) with some gluing rule the development of some convex polyhedron? One condition is Euler's famous formula $v + e - f = 2$, where v, e, and f are the numbers of vertices, edges, and faces of the given polygons with proper account taken of the identification on the borders[2] (for example, in figure 4a, $v = 5$, $e = 4$, $f = 1$). Another natural requirement is that the sum of the angles that converge after gluing at one vertex must not exceed 360°. In 1939 the outstanding Soviet geometer Alexander D. Alexandrov proved a theorem that made him a living legend in geometry: these two simple conditions are not only necessary but also sufficient for the existence of a convex polyhedron with the given development. He also proved the uniqueness of such a polyhedron, thus generalizing the Cauchy theorem.

a

b

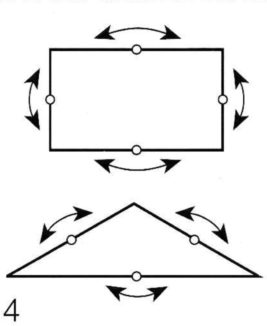


Figure 4

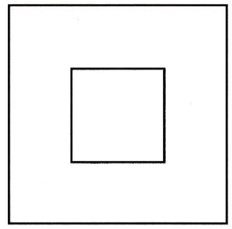


Figure 5

The next problems deal with restoring polyhedrons from their drawings.

8. Can a polyhedron have the head-on view shown in figure 5 and exactly the same top view?

9. Our artist claims that he drew the top views of two convex polyhedrons in figures 6 and 7 correctly, so that all their vertices, edges, and faces (except the far side) are seen. But it seems to us that the polyhedrons in these drawings are . . . well, somewhat askew. Has the artist made any mistakes? If you think there's a mistake in figure 7, can you set it right by displacing only one vertex X (and the edges issuing from it)?

Figures 6 and 7 would raise no doubts if we think of them not as parallel projections but as just the *networks* of some polyhedrons—that is, schematic drawings that show how a polyhedron's vertices are connected with its edges. Then figure 6 is the network of a triangular prism, while figure 7 is that of, say, a cube. Notice that each face of the polyhedron is represented as a polygon in figure 7 (which is not a plane network). In figures 6 and 8, which can be drawn in the plane, one of the faces does not correspond to a polygon inside the network, but rather to the polygon formed by the outside edges of the network.

[2]This formula was proven in "Topology and the Lay of the Land" in the September/October 1992 issue of *Quantum*.

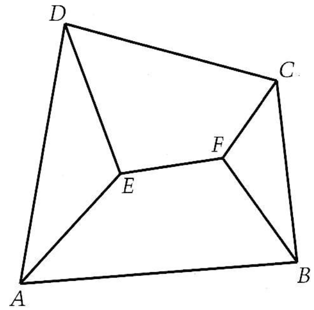


Figure 6

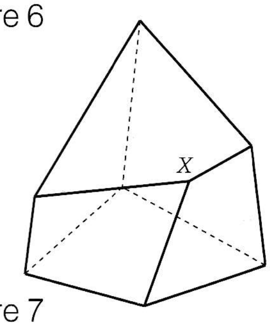


Figure 7

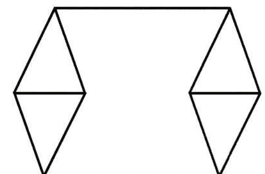


Figure 8

Ernst Steinitz, a German mathematician, proved in 1917 that any network satisfying some inevitable conditions (like "two faces can have no more than one common edge," "every vertex is an endpoint of at least three edges," Euler's formula, and so on) represents some convex polyhedron. But the network in figure 8 doesn't satisfy these conditions (why?). And yet . . .

10. Find a polyhedron with this network.

Finally, one more problem about polyhedrons having something in common with a cube.

11. Devise a polyhedron that has as many vertices, edges, and faces as a cube but (a) has two pentagonal faces; (b) has no quadrilateral faces.

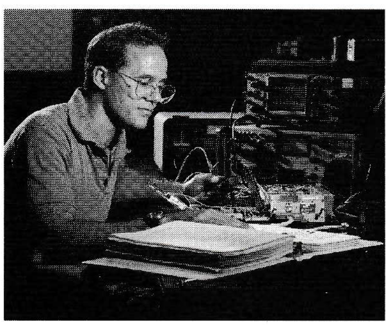
ANSWERS, HINTS & SOLUTIONS IN THE NEXT ISSUE

When we say, "Let's see if it will fly," we mean it!

Embry-Riddle engineering students use CADD resources to aid in the design of their two senior team projects. The curriculum includes a mix of theory and practical applications that fully prepare students for aerospace and aviation careers.

Over 67 international university teams entered the 7th Annual Aero Design Competition, co-hosted by the Embry-Riddle Society of Automotive Engineering Chapter. The objective of the competition is to lift as much weight as possible with a R/C aircraft designed and built by students.

Aerospace engineering students at Embry-Riddle don't have to wait until tomorrow to apply what they're learning today. The curriculum builds to two senior design projects that test your knowledge of aerodynamics, controls, materials, propulsion and structure as your team designs the internal and external details of an airplane or spacecraft. Both your imagination and abilities are put to the test, just as they will be in your professional life. We believe that's what education is all about.



Electrical Engineering students at Embry-Riddle receive a complete, broad-based education with full laboratory resources and equipment to complement academic study.

You may know about us, ***but did you know***:

- We offer over 20 different accredited aviation/aerospace majors at the Bachelor and Master's levels including aerospace engineering, electrical engineering, engineering physics, avionics engineering technology, aircraft engineering technology and aviation computer science.
- Our aerospace engineering enrollment at our Daytona Beach, Florida and Prescott, Arizona campuses is the largest in the United States. We are listed as one of the 18 best engineering schools in the nation by *U.S. News and World Report's 1992 College Guide* called "America's Best Colleges."
- Your university life is enriched through many social and professional clubs and organizations, like our SAE Chapter, that provide the opportunity for a variety of living and learning experiences, related to your major, outside the classroom.

Want to test your career design? Write to us today.

EMBRY-RIDDLE
AERONAUTICAL UNIVERSITY
Look at us now.

Mail to: Embry-Riddle Aeronautical University
Director of University Admissions
600 South Clyde Morris Boulevard
Daytona Beach, FL 32114-3900
Or Call 1-800-222-ERAU

Circle No. 1 on Reader Service Card

Please send me more information about Embry-Riddle.

Engineering Related Programs:	**Other Aviation/Aerospace Majors:**
☐ Undergraduate	☐ Undergraduate
☐ Graduate	☐ Graduate

Name ______________________________

Address ______________________________

City ____________________ State ________ Zip ____________

Phone (   ) ______________________________

MZ


TRR... UUU... TRRR.... UUU.... TRRR... UUU.... TRRR..... UUU....
СЕРГЕЙ БАРХИН 1993-I-I

The oceanic phone booth

Can a telephone receiver be as long as the equator?

by Andrey Varlamov and Alexey Malyarovsky

NOT SO LONG AGO—ABOUT 45 years ago, to put a number on it—scientists from the USSR and US discovered an amazing phenomenon. Sound waves propagating in the ocean could sometimes be detected thousands of kilometers from their source. In one of the more successful experiments, the sound from an underground explosion created by scientists off the coast of Australia traveled halfway around the globe and was recorded by another group of researchers in Bermuda, some 19,600 km away (a record distance for the propagation of pulsed sound signals). This means that the intensity of the sound didn't change as it traveled away from its source. What is the mechanism for such long-distance propagation of sound? It turns out that the ocean contains an acoustic waveguide—that is, a channel along which sound waves travel practically without attenuation (loss of strength).

Another example of an acoustic waveguide is the tube used on ships from time immemorial. The ship's captain uses the tube to give orders to the engine room from the bridge. It's interesting that the attenuation of sound traveling along a waveguide in air is so small that if we constructed a tube 750 km long, it could serve as a "telephone" for calls between Pittsburgh and Detroit. But it would be inconvenient to try to converse over such a line, because the person at the other end would need to wait a half-hour to hear your words.

Art by Sergey Barkhin

We should emphasize that the reflection of a wave from a waveguide's boundaries is a characteristic feature of the waveguide: it's because of this very property that the wave energy doesn't radiate in all directions but only along the given direction.

These examples would lead us to suppose that the propagation of sound over extremely large distances in the ocean is due to some sort of waveguide mechanism. But how is such a gigantic waveguide formed? Under what conditions does it arise, and what are the reflective boundaries that enable the sound waves to travel so far?

Since the ocean's surface can reflect sound fairly well, it can serve as the upper boundary of the waveguide. The ratio of the intensity of a reflected wave to that of a wave that passes through the interface between two media depends to a great extent on the densities of these media and the speed of sound in each of them. If these media differ greatly (for example, the density of air and water differ by a factor of a thousand, and their sound velocities differ by a factor of 4.5), then even when a sound wave falls perpendicularly on the flat water–air interface, practically the entire wave is reflected back into the water: the intensity of the wave that passes into the air is only 0.01% of the incident wave. The reflection is still stronger when the wave falls obliquely on the interface. But, of course, the ocean surface can't be perfectly flat because of the ever-present waves. This causes chaotic reflection of sound waves at the ocean surface and disturbs the waveguide nature of its propagation.

The results aren't any better when the sound waves reflect off the ocean floor. The density of the sediments at the bottom of the sea is usually within the range 1.24–2.0 g/cm^3, and the velocity of sound propagation through these sediments is only 2–3% less than that in water. So a significant amount of the sound wave's energy is absorbed by the ocean floor when it hits the bottom.

The ocean floor reflects sound weakly and therefore can't serve as the lower boundary of the waveguide. The boundaries of the oceanic waveguide must be sought somewhere between the floor and the surface. And that's where they were found. These boundaries turned out to be water layers at various depths in the ocean.

How do sound waves reflect off the "walls" of the oceanic acoustic waveguide? To answer this question, we'll have to examine the mechanism for sound propagation in the ocean.

Sound in water

Up to now, as we've talked about waveguides, our unspoken assumption has been that the speed of sound in them is constant. But the speed of sound in the ocean varies from 1,450 m/s to 1,540 m/s. The speed depends on the water temperature, salinity, hydrostatic pressure, and other factors. The increase in hydrostatic pressure $P(z)$ with depth z, for instance, leads to an increase in the speed of sound of 1.6 m/s for every 100 m of depth. An increase in temperature $T(z)$ also leads to an increase in the speed of sound. However, the water temperature, as a rule, decreases rapidly as one moves from the upper, well-warmed layers to the ocean depths, where the temperature is practically constant.

Due to these two mechanisms—hydrostatic pressure and temperature—the dependence of the speed of sound $c(z)$ on ocean depth looks like that shown in figure 1. Near the surface the overriding influence is that of the temperature, which drops rapidly. Here the speed of sound decreases with depth. As we plunge deeper, the rate of decrease in temperature slows, but the hydrostatic pressure continues to grow. At a certain depth these two factors balance: the speed of sound reaches its minimum. As the depth increases further, the speed begins to increase due to the rise in hydrostatic pressure.

We see that the speed of sound in the ocean depends on the depth, and this influences the nature of the sound propagation. To understand how "sound beams" move in the ocean, we'll turn to an optical analogy. We'll examine how a light beam propagates in a stack of flat parallel plates with varying indices of refraction. Then we'll generalize our findings for a medium in which the index of refraction varies smoothly.

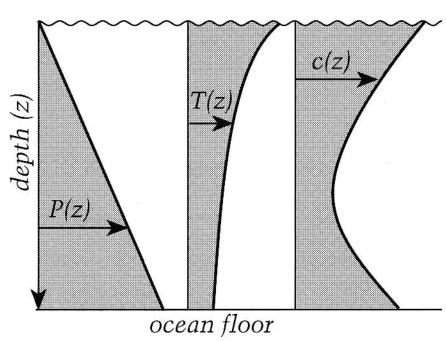


Figure 1

Light in water

Let's look at a pile of flat parallel plates with varying indices of refraction $n_0, n_1, \ldots, n_k$, where $n_0 < n_1 < \ldots < n_k$ (fig. 2). Assume that the beam is incident on the uppermost plate at an angle α_0 relative to the normal. After refracting at the 0–1 boundary, it leaves at the angle α_1, which is also the incident angle for the 1–2 boundary. Upon refracting at this boundary, the beam is incident on the 2–3 boundary at the angle α_2, and so on. According to Snell's law, we have

$$\frac{\sin\alpha_0}{\sin\alpha_1} = \frac{n_1}{n_0}, \quad \frac{\sin\alpha_1}{\sin\alpha_2} = \frac{n_2}{n_1},$$

$$\frac{\sin\alpha_{k-1}}{\sin\alpha_k} = \frac{n_k}{n_{k-1}}.$$

Remembering that the ratio of the indices of refraction of two media is inversely proportional to the ratio of the speed of light in these media, we'll write these equations in the following form:

$$\frac{\sin\alpha_0}{\sin\alpha_1} = \frac{c_0}{c_1}, \quad \frac{\sin\alpha_1}{\sin\alpha_2} = \frac{c_1}{c_2},$$

$$\frac{\sin\alpha_{k-1}}{\sin\alpha_k} = \frac{c_{k-1}}{c_k}$$

($c_0 > c_1 > \ldots > c_k$). Multiplying these equations by one another, we get

$$\frac{\sin\alpha_0}{\sin\alpha_k} = \frac{c_0}{c_k}.$$

Reducing the thickness of each plate to zero and increasing the number of plates to infinity, we'll approach the generalized law of refraction (Snell's law):

$$c(z) \cdot \sin\alpha(0) = c(0) \cdot \sin\alpha(z),$$

where $c(0)$ is the speed of a light beam at the point where it enters the medium and $c(z)$ is the speed of light at a distance z from the boundary of the medium. Thus, as light propagates through an optical medium with an increasing index of refraction, the light beam refracts more and more

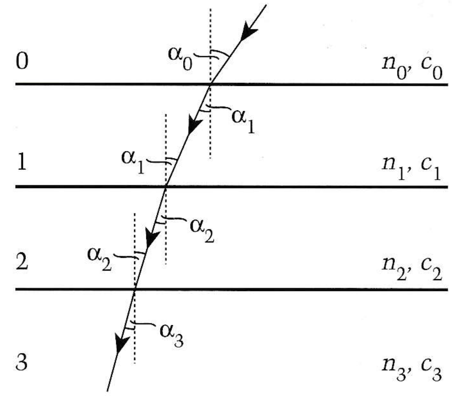


Figure 2

and gets closer and closer to the normal as the speed of light decreases (and the index of refraction increases).

If we know how the speed of light varies in a medium, we can use Snell's law to show how *any* beam travels in a heterogeneous medium. Sound beams propagating in a heterogeneous medium, where the speed of sound varies, deflect in exactly the same way. The ocean is an example of such a medium.

Watery waveguides

Now let's get back to the question of sound propagation in the oceanic acoustic waveguide. Imagine that the sound source is located at a depth z_m corresponding to the minimum sound velocity (fig. 3). How do the sound beams travel as they leave the source? The beam propagating along a horizontal line is straight. But the beams leaving the source at an angle with the horizontal will be bent because of sonic refraction. Since the speed of sound increases above as well as below the level z_m, the sound beams will bend in the direction of the horizontal. At a certain point the beam will be parallel to the horizontal and after being reflected it will turn back toward the line $z = z_m$ (see figure 3).

Thus, the refraction of sound in the ocean allows a portion of the sonic energy emitted by the source to propagate through the water without rising to the surface or dropping down to the ocean floor. This means that we have an oceanic acoustic waveguide. The role of "walls" in this waveguide is played

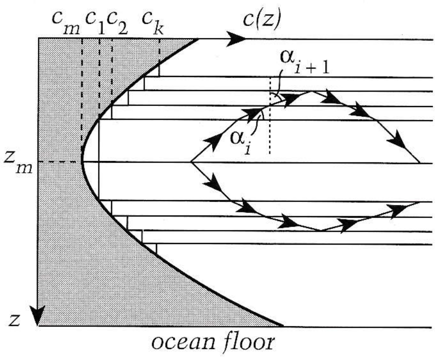


Figure 3

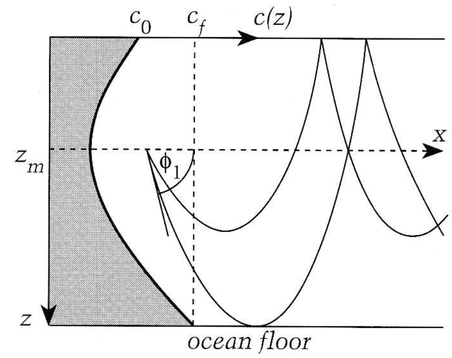


Figure 4

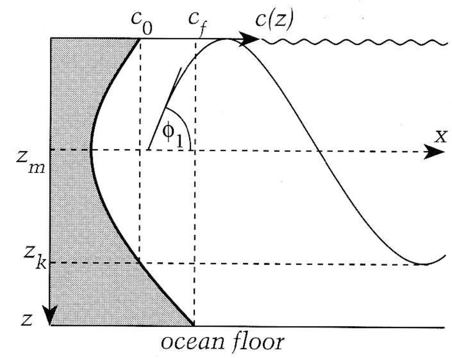


Figure 5

by the layers of water at depths where the sound beam reflects.

The depth z_m where the speed of sound reaches its minimum is called the axis of the waveguide. Usually these depths z_m are in the range of 1,000–1,200 meters, but in the lower latitudes, where the water is warmer at a greater depth, the axis can drop down to 2,000 m. On the other hand, in the higher latitudes the influence of temperature on the distribution of the speed of sound is noticeable only in the layer closest to the surface, and therefore the axis rises to a depth of 200–500 m. In the polar latitudes it rises still closer to the surface.

There are two different types of waveguide in the ocean. The first type occurs when the speed of sound near the surface (c_0) is less than that at the ocean floor (c_f). This usually occurs in deep water, where the pressure on the floor reaches hundreds of atmospheres. As we mentioned above, sound reflects well from the water–air interface. So if the ocean surface is smooth (dead calm), it can serve as the upper boundary of a waveguide. The channel then spreads through the entire layer of water, from the surface to the floor (see figure 4).

Let's see which portion of the sound beam is "captured" by the channel. We'll rewrite Snell's law as

$$c(z)\cos\phi_1 = c_1\cos\phi(z),$$

where ϕ_1 and $\phi(z)$ are the angles formed by sound rays with the horizontal at depths z_1 and z, respectively. It's clear that $\phi_1 = \pi/2 - \alpha_1$, $\phi(z) = \pi/2 - \alpha(z)$. If the source of sound is located on the axis of the channel ($c_1 = c_m$), the most extreme sound ray captured by the channel must have an angle $\phi(z) = 0$ with the ocean floor, as shown in figure 4. Therefore, all rays that leave the source at angles satisfying the condition

$$\cos\phi_1 \geq \frac{c_m}{c_f}$$

enter the channel.

When the water surface is rough, all the sound beams will scatter from it. The rays that leave the surface at larger angles will reach the floor and be absorbed there. Yet even in this case the channel can capture all the rays that fall just short of the rough surface because of refraction (fig. 5). The channel spreads from the surface to a depth z_k, which can be determined from the condition $c(z_k) = c_0$. It's clear that such a channel captures all sound rays with angles

$$\phi_1 \leq \arccos\frac{c_m}{c_0}.$$

The second type of waveguide is a feature of shallow water. It occurs only when the speed of sound near the surface is greater than that near the floor (see figure 6). It occupies the water layer from the floor to a depth z_k, where $c(z_k) = c_f$. It's as if we flipped the first type of waveguide upside down.

For certain types of dependence of the speed of sound on depth, the waveguide acts on sound beams like a focusing lens. If the sound source is located on the axis, the rays leaving it at different angles will periodically converge simultaneously at points along the axis. These points are called the foci of the channel. So if the speed of sound in the channel varies with depth according to a dependence that is close to being parabolic—$c(z) = c_m(1 + \frac{1}{2}b^2z^2)$—then for rays leaving the source at small angles with the horizontal, the foci will be at points $x_n = x_0 + \pi n/b$, where $n = 1, 2, \ldots$ and b is a coefficient whose dimension is inverse to depth (m^{-1}) (fig. 7). This type of curve for $c(z)$ is close to the actual dependence of the speed of sound on depth in deep oceanic acoustic waveguides. Deviations

CONTINUED ON PAGE 50

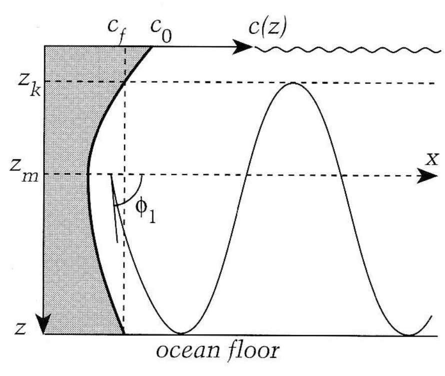


Figure 6

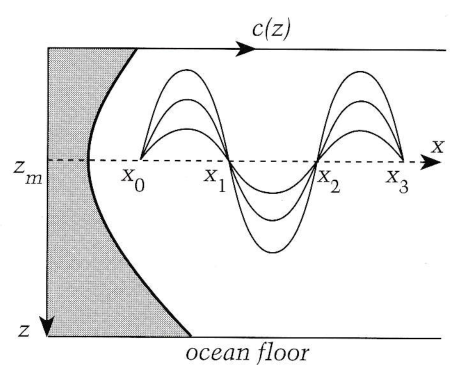


Figure 7

An act of Divine Providence

"Kepler, in his inquiries, asked questions that none before him, including Copernicus, has asked. . . . [They were] questions in physics—not in some preconceived geometrical framework."
—S. Chandrasekhar, Truth and Beauty

by Yuli Danilov

IT HAPPENED ON THE NINTH OF JULY IN 1595. Johannes Kepler, a young teacher at the Lutheran high school in the Austrian town of Graz, was solving a geometrical problem and drew an equilateral triangle with inscribed and circumscribed circles on the blackboard. At that very moment an idea hit him—an idea that seemed to be the key to solving the secret of the universe's structure: the ratios of the radii of the planetary orbits are determined by the ratios of the radii of inscribed and circumscribed circles of certain regular polygons. But Kepler encountered some difficulties on the path to discovering the Creator's intentions. The main problem was that he couldn't explain the number of planets. At that time there were six known planets (including the Earth), but there were infinitely many regular polygons—why should some be preferable to others? And then Kepler turned his attention to solid bodies. As you know, there are only five regular (convex) polyhedrons—just as many as there are intervals between the six planets. These are the so-called Platonic bodies: tetrahedron, cube, octahedron, dodecahedron, and icosahedron. And here, in Kepler's mind, was the solution to the "cosmographic mystery":

> The Earth is the measure of all orbits. Let us circumscribe a dodecahedron around its orbit. The sphere circumscribed about the dodecahedron is the sphere of Mars. Let us circumscribe a tetrahedron around the sphere of Mars. The sphere circumscribed about the tetrahedron is the sphere of Jupiter. Let us circumscribe a cube around the sphere of Jupiter. The sphere circumscribed about the cube is the sphere of Saturn. Let us insert an icosahedron in the sphere of Earth. The sphere inscribed in it is the sphere of Venus. Let us insert an octahedron in the sphere of Venus. The sphere inscribed in it is the sphere of Mercury.

All that remained was to adjust the thickness of the spheres, correct the remaining discrepancies, and the like. To this end Kepler needed observational data.[1] At that time only one person in Europe possessed such data: Tycho Brahe (1546–1601). But in vain did Kepler send the famous astronomer a copy of his *Mysterium Cosmographicum*. He even paid him a call at the Benatek castle near Prague. Tycho stubbornly refused to share his precious observations, compiled over many years. He harbored the dream of creating his own theory of the design of the universe (according to which the Sun revolves around the Earth and the planets revolve around the Sun). Nevertheless, he undoubtedly noticed his guest's talent and grasp of the facts. How else can one explain the sudden about-face? Tycho soon appointed Kepler his new assistant in the most difficult problem facing him: a theory of Mars, which had withstood the efforts of his other assistant Longomontanus.

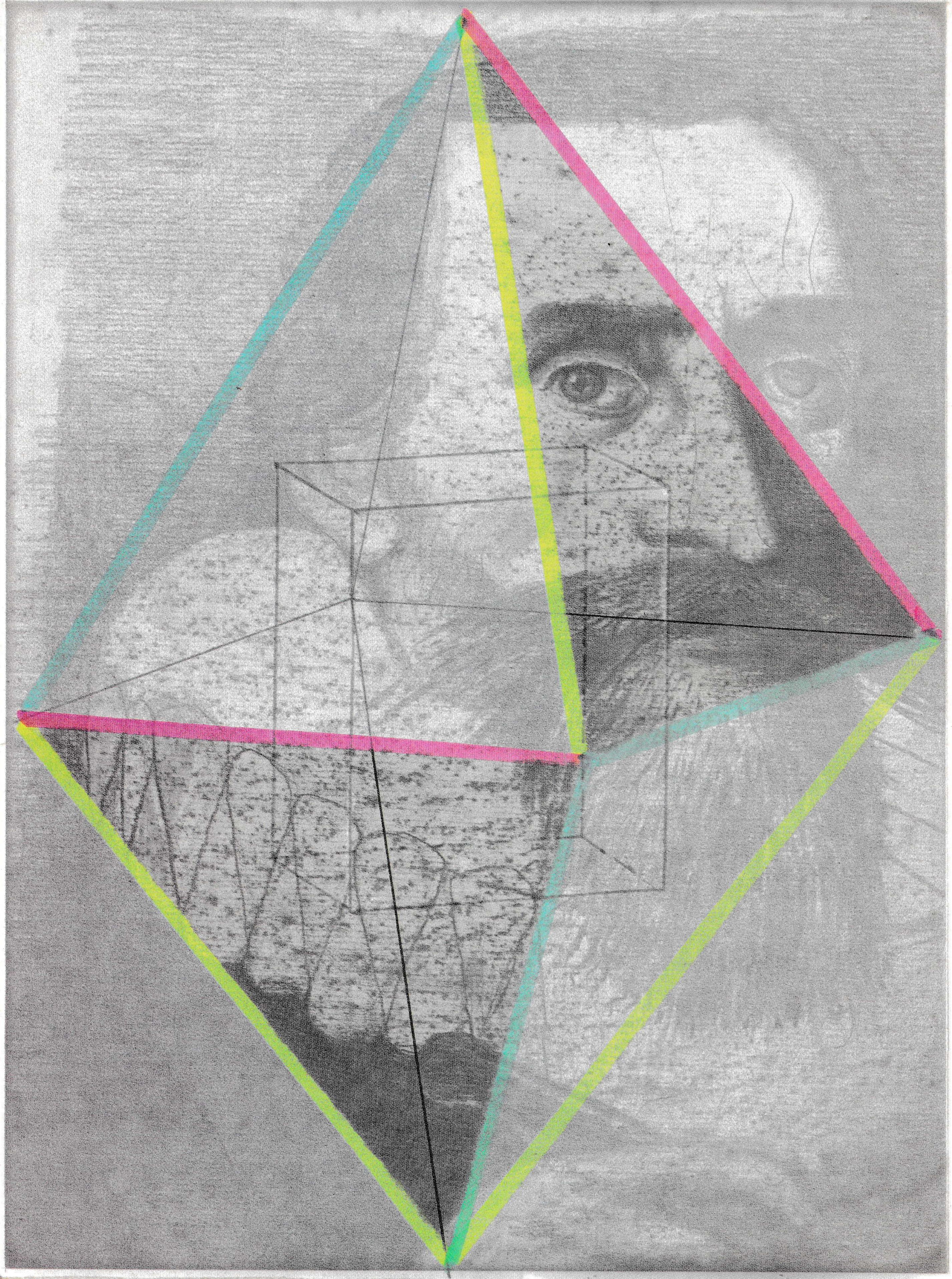
So various circumstances forced Kepler to leave Graz and settle at Benatek, where he became an assistant to Tycho. After his patron died he inherited the title (and duties) of Mathematician of His Imperial Majesty and—more to the point, perhaps—twenty precious volumes of the most exact astronomical observations.

"I think," wrote Kepler, "it was an act of Divine Providence that I arrived just when Longomontanus was busy with Mars. Only Mars gives us the opportunity to penetrate the secrets of astronomy, which otherwise would forever remain hidden from us."

Tycho gave Kepler the task of developing a theory of Mars and assigned Longomontanus the simpler problem of a theory of the Moon. The culmination of Kepler's work was his *Astronomia Nova*, a treatise written in Latin (as was customary at the time) and published in 1609.[2]

[1]Kepler never was able to make this scheme work.

[2]The title page reads in English: *A New Astronomy, Causally Justified, or Celestial Physics, together with Commentaries on the Movements of the Planet Mars in Accordance with Observations Made by the Eminent Tycho Brahe, by Decree and at the Expense of Rudolf II, Emperor of the Holy Roman Empire etc. Written in Prague during Many Years of Persistent Investigations by the Mathematician of His Most Holy Imperial Majesty, Johannes Kepler.*



When people speak about the *Astronomia Nova*, people usually emphasize Kepler's "beelike industry" (Einstein), overlooking or forgetting (Bertrand Russell) the audacity of his ideas and the resoluteness with which he broke with the age-old tradition of circular motion in astronomy. There is no doubt that such a step called for the highest sort of courage—courage of mind and spirit. Not without reason did Kepler, in his allegorical dedication to the emperor Rudolf II, liken the development of his theory of Mars to a battle with the terrible god of war himself. The victor was awarded a pair of trophies: two laws of planetary motion, now known as Kepler's first and second laws.

The excerpt from *Astronomia Nova* that follows will give you an idea of Kepler's battle with Mars and the importance of his victory.

Astronomia Nova

(Excerpt)

Introduction to this treatise

Difficult is the lot of anyone today who writes mathematical, especially astronomical, books. If one does not observe the necessary rigor in terms, explanations, proofs, and conclusions, the book will not be mathematical. If, however, one is rigorous, it is tiring to read such a book, especially in Latin, which lacks the charm of the written Greek language. That is why one can so rarely find appropriate readers nowadays; most people prefer to turn away from reading altogether. Are there many mathematicians who have made the effort to read all of the *Conics* by Appolonius of Perga?[1] This despite the presence of diagrams that make this work much easier to read than an astronomical treatise.

I consider myself a mathematician, yet when I reread my treatise, trying to reproduce in my mind the meaning of the proofs I had at one time inserted in the figures and text, all my mental faculties are strained. But if one tries to make the text easier to understand, inserting paraphrases here and there, this strikes me in mathematical matters as so much chatter, and to proceed in this way is to err in the opposite direction.

Indeed, a protracted exposition also hinders comprehension, to the same extent, in fact, as a short, compressed exposition. The latter slips away from the mind's eye, the former distracts it. In the first case there is not enough light, in the second too much; the eye either can see nothing or is blinded.

Therefore, I have decided to make it as easy as possible for the reader to understand my work by prefacing it with a detailed introduction.

I achieve this end in two ways. First of all I give a table, which provides an overview of all the chapters in the book. Because the subject of the book is unfamiliar to many readers, and because various special terms, and the various topics discussed herein, are similar to one another and are interconnected as a whole and in their details, this table, in my opinion, will be useful only if, comparing all the terms and all the topics, one is able to take them in at a glance and elucidate them by mutual comparison. . . .[2]

But even this overview will not be equally successful with all readers. To many the table I offer as a guiding thread for orientation in the labyrinth of my book will seem more tangled than a Gordian knot. Here at the beginning they are presented with many things in a summary form that might go unnoticed when the work is read straight through because they are scattered throughout the text. This will particularly be true of those who consider themselves physicists and reproach me, and Copernicus and the ancient authors even more, who assert that the Earth moves—they reproach us for disturbing the very foundation of science. For these readers I carefully list all the relevant propositions in the main sections in order to gather before their eyes the proofs underlying all those conclusions of mine that they find so detestable.

When they see that I have done this competently,

Translation and notes by Yuli Danilov. From Johannes Kepler, *Gesammelte Werke*, Bd. 3, *Astronomia Nova . . .*, München, 1937.

[1]Apollonius of Perga (ca. 262–ca. 190 B.C.), known by his contemporaries as "The Great Geometer," was an eminent representative of the Alexandrian school of Greek mathematics. He introduced such terms as parabola, hyperbola, ellipse, focus (of the hyperbola and ellipse), and asymptote. His major work was *Conics* (in eight books). The first four books have come down to us in the original Greek; books five through seven have survived in an Arabic translation; book eight was lost and is known only by references in other sources.

[2]The table is omitted here.

Art by Yury Vashchenko

they can either take on the heavy burden of reading and studying my proofs or trust that I, a professional mathematician, have correctly applied a pure geometrical method. In the latter case they can, according to the problems they have assigned themselves, turn to the foundations of proofs proposed here and test them in detail, bearing in mind that the proofs built on these foundations will be invalid if the foundations can be overturned. Thus, I proceed by mixing, as physicists are wont to do, the possible with the certain and erecting on this mixed foundation a probable conclusion. Because in this work I combine celestial physics with astronomy, it is not surprising that many hypothetical assertions are made. This is in the very nature of physics, medicine, and other sciences that use a priori suppositions along with obvious facts of unquestionable certainty.

As the reader probably knows, there are two schools of astronomers. One is headed by Ptolemy and is called the old school; the other is considered new but is really very old. The first school considers each planet in isolation, and for each planet it finds the causes of its motion along its path. The second school compares the planets and derives what is common in their motions from one and the same cause. This school is not unitary. Thus, Copernicus and old Aristarchus,[3] whom I join as well, think that the cause of the apparent rest and comprehensible motion of the planets is the Earth, where we reside; while Tycho Brahe looks for this cause in the Sun, near which, according to his supposition, the eccentric circles[4] of all five planets[5] are linked as if tied in a knot (immaterial, of course, but with a quantitative sense), and he forces this knot, so to speak, to revolve together with the Sun around the stationary Earth.

These three world views have other features that differentiate the different schools of thought. These individual features, however, can easily be altered and improved so that the three main views of astronomy or celestial phenomena become practically equivalent and can be reduced to one and the same thing.

The aim of my treatise is first of all to improve astronomical knowledge in all three forms, especially with regard to the motion of Mars—in particular, to bring the values calculated from the tables into agreement with celestial phenomena. This had yet to be done with sufficient precision. . . .

Having set myself such a goal and having achieved it, I now proceed to Aristotelian metaphysics—or, to be exact, celestial physics—and to investigate the natural causes of motion. On the basis of this investigation the truth of the Copernican teaching (with minor changes) and the falsity of the other two can be proven with utter clarity. . . .

There are many whose piety prevents them from agreeing with Copernicus. By asserting that the Earth moves and the Sun stands still, they fear they would be accusing the Holy Spirit, speaking in the Scriptures, of lying.

These persons should think about this: insofar as we obtain our most important and most numerous pieces of information visually, we cannot separate our speech from visual impressions. Every day we mostly speak on the basis of our visual impressions, even though we know quite well that things are not so. . . . We speak metaphorically about constellations rising and setting—that is, about things being lifted or lowered; when we say the Sun is rising, others are saying that it is setting. Thus, Ptolemy's adherents say that the planets stand still if they seem to stay near the same fixed stars for several days running, even though they consider that the planets are in fact moving straight toward or away from the Earth during this time. Many authors talk about the solstice, although they deny that the Sun really stands still.[6] There is hardly to be found a zealous follower of Copernicus who does not say that the Sun enters the constellation of Cancer or Leo, understanding this to mean that the Earth enters the constellation of Virgo or Aquarius. . . .

Thus, the Scriptures speak about normal things (without any intention of instructing people) in human language in order to be understandable. It employs expressions familiar to everybody in order to bring them Divine Revelation. . . .

People used to think that Psalm 103 is dedicated to the natural sciences because it refers to natural phenomena. It says there that God set the Earth on its foundations, which shall not move for all time. But the

[3]Aristarchus of Samos (fl. ca. 270 B.C.) was a Greek astronomer and mathematician, the "Copernicus of antiquity." Contrary to the commonly held geocentric conceptualization of his time, Aristarchus argued that the Sun stands still at the center of the universe and the Earth revolves around it. In his work *On the Sizes and Distances of the Sun and Moon* (his only extant work), Aristarchus used considerations of similarity to determine that the distance from the Sun to the Earth is 18 to 20 times the distance from the Moon to the Earth. (His estimate was low by a factor of 20.)

[4]The ancients thought that the planets, insofar as they are celestial bodies, could make only the most perfect (by ancient standards) motion: they must revolve uniformly around the Earth (the center of the universe) along circular paths. Astronomers of antiquity tried to explain away the observed irregularity of planetary motion by asserting that the revolution is regular when observed from some equalizing point—the *punctum aequans* (or simply equant)—displaced relative to the center. From this came the ancient term for orbit: the eccentric circle (or simply excenter).

[5]As noted above, in Kepler's time only the five planets (besides the Earth) visible to the naked eye were known: Mercury, Venus, Mars, Jupiter, and Saturn. Uranus was discovered in 1781 by Herschel; the position of Neptune was predicted by Leverrier (simultaneously by Adams), and the planet itself observed, in 1846; Pluto was discovered by Tombaugh in 1930.

[6]The Latin roots of the word solstice are *sol* ("sun") and *status* ("having come to a stop").

psalmist is a stranger to discussions of physical causes. For he is completely satisfied with the greatness of God, Who created all this, and sings the praises of the Creator, listing one after another all the things that can be seen with the eye. . . .

I also implore my readers not to forget about the goodness of God, which the psalm so insistently calls upon us to contemplate, when they return from the Temple and enter the School of Astronomy and glorify together with me the wisdom and greatness of the Creator. I show this convincingly as I lay out the picture of the universe, studying the reasons for errors in visual perception; and the reader will not only be able ardently to glorify God for the solidity and indestructibility of the Earth, as for a gift that constitutes the happiness of all inanimate nature, but to acknowledge the wisdom of Creator in the motion of the Earth—so mysterious, so unusual.

As for those who are too limited to understand astronomical science or too timid to believe Copernicus without harming their piety, I can only advise them to leave the school of astronomy, calmly condemn philosophical doctrines as they see fit, and devote themselves to their own affairs. They can denounce our ideas about motion in space, go home, and tend their gardens.

As for those who are too limited to understand astronomical science or too timid to believe Copernicus without harming their piety, I can only advise them to leave the school of astronomy, calmly condemn philosophical doctrines as they see fit, and devote themselves to their own affairs. They can denounce our ideas concerning motion in space, go home, and tend their gardens. Raising their eyes heavenward (for they see with their eyes only), let them wholeheartedly give thanks and praise the Lord, our Creator; let them remain convinced that they honor God no less than the astronomer whose God-given gift allows him to see more clearly with the eyes of wisdom and glorify God in his own way.

For this reason scientists can to some extent accept Tycho's opinions about the order of the universe. His notions lie somewhere in the middle. On the one hand, they liberate scientists as far as possible from the useless collection of innumerable epicycles;[7] they allow, as does Copernicus, causes of motion that were unknown to Ptolemy; and they leave room for physical investigations, placing the Sun at the center of the solar system. On the other hand, they are acceptable to the majority of educated people and eliminate a motion of the Earth that is difficult to believe—one that creates difficulties for astronomical theory and throws celestial physics into greater disarray.

That is all I have to say about the authority of the Sacred Scriptures. As for the opinions of the saints concerning natural phenomena, I will be brief: in theology, authorities carry weight; in philosophy,[8] rational foundations carry weight. . . . For me the truth is more sacred, and I, with all due respect to the Church fathers, prove on a scientific basis that the Earth is round and inhabited all over its surface, is insignificant and small, and flies through the constellations.

But that is enough on the truth of the Copernican hypothesis. Now we should return to the aim expressed at the beginning of this introduction. . . .

My tedious work came to an end only when I had passed through the fourth stage of physical hypotheses. By extremely painstaking proofs, after working through a great many observations, I found that the path of the planets in the heavens is not a circle but an oval—or, to be exact, an ellipse.[9]

Geometry teaches us that such an orbit emerges if we assigned the motive force behind each planet the following task: to bring the body into an oscillation along the straight line directed toward the Sun. . . .

[7]In order to explain the apparent planetary motion and its cessation, ancient astronomers presented the motion of planets as a combination of circular revolutions: a planet moves along a circle (the epicycle), the center of which moves along another circle (the deferent). More epicycles were introduced as needed to account for the perceived motion.

[8]By "philosophy" Kepler meant natural philosophy, or physics.

[9]Here, almost in passing, Kepler brings about a true revolution in astronomy, formulating his famous first law of planetary motion, which broke with a concept of circular motion that had survived for centuries. Kepler's first law states that the planets move along elliptical orbits having the Sun as one of the foci (common to all the orbits).

In his *Astronomia Nova* Kepler published also his second law of planetary motion: the line connecting the Sun and a planet sweeps out equal areas over equal intervals of time.

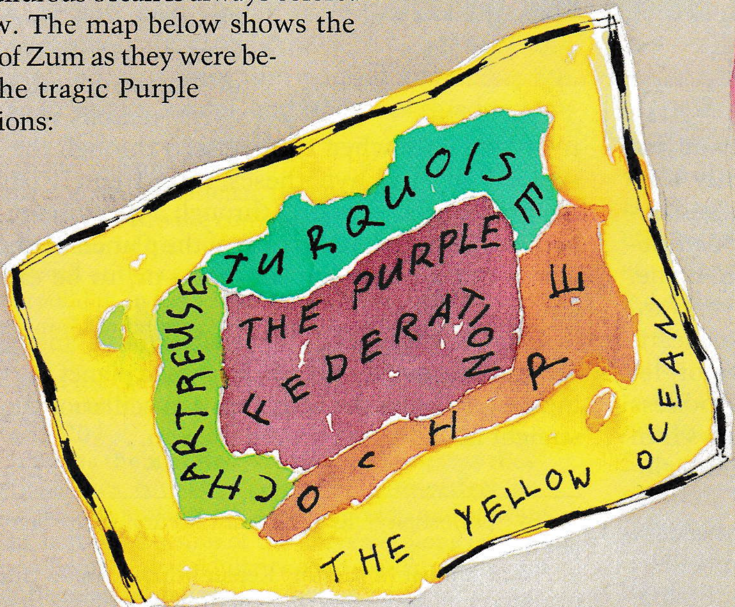
Kepler's third law, derived later and published in 1619 in his treatise *Harmonices Mundi* ("Harmonies of the World"), states that the squares of the periods of revolution of the planets are proportional to the cubes of their average distances from the Sun.

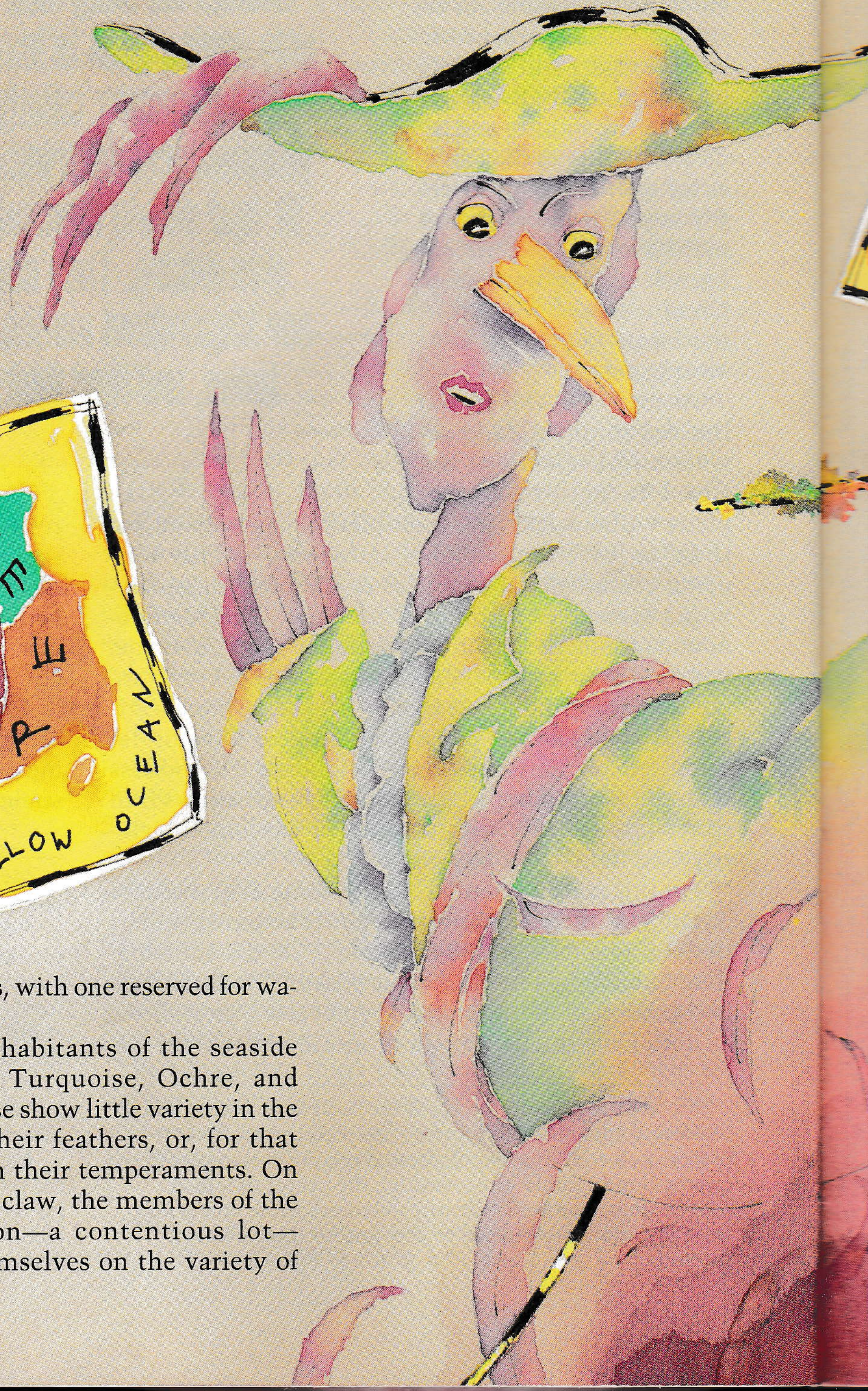
The mapmaker's tale

The Four Color Theorem guarantees that four colors are enough to distinguish countries on a map—doesn't it?

by Sheldon Lee Glashow

THE FEATHERED FOLK OF the faraway planet Zum nest on its one great continent. Long ago, Zumbirds of the same feather stuck together and founded its four ancient kingdoms: Turquoise, Ochre, Chartreuse, and the Purple Federation. Each name describes the plumage of the populace and the color in which it is to be shown on maps. For a touch of verisimilitude, the vast and sulfurous ocean is always colored yellow. The map below shows the lands of Zum as they were before the tragic Purple Partitions:





The Zumbian cartographer (who was also the leading Purple mathematician) knew that just four colors suffice to color any regions in the plane such that no abutters are colored alike. Such a map of Zum, however, would oblige the Purple Federation to be put in yellow. This would not do at all, so all maps of Zum (like those of Earth) once used five colors, with one reserved for water.

The inhabitants of the seaside states of Turquoise, Ochre, and Chartreuse show little variety in the tints of their feathers, or, for that matter, in their temperaments. On the other claw, the members of the Federation—a contentious lot—pride themselves on the variety of

Art by Sergey Ivanov

their plumage. The troubles began when the beloved Purple king was accidentally beheaded in an exhibition beak fight. A separatist party consisting of more brightly plumed Zumbirds demanded their own nestland, Fuchsia. The remainder of the Purple domain became the Republic of the Maroons with the motto: "Divided We Stand." The map below shows the result of the first partition of Zum:

The Maroonians, whose lands were unified by law if not by geography, insisted that their two provinces be colored the same. New maps were duly printed with each of the five nations shown in its proper color, but the mapmaker was puzzled. He found it impossible to make a map of the lands of Zum using only four colors. It was, of course, a purely academic question, since such a map would not have been politically correct.

Purple passions, once aroused, could not be contained. The new maps were hardly dry when Maroonians of a certain hue demanded autonomy as well. These Zumbirds lived in the extreme northwest and southeast of the old Federation. When the feathers stopped flying, the new nation of Magenta was born. The result is shown as a seven-color map:

The map-making mathematician was beside himself. There was no possible way to color the six countries with fewer than six colors! Was there something wrong with the Four Color Theorem? His faith in pure mathematics was shaken, but he had little time to worry about such matters. He was kept busy drawing and redrawing the map of Zum as new Purple lands proliferated.

Oddly, Zumbirds from Turquoise, Ochre, and Chartreuse can barely distinguish one shade of purple from another. They cannot tell bird from bird among the hundreds of ethnic ministates that sprout like weeds on the purple lands. They still use the ancient maps of Zum.

Moral: Purple Zumbirds and others should take care to read the hypotheses of a mathematical theorem. Four colors suffice to color any partition of the plane into disjoint and *connected* regions.

AT THE BLACKBOARD II

Formulas for sin nx and cos nx

Handy, simple, and easy to remember

by Dmitry Fuchs

IN THE COURSE OF YOUR SCHOOL WORK, MAYbe you've had to take the sines and cosines of "multiple angles"—that is, expressions like $\sin 7x$, $\cos 10x$, and so on. Sometimes—though not always—a reasonable way to solve an equation involving such expressions is to rewrite them as functions of $\sin x$ and $\cos x$. In principle, this can always be done automatically, using the trigonometric addition formulas (for $\cos(x + y)$ and $\sin(x + y)$). But when you have to apply even a very simple formula repeatedly, nothing can safeguard you from error. The method in this short article for writing out formulas for $\sin nx$ and $\cos nx$ is simple, easy to memorize, time saving, and reliable.

Pascal's triangle

There's a good chance you know what this is. Just in case, let me give you the definition. Pascal's triangle is the following number array:

1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1
1 9 36 84 126 126 84 36 9 1
1 10 45 120 210 252 210 120 45 10 1
.
.
.

Its nth row consists of $n + 1$ natural numbers; the numbers at the ends of each row are ones (so the first row has just two ones); any other number in the array is the sum of the two adjacent numbers immediately above it. The most remarkable property of Pascal's triangle is that the numbers in its nth row are the binomial coefficients—that is, the coefficients in the expansion of $(x + 1)^n$ (see "Combinatorics–polynomials–probability" in the last issue of *Quantum*):

$$\begin{aligned}
(x+1)^1 &= x + 1,\\
(x+1)^2 &= x^2 + 2x + 1,\\
(x+1)^3 &= x^3 + 3x^2 + 3x + 1,\\
(x+1)^4 &= x^4 + 4x^3 + 6x^2 + 4x + 1,\\
(x+1)^5 &= x^5 + 5x^4 + 10x^3 + 10x^2 + 5x + 1,\\
&\vdots
\end{aligned}$$

Perhaps you already know this from your precalculus course.

The formulas

Write the nth row of Pascal's triangle. (By the way, you don't have to count rows to find the nth one: it has the number n in the second place.) After each number in the row add $\cos^n x$, $\cos^{n-1} x \sin x$, $\cos^{n-2} x \sin^2 x$, ..., $\sin^n x$, respectively. Underline the terms in the even places—the second, fourth, and so on. Write the terms not underlined in one row and the underlined ones in another row below the first. Insert minus and plus signs alternately between the terms in each row (minus before the second term, plus before the third, and so on). And that's it—in the top row you get the expression for $\cos nx$, in the bottom row the expression for $\sin nx$.

Examples. The second row of Pascal's Triangle is 1, 2, 1. We write:

$$\cos^2 x \quad \underline{2\cos x \sin x} \quad \sin^2 x$$

$$\cos 2x = \cos^2 x - \sin^2 x,$$
$$\sin 2x = 2\cos x \sin x.$$

For $n = 3$ we write:

$$\cos^3 x \quad \underline{3\cos^2 x \sin x} \quad 3\cos x \sin^2 x \quad \underline{\sin^3 x}$$

$$\cos 3x = \cos^3 x - 3\cos x \sin^2 x,$$
$$\sin 3x = 3\cos^2 x \sin x - \sin^3 x.$$

Several formulas for cos nx and sin nx obtained by this method are given in figure 1. The circled numbers along the zigzag lines constitute the rows of Pascal's Triangle.

Exercise. Given the nth row of Pascal's Triangle, write the formula expressing tan nx in terms of tan x.

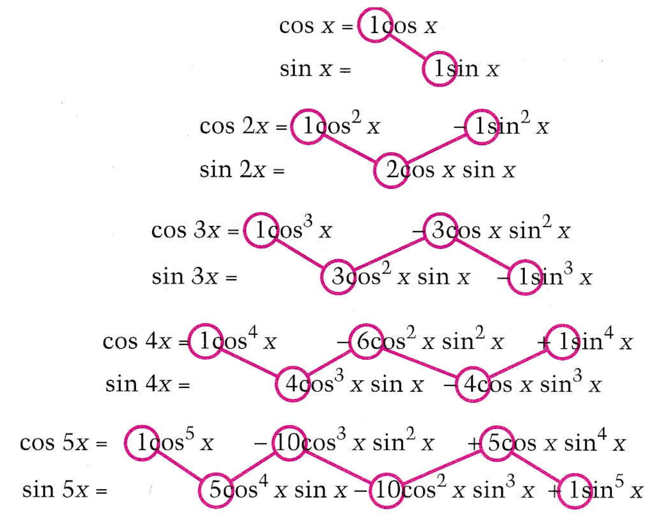


Figure 1

Proving the formulas

As a matter of fact, the above formulas are an immediate consequence of De Moivre's Theorem (for powers of complex numbers in polar form—don't worry if you don't know what this means). But basically the same proof can be rendered in a less advanced way. Here's a direct proof by induction over n.

For $n = 2$ the formulas are well known. Now let $a_1, a_2, \ldots, a_n, a_{n+1}$ be the nth row of Pascal's Triangle. Assume that the formulas

$$\cos nx = a_1 \cos^n x - a_3 \cos^{n-2} x \sin^2 x + \ldots,$$
$$\sin nx = a_2 \cos^{n-1} x \sin x - a_4 \cos^{n-3} x \sin^3 x + \ldots$$

have been already proven. By the addition formula for cosines

$$\begin{aligned}\cos(n+1)x &= \cos(nx + x) = \cos nx \cos x - \sin nx \sin x \\ &= (a_1 \cos^n x - a_3 \cos^{n-2} x \sin^2 x + \ldots)\cos x \\ &\quad - (a_2 \cos^{n-1} x \sin x - a_4 \cos^{n-3} x \sin^3 x + \ldots)\sin x.\end{aligned}$$

Multiplying out and collecting like terms, we get

$$\begin{aligned}\cos(n+1)x = a_1 \cos^{n+1} x - (a_2 + a_3)\cos^{n-1} x \sin^2 x \\ + (a_4 + a_5)\cos^{n-3} x \sin^4 x - \ldots .\end{aligned}$$

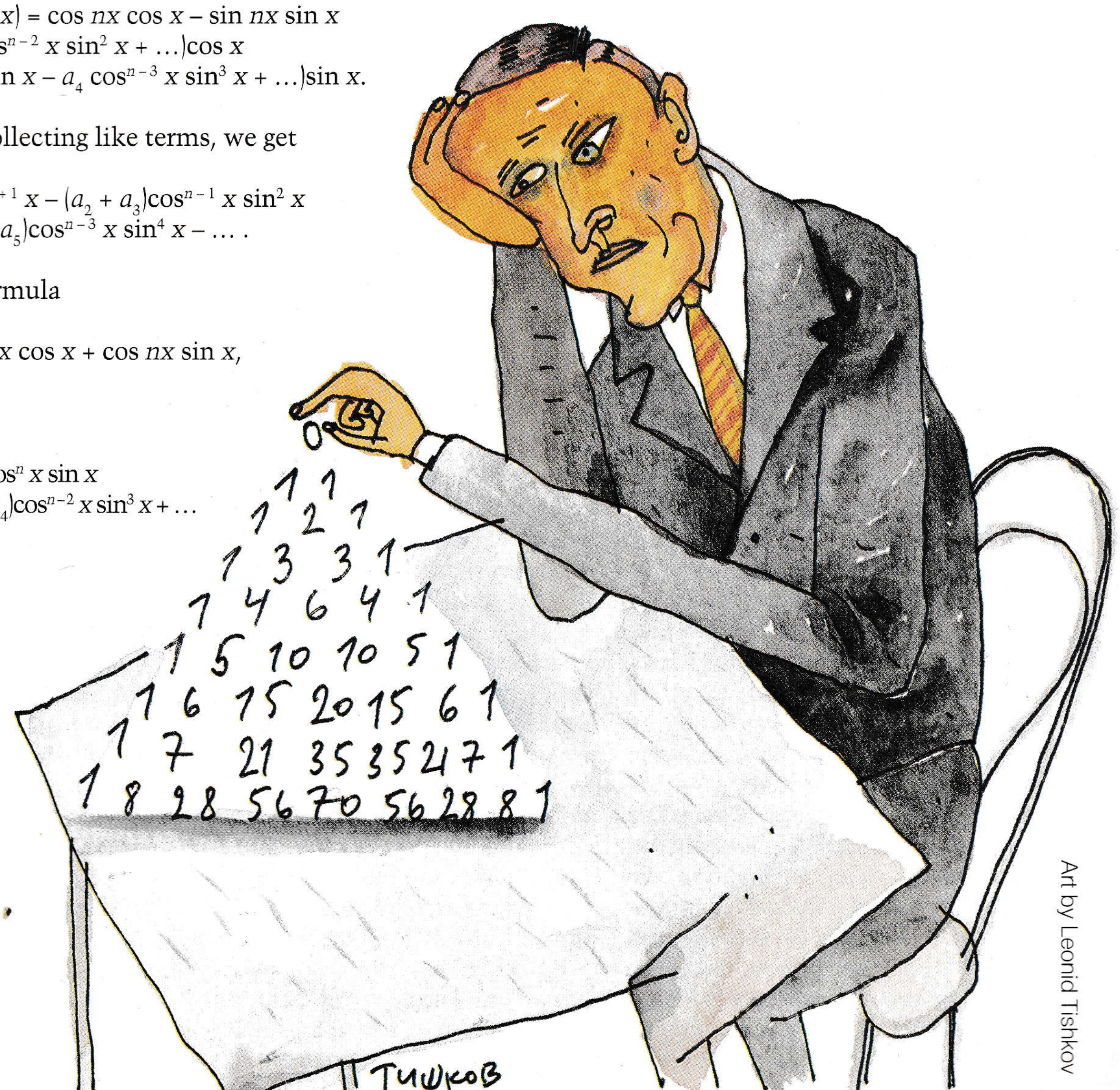
Similarly, using the formula

$$\sin(n+1)x = \sin nx \cos x + \cos nx \sin x,$$

we get

$$\begin{aligned}\sin(n+1)x = (a_1 + a_2)\cos^n x \sin x \\ - (a_3 + a_4)\cos^{n-2} x \sin^3 x + \ldots\end{aligned}$$

Since $a_1, a_1 + a_2, a_2 + a_3, a_3 + a_4, \ldots$ is just the $(n + 1)$st row of Pascal's triangle, the required formulas for cos $(n + 1)x$ and sin $(n + 1)x$ are thus proven.



Art by Leonid Tishkov

For the inquisitive only!

To conclude, I present a straightedge-and-compass construction of the regular pentagon. Let $x = 2\pi/5$. Then $5x = 2\pi$ and $\sin 5x = 0$. According to our formula for $n = 5$,

$$\begin{aligned}\sin 5x &= 5\cos^4 x \sin x - 10\cos^2 x \sin^3 x + \sin^5 x \\ &= \sin x[5\cos^4 x - 10\cos^2 x(1 - \cos^2 x) + (1 - \cos^2 x)^2] \\ &= \sin x(16\cos^4 x - 12\cos^2 x + 1).\end{aligned}$$

Therefore,

$$\sin x(16\cos^4 x - 12\cos^2 x + 1) = 0,$$

implying either $\sin x = 0$, or $\cos x = \pm\sqrt{(3 \pm \sqrt{5})/8}$. Since $\pi/4 < x < \pi/2$, the only suitable possibility is $\cos x = \sqrt{(3 - \sqrt{5})/8}$. Taking the circumradius of our pentagon as the unit length, we successively construct the segment of length $\sqrt{(3 - \sqrt{5})/8}$, the arc measuring $2\pi/5$, and finally the regular pentagon inscribed in a unit circle. The details of the construction are shown in figure 2.

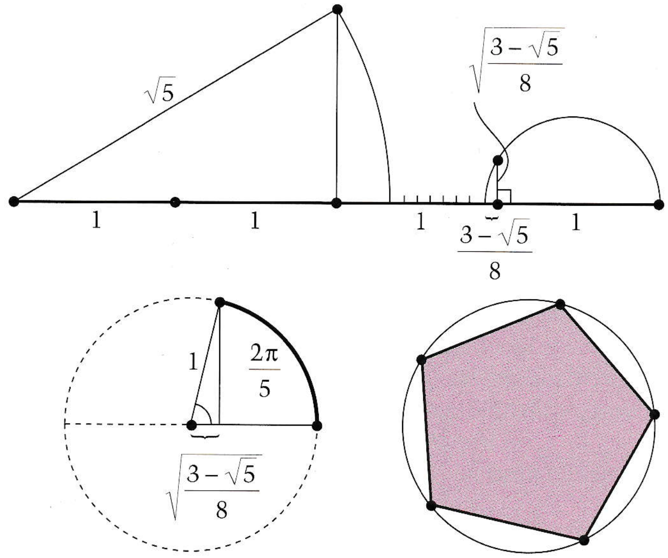


Figure 2

To be fair, this construction is of no practical use: with readily available drawing tools it's almost impossible to achieve the intended result with any precision. But the fact that a regular pentagon can be constructed by compass and straightedge—whereas, say, a regular heptagon (7-gon) can't—is interesting in itself. However, the challenging question of the constructibility of regular polygons lies far beyond the scope of this short article. ◘

"OCEANIC PHONE BOOTH" CONTINUED FROM PAGE 39

from an exact parabolic dependence in $c(z)$ cause the foci along the axis to be erased.

Applications?

Is it possible to send a sound signal along an oceanic acoustic waveguide and receive it at the point of origin, after it has completely circled the globe? The answer is a flat no. First and foremost, the continents present insurmountable obstacles, as do the deep valleys and cavities in the depths of the Earth's oceans. So it's impossible to choose a direction along which there would be only one waveguide around the entire globe. But that isn't the only reason. A sound wave propagating along an oceanic acoustic waveguide differs from sound waves in the "telephone" tubes on ships that we mentioned at the outset. The sound wave traveling from the bridge to the engine room is one-dimensional, and the area of its wavefront is constant at any distance from the source. Therefore, the strength of the sound will also be constant at any point along the tube (heat losses aren't taken into account). As for the oceanic acoustic waveguide, the sound wave doesn't propagate along a straight line but in all directions in the plane $z = z_m$. So the wavefront here is a cylindrical surface. Because of this, the strength of the sound decreases with distance—that is, the sound is proportional to $1/R$, where R is the distance between the source of the sound and the point where it's recorded. (Try to obtain this dependence and compare it with the law of attenuation for a spherical sound wave in three-dimensional space.)

Another reason for the attenuation of the sound is the damping of the sound wave as it travels through the waters of the ocean. Energy from the wave is transformed into heat due to the viscosity of the water as well as other irreversible processes. Also, a sound wave dissipates in the ocean because of various heterogeneities, such as suspended particles, air bubbles, plankton, and even the swim bladders of fish.

Before we close, we should point out that the underwater sound channel isn't the only example of waveguides in nature. Long-distance broadcasting from radio stations is possible only because of the propagation of radio waves through the atmosphere by means of giant waveguides. And we're sure you've heard of mirages, even if you've never seen one. Under certain atmospheric conditions, waveguide channels for electromagnetic waves in the visible range can form. That would explain the sudden appearance of a ship in the middle of the desert, or a city that springs to life in the middle of the ocean. ◘

The USA Computing Olympiad

The new kid on the block of international competitions

by Donald T. Piele

IT CAME AS A SURPRISE TO learn in November 1991 that an International Computing Olympiad had existed in Europe for three years and that the United States had never participated. The International Olympiad in Informatics (IOI) was created in 1989 at a meeting of UNESCO in Paris, and the first IOI was held in Pravetz, Bulgaria, in May 1989. It was patterned after the successful international olympiads in mathematics and the natural sciences, which were also conceived by UNESCO. IOI '90 was held in Minsk, Byelorussia (now Belarus), and IOI '91 took place in Athens, Greece.

IOI '92 was held in Bonn, Germany, and attracted over 170 students from 50 countries. I was fortunate enough to be able to quickly assemble a team, with the help of Harold Reiter and Patsy Hester in North Carolina and Barbara Larson in Virginia, to represent the United States in Bonn. Two of the team members, Nate Bronson from North Carolina and Shawn Smith from Virginia, came home with gold medals.

I was very impressed with the quality of IOI '92. From the minute our team stepped off the train in Bonn until we left for Frankfurt ten days later, we were treated to a highly organized series of events interspersed with two days of competition.

On one memorable day, we traveled by bus to Heidelberg and toured the Heidelberg Castle, the charming city center, and the old university. At noon, lunch was served on a boat as we rode down the Neckar River, and the day ended with a reception and talk on computer graphics, hosted by the home office of the science publisher Springer-Verlag. The excitement and enthusiasm expressed by all the team leaders and students convinced me that IOI was going to be an important international competition and that the United States must develop a democratic procedure for fielding the best US team.

With this goal in mind, I assembled an advisory committee composed of persons actively involved in computing and contests in the United States:

Dr. Bob Aiken, Temple University (Association of Computing Machinery)

Nate Bronson (Duke University—IOI '92 gold medalist)

Dr. Marc Brown (Director, American Computer Science League)

Patsy Hester (Enloe High School)

Barbara Larson (Thomas Jefferson High School of Science and Technology)

Mark Kantrowitz (CS graduate student, Carnegie Mellon University)

Dr. Rob Kolstad (USENIX—The Unix Users Group)

Dr. Harold Reiter (math professor, University of North Carolina at Charlotte)

Shawn Smith (Rice University—IOI '92 gold medalist)

Working together, we created the USA Computing Olympiad (USACO) in November 1992 and set the following goals:

1. Encourage students to study algorithmic computer problem solving.
2. Identify students with outstanding ability in computer problem solving and algorithm development.
3. Provide an opportunity for outstanding students to develop their ability with special training and international competitions.

To meet these goals, we established the following rounds in the USACO.

Qualifying Round. In this round, students are invited to solve a set of five problems within a one-week period in February. Those who solve at least three problems advance to the competition round. The five problems are printed in a brochure and widely circulated with the intention of motivating students to develop computer problem-solving skills.

Competition Round. The second round of USACO is open to those who have qualified in the first round. It is a controlled, five-hour competition, conducted by local coordinators on a specified day in March. The solutions are returned to the USACO and graded by a team of judges. The top twelve students advance to the final round.

Final Round. This round is conducted during a one-week training program held at the University of

Wisconsin–Parkside in June. Four students are selected as the team to represent the United States in the annual International Olympiad in Informatics. This year the IOI will be held in Buenos Aires, Argentina, from October 13 to 23.

Sponsorship

We were very fortunate to secure financial support for the USACO from the Center for Excellence in Education, a private nonprofit organization, under the direction of Joann P. DiGennaro. CEE was founded by Admiral H. G. Rickover to encourage the top high school students in the United States to excel in science and mathematics and to nurture international understanding among potential leaders. The center sponsors the Research Science Institute and other programs.

We sent our plan for the USACO to many national organizations to get their endorsement, including

ACM (Association of Computing Machinery),

ISTE (International Society of Technology in Education),

USENIX (UNIX Users Association),

FOCUS (The Federation of Computing in the United States),

ACSL (The American Computer Science League).

We published our own newsletter and sent articles to be published in *The Journal of Computer Science Education*, *ISTE Update*, *USA Mathematical Talent Search Newsletter*, and the *USENIX Newsletter*.

Summer Training Camp

At IOI '92 in Bonn I discovered that most of the other countries had established a training program where they selected their final four team members. With financial support from the Center for Excellence in Education, we were able to add this feature to the USACO.

The first summer training program for the USACO will be held June 13–20, 1993, at the University of Wisconsin–Parkside. Each student will be given a notebook 386 system for use during the week. Besides a heavy dose of computer problem-solving challenges, participants will have time for recreation, including swimming, tennis, jogging, weight training, bowling, soccer, and the like.

The twelve finalists will be divided up into three teams (Red, White, and Blue) and each assigned a staff member. Students will be moved between teams from time to time so that everyone gets to work with one another and with different staff members. The staff members will be Nate Bronson, Shawn Smith, and Rob Kolstad.

On Wednesday, June 16, the first challenge problem will be presented, and the students will have five hours to solve it. On Friday, June 18, the second and final challenge problem will be given. The rules used at the international olympiad will be followed to grade the results and pick the top four students. These four will constitute the team representing the US at IOI '93. The team will be announced at an awards banquet on Saturday, June 19, 1993. In October we will travel together to Argentina to participate in the fifth International Olympiad in Informatics (IOI '93).

Languages

The official languages for IOI '93 are Turbo Pascal, Borland Turbo C/C++, MicroSoft QuickBASIC, Logo, and Scheme. Pascal and C are likely to be the prevalent languages at the summer camp, since they were the most popular languages in the qualifying round.

Here is a sample problem taken from the set of five problems used in the qualifying round of the 1993 USACO. To qualify, students had to solve at least three of the problems.

Latin squares. A square arrangement of numbers

1	2	3	4	5
2	1	4	5	3
3	4	5	1	2
4	5	2	3	1
5	3	1	2	4

is a 5×5 Latin square because each whole number from 1 to 5 appears once and only once in each row and column. Write a program that will compute the number of $N \times N$ Latin squares whose first row is

$$1\ 2\ 3\ 4\ 5\ 6 \ldots N.$$

Your program should work for any N from 2 to 9. Test your program for $N = 4$ and $N = 5$.

Sample run:

ENTER A WHOLE NUMBER BETWEEN 2 AND 9: 4

THE NUMBER OF 4 X 4 LATIN SQUARES IS 24

USACO information

Those who have access to the Internet or Bitnet can use anonymous FTP to transfer information about the USACO and IOI from our files at the University of Wisconsin–Parkside.

On your system type

ftp ftp.uwp.edu

You will be asked to give your name. Respond

anonymous

When the system asks for your password, type your e-mail address.

Problem statements and information about IOI are located in the directory pub/contests/ioi. Similarly, problems and newsletters about the USACO are located in the directory pub/contests/usaco.

For example, to transfer all the information about IOI, type

```
ftp> cd pub/contests/ioi
ftp> dir
ftp> ascii
ftp> mget *
```

Answer (y) for each document you want to receive. Sign off with

```
ftp> quit
```

Getting involved

The success of the USACO depends heavily on the participation of local coordinators, who are usually mathematics and/or computer teachers in high schools throughout the United States. We welcome their enthusiastic support and applaud them for making this computer problem-solving challenge available in their area.

Students or teachers who would like to be placed on the USACO mailing list should send a request to

Dr. D. T. Piele, USACO Director
University of Wisconsin–Parkside
Box 2000
Kenosha, WI 53141-2000

Phone: 414 595-2231
E-mail: piele@cs.uwp.edu
Fax: 414 595-2056

Bulletin board

ARML Competition

The American Regions Math League (ARML) has announced its 1993 competition. This year it will be held on June 5 at two sites: Pennsylvania State University and the University of Iowa. The ARML competition is the largest on-site event of its kind in the country, drawing 15-member teams of high school students from every region. Teams are organized on a local basis. For information on organizing an ARML team or joining an existing team, write to Joseph Wolfson, Phillips Exeter Academy, Box 1172, Exeter, NH 03833, or Barbara Rockow, Bronx High School of Science, 75 West 205 Street, Bronx, NY 10468.

The ISEF Program

The International Science and Engineering Fair, the "World Series" of science fairs, is held annually with over 750 student contestants from affiliated fairs in the United States and a number of foreign nations. It culminates a yearlong selection process involving thousands of local, regional, and state fairs, their student participants, and their judges from science, medicine, and industry.

The ISEF is for students from grades 9 through 12, two of whom have been selected to represent each of the nearly 400 affiliated fairs. The fair takes place in May each year; this year's event will be held in Mississippi Beach, Mississippi, May 9–15.

The annual contest is for awards, but ISEF is more than just a competition—it's an educational experience as well. It is one of the few competitions in the world where the judges outnumber the contestants. Finalists say that one of the most enjoyable (though sometimes scary) aspects is the opportunity to be interviewed by the scientists, engineers, doctors, and mathematicians who form the judging panels. A number of tours are also organized for the students and their adult escorts to universities, research centers, industry, and places of cultural and historical interest.

For a list of project categories or information about the local and regional science fairs that are affiliated with ISEF, write to Science Youth Programs, Science Service, Inc., 1719 N Street NW, Washington, DC 20036, or call 202 785-2255.

Science program directory available

Searching for stepping stones to success in science? Interested in internships? On the prowl for hands-on research experience? The 12th edition of the *Science Service Directory of Student Science Training Programs* (SSTP), a comprehensive listing of 490 programs and internships for high school students who wish to pursue careers in science, mathematics, and engineering, is now available through Science Service, Inc.

Programs featured in the SSTP Directory offer detailed classroom instruction and hands-on laboratory research in a wide range of subjects. Workshops on career awareness, study skills, and motivation are included in the curriculum in many of the programs offered by universities, research institutions, and corporations throughout the United States and abroad. More than 130 of the programs in this year's edition of the SSTP Directory are specifically designed for minority students, women, and other groups who traditionally have been underrepresented in the sciences.

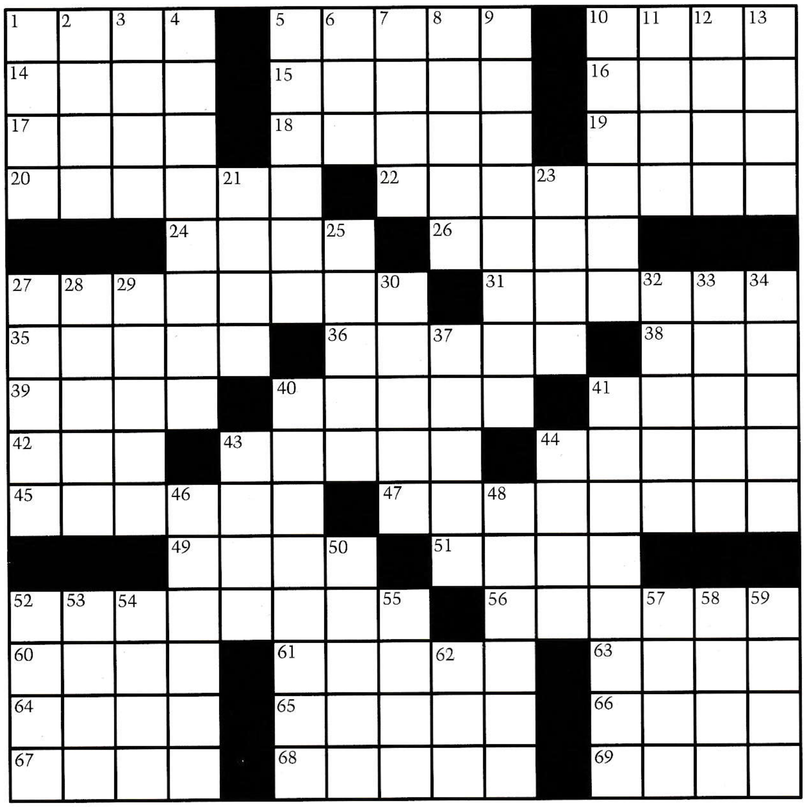
By participating in the classes, workshops, and internships featured in the SSTP Directory, students make valuable contacts and work directly with faculty members and professional research personnel. Students also have the opportunity to utilize the scientific resources of these institutions and universities to gain valuable knowledge in their fields of interest.

Each listing in the SSTP Directory includes the name of the sponsoring institution, a broad description of the activities such as courses, lab work, field trips, and guest speakers; program costs, availability of scholarships and financial aid, and addresses of key individuals to contact for additional program information or enrollment procedures.

Nearly 75,000 copies of the 12th edition of the SSTP Directory have been sent to schools, libraries, and guidance counselors. Additional copies are available for $3 each. To order copies for yourself or your school, write to Science Youth Programs, Science Service, Inc., 1719 N Street NW, Washington, DC 20036, or call 202 785-2255.

Criss × cross science

by David R. Martin



Across

1 Thick plate
5 Unit of length
10 Before coulomb or volt
14 ___ splitting (quibbling)
15 Disintegrate
16 *Brassica* genus plant
17 Ireland
18 Vetches
19 Church part
20 Solution
22 "One in _ ___"
24 Small cut
26 Confused
27 Alpha ___
31 Ankle bone
35 Distance/time
36 Ancient Jewish vestment
38 Pressure units (abbr.)
39 On _ ___ with (equal)
40 One rude person (2 wds.)
41 Invent (a new phrase)
42 Ancient Scand. horn
43 Construct
44 ___ rule (old calc.)
45 Toe preceder
47 A large northern constellation
49 Old automobiles
51 Units of yarn
52 Ear bones
56 Blood vessel
60 Prefix for 10^{-12}
61 Before nerve or axis
63 Fencing sword
64 "When I was _ ___"
65 Person of great size
66 Northern CA's Santa ___
67 ___ Star State
68 Units of distance
69 Fast jets

Down

1 *Butyrospermum parkii*
2 Reclined
3 Publicizes
4 ___ angle (of maximum polarization)
5 System of units
6 Geologic time division
7 Pentateuch (alt.)
8 Swelling in tissues
9 Circuit element
10 Directionless quantity
11 Hat in India
12 Too
13 Young person
21 City in Oklahoma
23 Radiation shield
25 Military freshman
27 Sacred song
28 Prop in Paris
29 Raises
30 Subdivision of a geologic period
32 Rot
33 Digression
34 Shortest paths between points
37 Lodging for travelers
40 Study of Mars
41 Crystal defects
43 Fundamental part.
44 Cicatrix
46 Old transistor
48 Produces a chemical change
50 Brown pigment
52 Hydrated amorphous silica
53 Missile storage structure
54 Examine electronically
55 Large ball of plasma
57 Epic poem
58 Remainder
59 Affirmative votes
62 Hoosier state (abbr.)

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE MARCH/APRIL PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M81

The rule for verifying that a coin is genuine is as follows. Put the coin in question aside; then put one half of the remaining 1992 coins on one pan of the balance and the other half on the other pan. If the balance reads an odd number of grams, the coin in question is counterfeit; otherwise, it's genuine.

Indeed, imagine we replace every lighter counterfeit coin on the scales by a heavier one. Each replacement will change the reading by 2 grams, so the total change will be an even number. In the end, when all the counterfeit coins on the balance are the heavy ones, it will read $l - r$ grams, where l and r are the numbers of counterfeit coins on the left and right pans, respectively. So the initial reading is of the same parity as $l - r$ or $l + r$ $(= (l - r) + 2r)$—that is, it's even if the coin in question is genuine $(l + r = 20)$ and odd otherwise $(l + r = 19)$.

M82

This problem can be solved in many ways. Perhaps the most artless is a vector solution. We can express vector $\overrightarrow{AP}$ in terms of $\overrightarrow{AB}$ and $\overrightarrow{AC}$ and compute the dot product $\overrightarrow{AP} \cdot \overrightarrow{PC}$ to show that it is equal to zero. This means that vectors $\overrightarrow{AP}$ and $\overrightarrow{PC}$ are perpendicular.

An elegant and, in fact, shorter geometric solution is based on the observation that the quadrilateral $AEPD$ (fig. 1) can be inscribed in a circle. Indeed, angles ADB and BEC are congruent, since triangles ADB and BEC are congruent: $AD = BE$, $AB = BC$, $\angle A = \angle B = 60°$; therefore, $\angle ADP + \angle AEP = \angle ADB + (180° - \angle BEC) = 180°$. The segment AE is the diameter of this circle, because its midpoint O is equidistant from A, D, and E: $AO = OE = \frac{1}{2}AE = \frac{1}{3}AB = AD$, so triangle ADO is equilateral and $OD = OA$. It follows that angle APE, subtended by the diameter, measures 90°.

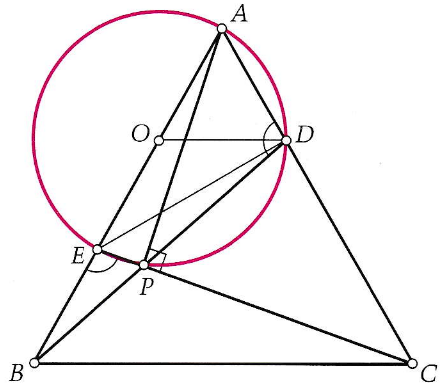


Figure 1

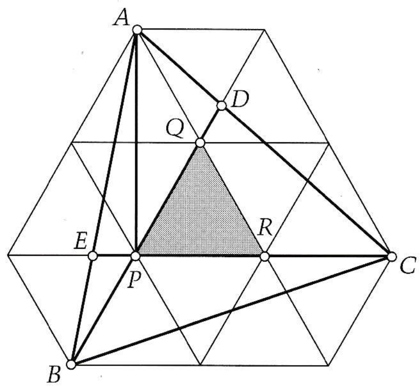


Figure 2

But perhaps the most beautiful solution is shown in figure 2. We start with an equilateral triangle PQR, surround it with a belt of congruent triangles, and join alternate vertices of the hexagon thus obtained to get triangle ABC. Obviously, this triangle is equilateral, and the extensions of the sides of triangle PQR divide the sides of ABC in the ratio 2 : 1 (for instance, the lines of the figure parallel to line CE divide AB into three equal parts, so $AE : EB = 2 : 1$). Thus, triangle ABC in figure 2 satisfies the condition of the problem, and it's equally obvious that $\angle APC = 90°$.

How can one come up with a solution like this, one that makes everything so obvious? Evidently, only after solving the problem in a dull laborious way. (V. Dubrovsky)

M83

(a) The statement is true, for instance, for $n = 3$. Each term of the sequence can be written as $k^{2m} + 1$ for some natural m, or as $(k^m)^2 + 1$. But the remainder of the square of an integer—in particular, of $(k^m)^2$—after division by 3 is 0 or 1 (because $(3l \pm 1)^2 = 3(3l^2 \pm 6l) + 1$), so $(k^m)^2 + 1$ is never divisible by 3.

(b) For any natural number n let's take $k = 2n - 1$ (so that k is odd). Then an arbitrary term of our sequence has the form $k^{2m+1} + 1$ for some natural m and, therefore, is divisible by $k + 1 = 2n$ (because $x^{2m+1} + 1 = (x + 1)(x^{2m} - x^{2m-1} + x^{2m-2} - \ldots + 1)$).

M84

Label the vertices of the given skewb as in figure 3 so that its diagonals are labeled AA_1, BB_1, CC_1, and DD_1; let the first three of them have a common point P.

Note that every two of the three edges AB, A_1B_1, and CD lie in one plane (because $ABCD$ and A_1B_1DC

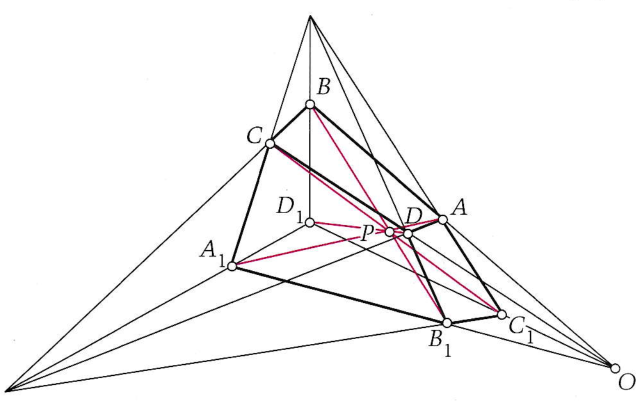


Figure 3

are simply faces of the skewb and edges AB and A_1B_1 join the endpoints of the intersecting diagonals AA_1 and BB_1). In other words, every two of the (extended) lines AB, A_1B_1, and CD are either parallel or intersecting. Suppose some two of them intersect; for instance, lines AB and CD meet at O. Then O belongs to both planes ABA_1B_1 and CDB_1A_1 and, therefore, to line A_1B_1, where they intersect. So in this case *all three lines have a common point*. If no two of the three lines intersect, *all three are parallel to each other*. Of course, this argument can be generalized to *any* three lines every two of which are coplanar (lie in one plane), but all three are not (we implicitly used the last assumption when we represented line A_1B_1 as the intersection of *different* planes ABA_1B_1 and CDB_1A_1; they're really different because otherwise the entire skewb would be flat).

Now, replacing CD by C_1D_1 in the argument above, we deduce that all four lines AB, A_1B_1, CD, and C_1D_1 either have a common point (if AB and A_1B_1 do so) or are parallel (if AB and A_1B_1 are). In any case, lines CD and C_1D_1 turn out to be coplanar, so *the diagonals CC_1 and DD_1 lie in one plane*. In exactly the same way, from intersecting diagonals AA_1 and CC_1 we find that *the diagonals BB_1 and DD_1 lie in one plane*. We also know by the condition that BB_1 and CC_1 intersect at P, so every two of three lines BB_1, CC_1, and DD_1 are coplanar, but all three are not (since the skewb is not flat). Since these three lines cannot be parallel (we know that BB_1 and CC_1 intersect), the argument of the last paragraph guarantees that they have a common point. This means that DD_1 passes through P, which completes the proof.

NOTE. In this problem "diagonals" are understood as infinite lines. If we think of them as segments (which would be more customary), the statement turns out to be wrong, as figure 4 shows. The nonconvex skewb in this figure can be constructed so that the planes ADD_1, B_1DD_1, and CDD_1 are its planes of symmetry. Then the *segments* AA_1, BB_1, and CC_1 will meet the *extension* of segment DD_1 at the same point. However, in a *convex* skewb two diagonals can meet only at an interior point, so the point P in the solution above will automatically lie on the *segments* AA_1, BB_1, CC_1, and DD_1.

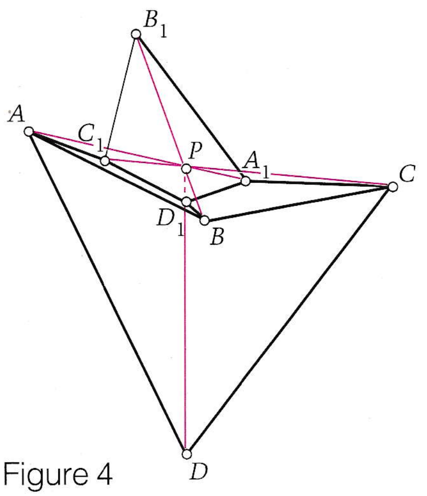


Figure 4

M85

The basic fact for the solution is that *if a polynomial of degree n has n distinct real roots, then its derivative has n – 1 distinct real roots.*

To prove this we note that the derivative is a polynomial of degree $n - 1$ and so has no more than $n - 1$ real roots. On the other hand, there is a root of the derivative in any interval between two roots of the given polynomial, because taking equal (zero) values at the endpoints of this interval, the polynomial must attain either the maximum or the minimum value on the interval at one of its interior points. The derivative at such a point is zero (fig. 5). It follows that the number of roots of the derivative is not less than $n - 1$. Since this is also its largest possible number of roots, the derivative must have exactly $n - 1$ roots.

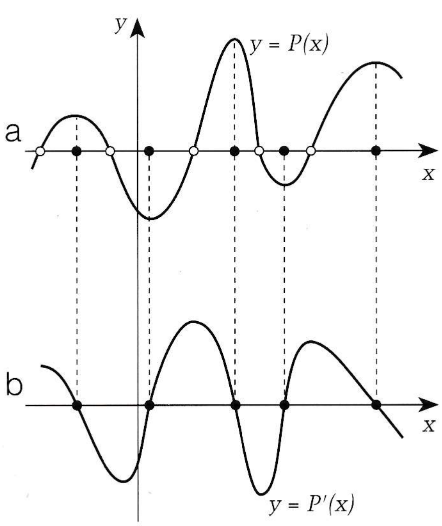


Figure 5

(a) The derivative of $x^4 + ax^3 + bx + c$ is equal to $f(x) = 4x^3 + 3ax^2 + b$ and must have three real roots. Rewrite the equation $f(x) = 0$ as

$$-b/x^2 = 4x + 3a,$$

and sketch the graph of the left side (for $b > 0$ it's the red curve in figure 6). The graph clearly shows that for $b > 0$ and $a \geq 0$ the equation has no nonnegative roots and exactly one negative root; for the case $b < 0$, $a \leq 0$ the situation is quite similar. For $b = 0$, the original expression for the derivative shows that there are exactly two roots, $x = 0$ and $x = -\frac{3}{4}a$. So three roots are possible only for $ab < 0$. We also see that the condition $ab < 0$ is not sufficient for three roots: in figure 6 the right branch of the red curve does not intersect the line $y = 4x + 3a$ even for some negative values of a.

It's easy to make this argument rigorous, but this isn't necessary, since we consider the general situation below.

(b) The statement can be proven by induction over n. For $n = 2$ it's easy: $a_2x^2 + a_0 = 0$ has two roots only for $a_2a_0 < 0$.

Now suppose it's true for all degrees less than some $n > 2$, and consider an equation of degree n without the kth power of x, $1 \leq k < n$:

$$P(x) = a_nx^n + \ldots + a_{k+1}x^{k+1} + a_{k-1}x^{k-1} + \ldots + a_0 = 0.$$

If it has n real roots, the derivative

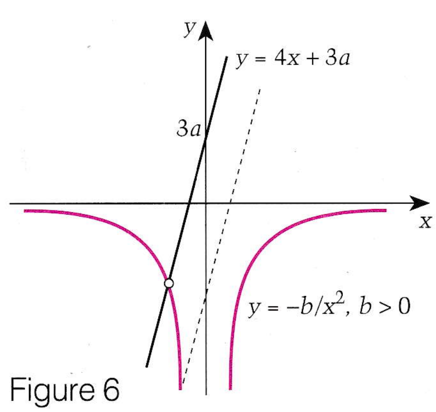


Figure 6

$P'(x) = b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + \dots + b_0$, where $b_m = (m+1)a_{m+1}$ has $n-1$ real roots. Since the term with x^{k-1} is lacking in $P'(x)$, by the induction hypothesis $b_k \cdot b_{k-2} < 0$. So for $k > 1$,

$$a_{k+1} \cdot a_{k-1} = \frac{b_k b_{k-2}}{(k+1)(k-1)} < 0. \quad (1)$$

The case $k = 1$ requires special consideration. Let $x_1, x_2, \dots, x_n$ be the roots of the polynomial $P(x)$. Then, as is well known,

$$\begin{aligned} P(x) &= a_n x^n + \dots + a_2 x^2 + a_0 \\ &= a_n(x - x_1)(x - x_2)\dots(x - x_n). \end{aligned}$$

Removing the parentheses on the right side and equating the coefficients of the same powers of x, we get

$$\begin{aligned} a_0 &= a_n x_1 x_2 \dots x_n, \\ 0 &= a_1 \\ &= -a_n(x_2x_3\dots x_n + x_1x_3\dots x_n + \dots + x_1x_2\dots x_{n-1}), \\ a_2 &= a_n(x_3x_4\dots x_n + x_2x_4\dots x_n + \dots + x_1x_2\dots x_{n-2}) \end{aligned}$$

(the sum in the last equation contains all possible products of x_i's taken $n-2$ at a time). Note that $a_0 \neq 0$, and so $x_i \neq 0$, $i = 1, \dots, n$ (otherwise, $P(x) = x^2(a_n x^{n-2} + \dots + a_3 x + a_2)$ would have at most $n-1$ roots, since the root $x = 0$ would have a multiplicity of at least two). So we can put $y_i = 1/x_i$. You can check that

$$\begin{aligned} 0 &= a_1 \\ &= -a_0(y_1 + \dots + y_n), \\ a_2 &= a_0(y_1y_2 + y_1y_3 + \dots y_{n-1}y_n). \quad (2) \end{aligned}$$

We can now finish the proof:

$$\begin{aligned} 0 &= a_1^2 \\ &= a_0^2(y_1 + \dots + y_n)^2 \\ &= a_0^2[y_1^2 + \dots + y_n^2 + 2(y_1y_2 + y_1y_3 + \dots + y_{n-1}y_n)] \\ &> 2a_0^2(y_1y_2 + y_1y_3 + \dots + y_{n-1}y_n) \\ &= 2a_0a_2, \quad (3) \end{aligned}$$

so $a_0a_2 < 0$.

In fact, the statement of this problem is a weakened version of a theorem of Newton's (a partial proof of which is found in an algebra textbook written by Lobachevsky—the complete proof wasn't published until 1866 by Sylvester): *if a polynomial* $P(x) = a_n x^n + a_{n-1}x^{n-1} + \dots + a_0$ *has n distinct real roots, then for all k,* $1 \le k \le n-1$,

$$a_k^2 > \frac{(n-k+1)(k+1)}{(n-k)k} a_{k+1}a_{k-1}. \quad (4)$$

This can be proven by induction, along the same lines as above. For $n = 2$, $k = 1$, we have to prove $a_1^2 - 4a_2a_0 > 0$, which is just the condition of the positiveness of the discriminant of the given quadratic equation. Further, for $n > 2$, $k > 1$, formula (4) immediately follows from the corresponding inequality for the coefficients b_k, b_{k-1}, b_{k-2} of the derivative of $P(x)$:

$$b_{k-1}^2 > \frac{(n-k+1)k}{(n-k)(k-1)} b_k b_{k-2}$$

(the degree of $P'(x)$ is $n-1$!), which is true by the induction hypothesis just as in the proof of formula (1) above.

Finally, for $k = 1$ we have to refine formula (3) above. Summing the obvious inequalities

$$y_i^2 + y_j^2 > 2y_iy_j \quad (5)$$

(which amount to $(y_i - y_j)^2 > 0$) over all pairs (i, j), $i < j$, we get

$$(n-1)(y_1^2 + \dots + y_n^2) > 2(y_1y_2 + y_1y_3 + \dots + y_{n-1}y_n),$$

because every y_m^2 enters into $n-1$ formulas (5). Now, dividing this by $n-1$ and using formula (2) (remembering that $a_1 \neq 0$), we arrive at

$$\begin{aligned} a_1^2 &> a_0^2\left(\frac{2}{n-1} + 2\right) \cdot \left(y_1y_2 + y_1y_3 + \cdots y_{n-1}y_n\right) \\ &= \frac{2n}{n-1} a_0 a_2, \end{aligned}$$

which is just formula (4) for $k = 1$.
(V. Dubrovsky)

Physics

P81

Velocity is usually described as a function of time t, but in our case it's a function of the coordinate x. This is an unusual way of looking at it, but perhaps it yields a well-known kind of motion. Let's find out.

Take a very small time period t and find the acceleration a of the body:

$$a = \frac{\Delta v}{\Delta t} = \frac{\Delta v}{\Delta x}\frac{\Delta x}{\Delta t}.$$

It's clear that the first factor is the derivative of the function $v(x) = A\sqrt{x}$:

$$\frac{\Delta v}{\Delta x} = \frac{1}{2}A\frac{1}{\sqrt{x}} = \frac{A}{2\sqrt{x}}.$$

The second factor is the velocity:

$$\frac{\Delta v}{\Delta t} = v(x) = A\sqrt{x}.$$

Thus,

$$a = \frac{A}{2\sqrt{x}} A\sqrt{x} = \frac{A^2}{2} = \text{constant}.$$

This means that the body simply moves with a constant acceleration and will be at a distance L from the initial equilibrium position after time

$$t = \sqrt{\frac{2L}{a}} = \sqrt{\frac{2L}{A^2/2}} = \frac{2}{A}\sqrt{L}.$$

P82

In the initial state we can neglect both the vapor pressure and the volume the water occupies. We'll consider that air at atmospheric pressure p_0 = 1 atm and temperature T_0 = 293 K occupies the entire volume V of the pressure cooker. In the final state the pressure inside the cooker $3p_0$ consists of the air pressure and the pressure of completely evaporated water. Denoting by ρ, v, and M the density, initial volume, and molar mass of water, respectively (ρ = 10^3 kg/m³, M = 18 g/mol), then for air pressure p_a and vapor pressure p_v in final state we get

$$p_a = \frac{p_0 T}{T_0}, \quad p_v = \frac{\rho v R T}{MV}.$$

By the statement of the problem

$$p_a + p_v = 3p_0.$$

From these formulas we can find the ratio of the water's volume to the pressure cooker's volume:

$$\frac{v}{V} = \frac{p_0 M(3 - T/T_0)}{\rho R T} \cong 10^{-3}.$$

P83

Since the capacitors are connected in series, the capacities add as reciprocals:

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2},$$

$$C = \frac{C_1 C_2}{C_1 + C_2}.$$

When connected to a voltage source V, the charge q on the combination will be

$$q = CV = \frac{C_1 C_2}{C_1 + C_2} V.$$

Since this charge appears on each capacitor in series, it must not exceed the maximum charge for either capacitor: $q_1 = C_1 V_1$ or $q_2 = C_2 V_2$. Let's assume that

$$q_1 < q_2.$$

Then the following relationship should hold:

$$\frac{C_1 C_2}{C_1 + C_2} V \le C_1 V_1,$$

which can be reworked as

$$V \le \frac{C_1 + C_2}{C_2} V_1.$$

If $C_1 V_1 = C_2 V_2$, then we get

$$V \le \frac{C_2 \frac{V_2}{V_1} + C_2}{C_2} V_1 = V_1 + V_2.$$

P84

The force acting on a ring in a magnetic field is proportional to the current in the ring. The current in turn is proportional to the electromotive force $\mathcal{E}$ induced in the ring as it moves, and this emf is proportional to the rate of change in the magnetic flux penetrating the plane of the ring, which means that it's proportional to the ring's velocity. When the force acting on the ring due to the magnetic field is equal to the force of gravity, the ring will experience no net force and will travel at its terminal velocity.

Since the ring's kinetic energy doesn't change while it's falling at its terminal velocity, the change in its gravitational potential energy must be equal to the heat produced by the current in the ring.

Let the ring's terminal speed be v. The emf induced by the moving ring is

$$\mathcal{E} = k\frac{\Delta\Phi}{\Delta t},$$

where Φ is the magnetic flux penetrating the ring and k is a proportionality constant. Then

$$\Phi = \frac{\pi d^2}{4} B = \frac{\pi d^2}{4} B_0(1 + \alpha H),$$

$$\frac{\Delta\Phi}{\Delta t} = \frac{\pi d^2}{4} B_0 \alpha \frac{\Delta H}{\Delta t}.$$

But $\Delta H/\Delta t = v$, so

$$\mathcal{E} = k\frac{\pi d^2}{4} B_0 \alpha v,$$

and the current flowing through the ring is

$$I = \frac{\mathcal{E}}{R} = \frac{k\pi d^2 B_0 \alpha v}{4R}.$$

Now let's write down the energy conservation law. Let the mass of the ring be equal to m. The amount of heat produced in the ring during time t is equal to $I^2 R t$. If during this time the height of the ring decreases by h, then

$$mgh = I^2 R t.$$

Since $h/t = v$, then

$$mgv = I^2 R,$$

$$mgv = \frac{k^2 \pi^2 d^4 B_0^2 \alpha^2 v^2}{16R}.$$

From this it follows that

$$v = \frac{16Rmg}{k^2 \pi^2 d^4 B_0^2 \alpha^2}.$$

P85

Consider each water drop an identical nontransparent ball of radius R. Such a ball has a cross-sectional area of $s = \pi R^2$. This means that n drops scattered evenly in 1 m^3 of air fill an area of (approximately) $S = \pi n R^2$. This estimate doesn't take into account any partial overlap of the drops, but this isn't essential for an approximation.

Since the visibility is 10 m, drops within a rectangular parallelepiped with a base of area 1 m^2 and a length of 10 m should cover an area of 1 m^2. There are $10n$ drops in this parallelepiped, which cover an area of

$$10\pi n R^2 = 1 \text{ m}^2.$$

From this follows the value of n:

$$n = \frac{1}{10\pi R^2}. \qquad (1)$$

This formula contains the unknown value R for the radius of a drop. Let's find it. Two forces act on a drop: the constant downward force of gravity and the upward force of air resistance, which varies with the velocity of the drop. There comes a moment during the fall when these forces counteract each other. After this the drop's velocity stops increasing and the drop falls with a constant velocity, which can be found by equating the forces of gravity and air resistance:

$$mg = 4.3Rv$$

($m = \frac{4}{3}\pi R^3 \rho$ is the mass of one drop, $\rho = 10^3$ kg/m^3 is the density of water). From this we get

$$\frac{4}{3}\pi R^3 \rho g = 4.3Rv$$

and

$$R^2 = \frac{3}{4}\frac{4.3v}{\pi\rho g} \cong \frac{v}{\rho g}\ \text{m}^2. \qquad (2)$$

Since the fog descends over the course of 2 hours and its initial height was 200 m, the velocity of the water drops is

$$v = \frac{h}{t} = \frac{200\ \text{m}}{2 \cdot 3600\ \text{s}} \cong 0.028\ \text{m/s}.$$

Plugging this into equation (2), we find the radius of the drops:

$$R = \sqrt{\frac{0.028}{10^4}} \cong 1.7 \cdot 10^{-3}\ \text{m}.$$

Now from equation (1) we can determine n:

$$n = \frac{1}{10 \cdot 3.14 \cdot (1.7)^2 \cdot 10^{-6}} \cong 1.1 \cdot 10^4.$$

Brainteasers

B81

Cut the quadrilateral along the diagonal BD, turn triangle ABD upside down, and put the piece back so that its vertices B and D change places (fig. 7). We get a triangle $A'BC$, because $\angle BDC + \angle BDA' = 180°$. This triangle is isosceles ($A'B = AD = BC$), so $\angle C = \angle A' = \angle A$.

B82

The answer is shown in figure 8. It's found simply by trial and error.

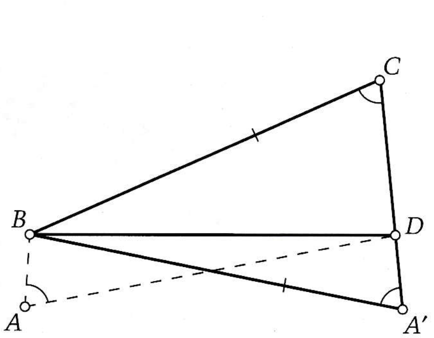


Figure 7

B83

It's clear from figure 9 that the Moon subtends a greater angle at the observer's position at its zenith than when it is on the horizon, so its angular diameter is greater in this position.

B84

A quick search shows that there is no solution in Arabic numerals (since bd is the number of a century, b is 1 or 2, but $b \neq 1$, because $a = 1$. So $bd = 20$; then, from the number of the year $abcd$, we get $b = 9$, which is a contradiction). So our letters must stand for Roman numerals. A further search yields the following unique answer: a = M, b = X, c = C, d = I, so the year is MXCI (or 1091), and the century is XI (the 11th).

B85

If there were k honest persons, the first k statements were wrong (and so were made by liars) and the last $12 - k$ statements were true (and so were made by honest persons). Therefore, $k = 12 - k$: there are $k = 6$ honest persons in the room.

Networks

(Solutions supplied by the editor)

1. If we draw the three force vectors, one after another, they'll form an equilateral triangle.

2–6. The solutions to these exercises can be found in the article "Botanical Geometry" in the September/October 1990 issue of *Quantum*.

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

Figure 8

7. The doubled number of segments of a shortest network is equal to $v_1 + 2v_2 + 3v_3 + 3f$ on the one hand, and $2(v_1 + v_2 + v_3 + f - 1)$ on the other. Equating the two expressions, we get $f = v_1 - v_3 - 2$.

8. This requirement was necessary to ensure that there is only one road from each village.

9. Use exercise 4.

10. The length of the shortest network is $4 + 3\sqrt{3}$.

11. According to the algorithm described in the article, the villages A and B are chosen under the assumption that there are no roads joining two villages. Since the number of villages that are the endpoints of only one road is still greater than the number of forks (exercise 7), at least two of them are connected to the same fork. We can choose them as A and B. In fact, if there is a village A with two or more roads starting at it, we can cut all these roads. Then the connection graph will be split into several disjoint parts, and the problem can be solved separately for the villages in each part plus village A. This allows us to reduce our search to the case when there is only one road from each village and this road leads to a fork.

12. Sometimes it's clear what triangle should be chosen, but it seems too difficult (if it is indeed possible) to formalize the rule of choice in the general situation. So perhaps the best thing one can do is to consider both possibilities.

13. The only essential feature of a "network" (not a "shortest network"!) is connectedness. And this property isn't violated when a connection graph is "shortened" according to the proposed algorithm.

14. The first question has to do with a very special arrangement of villages. However, one can conceive of such an arrangement and such a connection graph when it's impossible to avoid

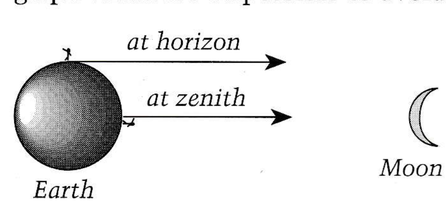


Figure 9

this unpleasant situation by, say, choosing another pair of villages A and B. We must confess that the algorithm doesn't work in this situation, and we couldn't find a way to make it work. Perhaps our readers will have better luck.

The situation described in the second question can simply be discarded: it can occur only if the connection graph that is being tested doesn't correspond to the shortest network.

Symmetry

1. (a) 1/6. Favorable outcomes are (6, 6), (6, 5), (5, 6), (5, 5), (6, 4), (4, 6). (b) 5/12. Favorable outcomes are (6, 3), (6, 2), (4, 2), and (n, n), $(n, 1)$ for all $n = 1, 2, \ldots, 6$.

2. We can assume that the moment when the passenger comes to the bus stop is uniformly distributed in the interval between the arrivals of buses of either line. Then the probability of waiting not less than t minutes for a bus of one line is $(10 - t)/10 = 1 - t/10$ if $t \le 10$ and zero if $t > 10$; for the other line, it's $1 - t/15$ if $t \le 15$. We're interested in the intersection of these two events. Since the two schedules are assumed to be independent, we can apply the multiplication rule, which yields the answer $(1 - t/10)(1 - t/15)$ for $t \le 10$ and 0 for $t > 10$.

3. Label the three parts of the segment with the numbers 1, 2, 3. Then any outcome of the experiment in the problem can be represented as a triple (k, l, m), where k, l, and m are the numbers of the parts hit by the first, second, and third point, respectively. There are $3^3 = 27$ outcomes, and all of them are equally likely. For favorable outcomes all three numbers must be different, and since there are 6 permutations of the numbers 1, 2, 3, the answer is $6/27 = 2/9$.

4. The points A, B, C, D can be chosen in two steps: first we choose four points on the circle, then we label them at random. After point A is labeled, there are three equally likely variants for labeling point C, exactly one of which (opposite to A) yields intersecting chords. So the answer is 1/3.

5. A chord of a circle of radius r is longer than the radius if and only if its midpoint is at a distance less than $(\sqrt{3}/2)r$ from the center (fig. 10). In part (a) the random point must hit the corresponding portion of the diameter, in (c) it must fall into the circle of radius $(\sqrt{3}/2)r$. So the answer in (a) is $\sqrt{3}/2$, and in (c) it is $\pi[(\sqrt{3}/2)r]^2/\pi r^2 = 3/4$. In part (b) one of the points can be fixed (point A in figure 10); this leaves 5/6 of the circumference (the greater arc BC) for the other point; so the answer is 5/6. The answers in part (d) are 1/2, 1/3, and 1/4, respectively.

6. (a) 1. Any triangle has an angle greater than 30°—otherwise, the sum of its angles would be less than 180°. Parts (b) and (c) can be answered by using the method applied in the first solution to the acute triangle problem. Figure 11 shows the sets of points (α, β) in the square $0 < \alpha < 2\pi$, $0 < \beta < 2\pi$, satisfying the conditions of these problems. The answers are: (b) 1/4 (fig. 11a); (c) 2/3 (fig. 11b).

7. The answer is 1/4. This problem is just another form of the acute triangle problem. Indeed, assume that the segment is 2π units long. Then the lengths of its three pieces can be thought of as the measures α, β, γ of the three arcs into which a unit circle is divided by three random points. The conditions that the triangle formed by these points is acute are $\alpha < \pi$, $\beta < \pi$, $\gamma < \pi$. But since $\alpha + \beta + \gamma = 2\pi$, these inequalities can be rewritten as $2\alpha < 2\pi = \alpha + \beta + \gamma$, or $\alpha < \beta + \gamma$, and, similarly, $\beta < \alpha + \gamma$, $\gamma < \alpha + \beta$, which are just the triangle inequalities for the pieces of the initial segment.

8. (a) $\pi/4$; (b) $\pi/2\sqrt{3}$. The coin covers a node if the distance from the coin's center to the node is not greater than 1/2 (the radius of the coin). So part (a) turns out to be equivalent to problem 6 in the article. In part (b) the unknown probability is the ratio of the area of the three shaded sectors in figure 12 to the area of the triangle.

9. Both problems are solved by using the technique in the second solution to the acute triangle problem in the article. The answer to part (a) is $1 - n/2^{n-1}$ (there are 2^n ways of choosing an endpoint of each of n diameters and $2n$ of these ways—the number of sectors formed by n diameters—

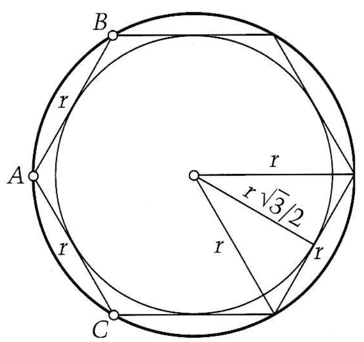


Figure 10

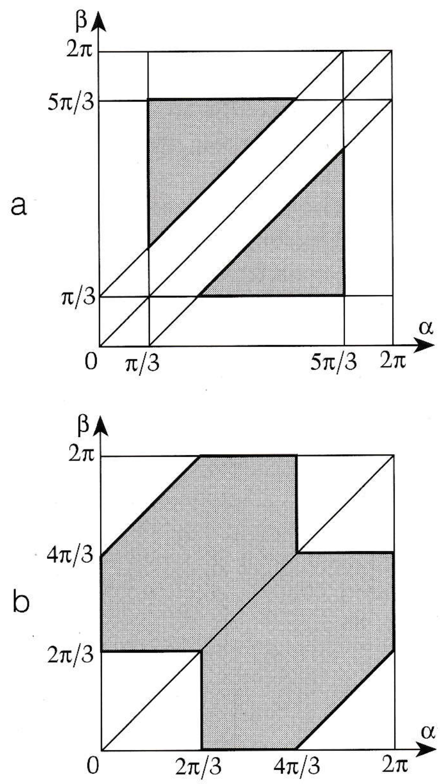


Figure 11

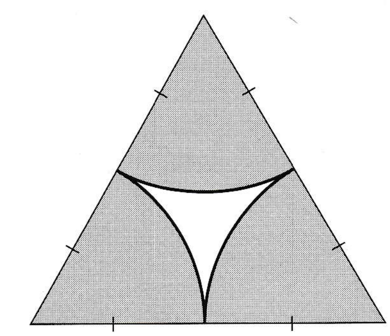


Figure 12

yield the n-gons that *do not contain* the center of the circle). The numerator $n^2 - n + 2$ in the answer to part (b) is the number of parts ("sectors") into which a sphere is divided by n great circles no three of which have a common point. This can be proven by induction: when a new great circle is added to k circles already drawn, it's cut by the "old" circles into $2k$ arcs; each of these arcs cuts a certain "old" piece of the sphere in two, thus adding $2k$ pieces. So for n circles the number of pieces is $2 + (2 + 4 + \ldots + 2(n - 1)) = n^2 - n + 2$.

IMO

(See the Happenings department in the March/April issue)

1. The main idea in this problem is that the ratio

$$A = \frac{abc - 1}{(a-1)(b-1)(c-1)}$$

is always greater than 1 and smaller than 2 for all sufficiently large values of a, b, and c, so it can be an integer only for a finite number of triples a, b, c, which can be examined simply by trial and error.

Letting $\alpha = 1/(a - 1)$, $\beta = 1/(b - 1)$, and $\gamma = 1/(c - 1)$, and noting that $x/(x - 1) = 1 + 1/(x - 1)$, we can write A as

$$\begin{aligned} A &= (1 + \alpha)(1 + \beta)(1 + \gamma) - \alpha\beta\gamma \\ &= 1 + \alpha\beta + \beta\gamma + \gamma\alpha + \alpha + \beta + \gamma. \end{aligned}$$

It follows that $A > 1$ and A decreases together with any of the arguments α, β, γ—that is, decreases as a, b, or c increases. Now let's consider three possibilities for a, the smallest of the numbers.

(1) $a \geq 4$. Then $b \geq 5$, $c \geq 6$, so

$$1 < A \leq \frac{4 \cdot 5 \cdot 6 - 1}{3 \cdot 4 \cdot 5} < 2,$$

and A cannot be an integer.

(2) $a = 3$. Then $b \geq 4$, $c \geq 5$, $A < 2\frac{1}{2}$. Therefore, $A = 2$, and we get the following equation for b and c: $3bc - 1 = 4(b - 1)(c - 1)$, or $(b - 4)(c - 4) = 11$, whose only integer solution satisfying $3 < b < c$ is $b = 5$, $c = 15$.

(3) $a = 2$. Similar reasoning yields two possible values for A: $A = 2$ and $A = 3$. In the first case, b and c must satisfy the equation $2(b + c) = 3$, which has no integer solutions. In the second case, we get the equation $(b - 3)(c - 3) = 5$, which has a unique solution for $2 < b < c$: $b = 4$, $c = 8$. So the *answer* is $(a, b, c) = (3, 5, 15)$ or $(2, 4, 8)$.

2. The answer is $f(x) = x$. This problem has many seemingly different solutions. But most of them are fundamentally alike. We'll give perhaps one of the shortest versions, based on an idea found by two or three Olympiad participants.

We can write the given equation as

$$y = f(x^2 + f(y)) - f(x)^2. \qquad (1)$$

In this form, it shows us that f must be a one-to-one function. For if $f(y_1) = f(y_2)$, then $y_1 = f(x^2 + f(y_1)) - f(x)^2 = f(x^2 + f(y_2)) - f(x)^2 = y_2$. Writing $f(x)^2 = f(x^2 + f(y))$ and substituting, we find that $f(x)^2 = f(-x)^2$. But f is one-to-one, so $f(-x) \neq f(x)$ only when $x = 0$. Hence $f(-x) = -f(x)$, and f is an *odd* function.

Suppose now there is some $z \neq 0$ such that $f(z) \neq z$. Let $y = z$ if $f(z) < z$ and $y = -z$ if $f(z) > z$. In both cases $f(y) < y$ (in the second case, $f(y) = f(-z) = -f(z) < -z = y$), so we can put $x = \sqrt{y - f(y)}$. Plugging these y and x into equation (1), we get

$$\begin{aligned} f(x^2 + f(y)) &= f(y - f(y) + f(y)) \\ &= f(y) = y + f(x)^2 \geq y, \end{aligned}$$

which contradicts $f(y) < y$. This contradiction ensures that $f(z) = z$ whenever $z \neq 0$.

Finally, setting $x = 0$, $y = 1$ in equation (1), we get $f(0 + f(1)) = f(1) = 1$ on the left side and $1 + f(0)^2$ on the right side, which means that $f(0) = 0$. This completes the proof that $f(x) = x$ for all x.

3. The answer is $n = 33$. The total number of edges joining 9 points to one another is $9 \cdot 8/2 = 36$. If exactly 33 of them are colored, we can choose three points so that each of the three uncolored edges has at least one of these points as its endpoint. Then all the edges joining the remaining 6 points are colored. But it's a well-known fact that in this case 3 of these 6 points are joined by edges of one color. (Here is the proof: take one of the six points, A; at least 3 of 5 edges joining A to the other points—say, AB, AC, and AD—are the same color—say, red; then either the triangle BCD is blue, or one of its sides is red and so forms a red triangle with two edges issuing from A.) So 33 colored edges always contain a one-color triangle and, therefore, $n \leq 33$.

On the other hand, there is an example of a coloring of 32 edges in two colors that has no one-color triangles. Such an example can be constructed step by step, using the following idea. We start with a set of points with all the line segments connecting them colored in one of two colors, but with no one-color triangles among them. We add a new vertex B to this graph by looking at one of the "old" vertices—say, A. As we join B to each old vertex X, we color edge BX using the same color as AX was. We leave edge AB uncolored. This construction cannot create one-color triangles, because if the edges of a triangle BXY are the same color, then this would have had to be true for triangle AXY too, which was not allowed for the original graph.

Now let's take the graph shown in figure 13 and extend it, using this construction four times. We'll obtain $5 + 4 = 9$ vertices and $10 + 4 + 5 + 6 + 7 = 32$ colored edges with no one-color triangles. So $n > 32$, and we're done.

4. The locus in question (see figure 14) is the ray l starting at the point B of circle C diametrically opposite the point A where C touches L. The

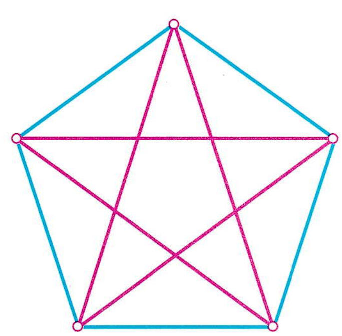


Figure 13

direction of the ray is defined by the condition that its extension beyond B cuts L at point D symmetrically to A with respect to M ($DM = MA$). In other words, l is parallel to MO, where O is the center of C; the point B itself is, of course, excluded.

Let PQR be an *arbitrary* triangle circumscribed about circle C with the base QR lying on line L. Draw the tangent Q_1R_1 to C through point B (see figure 14) and consider the dilation relative to the center P that takes triangle PQ_1R_1 into PQR. Clearly, this dilation takes B into D and the incircle C of triangle PQR into the "excircle" of this triangle—that is, the circle touching side QR (the image of Q_1R_1) from the outside and the extensions of sides PQ and PR. So D is the point of contact of the excircle and side QR, and A is the point of contact of the incircle with the same side. It's known that $QA = RD$ (both these segments can be expressed in terms of side lengths of triangle PQR—this will be explained below). So the condition $QM = MR$ is equivalent to $AM = MD$, which means that point P belongs to the locus in question if and only if it lies on the line DB beyond point B.

To prove the equality $QA = RD$, denote by p the distance from P to either of the points K or N where C touches sides PQ and PR (of course, $PK = PN$), and denote by q and r the

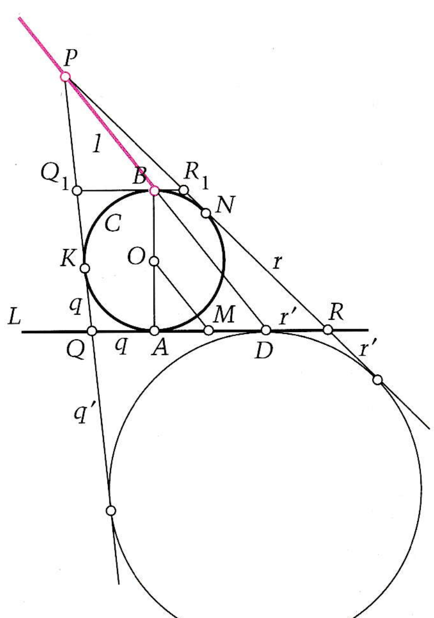


Figure 14

similar distances for vertices Q and R ($q = QA = QK$, $r = RA = RN$). Then

$$\begin{aligned} QA &= q \\ &= \frac{1}{2}\left[(q+r)+(q+p)-(p+r)\right] \\ &= \frac{1}{2}(QR+QP-PR). \end{aligned}$$

Similarly, if p', q', r' are the distances from P, Q, and R to the points of contact of the excircle with the corresponding sides of triangle PQR, then

$$\begin{aligned} RD &= r' \\ &= \frac{1}{2}\left[(r'+q')+(p'-q')-(p'-r')\right] \\ &= \frac{1}{2}(QR+QP-PR) \\ &= QA. \end{aligned}$$

If we allow that the circle C in the statement of the problem simply touches the lines PQ and PR (that is, is either the in- or excircle of triangle PQR), the unknown locus would consist of two rays—all the points of line BD outside circle C. The proof is virtually the same.

5. We'll prove the inequality by induction over the number of points $|S|$ in the set S.

For a one-point set ($|S| = 1$), it's trivial.

Now assume that the inequality is true for all sets S such that $|S| < n$, $n > 1$, and consider a set S such that $|S| = n$. Since $n > 1$, we can choose a plane parallel to one of coordinate planes—let it be the xy-plane—not passing through any point of S and dividing S into two nonempty subsets S_1 and S_2 (the projection of S onto one of the coordinate planes must contain more than one point; then the plane we need can be taken as perpendicular to this plane). Denoting the numbers of points in the projections of the sets S_1 and S_2 onto the coordinate planes by $x_1, y_1, z_1; x_2, y_2, z_2$ (so that $x_1 = |S_{1x}|$, $y_2 = |S_{2y}|$, and so on), we can write $z_1 \le |S_z|$, $z_2 \le |S_z|$, $|S_x| = x_1 + x_2$, $|S_y| = y_1 + y_2$, and so, by the inductive assumption,

$$\begin{aligned} |S|^2 &= (|S_1|+|S_2|)^2 \\ &\le \left(\sqrt{x_1y_1z_1}+\sqrt{x_2y_2z_2}\right)^2 \\ &\le |S_z|\left(\sqrt{x_1y_1}+\sqrt{x_2y_2}\right)^2 \\ &= |S_z|\left(x_1y_1+x_2y_2+2\sqrt{(x_1y_2)(y_1x_2)}\right) \\ &\le |S_z|\left(x_1y_1+x_2y_2+x_1y_2+y_1x_2\right) \\ &= |S_z|(x_1+x_2)(y_1+y_2) \\ &= |S_z||S_x||S_y|. \end{aligned}$$

Here we used the arithmetic mean–geometric mean inequality $\sqrt{ab} \le (a+b)/2$.

So the statement has been proven for $|S| = n$, and thus for all values of $|S|$.

The continuous version of the statement of this problem is also true (and is proven in much the same way): if V, A_x, A_y, A_z are the volume of some body and the areas of its projections onto the coordinate planes, then

$$V^2 \le A_xA_yA_z.$$

This version, as it turns out, was rather well known (in fact, it was even published in *Kvant*, *Quantum*'s Russian sister magazine), and some of the IMO participants used it in their proofs for the "finite case."

6. (a) Suppose $S(n) > n^2 - 14$ for some $n \ge 4$. Then n^2 can be represented as the sum of $k = n^2 - 13$ positive square integers:

$$n^2 = a_1^2 + a_2^2 + \ldots + a_k^2.$$

Subtracting k ones, we get

$$\begin{aligned} n^2 - k &= 13 \\ &= (a_1^2-1)+(a_2^2-1)+\ldots+(a_k^2-1), \end{aligned}$$

which means that 13 is the sum of several numbers of the form $a^2 - 1$—that is, the numbers 0, 3, 8, 15, ... (the addends may repeat). But this is not true. So $S(n) \le n^2 - 14$.

(b) To tackle this problem, we may note that $n^2 - 14 \ge 2$ whenever $n^2 - 14 \ge 0$, so n^2 must necessarily be representable as the sum of two squares. But it is known (see, for instance, "Genealogical Threes" in the November/December 1990 issue of *Quantum*, or any other text on Pythagorean triples) that in this case the number n itself is a multiple of the sum of two coprime squares. The first few possible values of n are $5 = 1^2 + 2^2$, $10 = 1^2 + 3^2$, $13 = 2^2 + 3^2$, $17 = 1^2 + 4^2$. The numbers 5 and 10 don't fit, because 25^2 and 10^2 can't be represented as the sums of three squares. But it turns out that 13 can, and that this is the value we seek for n. Let's prove it.

Note first that if n^2 is a sum of k squares at least one of which is even, then n^2 is the sum of $k + 3$ squares, because $(2m)^2 = m^2 + m^2 + m^2 + m^2$. If we split even squares in this manner, then, starting with

$$13^2 = 8^2 + 8^2 + 4^2 + 4^2 + 3^2,$$

we'll get representations of 13^2 as the sum of k squares for $k = 5, 8, 11, \ldots, 155$ (and even more—up to $64 + 64 + 16 + 16 + 1 = 161$); starting with

$$13^2 = 8^2 + 8^2 + 4^2 + 4^2 + 2^2 + 2^2 + 1^2,$$

we'll get the representations for $k = 7, 10, 13, \ldots, 154$ (in fact, up to $k = 169$); and

$$13^2 = 8^2 + 8^2 + 4^2 + 3^2 + 3^2 + 2^2 + 1^2 + 1^2 + 1^2$$

generates the representations for $k = 9, 12, 15, \ldots, 153$ (this is where we can't get longer sums: 153 is the number of terms after all possible splits). It remains to fill the gaps from 1 to 6:

$k = 1$: $13^2 = 13^2$;
$k = 2$: $13^2 = 12^2 + 5^2$;
$k = 3$: $13^2 = 12^2 + 4^2 + 3^2$;
$k = 4$: $13^2 = 10^2 + 8^2 + 2^2 + 1^2$;
$k = 6$: $13^2 = 12^2 + 3^2 + 2^2 + 2^2 + 2^2 + 2^2$.

It's not hard to show that 169 cannot be represented as a sum of 156 squares. First of all, none of these squares can be too large. In fact, if even one of them were as large as 16, the sum of the others would have to be at most $169 - 16 = 153$. But there are 155 of them, which is too many. So we would be representing 169 as the sum of, say, a 1's, b 4's, and c 9's. That is, $a + 4b + 9c = 169$, and $a + b + c = 156$. Subtracting these two equations, we find that $3b + 8c = 13$, which has no solution in integers.

Thus, 13^2 cannot be represented as a sum of 156 perfect squares, and $S(13) = 155 = 13^2 - 14$.

(c) Note first that any number $m \geq 14$ is representable as the sum of numbers of the form $x^2 - 1$—that is, of the numbers 0, 3, 8, 15, ... (Indeed, any $m \geq 14$ equals either $14 + 3k = 8 + 3(k + 2)$, or $15 + 3k$, or $16 + 3k = 8 + 8 + 3k$.)

Now let's prove that any $N \geq 28$ can be represented as the sum of k nonzero squares for all k, $N/2 \leq k \leq N - 14$. To this effect, rewrite $N = a_1^2 + \ldots + a_k^2$ as $N - k = (a_1^2 - 1) + \ldots + (a_k^2 - 1)$, and the inequalities above as $14 \leq N - k \leq k$. Since $N - k \geq 14$, we can represent $N - k$ as the sum of several nonzero terms of the form $x^2 - 1$. The number of these terms is not greater than $N - k \leq k$, so we can add as many terms $0 = 1^2 - 1$ as necessary to make the total number of terms exactly equal to k. We can then form the desired representation as a sum of squares by using the trick from the first paragraph of this solution.

Setting $N = n^2$, we see that what remains to prove is the existence of representations with $k = 1, 2, \ldots, n^2/2 - 1$ terms for infinitely many values of n. But it will be more convenient to prove directly that $S(n) = n^2 - 14$ implies $S(2n) = (2n)^2 - 14$, and so any value of n from part (b) (say, $n = 13$) generates an infinite sequence of integers $N = n, 2n, 2^2n, \ldots$, such that $S(N) = N^2 - 14$.

If $n^2 = a_1^2 + \ldots + a_k^2$, then, by splitting even squares in $(2n)^2 = (2a_1)^2 + \ldots + (2a_k)^2$, we can get representations of $(2n)^2$ as the sums of $k, k + 3, \ldots, 4k$ terms. Setting $k = 1, 2, \ldots, m$, we'll obtain representations of $(2n)^2$ with $1, 2, \ldots, m$, $m + 1 = (m - 2) + 3$, $m + 2 = (m - 1) + 3, \ldots, 4m - 8 = 4(m - 2)$, $4m - 7 = 4(m - 1) - 3$, $4m - 6$ terms (working a numerical example will make this clear). So

$$S(2n) \geq 4S(n) - 6.$$

Suppose that $S(n) = n^2 - 14$. Then, by the solution to part (b), $n \geq 13$. Since $(2n)^2 \geq 28$, there are additive representations of $(2n)^2$ with any number of terms from $(2n)^2/2$ to $(2n)^2 - 14$. On the other hand, $S(2n) \geq 4S(n) - 6 = 4n^2 - 62 > (2n)^2/2$. So the equality $S(2n) = (2n)^2 - 14$ is true.

Now the proof is complete.

There are numbers other than 13, $2 \cdot 13$, $2^2 \cdot 13$, ... such that $S(n) = n^2 - 14$. One of them is $n = 17$. Try to verify this.

Swinging techniques

"How I do like to go up in a swing,
Up in the air so blue . . ."
—Robert Louis Stevenson

by Alexey Chernoutsan

I FEEL PRETTY CONFIDENT IN asserting that the swing is not a human invention. Just look at how apes swing so adroitly among the vines. I imagine we humans inherited this form of entertainment from our ancestors on the evolutionary tree. The venerable age of this activity probably has something to do with its simplicity. This is no high-tech device! All you need to do is attach a rope at a height of a few meters, grab the other end, run, lift your legs—and off you go, back and forth on your homemade swing.

But is it as simple as all that? If you do nothing after taking off from the ground, the amplitude of the swing's oscillations will gradually damp until the swing stops oscillating. Look at figure 1. You see a swing with a person in the same pose at three sequential moments: at the extreme left position (point A); at the middle (point B); and at the extreme right (point C). The effect of friction at the swing's suspension point and air resistance result in a lowering of the center of gravity of the swing (together with the person) at the extreme positions (compare A with C). This is due to the gradual decrease in the mechanical energy of the system. How can we avoid this? What do we need to do to start the swing moving without touching the ground and to keep it swinging as long as we wish? All we need to do is lower our bodies' center of gravity a little at the extreme positions and lift it at the middle position. If you're standing on the swing, you just squat and straighten up at the right times. If you squat enough, the amplitude of the oscillations will increase (fig. 2).

What about if you swing sitting down? You've known the answer (if not the reason) since childhood: you bend and straighten your knees (causing your center of gravity to move up and down). In this case the increase in mechanical energy is due to the work performed by your muscles: when you raise your center of gravity at the middle position, you do more work than when you lower it at the extreme positions.

To understand why it's so, we'll leave the playground and enter our home laboratory. We imagine the simplest model of a person on a swing, consisting of a mass M attached to a thin thread. To imitate the way the center of gravity goes up and down, we'll run the upper end of the thread through a small hole at the swing's pivot point (fig. 3). As the need arises we can either pull the thread up, reducing the pendulum's length l, or let it out. It's

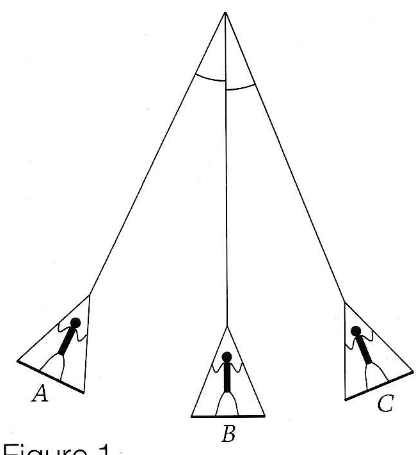


Figure 1

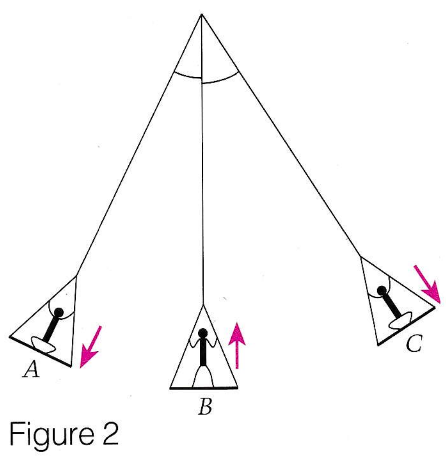


Figure 2

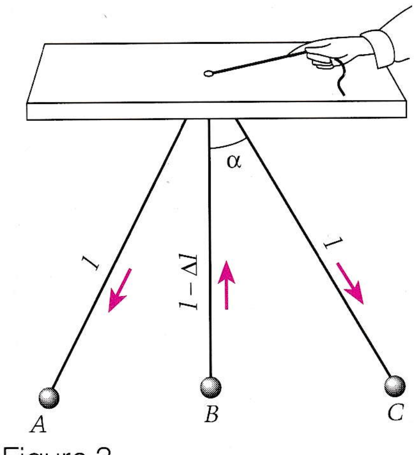


Figure 3

clear why the thread exerts more force at the middle position: as the weight moves along the arc, it has a centripetal acceleration $a = v^2/l$ and the force of the thread must exceed the force of gravity by $ma = mv^2/l$. A similar phenomenon is observed when you use a swing: at the middle position you're pressed against the seat of the swing (you've being affected by the centrifugal force—if you're familiar with noninertial systems, you'll understand what I'm talking about). Your leg muscles and feet are strained much more, and you perform even more work when you straighten your legs.

Let's get back to our simplified model (fig. 3) and do some simple calculations that will help us understand the conditions under which swinging occurs. The force of the thread at the low position is equal to $F_1 = mg + mv^2/l$, and at the extreme position it equals $F_2 = mg \cdot \cos\alpha$ (α is the maximum angle of deviation). The work done in one period equals

$$\Delta W = 2(F_1\Delta l - F_2\Delta l)$$
$$= 2\frac{\Delta l}{l}\left[mgl(1-\cos\alpha) + mv^2\right],$$

where Δl is the change in the pendulum's length (we assume that $\Delta l/l << 1$). The mechanical energy of the system is equal to the kinetic energy at the bottom of the swing, $E = mv^2/2$, or the gravitational potential energy at the extremes, $E = mgl(1 - \cos\alpha)$. Then

$$\frac{\Delta W}{W} = 6\frac{\Delta l}{l}.$$

Pay attention now: the relative increase in energy in a period doesn't depend on whether the pendulum is swinging strongly or weakly—that is, it doesn't depend on the amplitude α. This is very important. Why? If the pendulum isn't pushed, a certain amount of energy will be lost during each period because of the air resistance. Get it? The pendulum gains some energy here, loses it there. To increase the amplitude, the energy gained must be greater than the energy lost. And this condition remains the same regardless of the amplitude. So if the amplitude α decreases by 3% in one period, then the energy decrease is approximately 6%. (For small oscillations, $E \cong \alpha^2$. This can be understood by realizing that $\cos\alpha \cong 1 - \alpha^2/2$ for small α.)

To get a pendulum 1 m long to swing under these conditions, one merely needs to shorten it by 1 cm at the middle position and lengthen it at the extreme position. Once you begin swinging, you don't need to squat lower and lower—you just squat the same amount each time . . . and you'll go higher and higher! Not only that: even if the swing is stationary, if you start squatting low enough with a period that is half the swing's oscillation period, the swing will necessarily start oscillating. In the language of physics, not only can we amplify the oscillations, we can generate them as well.

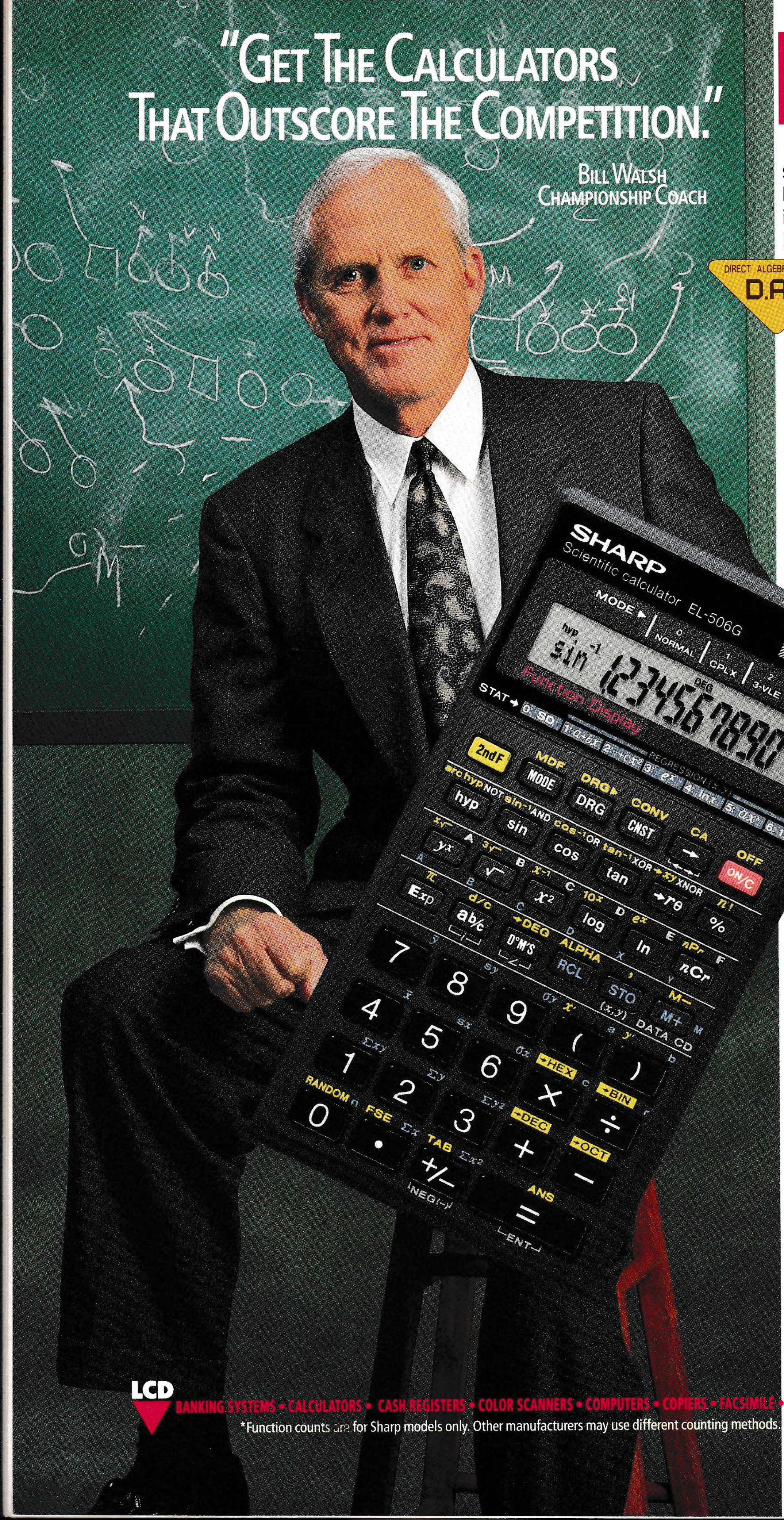
So, by looking closely at a favorite childhood pastime, we've managed to delve into many features of a significant physical phenomenon: *parametric resonance*.

Why is it called this? In order for something to start oscillating, one of the parameters that determine the oscillation period must change. It's not necessary to change the parameter twice in a period—you can do it just once, or even once over several periods, though in this case you must make a more pronounced change in the parameter.[1] It doesn't matter which parameter we change. For example, for an ideal pendulum the oscillation period $T = 2\pi\sqrt{l/g}$. Therefore, we can change not only l but g as well. (To do this you needn't try to change the force of gravity—all you need to do is hold the suspension point and give it the necessary vertical acceleration. Acceleration upward causes an increase in g in a noninertial system centered on the suspension point; acceleration downward causes a decrease.) For an object of mass m suspended from a spring, the oscillation period $T = 2\pi\sqrt{m/k}$. Therefore, we can change either the spring constant k or the mass of the object (I leave it to you to figure out how). In each of these cases, the parameter must be changed such that the total work is positive and the pendulum's energy increases.

We'll take a step further—after all, why should we restrict ourselves to mechanical systems? Let's look at an "electronic swing"—that is, an oscillating electric circuit. If the circuit contains an inductor L and a capacitor C in series, the circuit has a resonant frequency $\omega = 1/\sqrt{LC}$. Therefore, a periodic change in the capacitance should, by analogy, lead to an amplification of the oscillations. We'll increase the distance between the plates of the capacitor when its charge is maximum (corresponding to the extreme position of our swing—see figure 1). We'll move the plates back when the capacitor is not charged (the middle position). You'll notice right away that in the first case we do positive work (oppositely charged capacitor plates attract), but in the second case the work will be equal to zero. If the applied energy is greater than the heat losses, the oscillations must increase. In the 1930s a device like this was invented by the Soviet physicists Mandelshtam and Papaleksy (although they used another design: to change the capacitance they moved the plates laterally in opposite directions, thus changing the effective area of the plates). They named their device a *parametric generator*: not only does the oscillation amplitude in it increase due to mechanical work, oscillations are actually generated as well. All you need to do is change the distance between the plates with the required frequency—oscillations will arise all by themselves (there is always some random charge on the plates!).

You can see how much useful physical information can be gotten by just thinking about the underlying principle of a simple toy—in this case, a swing. So spend more time playing with your little sisters and brothers! ◘

[1]By the way, you may have noticed that children who swing standing up usually squat only as they're swinging forward—that is, once per period. No doubt it's just more comfortable that way.



ONLY SHARP'S D.A.L. SCIENTIFIC CALCULATORS GIVE YOU SUCH AN EASY WAY TO SIMPLIFY MATH.

Sharp introduces the most revolutionary scientific calculator advance in over a decade. D.A.L. It stands for Direct Algebraic Logic. And, it signifies a vast improvement in the algebraic logic currently used in scientific calculators.

Only Sharp makes this exciting innovation available on its five non-programmable models.

Developed in response to requests by educators, D.A.L. allows the user to enter equations in the exact order in which they appear on paper. This greatly simplifies entries and reduces errors.

EXAMPLE	D.A.L. KEYSTROKES	WITHOUT D.A.L.
2 log (3 + 2) =	2 log (3 + 2) =	2 × (3 + 2) log =

But most importantly, D.A.L. substantially reduces the time needed to learn a scientific calculator. It also makes teaching calculations faster and easier. These remarkable calculators provide up to 276* functions which frees students from tedious computations. And gives teachers an easier way to make math concepts come alive. What's more, the ETS has even approved usage of these calculators on several standardized tests (including PSAT and SAT).

All of Sharp's D.A.L. models have Sharp's leading LCD technology including a large 10-digit LCD display with 2-digit exponent. This high-contrast display is easy to read and includes 3 or 4 character function display to indicate type of calculation in progress. You get highly durable plastic keys on all models plus protective hard cases on most. Two even offer Twin Power solar/battery operation.

D.A.L. When it comes to simplifying math, nothing scores higher. Which puts Sharp's D.A.L. Scientific Calculators in a class by themselves.

For a free booklet on Sharp's D.A.L. Calculator line, call: 1-800-BE-SHARP.

Or see the following models at your Sharp retailer:

EL-509G, EL-531G, EL-506G, EL-520G, EL-546G

SHARP®
FROM SHARP MINDS COME SHARP PRODUCTS™

*Function counts are for Sharp models only. Other manufacturers may use different counting methods.

Circle No. 16 on Reader Service Card