QUANTUM

MARCH/APRIL 1994

$5.00

SPRINGER INTERNATIONAL

GALLERY Q

A NATOLY TIMOFEYEVICH Fomenko (b. 1945) is a mathematician at Moscow State University. Although he produces drawings of great emotional power and technical competence, he does not consider himself an artist. "In my mind, these are not just an artist's images. . . . I am a mathematician. To me, my drawings are like photographs of some strange and interesting mathematical world. For me it is not important to be an artist but to represent images of this world so that others can appreciate it. To penetrate this world, you must study mathematics at a reasonably high level, maybe even be a professional mathematician. If you study mathematics only for technical purposes and do not stop to think deeply about the ideas, then you really cannot understand this world. In that sense I differ from other artists."

Fomenko's drawing No. 229, "Deformation of the Riemann surface of an algebraic function," is but one of many beautiful and intriquing images to be found in his book *Mathematical Impressions* (Providence: American Mathematical Society, 1990). Here's his description:

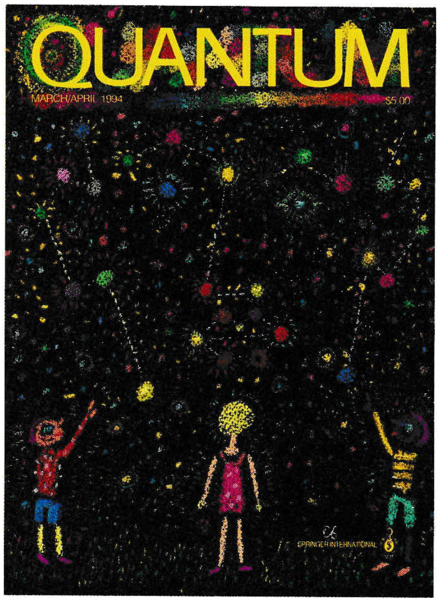
No. 229 (1983) by Anatoly Fomenko

Underlying this twisted deformation of space, where long tubes intertwine to weave a tortuous egg-like shape, is a certain three-dimensional model. The model shows a deformation of a Riemann surface of a special algebraic function, set in four-dimensional Euclidean space. This surface is also considered to be homeomorphic to a two-dimensional sphere with one handle as well as a two-dimensional torus. In terms of the theory of algebraic functions, we can construct this kind of Riemann surface by taking two copies of a two-dimensional sphere, making two cuts on each, and then gluing the corresponding cuts together. What we obtain is a torus, or donut-like object, represented as two spheres joined together by two tube-like cylinders (which are shown in this image). A curious feature of this form is that if we deform the underlying function, a polynomial, such that its roots coalesce, then so too does the corresponding Riemann surface follow. Vanishing cycles appear, singular points arise, and the surface loses its smoothness. In this image, two roots seek to coalesce into one, and, as a result, the upper sphere gets smaller while the lower one grows larger, a process somewhat visible through the object's cutaway sections.

If you need to catch your breath after visiting Fomenko's "parallel world," visit the Kaleidoscope and Toy Store in this issue. Riemann won't be there, but the torus will.

QUANTUM

MARCH/APRIL 1994 VOLUME 4, NUMBER 4



Cover art by Sergey Barkhin

"... How I wonder what you are ..." Well, according to the *Encyclopaedia Britannica*, you are "usually a self-gravitating mass of glowing gas, the range of masses and temperatures being comparable with those of the Sun, which is regarded as an average star." Not exactly what Jane Taylor had in mind when she penned her immortal verse back in 1806, but it's an answer to the question posed.

Happily, there always remain questions just waiting to be put into words. If we back up and recite the rhyme from the beginning, we're most likely lulled by its familiarity, if not irritated by its triteness. But if we let that *truly* magic word of childhood emerge—not "Please!" but "Why?"—whole new worlds open for exploration.

Beginning on page 22, Pavel Bliokh asks (and answers) the simple question: "Why do stars twinkle?"

FEATURES

DEPARTMENTS

PUBLISHER'S PAGE

Extremists of every stripe

Where is the middle ground in Russia?

THE RECENT ELECTIONS IN Russia have thrown American and European politicians into a tizzy. The Russian people, it would appear, do not accept the policies of rapid economic reform that the West wants to impose as a condition for our assistance. Not only that, they have elevated a ranting ultranationalist—the possibly dangerous, perhaps merely ridiculous Vladimir Zhirinovsky—to political prominence at home (if not abroad). Zhirinovsky's reckless rhetoric has frightened many in the West, with his (unfortunately plausible) talk of new bombs and his (fundamentally insane) threat to pile nuclear waste on the Russian border and install fans to blow the radioactivity into the Baltic states!

Perhaps we have overreacted. After all, we have seen citizens in our country cast "protest votes" for candidates whose extreme views and limited qualifications would frighten these same voters if their champion actually managed to get elected. There is evidence in the Russian press (for instance, recent letters and opinion polls in the weekly *Argumenty i Fakty*) that many, if not most, Russians are wary of Zhirinovsky. This is a nation that suffered horrendous losses in repelling Nazi expansionism. But it is history itself that cautions us to pay heed to embryonic dictators eager to capitalize on the uncertainty, resentment, and fear of a proud people. When a Russian politician speaks of regaining "lost territory," and other Russians vote for him, we all—Russians and Americans alike—must take precautionary measures.

But what kind of measures? As the former Soviet Union has steadily fallen apart, the United States has acted as if democratic and economic reforms are indistinguishable. The operating assumption seem to be: fix the Russian economy, and rehabilitation of the Russian body politic will follow. The appearance of Zhirinovsky and others of his ilk shows that, even were this assumption true, time may be running out. After several years of "shock therapy" largely imposed by the West, the banking system in Russia is still a shambles, production continues to decline, and a new "mafia" has taken charge of economic life. All shock, no therapy.

Why are we holding back?

So it's no surprise that some Russians are heeding the siren song of nationalism and iron rule from the top down. This is no time for legal niceties, they would argue. And if we were faced with such chaos, would we even try to refute them?

Sadly, this outcome could easily have been predicted, and in my view the United States is largely at fault for creating this situation. When the cold war ended, the United States was relieved of spending some $50 to $100 billion in defense-related funds to offset similar expenditures in the former USSR. Although our military-industrial complex was huge, the military component of the Soviet economy was staggering. Our conversion to a world without the threat of the USSR has merely produced recession and dislocation of many people. In the former Soviet Union, the effects have been pervasive and profound.

As a superpower, the polyglot USSR was held together largely by its well-controlled military machine, and its component "states" depended on one another for consumer goods and other commodities. The collapse of the union and the emergence of independent republics, all competing with one another, and each trying to emerge intact from the devastating economic consequences of the military reductions and the breakup of interdependence, have contributed to the drastic reduction in the standard of living in all of these republics. Ethnic and religious tensions merely compound an already volatile situation.

Now, what have we done to help? The United States, despite the much ballyhooed "peace dividend," has delivered a fraction of the modest amount of aid we have promised. We have provided less aid than several other nations like Germany and Japan. And the aid we were willing to provide was made contingent upon a set of economic restrictions that would have devastating effects on the lives of ordinary Russians. On top of it all, the worst kind of "free enterprise" has sprung up. The country is awash with bubble gum and VCRs, peddled by Russians and foreigners out to make a quick buck at the expense of the long-term health of the country. The economic "reforms" we want to impose as a condition of aid involve changes that are so dramatic and powerful that the entire safety net for the poor, unemployed, and medically dependent would be eliminated. These draconian measures would be inflicted on behalf of an unbridled free-enterprise system that has never been accepted here or anywhere else.

It's time to invest in democracy

In the meantime, we totally ignore the really important aspect of change in the former USSR: democratic reform. It is democracy and freedom that will ultimately create a great Russian nation. How can we support the transition to democratic processes in Russia? For starters, our government could acknowledge the wide range of political and economic views in Russia and not trivialize legitimate dissent from our current favorite in Moscow.

Let the US invest only a small part of the billions we no longer spend on the fraudulent Strategic Defense Initiative, planting seed money in Tula and Vilnius and Yerevan and Magadan rather than flying more bankers to Moscow and St. Petersburg! Our governments may be more comfortable dealing "at the highest levels," but democracy grows from the bottom up. Any aid to the central government should be designed to maintain the same sort of social safety nets that we depend on—programs that grew out of our own crisis of the Great Depression.

In short, let's show that we genuinely value the friendship of the Russian people and recognize their great strengths in natural and human resources. Let's make Russia one of our best friends and trading partners, not a laboratory for economic extremists! Transitional times call for gradualism and compromise, a meeting of minds in the broad middle ground.

—Bill G. Aldridge

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

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

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

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

Advertising:

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

Second class postage paid at New York, NY, and additional mailing ofices. **Postmaster:** send address changes to: *Quantum* Springer-Verlag New York, Inc., Journal Fulfillment Services Department, P.O. Box 2485, Secaucus, NJ 07094-2485.

Subscription Information: North America

Student rate: $15; Personal rate (nonstudent): $20; Institutional rate: $34; Single Issue Price: $5. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus, NJ 07094; or call 1-800-SPRINGER (in New York, call [201] 348-4033).

All Countries Outside North America

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

The pharaoh's golden staircase

Dynamic programming then—and now

by M. Reytman

THE PHARAOH'S PALACE was renowned for its luxury: pearl curtains, walls decorated with amber, gold dishes—one could hardly name all the treasures. But what was particularly striking for those who were allowed into the throne room of the palace was the gold staircase that led to the throne. Figure 1 shows a cutaway view of this staircase with dimensions translated from the ancient Egyptian measurements into decimeters. Archaeologists didn't have to rack their brains to understand why the steps start out so low. The explanation is pretty simple: the aged pharaoh didn't want his subjects to see how hard it was for him to climb up to the throne—they might lose their respect for him, and who knows where that could lead? That's why he ordered his court jewelers to make the steps of the staircase so low—no higher than 1.5 dm.

But time passed and the old pharaoh died. His young son assumed the throne.

The young pharaoh had already heard the courtiers poking fun at his father's naive slyness. And although he would fly up the stairs three steps at a time, the malicious talk continued. Finally the young ruler decided to put an end to the gossip by building up the steps of the staircase. He called for the court jewelers and the treasurer and said:

"My servants! I order you to build up the staircase in such a way that it has no more than four steps. Make them where you want, but there must not be more than four of them!"

"But, sire," the treasurer stepped forward timidly, "where can we find so much gold? Here's an estimate I've jotted down on this piece of papyrus." (And he showed the pharaoh the dotted outline of the staircase in figure 1.) "The staircase is 1 m, or 10 dm, wide. So to build it up we'll need $[3 \cdot 1.5 + 1 \cdot 1 + 1.2 \cdot (1 + 5) + 1 \cdot 3 + 0.8 \cdot (3 + 4) + 1.2 \cdot 11] \cdot 10 = 345$ dm^3 of gold! But, sire, we don't have that much in our treasury, which has been depleted by the war."

"Scoundrel! You want to ruin the country? I can see in an instant that you have never thought seriously about economics! Look!" And the pharaoh drew the red line in figure 1. "You're always ready to spend more

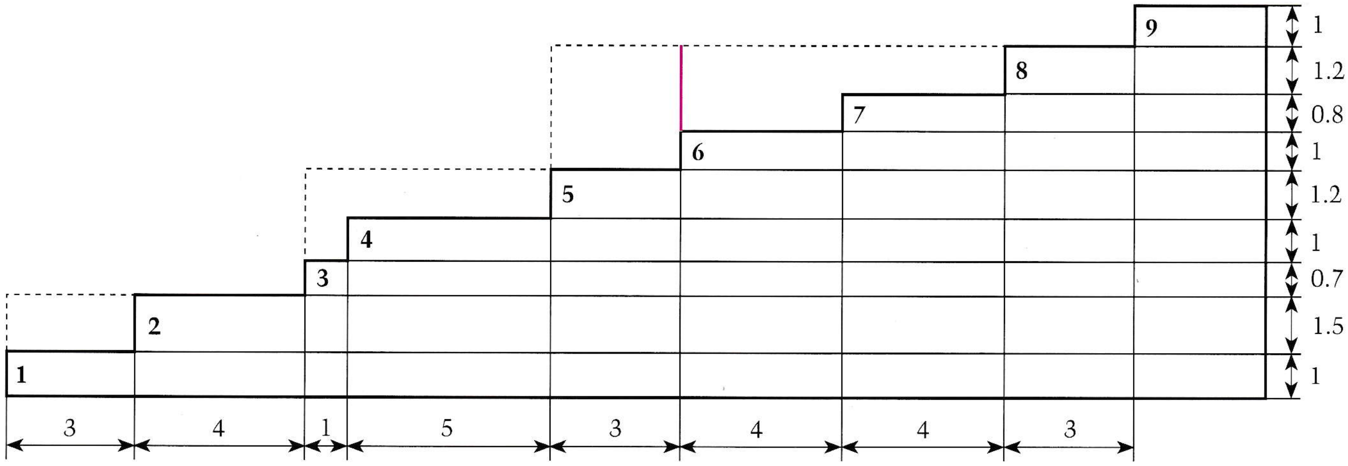


Figure 1

Art by Dmitry Krymov

gold than is actually needed!"

"O wisest of pharaohs! You are right, indeed. But this saves only 90 dm^3, while 345 − 90 = 255 dm^3. We still don't have enough gold."

"You despicable wretch! You cast a shadow upon my might! If the staircase is not ready in seven days, I'll drown you in the holy Nile myself and sell the jewelers into slavery. And don't even think of bringing me those ridiculous drawings again!"

The despondent treasurer left the palace and went to see a friend of his. Nobody was more skilful in the art of designing a pyramid or computing parcels of land. The friend was a priest in the capital's temple, and he was reputed to be a calculating and practical man. Hearing about his friend's misfortune, the priest asked:

"How much gold is left in the treasury?"

"Two hundred dm^3."

"Not much! But maybe you could get by with less, if the steps were well designed."

"I've tried that," the treasurer sighed. "But there are so many possibilities, and I barely have enough time."

"Well, my friend," the priest said, "let me think it over. Come back tomorrow."

The next day, the downhearted treasurer trudged over to see the priest, who met him with a contented smile.

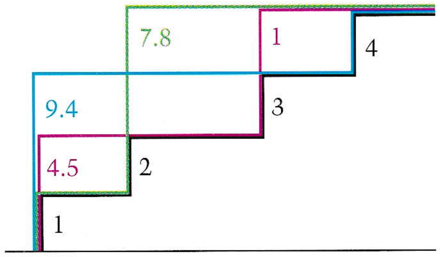


Figure 2

The staircase that requires the least gold is the red one: $f_2(4) = 4.5 + 1 = 5.5$.

"Tell me, my friend, can I count on receiving the remainder of the gold if I arrange it so you need a smaller amount?"

"O gods!" the treasurer exclaimed, not believing his luck. "You'll get a dozen of slaves in addition and as many measures of grain!"

"Then behold!" The priest held out a sheet of papyrus with another plan for building up the staircase. "This plan will require only 171 dm^3 of gold. And the remaining 29 dm^3 are mine!"

"But tell me, O greatest of calculators, how you managed to find this solution? I tried this and I tried that, but this variant didn't occur to me!"

"Then listen," said the priest. "First I examined the cases where a number of the lower steps of the staircase are replaced with one step. So that the numbers are simpler, I'll compute only the area f of the figure between the solid and dotted lines and introduce a constant factor—the width of the staircase—later. You don't mind, do you? To be consistent, I must begin with only one step, the lowest. There's nothing to build up in this case, because we have a single step from the very beginning, so I write

$$f_1(1) = 0.$$

"The formula

$$f_1(2) = 1.5 \cdot 3 = 4.5$$

shows that I need 4.5 dm^2 to replace the first two steps with one step. Similarly we find that replacing the first three steps will cost

$$f_1(3) = 4.5 + 0.7 \cdot (3 + 4) = 9.4 \text{ dm}^2.$$

Then,

$$f_1(4) = 9.4 + 1 \cdot (3 + 4 + 1) = 17.4,$$
$$f_1(5) = 17.4 + 1.2 \cdot (3 + 4 + 1 + 5) = 33,$$

and so on. Each new calculation adds an amount to the previous one."

"I understand all of this," the treasurer interrupted, "but what do I need one step for? The pharaoh wants four! Besides, the original staircase has many more than five steps!"

"Slow down!" the priest smiled. "Soon there will be more new steps—two, as a matter of fact (fig. 2). But before that, we have to find the areas that must be added to make a new step starting from various steps of the original staircase. These areas $g(i, j)$ depend on the numbers i and j of the old steps where the new step begins and ends, and we can compute them one after another. For

four steps	*three steps*	*two steps*	*one step*									
17.1	...	...	...	...	...	61.8	35.8	23.8	11.8	3	9	1
...	24.3	...	...	...	46	42.8	21.8	12.8	4.8	8		1.2
...	14.1	26	65	46.6	27.6	23.6	8.6	3.2	7			0.8
...	8.5	20.4	49	32.4	17.2	14	3	6				1
...	5.5	12.7	33	19.8	8.2	6	5					1.2
0	1	5.5	17.4	7.8	1	4						1
	0	2.8	9.4	2.8	3							0.7
		0	4.5	2								1.5
			1									1
			3	4	1	5	3	4	4	3		

Figure 3

The numbers $g(i, j)$, $1 \le i, j \le 9$, in the shaded portion of the table show how much gold is needed for one new step from the ith to jth old ones, and the numbers $f_k(j)$ in the left part (k = 2, 3, 4) show the minimum amount of gold required for a k-step staircase ending at the jth old step.

$f_2(3) = \min\{\underline{2.8}, 4.5\} = 2.8$
$f_2(4) = \min\{7.8, \underline{1 + 4.5}, 9.4\} = 5.5$
$f_2(5) = \min\{19.8, \underline{8.2 + 4.5}, 6 + 9.4, 17.4\} = 12.7$
$f_2(6) = \min\{32.4, 17.2 + 4.5, 14 + 9.4, \underline{3 + 17.4}, 33\} = 20.4$
$f_2(7) = \min\{46.6, 27.6 + 4.5, 23.6 + 9.4, \underline{8.6 + 17.4}, 3.2 + 33, 49\} = 26$

$f_3(4) = \min\{\underline{1}, 2.8\} = 1$
$f_3(5) = \min\{0 + 8.2, 2.8 + 6, \underline{5.5}\} = 5.5$
$f_3(6) = \min\{0 + 17.2, 2.8 + 14, \underline{5.5 + 3}, 12.7\} = 8.5$
$f_3(7) = \min\{0 + 27.6, 2.8 + 23.6, \underline{5.5 + 8.6}, 12.7 + 3.2, 20.4\} = 14.1$
$f_3(8) = \min\{0 + 46, 2.8 + 42.8, 5.5 + 21.8, 12.7 + 12.8, \underline{20.4 + 4.8}, 26\} = 25.2$

$f_4(9) = \min\{0 + 61.8, 1 + 35.8, 5.5 + 23.8, 8.5 + 11.8, \underline{14.1 + 3}, 24.3\} = 17.1$

Figure 4

instance, building up a new step from the second old one to the fourth requires

$$g(2, 4) = 2.8 + 1 \cdot 5 = 7.8 \text{ dm}^2.$$

"I wrote down all these numbers, which we'll use in our calculations later, in a special table that fits on a single papyrus sheet (fig. 3). The numbers $g(1, j) = f_1(j)$ in the column labeled "one step" have already been calculated, and further calculations are easily done row by row, moving to the right and upward: $g(2, 3)$, $g(3, 4)$, $g(2, 4)$, $g(4, 5)$, … .

"Now I can find the minimum areas for two steps. Since we can't replace one step with two, $f_2(1)$ is meaningless. To obtain two new steps instead of two old steps (the two lower ones), we don't have to add any gold:

$$f_2(2) = 0.$$

"But to find $f_2(3)$ we need to do some calculations—we have to consider two possibilities and choose the best one. Either we build up the second step to the level of the third, which adds the area of $g(2,3)$, or we build the first step to the level of the second ($g(1, 2) = f_1(2)$). What we need is the minimum of the two areas:

$$f_2(3) = \min\{g(2, 3), f_1(2)\} = \min\{\underline{2.8}, 4.5\} = 2.8.$$

"To find $f_2(4)$ we'll have to find the smallest of three numbers (in figure 2 they're shown in different colors):

$$f_2(4) = \min\{g(2, 4), f_1(2) + g(3, 4), f_1(3)\} = \min\{7.8, \underline{4.5 + 1}, 9.4\},$$

and so on." (The complete calculations are given in figure 4.)

"Wait. I've understood how to find the best staircase with two steps, but then the number of possibilities will quickly grow: I'm scared to think how many three-step staircases we have to look at, and we need four steps! Did you actually examine all the possibilities?"

"There was no need to, as you'll see now. Let's find, for example, $f_3(4)$—the staircase with three steps that ends level with the fourth old step and has the minimum area. Let's see where its second step might end. It ends up flush with the second or third old step, naturally. That is, either we must build a new step from the third to the fourth old one, at a cost of $g(3, 4)$, or we replace the first three steps with two. But we already know exactly which of the two-step staircases is the best. It costs $f_2(3) = 2.8$. So,

$$f_3(4) = \min\{g(3, 4), f_2(3)\} = \min\{\underline{1}, 2.8\} = 1."$$

"Wait a minute, I want to make sure I've understood how to find $f_3(5)$. Of three new steps that end flush with the old fifth step, the second can end on the old second, third, or fourth. Therefore,

$$f_3(5) = \min\{g(3, 5), f_2(3) + g(4, 5), f_2(4)\} = \min\{8.2, 2.8 + 6, \underline{5.5}\} = 5.5."$$

"Verily, the wise look after the state treasury," the priest said flatteringly. "In the same way we can compute f_3 of 6, 7, and 8, and then proceed to the case of four steps. Now, we don't have to consider $f_4(4)$, …, $f_4(8)$, because we know for certain that the fourth step must end flush with the old ninth step (otherwise the staircase would have more than four steps). So it remains for us to find $f_4(9)$ as the minimum of the sums $f_3(j) + g(j, 9)$, where j is one of the numbers 3, 4, …, 8. This minimum is equal to $f_3(7) + g(8, 9) = 14.1 + 3 = 17.1$. And this is the best of all four-step staircases!

"This means that the new staircase will require $17.1 \cdot 10 = 171$ dm^3 of gold in all. Now let's see how the new steps should be arranged. We underlined all the optimal choices. The last optimal solution includes the area (which we can think of as the cost) of three new steps $f_3(7) = 14.1$, so the third step should end up level with the old seventh step. Now let's take one step back to the definition of $f_3(7)$ to see that it includes the cost $f_2(4) = 5.5$ of the first two steps. It follows that the second new step must be level with the fourth old step. Finally, in the expression for $f_2(4)$ we see the term $f_1(2) = 4.5$. This means that the first step of the new staircase must end flush with the second step of the old staircase."

The resulting plan for the cheapest new staircase is shown in figure 5.

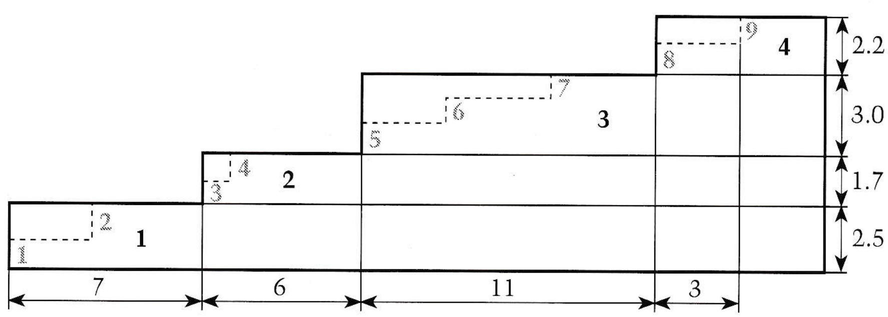


Figure 5

The theory of dynamic programming

I don't know whether the crafty treasurer kept his promise to the priest. I don't even know whether any of this actually took place or if the whole thing is made up. But I do know that it could very well have happened, because we used only two arithmetic operations, addition and multiplication, that were well known in ancient Egypt (although the notations of numbers and operations we use were created much later). And, of course, we used a bit of common sense, too, but this human ability can be found in sufficient quantity anyplace and anytime. However, the mathematical technique applied here gained currency only about forty years ago. It's called *dynamic programming*.

Let's give some thought to how we managed to find the optimal solution without searching through every possible way of building up the staircase (and there are a lot of them—the priest really wouldn't have had enough time to examine each one thoroughly). The idea was to split the general problem at each stage into a number of simpler ones.

Suppose we're going to build up a staircase with N steps so as to make an n-step staircase, where N is greater than n. Assume we already know the optimal arrangement of the steps in the new staircase in the case when their number is fewer than n. Let h_k be the number of the old step level with the kth new step. It's clear that this kth step begins on the $(h_{k-1} + 1)$th old step. Denote the additional area, shown in figure 6, by $g(h_{h-1} + 1, h_k)$. Then the general problem consists of finding the values h_1, h_2, ..., h_{k-1} that minimize the sum of k values $g(1, h_1) + g(h_1 + 1, h_2) + \ldots + g(h_{k-2} + 1, h_{k-1}) + g(h_{k-1} + 1, h_k)$ (for a given $k = n$ and $h_k = h_n = N$). Denote this minimum value $f_n(h_n)$. Notice that to find the minimum value $f_{n-1}(h_{n-1})$ of the first $n - 1$ terms, where h_n doesn't enter, we must solve the same problem for the number of steps minus one. This leads to the following remarkable formula found by the American mathematician Richard Bellman (1920–1984):

$$f_k(h_k) = \min \{f_{k-1}(h_{k-1}) + g(h_{k-1} + 1\ h_k)\},$$

where the minimum is taken over all possible values of h_{k-1} $(k - 1 \le h_{k-1} < h_k)$. This formula allows us to calculate successively the numbers $f_k(h_k)$ for all $k \le n$ and $h_k \le N$. It should be understood as follows: to find the minimum, we must assign all possible values to k, starting with $k = 1$, and for each of them determine and remember the value h_{k-1} that minimizes $f_k(h_k)$. Then, when we reach the last value $k = n$, we have to move backward, picking up all the optimal values h_{n-1}, h_{n-2}, ..., h_1 on the way (see the red arrows in figure 3).

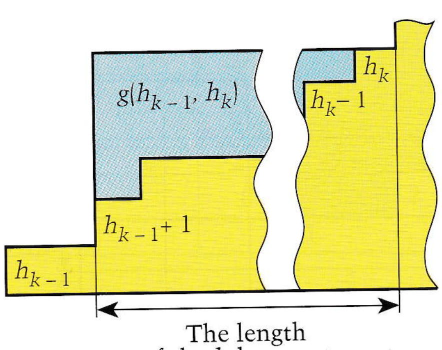


Figure 6

Discussion and a bit of history

An important feature of the aforementioned method is that it's applicable without any special conditions imposed upon the function g. (In this respect dynamic programming compares favorably with so-called linear programming, which requires that all the relations considered be linear and that the variables change continuously.) It's essential only that the function f to be minimized be represented as the sum of terms each of which depends on fewer variables than the next one. So Bellman's equation and the method described above are currently much used, though not for satisfying a pharaoh's whims. Try to get along without it when you are looking for the optimal proportion between the weights of the stages of a rocket, or determining the best way of putting pieces together in a manufacturing process, or creating a schedule with the least possible downtime! You'll find yourself in the desperate situation of the treasurer depressed by the plethora of possibilities. Of course, the calculations required for dynamic programming are hardly ever done "by hand." But this method is well suited for computer processing and is free from the sorts of pitfalls that unfortunately are encountered in linear programming and other methods of locating extrema of functions of many variables.

The idea of dynamic programming emerged a long time ago. We can trace it back to the works of C. Maclaurin (1698–1746), and Archimedes applied it, in a certain sense, much earlier. But it took its final form in Bellman's works and is forever linked with his name.

We saw here how dynamic programming works with a very simple question. To what extent is it applicable to more complex problems? This question must be answered with certain reservations. The problem is that a particular difficulty, called "the curse of dimension," can arise. One has to search through so many variants and memorize so many intermediate results that the method loses its advantageous features. But it has gotten a second wind thanks to parallel computation, which receives more and more attention from programmers and computer designers (for instance, those who work on the Connection Machine at Boston University).

Dynamic programming in molecular biology[1]

Not long ago dynamic programming found a new application—in molecular biology, where data bases contain biological sequences (such as molecules of DNA, RNA, and proteins) numbering many millions of "letters" that cannot be analyzed without computers. Molecules of DNA carrying genetic information

[1]This section was written by Nikolay Vasilyev.—*Ed.*

C	1	2	3	3	4	5	5	6	7
T	1	2	2	3	4	5	5	6	7
G	1	2	2	3	4	4	5	6	6
G	1	2	2	3	4	4	5	5	5
A	1	1	2	3	4	4	4	4	4
T	1	1	2	3	3	4	4	4	4
A	1	1	2	3	3	3	3	3	3
C	0	1	2	2	2	2	2	2	2
G	0	1	1	1	1	1	1	1	1
	A	*G*	*C*	*A*	*A*	*T*	*G*	*G*	*T*

Figure 7

The red squares mark one of the longest subsequences (of length 6).

can be thought of as long series of the four letters A, C, G, T (standing for adenine, cytosine, guanine, and thymine—the chemical groups that form these molecules). In the process of evolution, mutations alter these sequences: an old letter may be replaced by another, it may drop out, or a new letter may appear. How similar are two DNA fragments? What is the smallest number of transformations that turn one of them into the other?

Here is a more precise formulation of the problem. We call a string of letters A, C, G, T a *word*. A *subsequence* of a word is a new word made by striking out some of the letters of the original word. Given two words of length m and n, our problem is to find the longest subsequence that appears in both. A simple procedure proposed in the 1970s solves this problem (biologists call it the Nudelman–Wunsh algorithm).

Write one of the given words—$x = X_1X_2...X_m$—from the bottom up, and the other—$y = Y_1Y_2...Y_n$—from left to right along the sides of a $m \times n$ table (fig. 7). In this example, $m = n = 9$, x = GCATAGGTC, y = AGCAATGGT. We create a function that establishes identity or nonidentity:

$$d_{ij} = \begin{cases} 1, & \text{if } X_i = Y_j, \\ 0, & \text{if } X_i \neq Y_j. \end{cases}$$

$$\begin{aligned} x &= \_\ \boxed{GCA}\ T\ \boxed{A}\ \_\ \boxed{GGT}\ C \\ y &= A\ \boxed{GCA}\ \_\ \boxed{A}\ T\ \boxed{GGT}\ \_ \end{aligned}$$

Figure 8

One of the best alignments of the two "words" in figure 7.

(squares with $d_{ij} = 1$ are colored blue in the table). Let a_{ij} be the length of the longest subsequence we can choose from the opening segments of our words: $X_1...X_i$ and $Y_1...Y_j$. All the numbers a_{ij} ($i = 1, ..., n$, $j = 1, ..., m$) can be found successively by using the following formula (which resembles Bellman's formula):

$$a_{ij} = \max\{a_{i-1,j}, a_{j,i-1}, a_{i-1,j-1} + d_{ij}\}$$

(for $i = 1$ or $j = 1$ the formula is simplified: $a_{11} = d_{11}$, $a_{i1} = \max\{a_{i-1,1}, d_{i1}\}$, $a_{1j} = \max\{a_{1,j-1}, d_{1j}\}$). In other words, a number in each square of the table is the largest of three numbers: the one at the left, the one below, and, if the square is blue, the number at the bottom left plus 1.

Thus, by computing each time the maximum of two or three numbers—and remembering, as in the problem with the pharaoh's staircase, the intermediate results—we'll fill out the entire table in $m \cdot n$ steps.

Verify that the last number a_{mn} will be equal to the length of the longest subsequence common to both given words. The subsequence itself can be read as we move in the reverse direction (sometimes the path might not be unique, however). This reverse path must proceed in "diagonal steps" that increase by one the numbers in the table when we move directly; the squares framed in red represent one of the longest subsequences. Biologists usually depict the result as an "alignment" of the given words (fig. 8). Fast algorithms like this, that don't require much memory, can be found for similar problems. For instance, we can assign a different "weight" to the replacement of a letter compared to its removal (or insertion). With protein sequences (words of 20 amino acid "letters"), it's reasonable to assign each pair of letters a "weight" indicating their degree of likeness (measured, for example, by the frequency of replacements of one letter by another). Another very important problem is that of "multiple alignment" of several words at a time, but the number of operations in this problem quickly grows with the number of words, and more complex heuristic algorithms have to be applied.

Exercises

1. Complete the calculations for the pharaoh's staircase to find the most economical five-step staircase. How much gold will it require?

2. Some years ago a school decided to buy nine different lathes for its workshop at a cost of 10, 20, 40, 60, 75, 100, 130, 160, and 190 dollars. Each more expensive lathe of these nine can replace any cheaper model. Since each lathe needs its own spare parts, the school's principal suggested that nine lathes of only four types be bought, on the condition that they be able to do everything the original nine lathes can. So the math teacher was asked to choose the lathes to be purchased such that the cost would be lowest. What was the math teacher's choice, and how much did the school pay? (This problem is an example of *unification problems*.) ◘

Drops for the crops

Making rain while the sun shines

by Yuly Bruk and Albert Stasenko

A LITTLE GIRL WITH AN INquiring mind asked: "Why aren't raindrops ever as big as my head?" Her friend replied: "Well, if they were, it would be rather awkward to walk with an umbrella."

But the question remains. It certainly might be that the raindrops falling out of the clouds don't have enough time to grow this big from the condensation of water vapor or by combining with other drops when they bump into each other. But if we rigged up a meteorological balloon to dump a bucket of water from a cloud's height, the water won't hit the ground in one piece but in the form of separate drops. (If you don't believe us, you could repeat this experiment from the roof of a tall building.) So something else must determine the final size of raindrops.

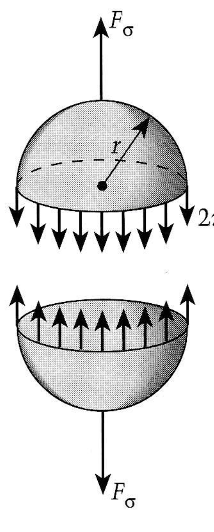


Figure 1

This "something else" is surface tension. In order to tear a drop in half at its "equator" (fig. 1), we need to apply a force F_σ equal to the product of the equator's length $2\pi r$ and a factor σ called the coefficient of surface tension. Its dimensionality is $[\sigma] = \text{N/m} = \text{J/m}^2$. For water, $\sigma \sim$ 0.05–0.08 N/m.

Think of the rubber bladder inside a soccer ball. The tension of the bladder, which at every point is tangential to its surface, produces radial forces directed toward the center of the ball that counterbalance the extra pressure of the gas. In the case of a liquid drop, surface tension results in an extra internal pressure, known as Laplacian pressure. (Of course, this analogy isn't precise. The surface tension in a drop, unlike that in the ball, is practically unrelated to its radius—at least in the range of sizes we'll be using in this article.)

Because of these forces a drop of liquid in weightless conditions (say, in an orbiting space vehicle) retains the shape of a ball. On the other hand, a drop placed on the dry surface of a table flattens out. You can observe a similar phenomenon in nature: a drop of dew on a leaf or flower petal. In exactly the same way a raindrop falling with a uniform velocity is flattened by the force of air resistance and assumes the shape of a hamburger bun.

In short, air resistance deforms the drop, and the other force—surface tension—tries to reshape it into a sphere. Obviously the equality of these two forces gives us an approximation (if only the order of magnitude) of the maximum size of a drop falling with a uniform velocity. The uniform motion of the drop means that the force of air resistance is equal to the drop's weight. Thus,

$$2\pi\sigma r \sim \frac{4}{3}\pi r^3 \rho_w g,$$

where ρ_w is the density of water. Solving for r we obtain

$$r_{\max} \sim \sqrt{\frac{\sigma}{\rho_w g}}.$$

(Naturally we have neglected the factor $\sqrt{6/4}$, taking it to be of the order of 1.) Substituting the numerical data g = 10 m/s^2, $\sigma \sim$ 0.06 N/m, ρ_w = 10^3 kg/m^3 yields

$$r_{\max} \sim \sqrt{6 \cdot 10^{-6}} \cong 2.5 \cdot 10^3 \text{ m} = 2.5 \text{ mm},$$

which seems quite realistic, as you can verify yourself during a summer shower.

Now let's estimate the velocity of the drops. As we mentioned above we should set the weight of the drop equal to the force of air resistance. The aerodynamic force

Art by Pavel Balod


БАЛОД
93

acting on a ball of radius r moving with a velocity v in a medium of density ρ_a is known to be of the order of $F_a \sim \rho_a r^2 v^2$. This formula can be obtained by using dimensional analysis.[1] If we want to change the sign ~ ("of the order of") for an equal sign, we need to determine experimentally the dimensionless factor in this formula. And that's what wind tunnels are for. However, rough estimates are sufficient for our present purposes.

Thus, for the uniform fall of a drop we have

$$F_a \sim \rho_a r^2 v^2 \sim \frac{4}{3}\pi r^3 \rho_w g,$$

from which we get

$$v \sim (4\rho_w r g/\rho_a)^{1/2}.$$

Substituting the maximum radius of the drop obtained above into this estimate yields

$$v_{max} \sim (4 \cdot 10^3 \cdot 2.5 \cdot 10^{-3} \cdot 10/1)^{1/2} \sim 10 \text{ m/s},$$

which also seems realistic, as you can verify by estimating the angles of the tracks left by falling drops on the window of a moving car or train, if you happen to know the vehicle's velocity.

But there is yet another source of resistance to a moving body: the viscosity of the medium. Its existence is clearly seen when you drop a marble in honey (the colder the honey, the slower the descent). As it falls, a certain kind of slow, "creeping" movement appears around the marble. Now, of course, air isn't honey, but if the drops are small, the phenomena are similar. In these cases the surrounding medium (honey or air) can be thought of as sticking to the moving body: layers of liquid farther from the body slide more quickly, so that a tangential force of friction F_μ arises between them, like when you rub your palms together.

As a result, a force acts on the moving body that is proportional not to the square of its radius and velocity but simply to its radius and velocity. The formula for this force was obtained in the last century by Sir George Gabriel Stokes:

$$F_\mu = 6\pi\mu r v,$$

where μ is the coefficient of viscosity of the medium. For air, $\mu \cong 2 \cdot 10^{-5}$ N · s/m². Comparing $F_a \sim (rv)^2$ and $F_\mu \sim rv$, it's easy to see that as the product rv decreases (that is, as either or both factors decrease), the aerodynamic force F_a decreases more quickly than the viscous force F_μ, so that the second becomes prevalent for small bodies and slow movement. Equating the viscous force and the weight yields the velocity of the slow descent for small drops:

$$v = \frac{2}{9}\frac{\rho_w g r^2}{\mu} = \alpha r^2,$$

where $\alpha = (2/9)(\rho_w g/\mu)$ is a constant.

Figure 2 shows both relationships: a parabolic function for small drops, a square-root function for larger ones. Obviously both curves have a common point at a certain value of the radius $r = r_*$, thus providing a physical basis for using the term "droplet" for drops whose radius $r < r_*$ and the term "drop" for drops characterized by $r > r_*$. Both forces are equal where the lines intersect: $F_a/F_\mu = 1$. The relation on the left-hand side is known as the Reynolds number Re, which plays an important role in aerodynamics. It's clear that microbes and droplets of fog have Re < 1, while airplanes, ships, and raindrops in a downpour have Re > 1.

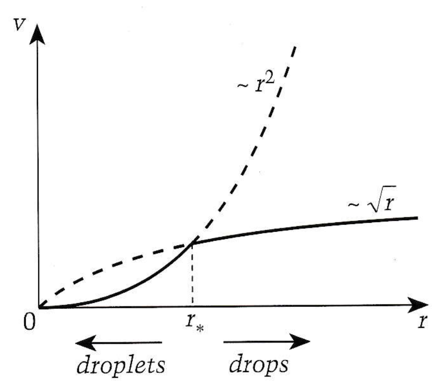


Figure 2

But is there any practical use in all these cogitations? Remember the words of Winnie-the-Pooh: "The only reason for being a bee that I know of is making honey. And the only reason for making honey is so as *I* can eat it."

So, what are raindrops for? For watering our gardens! At least, that's the only reason that *we* know of.

But we can't always count on rain everywhere. Sometimes we have to take matters into our own hands. We can do it with pails and watering cans, of course—"on foot," so to speak. Or we can drive a tractor between the rows, dragging a tank of water and a sprayer. But if we need to water huge expanses of land, it's more practical to use the wind that blows across the great open expanses of land almost constantly.

Let's set a horizontal pipe with sprayers at a height h above the ground (fig. 3). Water is supplied under pressure and forced out in the form of drops flying in all directions. The radii of the drops may differ, but we'll look at only those drops that can be considered small by the criteria we established above (that is, flattened according to Stokes's law) because only small droplets can be carried by the wind over large distances.

Let the wind blow with a uniform velocity u perpendicular to the pipe. It can be shown that for any initial velocity of a droplet, the droplet is quickly drawn into a uniform horizontal movement with the velocity of the wind (a physicist would say that the relaxation time is short for small droplets). Therefore, each droplet has a constant velocity u in the horizontal direction and a con-

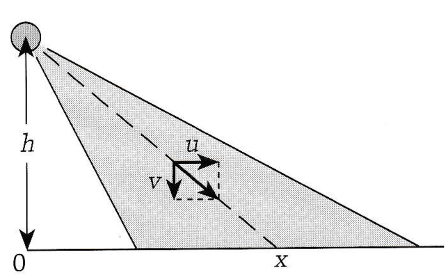


Figure 3

[1]See "The Power of Dimensional Thinking" in the May/June 1992 issue of *Quantum*.—*Ed.*

stant velocity $v = \alpha r^2$ in the vertical direction.

Since both components of the velocity are constant, all the droplets move in straight lines. It's clear that a smaller droplet will travel farther along the x-axis, but will carry less water to irrigate the soil. If we want to irrigate a field uniformly, which is usually the case, then the number of small droplets should be greater than the number of large droplets. And here a reasonable question arises: what is the relationship between the numbers of droplets of different sizes? In the language of physics, what is the distribution function of the droplets according to size that results in uniform irrigation?

All these words are more or less clear by intuition, but we'll try to formulate the problem in a more precise way. What does "uniform irrigation" mean? It means that any small patch in the field of length L (which is also the length of our pipe) and of width Δx (therefore, of area $L\Delta x$) gets an amount of water ΔM such that the ratio $G = \Delta M/L\Delta x$ is constant. We'll call the constant G the irrigation surface density.

And what does "distribution of droplets according to size" mean? Let's take some interval in radius Δr and denote by Δn the number of droplets with radii in this interval. The ratio

$$f(r) = \frac{\Delta n}{\Delta r}$$

is the droplet size distribution function we're looking for.

In the same way one can find the distribution function for students in a college according to age, height, or eye color (if it's expressed as a wavelength of visible light); the molecular velocity distribution function for a gas (the formulas found by Boltzmann and Maxwell); the distribution function for the quanta of solar radiation according to their frequencies (Planck's law); the distribution of electrons in a solid according to their energies (Fermi statistics); and so on.

But let's get back to the problem of uniform irrigation. The time it takes a droplet of radius r to fall to the ground from a height h is $t(r) = h/v(r) = h/\alpha r^2$. The horizontal displacement of the droplets is

$$x = ut = \frac{uh}{\alpha r^2}.$$

What is the width Δx of the area irrigated by droplets whose radii deviate from the given value r by a small amount Δr? This can be obtained by differentiation or by taking the difference in x-values for values of r and $r + \Delta r$:

$$\Delta x = -\frac{2uh}{\alpha r^3}\Delta r.$$

Here the minus sign reflects a fact mentioned earlier: that a decrease in the droplet radius ($\Delta r < 0$) results in a greater travel distance ($\Delta x > 0$), and vice versa.

Now let's see how much water lands on the area $L\Delta x$. The number of droplets with radii varying from r to $r + \Delta r$ is $\Delta n(r)$. We can make Δr so small that the masses of the droplets in this group can be considered identical and equal to $m(r) = \frac{4}{3}\pi r^3 \rho_w$. Then the mass of water carried in these droplets is

$$\Delta M = m(r)\Delta n(r) = \frac{4}{3}\pi r^3 \rho_w \Delta n(r).$$

From this it follows that

$$G = \frac{\Delta M}{L\Delta x} = -\left(\frac{2\pi\rho_w\alpha}{3Luh}\right)r^6\frac{\Delta n}{\Delta r}.$$

All the constant parameters in the problem are collected in the parentheses. But the ratio $\Delta n/\Delta r$ is the desired distribution function for the droplets. Thus,

$$f(r) = \frac{\Delta n}{\Delta r} \sim \frac{1}{r^6}.$$

This is a very steep function. Let's consider two equal intervals of droplet radius Δr near two values r_1 and $r_2 = 2r_1$. Then we can draw the following conclusion from the last formula: within this interval Δr the number of small droplets $\Delta n(r_1)$ must be $(r_2/r_1)^6 = 2^6 = 64$ times the number of large droplets $\Delta n(r_2)$.

So we have found the distribution function that a designer should use to make a proper choice of sprinklers and a procedure for using them. And it doesn't just apply to irrigation but to other applications as well—for instance, the uniform distribution of pesticides.

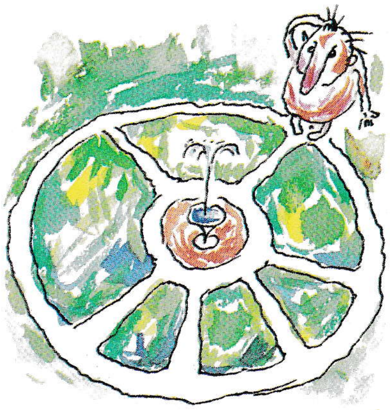
Now, what have we failed to take into account in our solution? Lots of things. For instance, the fact that the wind doesn't always blow with a uniform velocity, or that its velocity changes with height above the ground, or that the droplets (particularly the smaller ones that travel farther) can change due to evaporation; and so on. Any physical model has its limits within which it is true. Outside those limits it requires corrections and complications. And sometimes it stops working altogether.

Then again, look at how many fine physicists came to mind just because we wondered how best to water our garden. Now there's a rich harvest for you! ◘

Just for the fun of it!

B106

Archeometics. Each square in the row in the figure had a number inscribed in it such that the sum of any three numbers in succession was fifteen. Then all the numbers but two were erased. Restore the erased numbers.

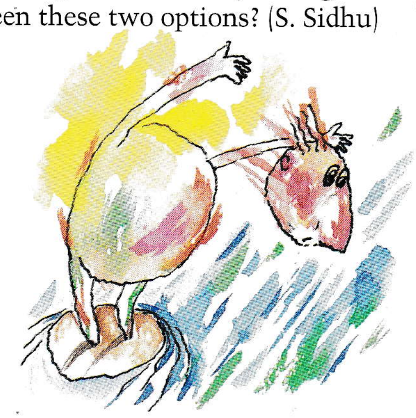


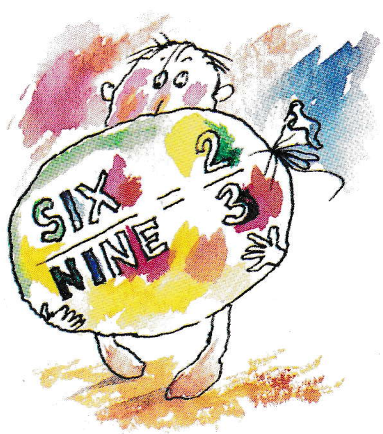
B107

Around the garden. A garden has two concentric circular paths with many radial walkways connecting them. You are standing on the outer circle and want to go to another point on the same circle. If the latter point is within the same quadrant, it's apparent that walking along the outer circle is the fastest way to get there. But for a destination that's almost directly across from where you are, it seems advantageous to first get to the inner circle along a radial path, walk along the inner circle, and get back to the outer circle by the nearest radial path. Assuming that there is radial path wherever you might need one, where is the "dividing line" between these two options? (S. Sidhu)

B108

Circles on water. A stone thrown into still water generates ripples that propagate outward as circles. What shape will the ripples take if the stone is thrown into the flowing water of a river? (S. Krotov).



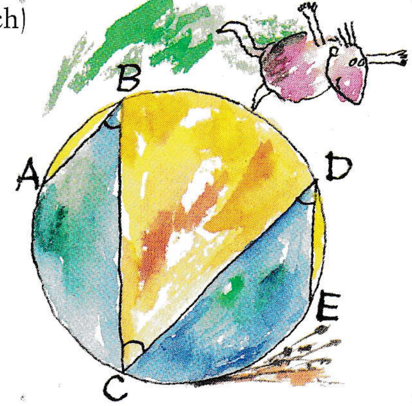


B109

Wonderful simplification. Solve this number rebus:

$$\frac{SIX}{NINE} = \frac{2}{3}.$$

(Different letters stand for different digits, identical letters stand for identical digits.) (P. Filevich)



B110

Circle-halving zigzag. All the vertices of a polygonal line $ABCDE$ lie on a circumference (see the figure), and the angles at the vertices B, C, and D are each 45°. Prove that the area of the blue part of the circle is equal to the area of the yellow part. (V. Proizvolov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 56

Art by Pavel Chernusky

Light at the end of the tunnel

Some things that come to an end

by Dmitry Fomin and Lev Kurlyandchik

IF YOU LIKE MIXING BUSINESS with pleasure (in this case, the business of learning mathematics), then get yourself a bag of candies and invite a friend to play this game. Make two piles of candies—12 pieces in one pile and 13 in the other. Each player in turn makes a move consisting either of eating two candies from either of the piles or moving a candy from the first pile to the second. The player who can't make a move loses.

If your resources allow you to play this wonderful game long enough, you'll observe a strange regularity: *the player who begins the game always loses*! However, the reason for Lady Luck's bias is easy to understand. Each move changes by 2 the difference between the numbers of candies in the two piles. So the remainder of this difference upon division by four will change in a strictly defined way: 1 (= 13 – 12), 3, 1, 3, 1, 3, … . We see that before every move made by the second player the remainder is three. But the game stops only when the candies are eaten up (the remainder is zero), or there is exactly one candy left and it's left in the second pile (the remainder is 1). So the second player is never exposed to the danger of losing.

Maybe you're already familiar with this line of reasoning. Remember "Some Things Never Change" (in the September/October 1993 issue of *Quantum*)? *An invariant is something that doesn't change.* And in our game every other remainder is the same. However, the problem is not yet *completely* solved. We've only proved that the second player can't lose. But does she or he necessarily win? In other words, does the game necessarily end?

It's not hard to come up with the answer to these particular questions (which is, of course, yes). But there are much more interesting problems of this sort. In this article we'll describe one method for solving them.

To start us off, here's an old problem (proposed at the First All-Russian Math Olympiad in 1961).

Problem 1. (A. Schwarz) *Real numbers are written in an $m \times n$ array. It's permissible to reverse the signs of all the numbers in any row or column. Prove that after a number of these operations we can make the sum of numbers along each line (row or column) nonnegative.*

Let's see what happens to the sum of all the numbers in the array after one operation. If the sum along the chosen line (where the signs have been changed) was negative, the total sum increases; if the sum along the line was positive, the total decreases; and if the sum along the line was zero, the total remains the same. So, if there is a line with a negative sum, we can increase the total sum by applying our operation to this line.

But is it possible that the total sum increases infinitely many times?

Of course not! Indeed, our operations can produce only a finite number of tables, because each of the mn entries can take only two values (differing in their sign), and so all in all there could be no more than 2^{mn} different tables.[1] Therefore, the total sum of the numbers in our array can take only a *finite* number of different values.

Now let's look at the original array. Choose a line with a negative sum (if there are no such lines, we're already done). Apply the reverse-sign operation to this line. Again find a line with a negative sum in the array thus obtained, and so on. Note that the total sum (1) increases every time we apply our operation; (2) takes only a finite number of values. Therefore,

[1]In fact, the number of tables obtainable from a given table by our operations is significantly smaller. Since any two applications of our operation to the same line cancel out, any attainable table can be obtained by reversing signs in each line no more than once. There are $m + n$ lines, so the number of tables is not greater than 2^{m+n}. Actually, it's even smaller, because different sets of operations can produce the same result. You may want to verify that the exact number of tables produced from a given table is 2^{m+n-1}.—*Ed.*

Art by Alexander Gazhur

sooner or later the process must come to an end. This means that at some point we'll be unable to find a line with a negative sum—that is, all the sums along lines will have turned nonnegative, so we will have arrived at a required table.

The next problem has nothing in common at first glance with the one we've discussed. However . . .

Problem 2. *Given are n points, no three of which are collinear, and n lines, no two of which are parallel, in the plane. Prove that we can drop a perpendicular from each point to one of the lines, one perpendicular per line, such that no two perpendiculars intersect.*

Let's start by drawing perpendiculars—one from each point and one to each line—in an arbitrary order. If no two of them intersect, they comply with the requirement in the problem.

Otherwise, take two intersecting perpendiculars AA_1 and BB_1 dropped from points A and B to lines α and β, respectively (fig. 1). Let P be their common point. Replace AA_1 and BB_1 with the perpendiculars AA_2 and BB_2 dropped from A and B to β and α, respectively. Then the sum of the lengths of the perpendiculars will decrease. Indeed, it's clear that $AA_2 \le AB_1 \le AP + PB_1$, and both these inequalities become equalities only if AB_1 is perpendicular to β. Similarly, $BB_2 \le BP + PA_1$. Adding these inequalities and taking into account that AB cannot be perpendicular to both lines α and β at the same time, we get $AA_2 + BB_2 < (AP + PA_1) + (BP + PB_1) = AA_1 + BB_1$.

Now let's do the same thing as in the first problem. Take the initial arbitrary set of perpendiculars. Choose two intersecting perpendiculars (if there are any) and apply our operation: replace them with two perpendiculars having a smaller sum of lengths. Find two intersecting perpendiculars in the new figure, apply our operation again, and so on. Note these two properties of the sum of the lengths of all perpendiculars: (1) the sum decreases at every step of our process; (2) the sum can take only a finite number of values (why?). It follows that our unbraiding process can't go on forever. But once it stops we'll have arrived at the required configuration, because it will be impossible to find any more intersecting perpendiculars in it.

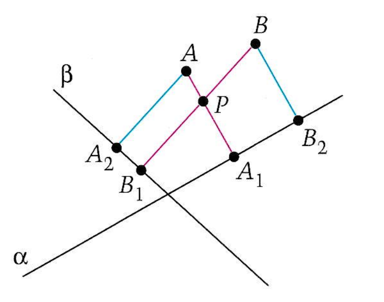


Figure 1

Let's analyze the two solutions. We took the same approach both times: we introduced a certain value (the sum of all the numbers in the array in problem 1, the sum of the lengths of all perpendiculars in problem 2) and an operation whose application each time changes the value in the same way (increases it in the first problem, decreases it in the second). The solution was based on the fact that the value we introduced could take only a finite number of values. Consequently, the operation could be applied only a finite number of times, after which we inevitably arrived at a required situation.

From this point of view, the second problem is more difficult than the first, because to solve it we not only had to invent the unknown value to keep track of, we had to invent the operation as well.

Also, the second problem is an excellent example of how easy it is to get off track: the new perpendiculars AA_2 and BB_2 don't intersect, so it seems reasonable to consider the total number of intersections of all n perpendiculars in our configuration, which appears to decrease upon each application of the described operation. However, this is not so (try and come up with a counterexample). It's far from easy to learn how to create an appropriate pair "operation–value"—it requires experience and insight.

There is no standard mathematical term for a quantity that changes monotonically and takes a finite number of values. We'll call it a *monovariant.*

The next problem can also be considered widely known.

Problem 3. *Prove that any 2n points in the plane can be viewed as the endpoints of n disjoint segments.*

First, we draw an arbitrary n segments joining n pairs of the given points. If any two of them are disjoint, we're done. Otherwise, take a pair of intersecting segments AB and CD (fig. 2a). The operation will consist of replacing these segments by the disjoint segments AC and BD.

It remains to find a monovariant—some value that is changed monotonically by our operation. The sum of diagonals AB and CD of a convex quadrilateral $ACBD$ is greater than the sum of opposite sides AC and BD: in figure 2a, $AB + CD = (AO + OB) + (CO + OD) = (AO + OC) + (BO + OD) > AC + BD$. The reader can check that this inequality remains true in the degenerate case of figure 2b as well. So we can take as a monovariant the sum of the lengths of all n segments. This sum can take only a finite number of values, because there is a finite number of possible segments. Just as in the previous problem, these operations will lead us to a set of n disjoint segments.

Originally in this problem we had neither an operation nor a monovariant. But once we came up with an operation, it was no problem to pick

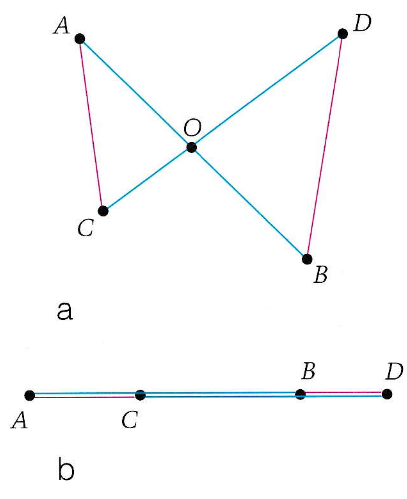


Figure 2

out a proper monovariant.

Problem 4. (V. Alexeyev) *Several numbers are arranged around a circle. If four consecutive numbers a, b, c, and d satisfy the inequality* $(a-d)(b-c) > 0$, *we can exchange b and c. Prove that we can perform this operation only a finite number of times.*

Here the operation is given from the very start. Although it involves only two numbers, it's convenient to think of it as an operation on the whole set of numbers by assuming that all the numbers except b and c remain unchanged: in accordance with our "solving philosophy" any monovariant we introduce must depend on the whole set of numbers.

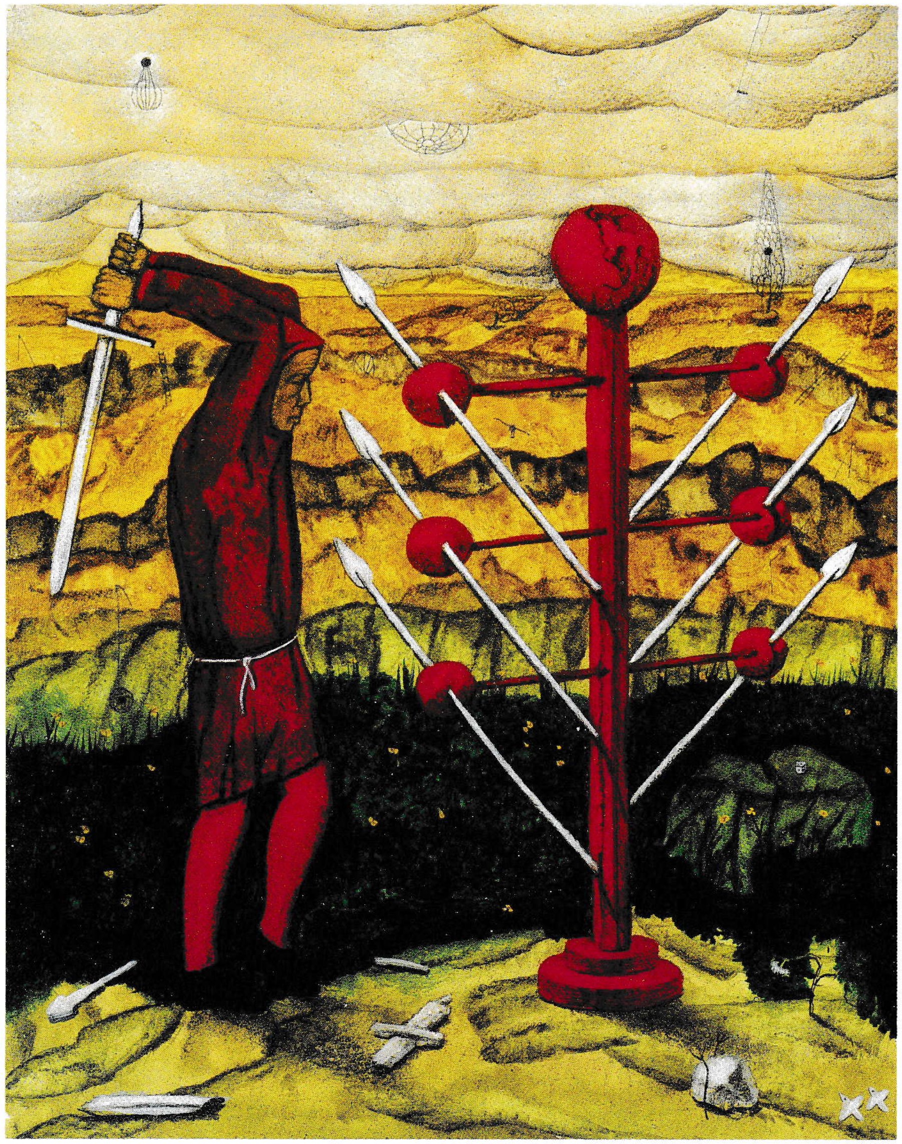
So, suppose we can apply our operation to the numbers a, b, c, and d—that is, $(a-d)(b-c) > 0$ or $ab + cd > ac + bd$. The operation turns the quadruple a, b, c, d into a, c, b, d and, by the above inequality, decreases the sum of the products of neighboring numbers: $ab + bc + cd > ac + cb + bd$.

Now it's clear that the required monovariant can be defined as the sum of products of all pairs of neighboring numbers around the circle. Our operation decreases this monovariant, and since our sum of products can take only a finite number of values (why?), the operation can be applied only a finite number of times.

The next problem was borrowed from *Problems in Plane Geometry* by V. Prasolov (in Russian).

Problem 5. *A nonconvex polygon is subjected to the following operation: if it lies on one side of a line AB joining its nonadjacent vertices A and B, then one of the parts into which the perimeter of the polygon is divided by A and B is reflected about the midpoint of the segment AB—that is, it's rotated 180° about the midpoint (fig. 3). Prove that after a number of such operations the polygon becomes convex.*

Here again, the operation is specified by the statement of the problem. The monovariant quantity is equally obvious—it's the area of the polygon. The fact that the area can take only a finite number of values is less obvious. To prove it, consider the vectors joining consecutive vertices of the polygon in, say, the clockwise direction. Our operation preserves this set of vectors, although it rearranges them around the perimeter. We can see that every polygon that may arise, starting from the given one, is completely defined by the order of the vectors. But there is a finite number of different orders, and therefore, of polygons and their possible areas.[2]



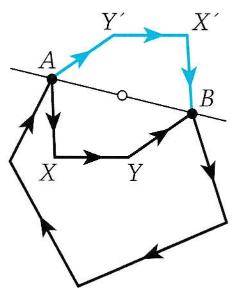


Figure 3

Now here's a problem proposed for eighth-graders at the All-Union Olympiad in 1979. Only three participants managed to solve it.

Problem 6. *Each member of a parliament has no more than three enemies among other MPs.[3] Prove that the parliament can be split into*

[2]It's interesting that this proof doesn't work for reflection about the *line AB* instead of the midpoint of *AB*—simply because the above set of vectors is changed by line reflection. —*Ed.*

[3]In this and the next problem it's assumed that if B is an enemy of A, then A is an enemy of B.

two houses such that each MP has no more than one enemy in the same house.

As we did before, let's first divide the parliament into two houses in an arbitrary way. If each member of parliament has no more than one enemy in the house where he or she belongs, it's the required partition. Otherwise, there's a member of parliament A who has at least two enemies in his or her house. Then A has no more than one enemy in the other house. So if we transfer A to the other house, *the number of pairs of enemies that belong to the same house* will decrease. This means that we can take this number as a monovariant. The proof can now be completed by the ritual phrase, "The monovariant can take only a finite number of values."

By the way, in this problem we can make sure of the latter statement almost without looking into the situation in question: it suffices to note that our monovariant is a positive integer (a sequence of positive integers cannot decrease forever).

The following problem was offered at the 1964 Moscow City Olympiad. It proved to be so difficult that none of the competitors could solve it.

Problem 7. *King Arthur summoned 2N knights to his court. Each knight has no more than $N - 1$ enemies among the knights present. Prove that Merlin can seat the knights at the Round Table in such a way that no two enemies will sit next to each other.*

Let's seat the knights at the Round Table arbitrarily. If the requirement of the problem is not yet satisfied, we can find two enemies, A and B, sitting side by side. For definiteness, let's assume that B sits to the right of A (fig. 4a).

The problem itself suggests the value to be tried as a monovariant: the number of pairs of enemy-neighbors at the table (compare this with the previous solution).

Consider the friends (that is, knights who aren't enemies) of A. One of them, C, must have B's friend as his right-hand neighbor—we'll call him D. (Otherwise, B would have more than $N - 1$ enemies, because A has no less than N friends.) Now let's reverse the order of knights sitting between A and D to the right of A (fig. 4b). Then C becomes the neighbor of A, B becomes the neighbor of D, while all the other pairs of neighbors remain the same (fig. 4c). Therefore, our hypothetical monovariant really does decrease, because the pairs (A, B) and (C, D)—at least one of which, (A, B), consists of enemies—are replaced with two "friendly pairs" (A, C) and (B, D). Now we can complete the proof in the usual way.

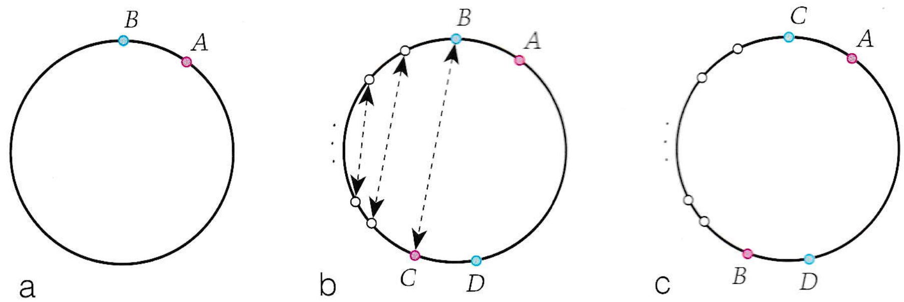


Figure 4

A natural question arising in a problem on monovariants is how long the process considered in the problem will continue. Examine all the problems we've discussed from this point of view. You'll see that it's usually not that easy.

We'll give just one example.

Problem 8. *Written on the blackboard is a string of N numbers, each equal to +1 or −1. A "move" consists of reversing the signs of several consecutive numbers. What minimum number of moves is sufficient to obtain a string of +1's from any initial arrangement of signs?*

The operation here is given by the condition. Let's count the pairs of neighboring numbers with opposite signs and take the result as the monovariant. Our operation changes it by no more than two.

Let's prove that the string of N numbers $-1, +1, -1, +1, \ldots$ cannot be turned into $+1, +1, +1, \ldots$ in less than $n = [(N + 1)/2]$ moves, where $[x]$ is the greatest integer not exceeding x. After performing k moves, we change our monovariant by no more than $2k$. The change will be even less if at least one of the moves involves any of the numbers at the ends of the string, because such a move changes the monovariant by one (or even makes no change at all, if it reverses the signs of all the numbers). Turning the given alternate string into the constant string of +1's, we necessarily make such a "small change" at least once—when we change the sign of the leftmost −1. So if we manage to do this transformation in $k < n$ moves, we'll have changed the monovariant by *less* than $2k \le 2(n - 1) \le 2((N + 1)/2) - 2 = N - 1$. But this change must be *exactly equal* to $N - 1$, because the initial value of the monovariant is $N - 1$ and the final value is zero. So the minimum number we seek is *not less than n.*

In fact, it's equal to n. To show this, take any initial string and pick out groups of consecutive −1's. If k is the number of these groups, then there are at least $k - 1$ numbers +1 between the groups, so $k + (k - 1) = 2k - 1 \le N$, and the number of the groups $k \le (N + 1)/2$. Since k is an integer, $k \le [(N + 1)/2] = n$. Now, successively changing the signs in the groups, we'll make all the numbers positive in $k \le n$ moves.

So n moves always suffice, while fewer than n moves might not suffice.

Turning back to our candy game, we notice that it also has a monovariant—for instance, the expression $2a + b$, where a and b are the numbers of candies in the first and second piles, respectively. In particular, this means that the game can't go on forever.

Exercises

Most of these exercises are from All-Union and Leningrad Math Olympiads, 1973–87.

1. N red and N blue points lie in the plane, and no three of them are

collinear. Prove that one can draw N non-intersecting segments joining red points to blue points.

2. A number of red and blue points are given, some of them joined with segments. We'll call a point *special* if more than a half of the points joined to it are a different color. We choose a special point (if there are any) and paint it the other color. Prove that after a number of such steps there will not be any special points left.

3. N points lie in the plane, some of them joined by segments. There are no more than 11 segments issuing from any point. Prove that these points can be painted in four colors in such a way that there are no more than N segments with endpoints of the same color.

4. Each face of a cube has a number written on it, and not all the numbers are the same. Each of the numbers is replaced by the arithmetic mean of the numbers written on the four adjacent faces. Is it possible to obtain the initial numbers on the faces again after a number of these operations?

5. N positive integers are written around a circle. The greatest common divisor of every two neighboring integers is written in between them. Then the former numbers are erased, and the newly written numbers are subjected to the same operation. Prove that after a number of steps all the numbers will be equal.

6. A 1 and nine 0's are written on the blackboard. Each of any two of these numbers can be replaced by their arithmetic mean. What smallest number can appear in place of the number 1 after a series of such operations?

7. Finitely many squares of an infinite square grid drawn on white paper are painted black. At each moment in time $t = 1, 2, 3, \ldots$ each square takes the color of the majority of the following three squares: the square itself and its top and right-hand neighbors. Prove that some time later there will be no black squares at all.

8. Several children are standing around a circle, each holding a handful of candies. On a signal from the leader each player hands over a half of his or her candies to the neighbor on the right (if the number of candies is odd, the leader gives one additional candy to the player). Prove that after a certain number of rounds all the children will have the same number of candies.

9. All the volumes of the *Encyclopaedia Britannica* stand in an arbitrary order on a special shelf in a library. Every minute, a robot-librarian takes an arbitrary volume that is not in its proper place and puts it in its proper place—that is, if the number of the volume is k, the robot puts it into the kth place. Prove that sooner or later all the volumes will be standing in their proper places.

Editor's note: You can find more about some special methods of solving various types of problems in these *Quantum* articles: "Pigeons in Every Pigeonhole," January 1990 (the pigeonhole principle); "Going to Extremes," November/December 1990 (the "extremity" rule); "Off into Space," January/February 1992 (looking at plane geometry problems from a three-dimensional viewpoint); "Jewels in the Crown," July/August 1992 (mathematical induction); "Some Things Never Change," September/October 1993 (invariants).

The problems involving monovariants are related to the problems involving invariants mentioned at the beginning of this article: both kinds of problems deal with the result produced by a sequence of certain operations. But the problems considered in this article are even closer to those studied in the article "Going to Extremes" mentioned above. For instance, in problem 3 we could choose to apply the "extremity" rule introduced in that article and consider the set of segments with the minimum total length; then the same basic argument as the one we used ensures that the segments of this set do not intersect. Something similar can be done with other problems that do not directly refer to any operations. The advantage of monovariants is that they not only allow us to prove that a certain configuration exists, they also show how to construct it.

ANSWERS, HINTS & SOLUTIONS ON PAGE 58

What little stars do

And the big old planets don't

by Pavel Bliokh

TRY THIS LITTLE PARLOR trick sometime: you say a word and your friend says an action that comes to mind. Then you show a sheet of paper with the responses already written on it, and in most cases they're the verbs your friend has uttered—you're a mind reader! You say "a dog," your friend says "barks," you say "a bird," she says "flies," you say "a fish," she says "swims." And if you say "a star," then most likely you will hear "twinkles," because the beautiful sight of twinkling stars, or sparkling lights off in the distance here on Earth, is embedded in our consciousness from the time we were babies: "Twinkle, twinkle, little star . . ."

The firmament with its stars whose sparkle constantly changes and pulses with all the colors of the rainbow is indeed a beautiful sight. We can admire it unquestioningly for a long time. But sooner or later the thought occurs to an inquisitive mind: "Why *do* they twinkle?"

Twinkling facts

A star can change its brightness in two ways: either the amount of light emitted by the star varies (just as candlelight varies on a windy day) or the light changes during its long journey to Earth. Let's look at some facts related to the twinkling of stars (you can verify most of them yourself).

1. The twinkling of stars depends both on the weather and on your location. In good weather in the mountains, the stars hardly twinkle. This is all the more true when you look at the stars from an airplane flying at high altitude, and the brightness of stars seen by astronauts doesn't vary at all.

2. Stars near the zenith (straight up) twinkle noticeably less than those near the horizon. The brightness of far-off lights here on Earth also varies to a considerable extent, but these same lights give off a steady glow (without twinkling) when they are nearby.

3. The intensity of the twinkling depends on whether we're looking at the heavens with the naked eye or through a telescope. If the diameter of the objective is more than a few dozen centimeters, the brightness of a star does not seem to change, although its image quivers—that is, oscillates chaotically around a certain average position.

4. If we look hard at the heavenly bodies, we notice certain "stars" that do not twinkle at all. These are not stars but planets. The best way to convince ourselves that planets don't twinkle is to look at Venus—the most brilliant object in the heavens after the Sun and Moon. Venus can be seen either in the morning in the east or in the evening in the west, but it's best to know where to look ahead of time. You can find this information in an astronomical handbook.

5. When the stars are near the horizon, you can see chromatic twinkling: more or less rapid changes not only in brightness but in color as well.

The features noted in items 1 and 2 show that the cause of the twinkling should be sought not in the physical characteristics of the stars but in the optical properties of the atmosphere. The blanket of air is very thin compared to the Earth's radius $R_{\oplus} = 6{,}500$ km. About 50% of the mass of the air lies below an altitude of 6 km, and at 30 km this rises to almost 99%.

A beam of light coming from a star near the horizon travels a far greater distance through the lower, dense layers of the atmosphere than a beam traveling from straight overhead (fig. 1 on page 24). The longer path of light in air is a strong argument in favor of the atmospheric nature of the twinkling, although it explains nothing by itself. After all, pure air is completely transparent, but twinkling is observed in clear weather.

Art by Sergey Barkhin

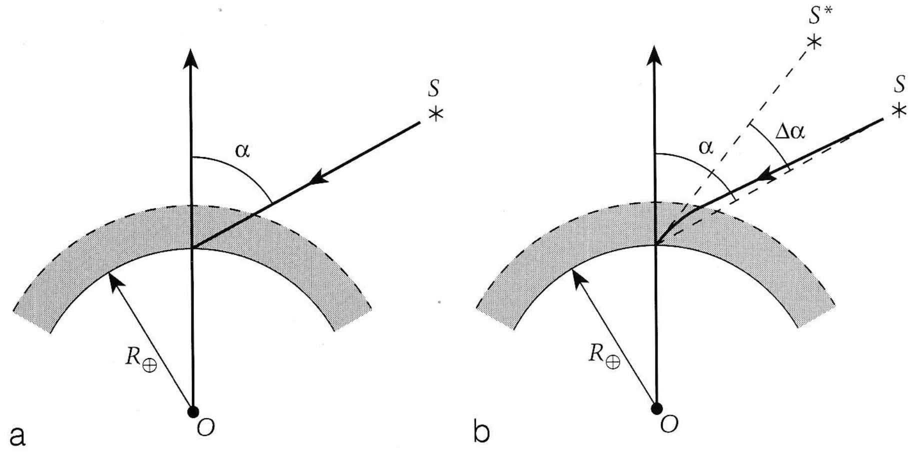


Figure 1

Light beam in the Earth's atmosphere (a) without refraction and (b) with refraction. S is the light source and S is the apparent position of the source.*

Atmospheric refraction

The optical properties of the atmosphere are characterized by the refractive index n, which is always a trifle greater than one and depends on the density of the air ρ:

$$n = 1 + k\rho, \tag{1}$$

where k is a coefficient of proportionality. The quantity $k\rho$ is very small—of the order of 10^{-4}—but it is these ten-thousandths that distinguish "absolutely transparent air" from an empty medium: the vacuum.

Let's recall that the refractive index shows how much the speed of light c_n is reduced in the given medium in comparison with its velocity in a vacuum c: $c_n = c/n$. The difference between c_n and c is so small that it would seem it can only be measured in especially subtle experiments that make it possible to detect a negligible decrease in the speed of light.

And that would be the case if the air were absolutely homogeneous. But as we mentioned earlier, the density ρ decreases with altitude, and the refractive index decreases with it. Near the Earth's surface $n \cong 1.0003$, but at a height of 10 km $n \cong 1.0001$. So the Earth's atmosphere is an optically heterogeneous medium that bends light rays.

This phenomenon (refraction) wasn't taken into consideration when we drew the rays in figure 1a. It *is* taken into account in figure 1b. Clearly the path of the light beam in the atmosphere increases slightly, and—what is especially important—the angle at which the beam arrives at the observer changes.

Due to refraction all the heavenly bodies appear to be elevated above the horizon. The angle of refraction $\Delta\alpha$ is greater when the light source is closer to the horizon. The maximum value of $\Delta\alpha$ is about 30′, which is not insignificant. For example, the angular diameter of the Sun is 30′. Thus, when the Sun sets, we watch it until the upper edge of the disk drops below the horizon by a half-degree. The same thing happens at sunrise: the Sun becomes visible a little before the "true" sunrise. As a result, daytime is prolonged somewhat (by 8–13 minutes in the middle latitudes), and the polar night is reduced by several days.

We've been talking about the so-called regular refraction, but this doesn't explain the twinkling of stars.

Random variations

The simplest meteorological instruments—the thermometer and the barometer—can be found in many households. As you know, their readings vary from day to day. Sometimes these variations occur within a few short hours, indicating a sudden change in the weather. If we use more sensitive and "non-inertial" devices that can record small, quick changes, we see that the temperature and pressure vary almost constantly. Not only that, they have different values in different places even if the readings are taken at points that are not that far from one another.

Air density varies with temperature and pressure, and the same holds for the refractive index, which is thus a function of the coordinates $\mathbf{r}$ and time t. This function can be written as the sum of two components:

$$n(\mathbf{r}, t) = n_0(z) + \delta n(\mathbf{r}, t). \tag{2}$$

The first item describes the regular variation of n with height z and, generally speaking, it also depends on time. But we aren't interested in such slow changes (for example, from day to night), and so we don't write the argument t in $n_0(z)$. The regular refraction mentioned above is linked with $n_0(z)$. The second item $\delta n(\mathbf{r}, t)$ is characterized by changes that are quick (measured in seconds or even fractions of seconds) and small-scale (from a few dozen meters to millimeters).

But the differences between n_0 and δn can't be ascribed entirely to the rate and range of their changes. Very important features of these values are expressed by the phrases "determinate function" (for n_0) and "stochastic function" (for δn). *Determinate* means that in principle the function $n_0(z)$ can be calculated for any height z. On the other hand, the quantity $\delta n(\mathbf{r}, t)$ depends on a multitude of random circumstances (hence the name *stochastic*) and it's impossible to calculate. We can only indicate the probability that this function assumes a certain value. Not as comprehensive as the probability function but still a very important characteristic of the random value is its mean value, which we denote by a horizontal line above it: $\overline{\delta n}$.

Unlike the function itself, its mean value is often a constant. To tie things down a bit, we'll consider that $\overline{\delta n} = 0$. This means that δn assumes both positive and negative values with equal probability. However, the square of the fluctuations $(\delta n)^2$ is another story entirely. This value is always positive. Its mean value $\overline{(\delta n)^2}$ characterizes the range of the fluctuations and is known as the *dispersion* $\sigma_n^2 \equiv \overline{(\delta n)^2}$. The root mean square (rms) value is also used: $\sigma_n = \sqrt{\overline{(\delta n)^2}}$. In the lower layers of the atmosphere, $\sigma_n \sim 10^{-6}$–10^{-7}—that is, the fluctuations in the refractive index are very small even compared to the small deviation of n_0 from unity mentioned above ($n_0 - 1 \sim 10^{-4}$).

Fluttering

How does the stochastic heterogeneity δn influence the refraction of light beams? It's easy to picture how in addition to the regular refraction (angle $\Delta\alpha$ in figure 1b) there may be stochastic deviations $\delta\alpha$ that can be directed to any side with equal probability—that is, $\overline{\delta\alpha} = 0$. In a way similar to that for δn, we introduce the notions of the dispersion $\overline{(\delta\alpha)^2} \neq 0$ and the root mean square refraction $\sigma_\alpha = \sqrt{\overline{(\delta\alpha)^2}}$. Astronomical observations give us the estimate $\sigma_\alpha \sim 1''$. This means that the image of a star in a telescope wanders randomly within a circle with a radius of approximately one second of arc. Obviously such fluttering gets in the way of astronomical observations. This is why astronomers choose the locations of their large optical instruments with great care, preferring mountainous regions with good "seeing"—that is, σ_n and σ_α are very small.

The fluttering of the image in a telescope often occurs so rapidly that only the eye can make out the oscillations. When a photograph is taken, the result depends on the exposure time, which in turn is determined by the sensitivity of the film and the brightness of the star. To obtain images from weak sources, the first astronomical photographs were taken with exposures of many minutes or even several hours. The fluttering of the image was averaged over this long period and each "point" source of light produced an image of a diffuse circle with angular size $\sim \sigma_\alpha$.

Along with the rapid shifts of an image, there are also slow ones with a period of about 1 minute. They have larger amplitudes and can be decreased to a considerable extent if the astronomer compensates for the shift by adjusting the direction of the telescope or the position of the film. It's hard to believe now, but in the beginning of this century astronomers had to work through the night, simultaneously using their eyes, their hands, and even their mouths. Their eyes watched the image of a star, their hands adjusted the photographic plate in two perpendicular directions, and their mouths held the cable release for the camera's shutter. As you might have guessed, nowadays this work is fully automated.

Fluttering ≠ twinkling

Our intuition tells us that the random shifts of a star's image and the chaotic variations of its brightness are related somehow, and yet they are different phenomena. At first glance twinkling can easily be explained not only by fluctuations in the air's refractive index but by fluctuations in the air's transparency. Certainly this is a nice, simple explanation, but why then do some heavenly bodies twinkle (stars) and others don't (planets)? Also, direct optical measurements show that the transparency of air is very high and does not vary over short periods of time. So we need to look for other explanations.

Let's consider in more detail the influence of the heterogeneities δn on light beams. Each heterogeneity acts like a small lens, which can be either converging (when $\delta n > 0$) or diverging (when $\delta n < 0$). For definiteness the case of $\delta n > 0$ is shown in figure 2, and the beams passing through the heterogeneity converge.

The flux of light energy E falling on a lens with characteristic dimensions l is proportional to its cross-sectional area S_0,

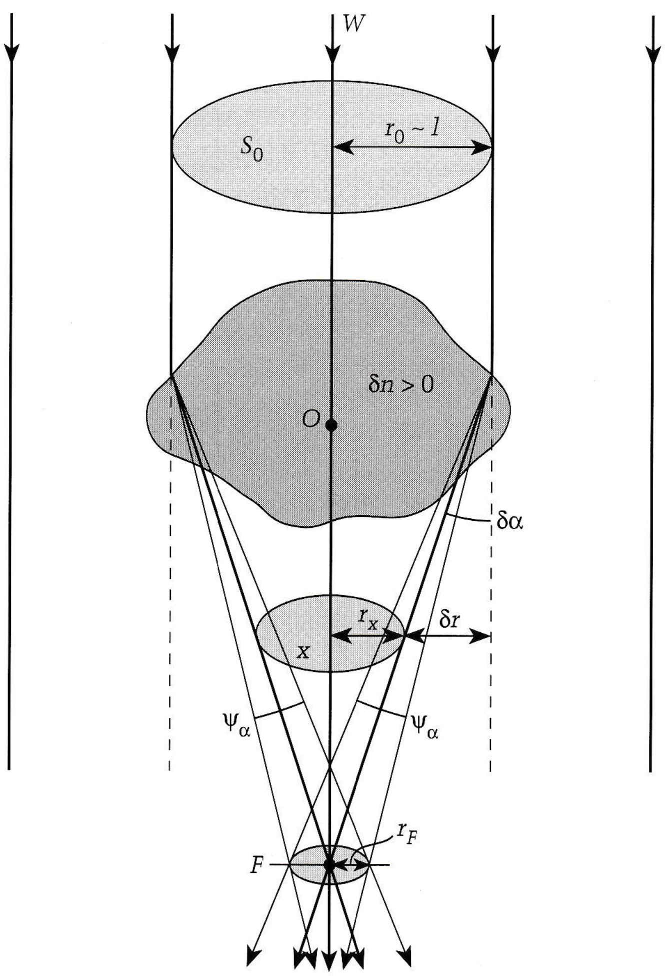


Figure 2

Refraction of rays caused by random variations in the air's refractive index.

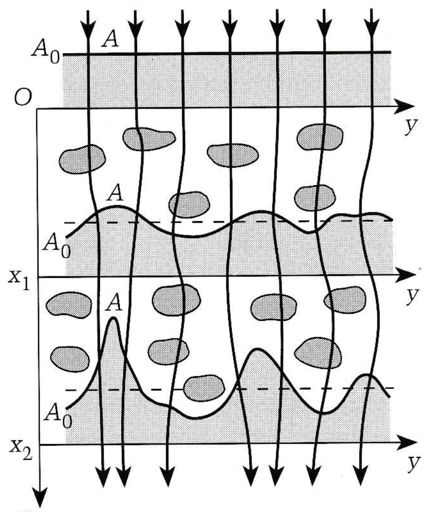


Figure 3

Refraction of rays in random atmospheric heterogeneities, which leads to a redistribution of the intensity in the transverse direction. Where the beams converge, the amplitude increases, and vise versa.

which we can estimate to be approximately the area of a circle of radius $r_0 \sim l$: $S_0 \cong \pi l^2$. Therefore, $E \cong J_0\pi l^2$, where J_0 is a factor characterizing the light intensity. After passing through the lens the radius of the cross section of the light cone decreases by $\delta r \cong x\delta\alpha$ and becomes equal to $r \sim l - x\delta\alpha$. Here $\delta\alpha \cong \delta n$ is the deflection angle caused by a single heterogeneity, and x is the distance from the lens. The law of conservation of energy yields $E \cong J_0\pi l^2 = J_x\pi(l - x\delta\alpha)^2$, from which it's easy to find the intensity of the focused beam:

$$J_x \cong \frac{J_0 l^2}{(l - x\delta\alpha)^2}. \qquad (3)$$

The value of $\delta\alpha$ is always small, but the product $x\delta\alpha$ can become comparable to l, when x is sufficiently large. We need to pay special heed to this, because when $x = F = l/\delta\alpha$ (where F is the focal length of the "lens"), the denominator in equation (3) become 0, while $J_x \to \infty$.

In actuality the increase of J_x is always constrained by the wave nature of light. As you probably know, light is electromagnetic radiation that is characterized by a wavelength λ and an amplitude A. The intensity of the light is proportional to the square of the amplitude: $J_0 \propto A_0^2$, $J_x \propto A^2$. The wavelength determines the angle of diffraction: when the wave encounters an obstacle with dimensions l, the wave passes around it and spreads out within the angle $\psi_\alpha \sim \lambda/l$ (fig. 2 on the previous page). As a result, even an ideal lens of diameter l concentrates the light not into a point (as geometrical optics predicts) but into a circle of radius $r_F \sim x\psi_\alpha$ lying in the focal plane $x = F$ (here we take into account that the angles $\delta\alpha$ and ψ_α in figure 2 are very small).

We also recall that the wavelengths perceptible to the human eye range from ~0.4 μm (violet) to ~0.76 μm (red). It is atmospheric dispersion that is partially responsible for multicolored (chromatic) twinkling. This is because the factor k in equation (1) depends on λ. Generally speaking, σ_n^2 also depends on λ. Chromatic twinkling can be seen at small angles to the horizon and is a very pretty sight.

Concluding remarks

Why does the intensity of light vary in absolutely transparent air? Because the energy is redistributed in the *transverse* direction relative to the beam (fig. 3). Up to now we've only addressed *spatial* variations δA, but twinkling is a change in a star's brightness that varies with *time*. Were the air absolutely motionless, no twinkling would be observed. The same star would simply be brighter in one place (say, at point x_1y_1) than in another (x_2y_2).

In reality, though, air is constantly moving, and with it move the heterogeneities of the coefficient of refraction. The leading role here is played by wind that is transverse to the light beam, which "carries" the curve $A(y)$ as a whole along the y-axis with a velocity $v_\perp$. As a result, a *stationary* observer sees changes *in time* in a star's brightness—that is, twinkling (fig. 4). The average distance between peaks of the intensity is about the size of a heterogeneity l. Thus, the characteristic time for a change in brightness (the "period"

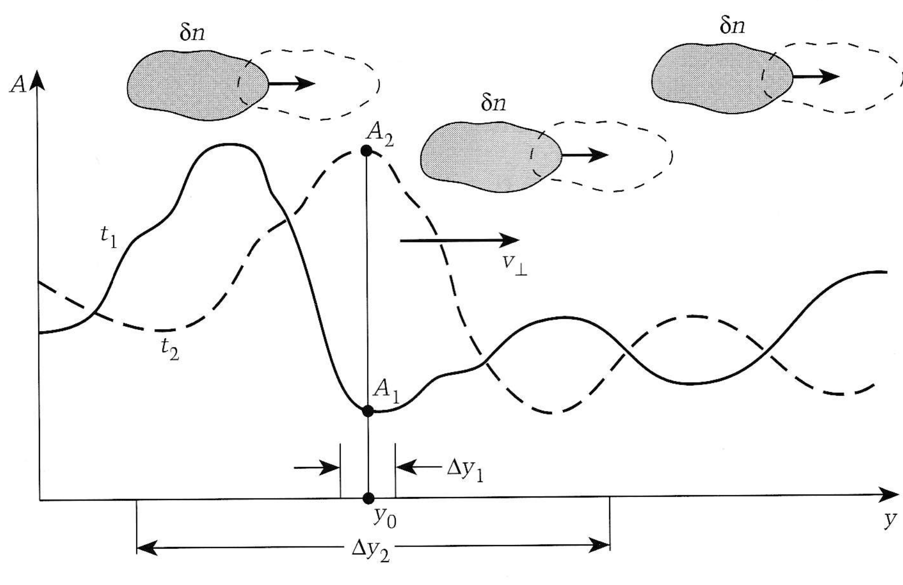


Figure 4

Translation of atmospheric heterogeneities by the wind, which leads to temporal variations in the intensity of light radiation (twinkling). Δy_1 is the size of the pupil of the eye and Δy_2 is the diameter of an objective (y_0 is their center).

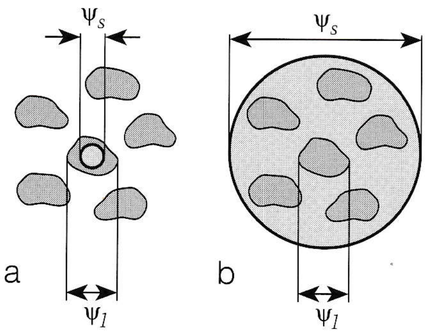


Figure 5

Sources of (a) small and (b) large angular size. In the first case, twinkling is observed; in the second, it is absent.

of twinkling) is $\tau \cong l/v_\perp$. Let's assume, for the sake of our estimate, that $l \cong 10^{-1}$ m and $v_\perp \cong 1$–10 m/s. Then we get a period $\tau \cong 1$–10^{-1} s. If we take into account that the air contains heterogeneities of different sizes, we can understand why twinkling is characterized by such a wide range—from tenths of a hertz to 10 Hz or more.

This scenario helps us understand other properties of twinkling mentioned at the beginning of this article. When we look at a star with the naked eye, a light beam a few millimeters in diameter enters the pupil. But the distance between the oscillations of intensity in air is much larger, so the eye doesn't average the variations of $A(y)$. The opposite occurs with a telescope. If the diameter of its objective $d > l \cong 10^{-1}$ m, it takes in both the stronger and weaker portions of the light flux, which results in an averaging of the intensity and a damping of the twinkling.

Now all that's left is to figure out why the planets don't twinkle. Let's imagine that the light source isn't a "point" but a luminous body with a rather large diameter D—so large, in fact, that its angular size $\psi_S = D/R_S$ (R_S is the distance from the source) exceeds by far the angular dimensions of the atmospheric heterogeneities $\psi_l \cong l/x$. This means that when we look at a luminous body many heterogeneities are projected onto its area. Figure 5 compares two sources: a point source ($\psi_S << \psi_l$) and an extended one ($\psi_S >> \psi_l$). In the first case, the light beam hits either one or another heterogeneity (remember the wind!) and its intensity constantly changes. In the opposite case, the extended source can be thought of as a collection of a large number of "points." The brightness of each point varies as in the first case, but they fluctuate independently, because the rays of light from different parts of the source pass through different heterogeneities. As a result, if the brightness of one part increases, it decreases in another. The total intensity of light emitted by the source as a whole practically doesn't change.

So, we've answered (though not exhaustively) the questions raised at the beginning of the article and, we hope, showed once more how an understanding of physics helps us appreciate even more deeply the beauty of Nature.

QUANTUM SMILES

Three metaphysical tales

The Meaning of Life for lines, light, and cosmic bodies

by A. Filonov

Ways of looking at the plane

Many lines passed through the plane: straight lines, curved lines, broken lines. And each of them had its own way of looking at the world.

"Everything in the world is either raised up or sunk down"—that was one straight line's slice of the Truth.

"No," another cut her off, "everything in the world is either right or not right."

"Stop bickering, girls," interjected a curve, bending ever so smoothly. "Everything in the world is dialectical: here you're right, but there you're sunk."

The broken lines were too embarrassed to express their opinion, but the circle formulated its view thus: "The entire world is either inside, or it's everything else. I, uh, don't have much to say about all the rest, but my inner world is very rich indeed. Only all-round types like me . . ."

"Two-dimensional personalities always have so much to say," interrupted a spiral. "As far as the world is concerned, it's just a layer between coils that keeps them from getting tangled . . ."

And no one solicited the opinion of the inconspicuous little point—the only common point of the plane and a line passing outside this plane.

Universal gravitation

At one time the Earth had no satellites. When a satellite did appear, is turned out to be the defenseless little Moon.

And every meteorite, every cosmic speck of dust injured her. After all, the Moon has no atmosphere to protect her.

The Earth inclined its axis, deep in thought. What business was it of hers? She had an atmosphere, water even—why should she worry? But the Earth did worry: there was a planet nearby that needed protection.

The concerned Earth offered the Moon her air, her water: "Take them, Moon, I have plenty to spare." The Earth drew itself toward the Moon. The Moon drew itself toward the Earth.

That's how tides arose, which often sweep us up. And that's the origin of what people later called "selfless love."

Attraction to powerful sensations

Light travels in a straight line—if it's in a vacuum. But what if it's near a big, bright star? It bends its path ever so slightly so it can pass closer.

"Well, just the tiniest bit, hardly at all, really!" light says defensively. "I was drawn to take a look!"

But then it comes upon a black hole, which doesn't let the least scrap of light escape its tight embrace. Like a miser, it hides its own light from others, and also tries to grab any other light that happens by. And that light bends toward the black hole out of curiosity.

"Just one last time!" it says.

For the very last time.

So don't deviate from the straight and narrow! (Even if you're really drawn to.)

MATH INVESTIGATIONS

The Pizza Theorem—Part II

And when you're done with the pizza, try the calzone!

by George Berzsenyi

AS INDICATED IN PART I OF this account, I am most indebted to Stan Wagon (Macalester College) for my initial information about this problem area, which seems to have originated with Problem 660, proposed by L. J. Upton, in *Mathematics Magazine*. Its solution appeared in 1968 on page 46 with a comment by Michael Goldberg that the result can be generalized to $2n$ equally spaced chords. A closely related problem was later posed by Stanley Rabinowitz as Problem 1325 in *Crux Mathematicorum*; for two different solutions to it, see vol.15 (1989), pp. 120–22. In particular, it was noted in the solution that M. S. Klamkin generalized the problem to n chords through an arbitrary point P with equal angles π/n between successive chords. One of the two solutions given is geometric, while the other one uses calculus. The third method of solution (for the special case of $n = 4$ referred to in part I) is by dissection; it was discovered by Stan Wagon and Larry Carter (IBM) and will appear in *Mathematics Magazine*. My first challenge to my readers is to **find a proof by dissection to the Pizza Theorem**. Dissection proofs for $n = 6$ and 8 were later discovered by Allen Schwenk (Western Michigan University), who feels that such proofs exist in general, though it's hard to see how one could prove this in a uniform way. Incidentally, upon reading part I, Stan Rabinowitz sent me a wonderful 37-page manuscript entitled *A Survey of Interesting Results about Regular Polygons*, which includes nearly everything known about the Pizza Problem. You may wish to request a copy by writing to Stan (MathPro Press, 12 Vine Brook Road, Westford MA 01886). If you do so, please also inquire about (and purchase!) his *Index of Mathematical Problems, 1980-1984*, which is a *must* in every problemist's library.

The Pizza Theorem: *If a circle is divided into eight parts by chords through an arbitrary point inside or on the boundary of the circle, if the resulting "pseudoradii" form equal angles with one another, and if the resulting "pseudosectors" are colored alternately black and white, then the sum of the black areas is equal to the sum of the white areas.*

In reaction to my previous note I also heard from Murray Klamkin, informing me that an expository article of his on three kinds of "equi-area partitions" is now in preparation. In particular, he found that the restriction of P to the interior of the circle is not necessary if one uses signed areas. Moreover, he generalized the problem to kn points distributed regularly on the circle: if one sums every kth sector, one obtains $1/k$ of the circle's area. My second challenge is: **verify Klamkin's claims**.

My third challenge concerns a three-dimensional extension communicated to me by Tom Banchoff. It was found by a student of his, Michael Nathanson, while a junior at Brown University. They refer to it as the Calzone Theorem: *Choose any point P inside or on the boundary of a sphere (calzone), any line through this point, and four planes through this line making eight equal 45° angles at P. Then these planes, together with the plane perpendicular to this line, divide the calzone into 16 pieces, which can be colored alternately black and white, so that the total volume of the black pieces will be equal to the total volume of the white pieces.* The proof can be obtained by using Cavalieri's Principle; my third challenge is: **develop such a proof** or **find a proof by dissection**.

In closing, I should mention that the aforementioned manuscript of Rabinowitz also deals with the divisions of regular polygons and affine images of them (and of circles), and that he suspects the presence of a duality principle between equally placed points on the perimeter and equal angles between the "pseudoradii" emanating from P. I will inform you when his paper, Klamkin's article, and the findings of Wagon and Carter appear in print. In the meantime I wish you a happy journey in this wonderful problem area. For the related *Kvant* material mentioned in part I, see brainteaser B110 in this issue of *Quantum*.

Challenges in physics and math

Math

M106

Economical omnipresence. (a) What smallest number of points is it sufficient to mark inside a convex pentagon so that at least one of these points lies inside the triangle formed by any three vertices of the pentagon? (b) Answer the same question for a convex n-gon.

M107

Finding Pa. From point O inside triangle ABC perpendiculars OM, ON, and OP are drawn to sides AB, BC, and CA, respectively. If $AM = 3$, $MB = 5$, $BN = 4$, $NC = 2$, and $CP = 4$, find PA. (E. Tsinovi)

M108

Inequality on the unit sphere. For any three nonnegative numbers x, y, z satisfying the equation of the unit sphere $x^2 + y^2 + z^2 = 1$, prove the inequality

$$\frac{x}{1-x^2}+\frac{y}{1-y^2}+\frac{z}{1-z^2} \geq \frac{3\sqrt{3}}{2}.$$

(V. Matizen)

M109

Lots of right angles. From the vertex A of a square $ABCD$ two rays are drawn inside the square. From vertices B and D, perpendiculars are dropped to the two rays: BK and DM are dropped to one of them, and BL and DN are dropped to the other. Prove that the segments KL and MN are congruent and perpendicular. (D. Nyamsuren [Mongolia])

M110

Don't get around much anymore. The dormitory of the Easy Listening School of Music has an infinitely long hallway with an infinite row of rooms on one side. The rooms are numbered in order by integers, and each room has a grand piano in it. A finite number of students live in these rooms (several students might live in the same room). Every day two students from adjoining rooms—the kth and $(k + 1)$st—get annoyed with each other's playing and move apart—to the $(k - 1)$st and $(k + 2)$nd rooms. Prove that the moving ends in a finite number of days. (V. Ilyichov).

Physics

P106

A flying leap. Estimate the minimum size of a spherical asteroid that an astronaut couldn't leave by jumping off. (G. Meledin)

P107

Spilled milk. Pouring milk into a glass, you spill some of it on the tablecloth, which happens to be a piece of oilcloth. You discover that the design of the oilcloth is just barely visible through the film of milk. Considering that milk is a suspension of little fatty balls in water, estimate their size. (P. Zubkov)

P108

A watched pot. One liter of water in a pan cannot be boiled by using a 100-watt electric heating element. Find the time it takes the water to cool 1°C when the heating element is turned off. (K. Sergeyev)

P109

One more capacitor. A capacitor C and an inductor L are connected in series to a battery of voltage V_0. The coil is then connected in parallel to another capacitor with the same capacitance C. What is the maximum charge of this capacitor? Neglect the resistance of the wires and the internal resistance of the battery. (A. Zilberman)

P110

Disappearing fence. Watching tennis players through a chain link fence, we can observe two phenomena: first, if we move farther away we can see the players better; and second, if we walk quickly along the fence, the fence seems to disappear. Explain why. (S. Krotov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 54

Tori, tori, tori!

From bagels to tokamaks to topological mind-games

WHEN DID YOU SEE YOUR first torus? Was it the doughnut you gnawed on as a baby? Or the inner tube you used to float down a stream? The wedding band on your father's finger? Or maybe the delicate ring blown by a stogie-smoking uncle?

The mathematically perfect torus is defined as a solid body formed by rotating a circle about a line in the circle's plane that has no common points with the circle (fig. 1). The volume of such a ring was calculated by Johannes Kepler. This is what he wrote in his famous book *The New Stereometry*[1] *of Wine Barrels*:

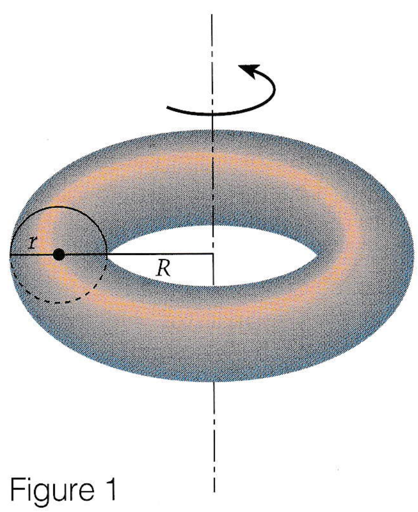


Figure 1

"Any ring of circular or elliptical section is equal in volume to the cylinder whose height equals the length of the circumference described by the center of the rotating figure, and the base equals the section of the ring."

The complete title of this work, according to the style of that time, was rather verbose: *The New Stereometry of Wine Barrels, Mostly Austrian, As Having the Most Advantageous Shape, and the Remarkably Convenient Use of the Cubic Ruler with Them, with an Addendum on Archimedean Stereometry. A Work by Johannes Kepler, the Mathematician of Emperor Caesar Mattheus I and His Faithful Ranks of Upper Austria with the Caesarean Privilege for 25 Years.*[2]

In modern notation, Kepler's formula for the volume V of a torus with a generating circle of radius r and a distance R from its center to the rotation axis is

$$V = 2\pi^2 r^2 R.$$

The surface area of the torus is $4\pi^2 Rr$—that is, it's equal to the product of the lengths of two circumferences: the one generating the torus and the one described by its center as it rotates about the axis. Similarly, the volume can be interpreted as the product of the *area* of the rotating circle and the length of the circumference described by its center. Both interpretations are particular cases of the so-called Guldin formulas for the volume and surface area of rotational solids.

The term "torus" is used not only for the solid but also for its surface. This surface is a particular favorite of topologists—mathematicians who study the properties of figures that are preserved under continuous deformations. From the topological point of view, the torus is the simplest surface after the sphere. (What distinguishes it from the sphere is, of course, the hole.) This visible geometric difference can be expressed algebraically. Leonhard Euler noticed that any polyhedron topologically equivalent to a sphere obeys the formula

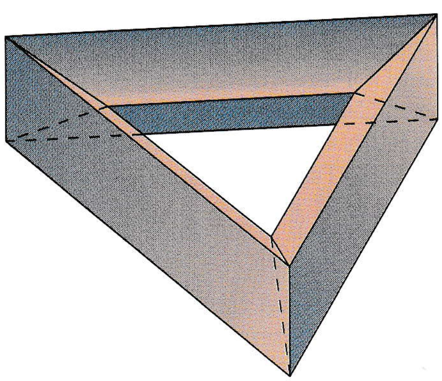


Figure 2

$$V - E + F = 2,$$

where V, E, and F are the numbers of its vertices, edges, and faces, respectively.[3] For instance, a cube has 8 vertices, 12 edges, and 6 faces; accordingly, $V - E + F = 8 - 12 + 6 = 2$.

[1] "Stereometry" is the old name for solid geometry.—*Ed.*

[2] See "The Secret of the Venerable Cooper" in the May 1990 issue of *Quantum* to learn more about the main subject of this curious work.—*Ed.*

[3] You can find a proof of this theorem in "Topology and the Lay of the Land" in the September/October 1992 issue of *Quantum*.—*Ed.*

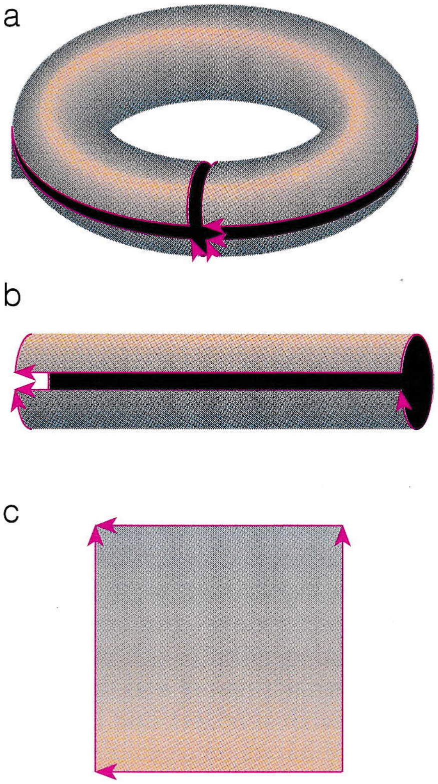


Figure 3

Or take an n-gonal pyramid: it has $n + 1$ vertices, $2n$ edges, and $n + 1$ faces, which gives us $V - E + F = (n + 1) - 2n + (n + 1) = 2$ again. But if you compute the same expression (called the *Euler characteristic*) for a toroidal (that is, torus-shaped) polyhedron, you'll find that it's equal to 0, not 2. For instance, for the "triangular torus" in figure 2, $V - E + F = 9 - 18 + 9 = 0$.

Now imagine we cut a torus (the surface, not the solid!) across it and along it (fig. 3a)—that is, around the rotating circle and the circle described by any of its points—and develop it. When we unbend the ring after the first cut (fig. 3b), we get a cylinder with the second cut running along it. So in the end we'll get a rectangle whose opposite sides—the edges of the cuts—can be thought of as identical (glued to each other). Since dimensions are of no importance to topologists, they usually draw the resulting figure as a square (fig. 3c). The arrows on the sides are drawn to make sure they are glued together properly—arrow to arrow. (Reversing an arrow in one or both parallel pairs of sides leads to other interesting surfaces.) This representation is very convenient for exploring and explaining many properties of the torus. For instance, figure 4 illustrates a map on the torus in which every two countries have common border. Don't forget that the opposite edges of this map must be glued together, so the four apparently disjoint pieces at the corners of the map are actually parts of the same country. Other pieces of the same color at the edges belong to the same countries as well. Another version of the same map (fig. 5), obtained by cutting the square map into pieces and gluing them together along the edges, clearly shows that there are seven hexagonal countries, each surrounded by the other six. (In the figure we see only one country surrounded by the others, but in fact it's true for all the countries. Try to show this by recarving the map in an appropriate way.) This map illustrates the fact that we need at least seven colors to paint any map on the torus so that neighboring countries are a different color. It can also be proved that seven colors suffice to paint *any* map. The same problem for the sphere wasn't solved until several years ago, when it was definitively proved that four colors will always suffice for a spherical map.

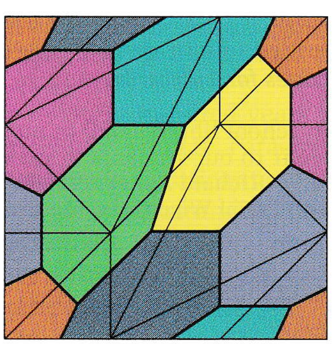


Figure 4

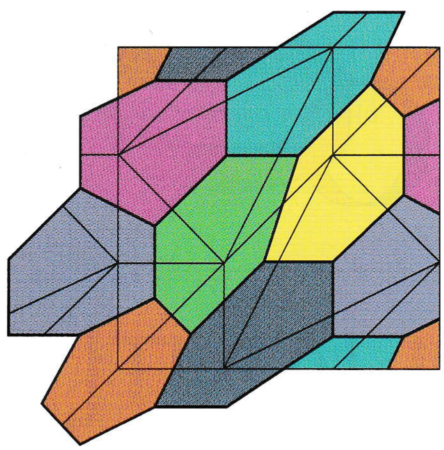


Figure 5

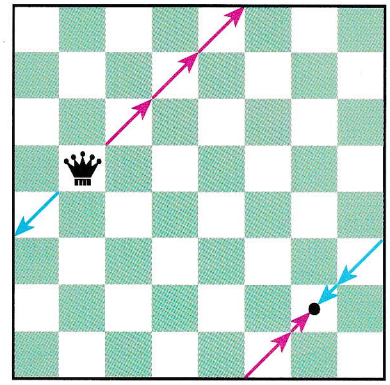


Figure 6

The thin black lines in figure 4 constitute a triangular grid on the torus (the vertices of our hexagonal map are the centroids of the network triangles). This network has seven vertices every two of which are connected with an edge. The edges form 14 triangular faces. The Toy Store in this issue explains how to construct a model of a polyhedron whose development is topologically equivalent to this network. And this is the simplest toroidal polyhedron in the sense that it has the smallest possible number of vertices.

Imagine we glue together the opposite edges of a chessboard. Then the "edge" and "corner" squares will disappear—all the squares will become geometrically equivalent. What a plethora of opportunities for bishops, rooks, and queens! A bishop or queen can get from b5 to, say, g2 in one move, and even by moving in two opposite directions (fig. 6). A king and a rook or a queen can't checkmate a solitary opposing king. (See the Toy Store in this issue for some toroidal chess problems.)

Other games on a square board can also be transferred to a torus. For instance, try to play tic-tac-toe on the torus. It's well known (and easy to show) that on an ordinary 3×3 board

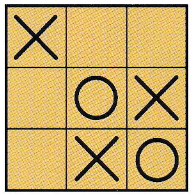


Figure 7

either player can force a draw. Who wins on the torus? What should the first move be? Notice that in the position shown in figure 7 the first player has already formed a line of three X's and, therefore, has won!

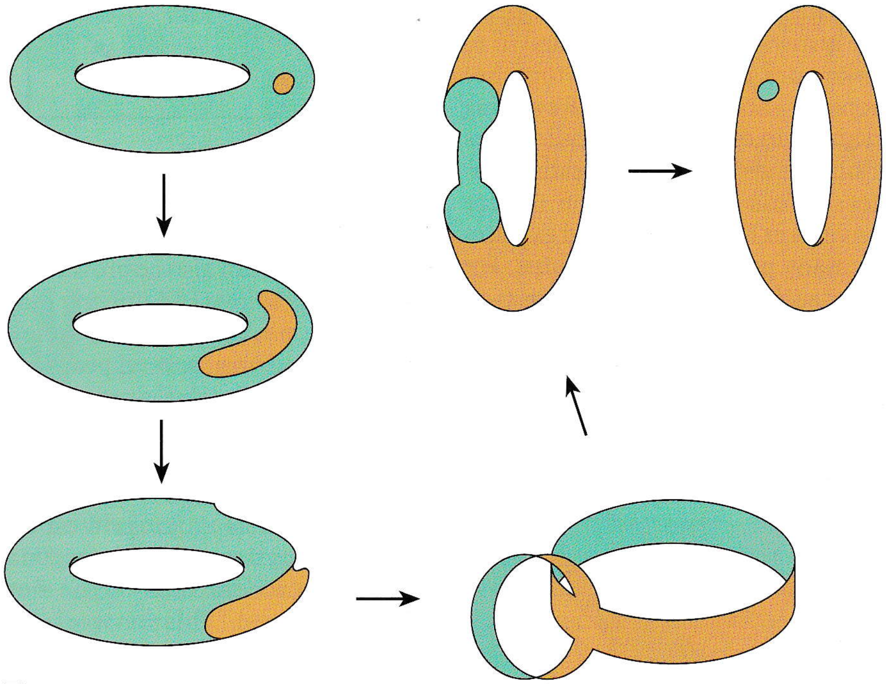


Figure 8

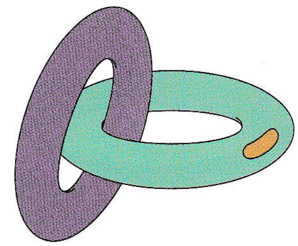


Figure 9

To close, let's turn the torus inside out. It's easy to turn a balloon inside out through the hole you use to blow it up, if the rubber is elastic enough. It might seem that this operation is impossible with the inner tube of a bicycle tire. But we can do it, as figure 8 shows. And now a topological question to test your imagination: think of a second torus linked with the first one initially (fig. 9). What happens to it after we perform our barbarous act?

ANSWERS, HINTS & SOLUTIONS ON PAGE 57

Phlogiston and the magnetic field

A brief tour of the junkyard of science

by Stephanie Eatman, Fraser Muir, and Hugh Hickman

SOMETIMES IN SCIENCE, OLD ideas have to be replaced by new ideas. Phlogiston, caloric, and the luminiferous ether were all legitimate scientific concepts, once believed to be real. One by one they were discarded when new information revealed that they were no longer needed to explain the phenomena they were created to explain. Could it be that the magnetic field will one day join them in the scientific scrap pile?

Electricity and magnetism were both known to the ancient world, but it wasn't until 1820 that Hans Christian Oersted established a definite connection between the two phenomena. Oersted discovered that a wire carrying an electric current was able to deflect a compass needle when the wire was placed parallel to the original direction of the needle. Since an electric current consists of a line of moving charge, and compass needles are known to be deflected by magnetic fields, the connection became clear—lines of moving charge produce magnetic fields.

For the next hundred years great scientists like Ampere, Faraday, and Maxwell made significant contributions to our understanding of electromagnetism. In the final analysis, however, the electric field and the magnetic field were still perceived to be two different things. It's easy to see why. The two fields seem to produce quite different effects. For example, an electric field will exert a force on an electric point charge whether the charge is moving or standing still. But for a magnetic field to exert a force on the same point charge, the charge must be moving, and it must be moving in a direction that is not parallel to the direction of the magnetic field.

This perception of the two fields as two distinct entities was seriously challenged soon after 1905, when Albert Einstein published his special theory of relativity. For the first time it became possible to demonstrate that the magnetic force on a moving point charge actually comes about because of the

Art by Sergey Ivanov

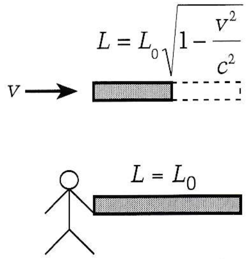


Figure 1

relativistic transformation of the electric field. Let's see how.

According to Einstein, the length of a moving object, measured by a stationary observer, becomes contracted in the direction of the object's motion (fig. 1). The faster the object moves, the shorter it becomes compared to its stationary length. Mathematically, this is expressed as

$$L = L_0\sqrt{1-\frac{v^2}{c^2}}, \qquad (1)$$

where L is the contracted length, L_0 is the stationary length, v is the speed of the object relative to the observer, and c is the speed of light (approximately $3 \cdot 10^8$ m/s in a vacuum).

The idea that the length of a moving object becomes contracted in the direction of its motion has been well established experimentally. On the other hand, both experiment and theory indicate that the amount of electric charge possessed by a moving object does not change due to the object's motion.

Suppose we have a uniform line of positive charge situated in air. The electric field around an infinite, stationary, positive line charge is given by

$$\mathbf{E} = \frac{\lambda}{2\pi\varepsilon_0 r}\hat{\mathbf{r}}, \qquad (2)$$

where the boldface $\mathbf{E}$ indicates that the electric field is a vector quantity, having both magnitude and direction (on the right-hand side, $\hat{\mathbf{r}}$ is a unit vector, meaning that the field always points directly away from the line); ε_0 is the permittivity of free space (permittivity is a measure of how much the medium surrounding the line charge affects the field strength); r is the perpendicular distance from the line out to the point of measurement; and λ is the linear charge density:

$$\lambda = \frac{Q}{L}. \qquad (3)$$

If the line moves in the direction of its own length, a stationary observer sees L become $L_0\sqrt{1-v^2/c^2}$, but she does not see a change in Q. So equation (3) becomes

$$\lambda = \frac{Q}{L_0\sqrt{1-v^2/c^2}} = \frac{\lambda_0}{\sqrt{1-v^2/c^2}}, \qquad (4)$$

where λ_0 is the stationary charge density Q/L_0. Substituting equation (4) into equation (2) leads to

$$\mathbf{E} = \frac{\lambda_0}{2\pi\varepsilon_0 r\sqrt{1-v^2/c^2}}\hat{\mathbf{r}} = \frac{\mathbf{E}_0}{\sqrt{1-v^2/c^2}}, \qquad (5)$$

where $\mathbf{E}_0$ now stands for the electric field of the stationary line charge.

Equation (5) is already quite revealing. According to this equation the strength of the electric field that emanates from a moving line charge depends on the relative velocity between the observer and the line. The faster the line is observed to move, the stronger the $\mathbf{E}$ field. Interpreted within the framework of special relativity, it is this "variability" of $\mathbf{E}$ that actually gives rise to the force that for centuries has been associated with the existence of the magnetic field $\mathbf{B}$.

In order to see how this works, imagine you're holding a positive point charge and standing close to the moving line charge (fig. 2). Any charge q located in an electric field always experiences a force given by $\mathbf{F} = q\mathbf{E}$. In this case,

$$\mathbf{F} = q\mathbf{E}_0\left(\frac{1}{\sqrt{1-v^2/c^2}}\right), \qquad (6)$$

since the line is moving with velocity v. So your positive point charge is being pushed away from the positive line charge, and you have to push

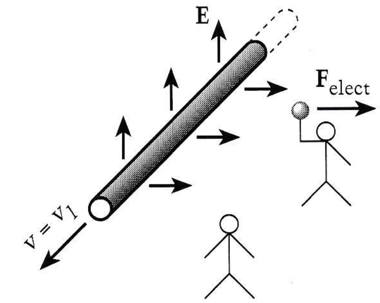


Figure 2

back in order to hold it in place.

Now, the moving line charge is also supposed to produce a $\mathbf{B}$ field (remember Oersted's experiment), but that field will not affect your point charge, since a $\mathbf{B}$ field can only exert a force on a moving charge and your charge doesn't move. (Relative to you, it's stationary as long as you keep holding it.)

Suppose you now start running parallel to the line, but in the direction opposite to the motion of the line (fig. 3). (Maybe you could hold the charge over your head like an Olympic torch bearer.) From your viewpoint the line goes by faster because the relative velocity between the line and the charge is now

$$v = v_l + v_q. \qquad (7)$$

(Actually, this equation is wrong. Galileo, who first proposed equation (7), didn't know that the relative velocity between two moving objects can never exceed the velocity of light. Einstein had to replace equation (7) with

$$v = \frac{v_l + v_q}{1 + v_l v_q/c^2} \qquad (8)$$

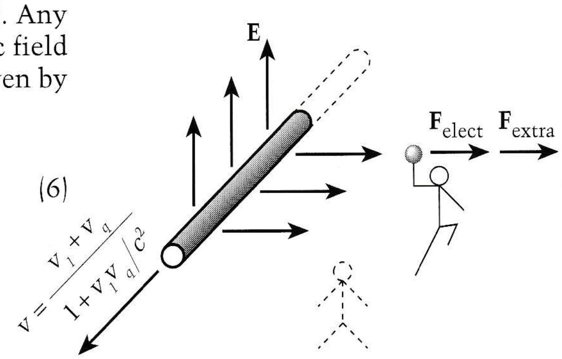


Figure 3

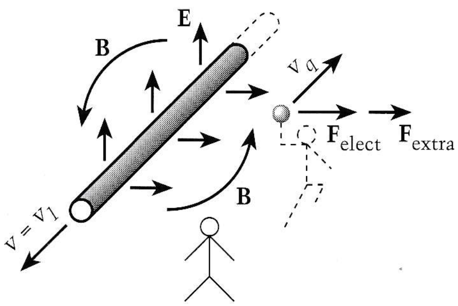


Figure 4

in order to preserve the velocity of light as the maximum possible speed in our universe. If $v_l = v_q = c$, then Einstein's equation results in

$$v = \frac{c+c}{2} = c.$$

If v_l and v_q are small, such that $v_l v_q << c^2$, then Einstein's equation reduces to Galileo's equation. But we're getting away from our story.)

The point is, when the charge moves parallel but opposite to the direction of line motion, the relative velocity between the line charge and the point charge *increases*. So to an observer traveling with the point charge, the point charge appears stationary, but the line charge goes by faster and becomes even more contracted. The extra contraction results in an increase in the electric field strength. (Increasing v in equation (6) causes an increase in **E**.) And the new, total force on our moving point charge could be written as

$$\mathbf{F}_{total} = \mathbf{F}_{elect} + \mathbf{F}_{extra}, \qquad (9)$$

where $\mathbf{F}_{elect}$ is the electric field force experienced by the point charge before it began to move, and $\mathbf{F}_{extra}$ is the component of force that comes about because the point charge is now moving.

Of course, you are moving with the point charge, so the origin of $\mathbf{F}_{extra}$ is perfectly clear to you. $\mathbf{F}_{extra}$ is simply the increase in electric field force caused by the extra line contraction that you see as you run along. But suppose a friend of yours is standing still, watching you run by. He sees the same line that you saw before you started to run, and there is no extra contraction. So how does he explain $\mathbf{F}_{extra}$?

Before you answer this question, let's recap the two perspectives. You are running with the point charge. To you the point charge is stationary, but the line charge goes by quickly, so it looks highly contracted. You see the extra force as being caused by an increase in the line's electric field. No problem. Your friend is standing still. To him, the point charge is moving, but the line charge goes by slowly, so it looks just like it always did—not so highly contracted. He sees an extra force, but he has no clue as to the origin of the extra force. Big problem.

Your friend can't explain the extra force using just the electric field from the line, because he never saw that field change. Instead he has no choice but to postulate the existence of a second field. He believes that the second field originates in the moving line charge (since where else could it come from?), and he sees that it only acts on *moving* point charges. He calls it **B** (fig. 4).

Our friend postulated the existence of **B** in order to explain a phenomenon that he couldn't otherwise explain from his (stationary) point of view. Phlogiston (the "element" of combustion), caloric (the "fluid" of heat), and luminiferous ether (the medium through which light was supposed to propagate) were all postulated for exactly the same reason. But all three concepts were later abandoned as scientific knowledge advanced.

Is it really possible that science could do without the concept of a magnetic field? Most physicists would say no. They would agree that a purely electric or magnetic field in one coordinate frame will appear as a mixture of electric and magnetic fields in another coordinate frame. But they would argue that **E** and **B** are both elements of something called a second-rank tensor, and that one should properly speak of the electromagnetic field $F^{\alpha\beta}$ rather than **E** or **B** separately. **B**, they would say, is too deeply ingrained in the theory to discard.

Still, there's no doubt that the development of electrodynamics would have proceeded quite differently if special relativity had been introduced 100 years earlier.

Ptolemy's epicycles formed the backbone of celestial mechanics for 1400 years. Even Copernicus used them in his original heliocentric theory. Epicycles vanished with the acceptance of Kepler's conjecture that planetary orbits are elliptical and not circular. The magnetic field is already shaky. Could it be that a new theory of broader scope will eradicate **B** entirely?

Stephanie Eatman *and* **Fraser Muir** *are students at Hillsborough Community College in Tampa, Florida. Dr.* **Hugh Hickman** *teaches physics at the same institution.*

PHYSICS CONTEST

Fun with liquid nitrogen

"Some say the world will end in fire, Some say in ice."—Robert Frost

by Arthur Eisenkraft and Larry D. Kirkpatrick

IRON IS SOLID, MERCURY IS liquid, and nitrogen is a gas. We gain our familiarity with substances at ambient temperature and tend to think of them in that context. Over millennia our technology has found ways to heat iron so that it becomes a liquid and to cool mercury so that it becomes solid. Cooling nitrogen to form a liquid was first achieved 117 years ago. And the world of liquid gases—liquid hydrogen, oxygen, and nitrogen—could not be stranger.

Liquid nitrogen is used to perform lots of interesting experiments. It's also fun. Demonstrations exploiting the extreme cold of liquid nitrogen provide entertainment for children of all ages. Since liquid nitrogen boils at a temperature of 77 K at atmospheric pressure, we keep it cold in a dewar. If we pour some of the liquid nitrogen on the floor, the liquid forms droplets that scoot around the floor like droplets of water on a hot skillet. Of course, the floor *is* like a hot skillet to the very cold liquid nitrogen droplets!

The liquid nitrogen can also be used to superfreeze common materials. In another demonstration we take a rod-shaped piece of rubber sharpened on one end and drop it into the liquid nitrogen. We then remove it with tongs and hammer it into a board. Frozen rubber is as good as a nail—until it thaws. Some things become brittle at these cold temperatures. It's rather spectacular to shatter a frozen banana or a flower with a hammer. It's as if they were made of glass. One class of kindergartners remembered this demonstration five years later.

The expansion and contraction of gases also seem spectacular when liquid nitrogen is used. A blown-up balloon inserted into the liquid nitrogen shrinks down to essentially zero volume, showing that the ideal gas law is not valid for these conditions. The gases in your breath liquify or freeze at these temperatures. In fact, the balloon makes a good rattle if shaken as it warms up. Usually the balloon expands back to its original volume as it warms, but not always!

In this contest problem we want to measure the latent heat of vaporization of liquid nitrogen. The latent heat of vaporization is the amount of heat required to convert a unit mass of liquid to vapor at the boiling point of the substance. This is based on one of the two experimental problems given at the XXIV International Physics Olympiad held last summer in Williamsburg, Virginia.

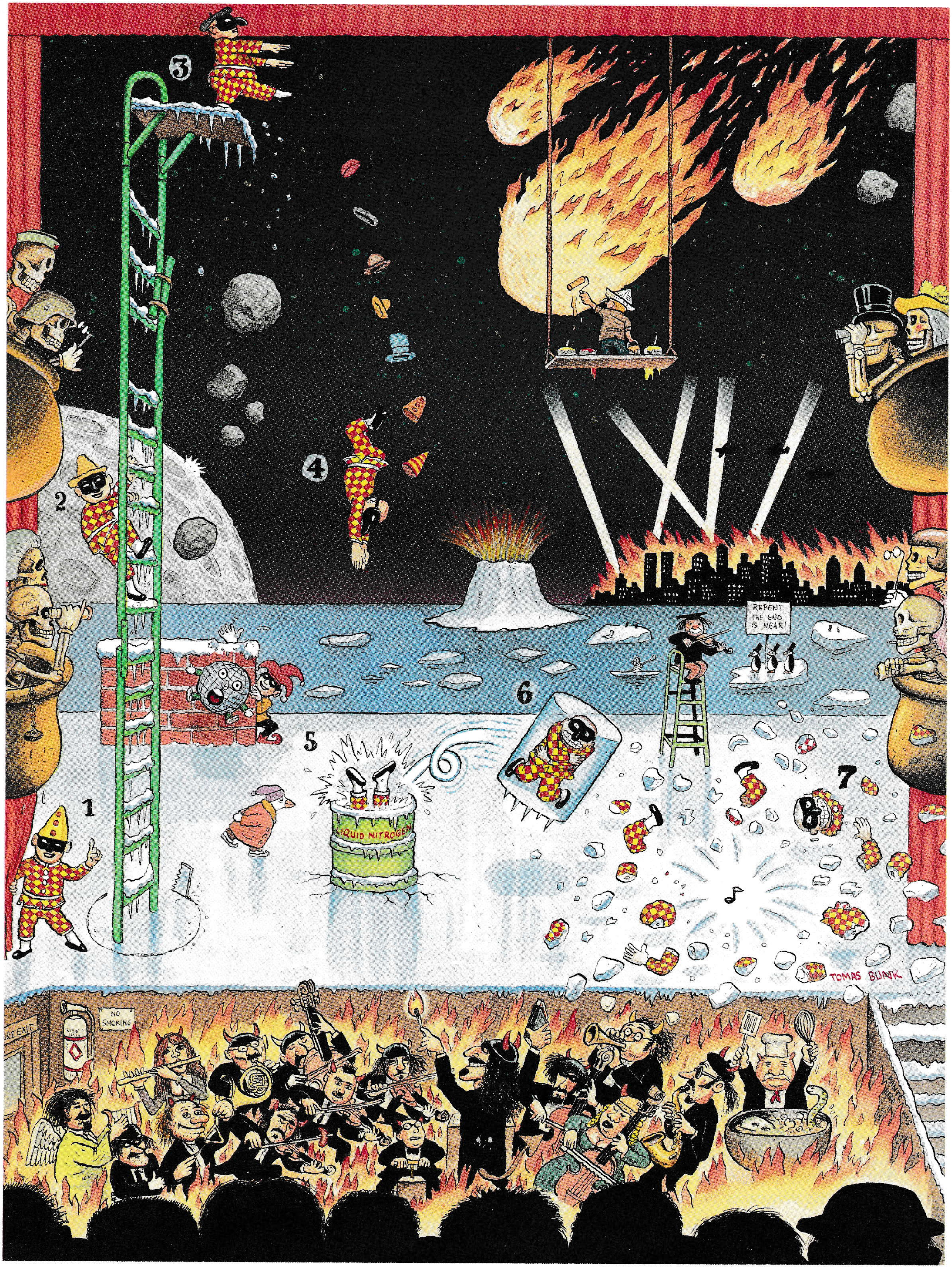
Our method is a variation of the thermal experiment that many of you have performed in your school laboratory. We usually use the known thermal properties of water to measure the specific heat of a block of metal. The specific heat of water c_w is the amount of heat required to raise the temperature of a unit mass of the metal by 1 degree. We usually assume that c_w is constant with a value of 1 cal/g · C° = 4.186 J/g · K. The heat lost by the metal block is equal to $mc\Delta T$, where ΔT is the change in temperature of the metal. Setting this equal to a similar expression for the heat gained by the water allows us to solve for the specific heat of the metal.

Let's begin this contest problem with an analysis of this common experiment. We then move on to the more challenging Olympiad experiment.

A. Calculate the specific heat of aluminum given the following data: the aluminum block has a mass of 36.2 g and an initial temperature of 100°C. You have an ideal calorimeter (that is, one that loses no heat to the surroundings and does not absorb any heat) containing 100 g of water at an initial temperature of 17°C. After the block is placed in the water, the temperature rises to 23°C.

In the Olympiad experiment we will place a "hot" block of aluminum into liquid nitrogen and determine the latent heat of vaporization of the liquid nitrogen from the amount of liquid that is vaporized. Of course, in a real experiment the

Art by Tomas Bunk


3
2
4
1
REPENT THE END IS NEAR!
6
5
LIQUID NITROGEN
7
TOMAS BUNK
NO SMOKING

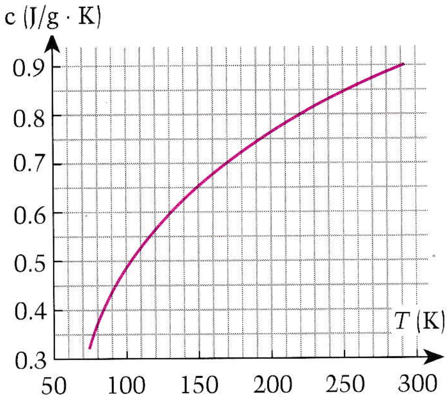


Figure 1
Specific heat of aluminum.

calorimeter is not ideal and heat is exchanged with the surroundings whenever there is a temperature difference. In working with liquid nitrogen, there will be a large temperature difference, and the calorimeter will continually allow heat to enter the system. We also cannot assume that the specific heat of the aluminum is a constant. In fact, it varies a lot, as shown in figure 1.

B. Calculate the latent heat of vaporization of liquid nitrogen given the following data: the aluminum has a mass of 19.4 ± 0.1 g and is initially at a room temperature of 20.0° ± 0.1°C. The total mass of the system is monitored as a function of time and gives the following data:

total mass (g)	time (s)
153	0
152	37
151	79
150	121
149	161
148	203
Aluminum block added	
150	332
149	382
148	457
147	489
146	541
145	595

During the Olympiad the students had to measure the changing mass using a triple beam balance. This is why the time was recorded for specific decreases in the mass rather than the mass recorded for specific time intervals.

C. Because every good experimenter gives an uncertainty for each experimental value, estimate the uncertainty in your value for the latent heat.

The actual experimental problem during the Olympiad required students to measure the latent heat of vaporization by two independent methods. The evolution of the experiment began with the decision by Professor Anthony P. French of MIT (chair of the examinations committee) to make use of the ample supplies of liquid nitrogen that the College of William & Mary (the Olympiad host institution) maintains for research. Peter Collings of Swarthmore College accepted the challenge and devised the Olympiad experiment. As with many good ideas, this one was independently created by Gerhard Salinger (National Science Foundation) and published in *The Physics Teacher* in September 1969.

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

Thrills by design

We hope that *Quantum* readers enjoyed analyzing and creating amusement park rides. The two best solutions to this problem came from Stephan Chan of Ontario, Canada, and Chao Ping Iris Yan of Rio de Janeiro, Brazil. *Quantum* is truly an international magazine!

The example provided in our column explained the physics of the rotor—a hollow cylinder that spins and then "pins" the riders against the wall as the floor drops out. The ride that we hoped readers would design used a rotating hemisphere. The physics of the rotating cylinder and rotating hemisphere are similar in that a centripetal force must be furnished to keep the passenger moving in a circle. In the rotor, the normal force supplied this centripetal force. In the hemisphere, the horizontal components of the normal force and the frictional force must provide the centripetal force. Another difference in the analysis of the two rides is in the measurement of the radius. In the rotor, the radius of circular movement is equivalent to the radius of the cylinder. In our hemisphere, the radius of circular movement is equal to a component of the radius of the hemisphere—$R \sin \alpha$. The final difference is that in the cylinder, we recognized that the frictional force keeps the rider from slipping down. In the hemisphere, the frictional force may keep the rider from sliding down or from sliding up!

We wish to find the minimum coefficient of friction required to keep the rider from sliding down when the angular velocity ω is small (5 radians per second). As with many physics problems, the first step is a carefully drawn diagram and vector analysis (fig. 2)

The sum of the horizontal components must equal the centripetal force; the sum of the vertical components must equal zero:

$$\sum F_x = F_N \sin\alpha - F_f \cos\alpha = m\omega^2 R \sin\alpha,$$

$$\sum F_y = F_N \cos\alpha + F_f \sin\alpha - mg = 0.$$

Since the frictional force is (less than or) equal to the coefficient of friction μ multiplied by the normal force, we

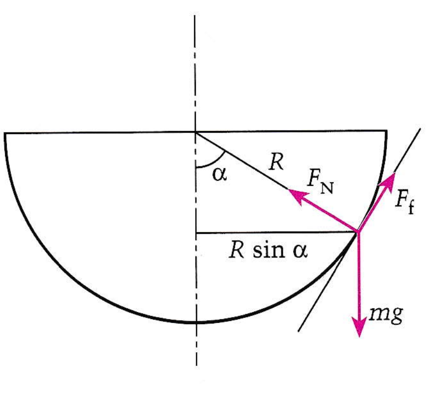


Figure 2

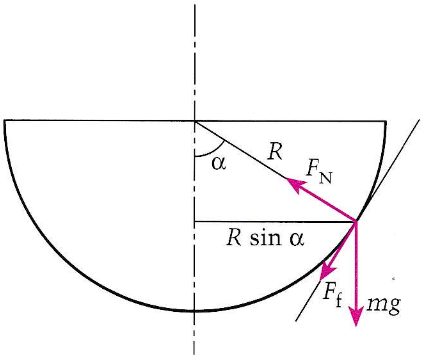


Figure 3

can solve the simultaneous equations for μ:

$$\mu \geq \sin\alpha \frac{1-\omega^2 R\cos\alpha/g}{\cos\alpha + \omega^2 R\sin^2\alpha/g}.$$

For the values given (ω = 5 rad/s, R = 0.5 m, and α = 60°), we get

$$\mu \geq \frac{3\sqrt{3}}{23} = 0.23.$$

Part B asked for the coefficient of friction required when ω = 8 rad/s. The analysis is similar except that the vector diagram in figure 3 now shows that the frictional force is preventing the object from slipping up the hemisphere. Solving the simultaneous equations again for μ, we get

$$\mu \geq \sin\alpha \frac{\omega^2 R\cos\alpha/g - 1}{\cos\alpha + \omega^2 R\sin^2\alpha/g}.$$

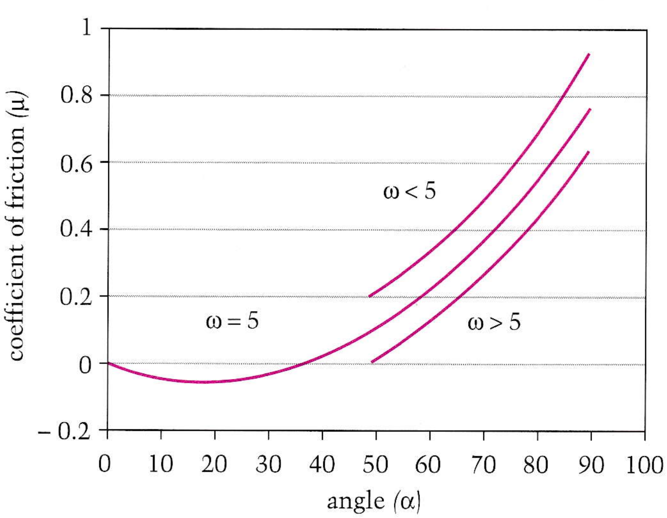


Figure 4

Substituting the values given (ω = 8 rad/s, R = 0.5 m, and α = 60°) gives us

$$\mu \geq \frac{3\sqrt{3}}{29} = 0.18.$$

In part C of the problem, we wanted to analyze the stability of the mass in the hemisphere ride for small variations of the position of the block and for small variations of the angular velocity of the block.

Using a graphing calculator, a spreadsheet program, or sketching, we can look at a graph with three curves (fig. 4). The main curve shows the relationship between μ and the angle. This μ is the friction required to remain at that angle. The two other curves show that relationship for different values of ω. If the object moves to a higher angle, the minimum friction required to stay at that height is greater. The object does not have that much friction and it slides back down to the original position, where the friction is sufficient to have it remain at that height. If the object moves to a smaller angle, the friction required to remain at that position is less. The object is able to remain at this height. If ω increases, the object will remain where it is (since less friction is required at that new ω); if ω decreases, the object will not have the friction required to maintain its position and will start to slide down to an angle where the friction provided is sufficient for this ω.

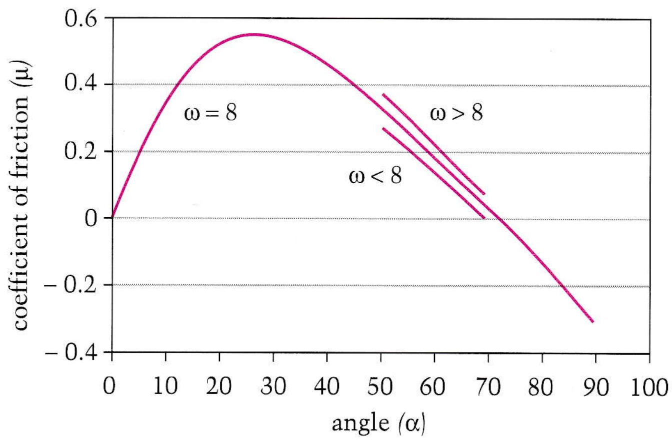


Figure 5

At a higher initial ω, the graph reveals a different situation (fig. 5). If the object moves to a higher angle, it stays there; if the object moves to a smaller angle, it will return. If ω increases, the block slides up; if ω decreases, the block will maintain its position.

In part D readers were asked if this hemisphere problem could be a ride for an amusement park and what problems might arise. Chan was able to show that the person would experience an acceleration of approximately 1.4g during this ride. He thinks that the person would enter the ride from the bottom, and as the ride spins the rider would slide up against the wall. He doesn't see it as an exciting ride as it stands—he suggests that we increase the speed to make the g forces greater.

Yan thought that getting on the ride could be accomplished with a floor at a height equal to $R/2$. The floor would then rotate out of the way during the ride. This would limit the riders to only one side of the hemisphere. Yan suggests that the velocity be increased and decreased during the ride. Yan concludes that the ride may be too scary and people would probably be more secure on the roller coaster.

AT THE BLACKBOARD

The beetle and the rubber band

A problem with an unexpected solution

by Alexander A. Pukhov

A RUBBER BAND OF LENGTH L is attached to a wall. Its loose end begins to move with a velocity v, which stretches it. Simultaneously, a beetle begins to crawl from the wall along the rubber band with a velocity $u < v$. The rubber band is assumed to be infinitely extensible. Will the beetle ever reach the loose end of the rubber band? If so, how long will it take?

What the greats say

I came across this problem in an issue of the journal *Priroda* ("nature" in Russian) dedicated to the memory of Andrey Dmitryevich Sakharov (1990, No. 8, p. 119). Among other things it described an episode that took place at a conference on elementary-particle physics. During one break the participants were offered the beetle problem as an intellectual test. Some physicists took only 15 minutes to solve it, others needed up to an hour, and there were some who couldn't solve it at all, concluding that the beetle would never reach the end of the rubber band. When the problem was posed to Andrey Dmitryevich, it took him only one minute to find the correct answer—that's how long it took him to jot the solution on the back of the conference program. Let's try to solve it ourselves.

At first glance

A brief acquaintance with the problem leads one to think that

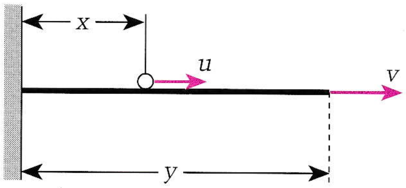


Figure 1

the problem can be most easily solved in the stationary coordinate system associated with the wall (fig. 1). The distance between the beetle and the wall $x(t)$ increases with time t as $x(t) = ut$, and the rubber band's length $y(t)$ changes during the same interval as $y(t) = L + vt$. Since $v > u$, $y(t)$ seems to be always larger than $x(t)$, which means that the beetle will never reach the end of the rubber band. Right?

Not so fast! We haven't taken into account that the rubber band is being stretched. This results in an additional contribution to the velocity dx/dt of the beetle as it crawls from the wall. What is the value of this contribution? The rubber band stretches linearly, so the velocity of the point where the beetle is crawling is proportional to x and inversely proportional to y. Logically, the beetle's total velocity is $dx/dt = u + x \cdot v/y$. This relationship is a differential equation with the solution $x(t)$ determining the graph of the beetle's path. If we manage to find the solution, then by comparing it with the movement of the loose end $y(t) = L + vt$, we'll settle the issue. So the equation we need to solve is

$$\frac{dx}{dt} = u + \frac{v}{L + vt}x. \qquad (1)$$

How do we solve the equation?

Equation (1) seems "frightful." Let's try to deal with it, though. We begin with something more pleasant. Surely we can solve the equation

$$\frac{dx}{dt} = \frac{v}{L + vt}x. \qquad (2)$$

We get

$$\int \frac{dx}{x} = \int \frac{vdt}{L + vt},$$

$$\ln x = \ln(L + vt) + \text{const},$$

$$x = C(L + vt).$$

The constant C always appears in the solution of a differential equation, and in this case it is determined unambiguously by the beetle's initial position. Unfortunately this method, known as "separation of variables," falls short with equation (1). Let's try something else: "substitution of variables." We'll search for a solution of the form $x(t) = C(t)(L + vt)$, where $C(t)$ is a new unknown function. Substituting for

At right: "Will it catch him or not?"

Art by Leonid Tishkov


Догонит? не догонит?
ТИШКОВ 1993

this $x(t)$ in equation (1) results in the equation for $C(t)$,

$$\frac{dC}{dt} = \frac{u}{L+vt}, \tag{3}$$

which we can readily solve:

$$\int dC = \int \frac{udt}{L+vt},$$

$$C = \frac{u}{v}\ln(L+vt) + \text{const.}$$

By dint of the initial condition $x(0) = 0$, we finally obtain the result

$$x(t) = \frac{u}{v}(L+vt)\ln\left(1+\frac{vt}{L}\right).$$

Back to the beetle

Notice that as $t \to \infty$, the distance $x(t)$ increases faster than the first power of t. This means that the beetle will always reach the loose end however small its velocity $u << v$ (fig. 2). This will occur at the moment T when $x(T) = y(T)$—that is, when

$$T = \frac{L}{v}\left(e^{v/u} - 1\right). \tag{5}$$

So, the problem is solved. As we admire our solution, though, an uneasiness comes over us. Isn't there some arithmetical error in equation (5)? Let's test the solution. If $u = 0$, the beetle doesn't move and will never reach the loose end. Indeed, $T \sim e^{\infty} \sim \infty$. On the other hand, if $v = 0$, the rubber band is not stretched and $T = L/u$. The same value follows from equation (5), since as $v \to 0$,

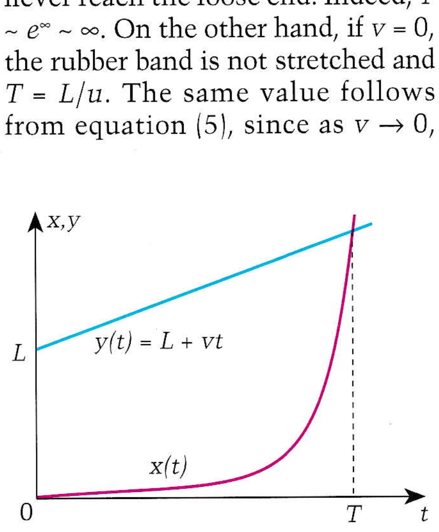


Figure 2

$e^{v/u} = 1 + v/u + 0(v/u)$. Everything's okay, and only one question remains.

Is there a shortcut?

Yes! It's based on a very simple idea. Consider the fraction of the total length $z(t) = x(t)/y(t)$ that the beetle manages to crawl in the time t. The rubber band's length increases, but no matter how much we stretch it, the fraction z of its length remains constant. Therefore, z is an additive value, and the "increment of the fraction" is equal to the "fraction of the increment." The increment dz taken for the time dt results in

$$dz = \frac{udt}{y(t)}. \tag{6}$$

If youhave any doubts about the correctness of equation (6), just substitute $z = x/y$. What do you get? Why, equation (1).

It's much easier to solve equation (6) than equation (1). Taking into account that $z(0) = 0$, we get

$$z(t) = u\int \frac{dt}{L+vt} + \text{const}$$

$$= \frac{u}{v}\ln\left(1+\frac{vt}{L}\right).$$

The beetle will arrive at the loose end when $z(T) = 1$, which results in equation (5), obtained previously for the time value T. Nice! In fact, this is the elegant solution that Andrey Sakharov found lickety-split.

The idea of using the fraction of the rubber band's length appears particularly attractive in the discrete variant of the beetle problem.

A slight modification

Let's assume that $L = 1$ km and $u = 1$ cm/s, and let's suppose that the rubber band is stretched stepwise: after the first second it's stretched 1 km, after another second again by 1 km, and so on. This variant of the problem was given in Martin Gardner's book *Time Travel*. So, after the first second the beetle has crawled $1/10^5$ of the way; after the second, only $1/(2 \cdot 10^5)$; and so on. The beetle will arrive at the other end after the nth second, when

$$\frac{1}{10^5}\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}\right) = 1. \tag{7}$$

The harmonic series $1 + 1/2 + 1/3 + \ldots$ diverges, which means that the relationship is valid for sufficiently large n. The series can be evaluated by the integral

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n} \cong \int_1^n \frac{dn}{n} = \ln(n),$$

and the larger n is, the more precise the estimate. The beetle will crawl the entire length of the rubber band in the time $n = \exp(10^5)$ s, which coincides with equation (5).

This variant of the problem provokes further thought. What will happen if the length of the rubber band *doubles* every second? By the end of the nth second the beetle will have crawled

$$\frac{1}{10^5}\left(1+\frac{1}{2}+\frac{1}{2^2}+\cdots+\frac{1}{2^n}\right)$$

of the length of the rubber band. The series in parentheses is a geometric series that converges to 2. This means that as $n \to \infty$, the beetle will approach the point marking $2/10^5$ of the continuously stretching rubber band. It's clear that the beetle will never reach the loose end of the rubber band.

And so, the beetle's fate depends on how the rubber band is stretched and how fast the beetle is crawling. This circumstance suggests that we try to generalize the beetle problem.

A digression before the generalization

We'll assume that the velocities of the beetle $u(t)$ and the end of the rubber band $v(t)$ are both arbitrary functions of time. How can we know whether the beetle will reach the loose end? Recall equation (6), which is universally valid. Integrating it, we obtain the equation for the duration T of the beetle's journey:

$$\int_0^T \frac{u(t)dt}{y(t)} = 1, \tag{8}$$

where the current length of the rubber band is

$$y(t) = L + \int_0^t v(t')dt'. \qquad (9)$$

The relation (8) is the criterion for determining whether the beetle will reach the other end of the rubber band. The answer is simple: it will occur only when equation (8) is valid for some value of T. As we saw above, the beetle is far from being successful in every case. Thus, by substituting different functions for $u(t)$ and $v(t)$ in equation (8), we can quickly learn the beetle's fate.

A few examples

1. *Stretching the rubber band with constant acceleration.* Substituting $u(t) = u = \text{const}$ and $v(t) = at$ in equations (8) and (9) yields

$$\arctan\left(\frac{aT^2}{2L}\right)^{1/2} = \left(\frac{aL}{2u^2}\right)^{1/2}. \qquad (10)$$

From this it follows that the beetle will not complete its journey if $aL/u^2 > \pi^2/2$. In the opposite case, when $aL/u^2 < \pi^2/2$, the beetle's trip will end successfully by the time $T = \sqrt{2L/a} \cdot \tan\sqrt{aL/2u^2}$. The condition obtained has a simple physical meaning: the beetle has enough time to crawl over the entire rubber band if during this time $\sim L/u$ the band's loose end does not acquire a velocity comparable to that of the beetle—that is, if $L/u \le u/a$.

2. *Constant acceleration of the beetle.* If the beetle also moves with a constant acceleration $u(t) = a_1 t$, then it will make it across the rubber band however small its acceleration $a_1 << a$. Equations (8) and (9) yield the duration of the trip

$$T = \left[2L\left(\exp\frac{a}{a_1} - 1\right)\Big/ a\right]^{1/2}.$$

3. *Stretching the rubber band n-fold per second.* In this case the length of the rubber band changes according to $y(t) = L \cdot \exp(t/\tau)$. For example, when $\tau = 1/\ln 2$ s, the length doubles each second. Then, with $u(t) = u = \text{const}$, condition (8) becomes

$$1 - e^{-T/\tau} = \frac{L}{u\tau}.$$

From here it follows that when $L < u\tau$, the beetle's trip will end after a time $T = \tau \cdot \ln [u\tau/(u\tau - L)]$. In the opposite case, when $L > u\tau$, the beetle will crawl forever. This condition means that the length of the rubber band must not increase significantly in the amount of time the journey takes ($\sim L/u$)—that is, $L/u \le \tau$.

4. *Exponential increase in the beetle's velocity.* If the beetle's velocity also increases exponentially $u(t) = u \exp(t/\tau_1)$, then if $1/\tau < 1/\tau_1 + u/L$, the trip will end successfully after a time $T = (1/\tau_1 - 1/\tau)^{-1} \ln [1 + L(1/\tau_1 - 1/\tau)/u]$. In the opposite case, the beetle will never complete its journey.

Trajectory of the beetle's success

The examples above show that there is a kind of competition between the functions $u(t)$ and $y(t)$ that are integrated in equation (8). If $u(t)$ gains the upper hand in the competition with $y(t)$, the total integral increases in such a way that it will reach 1 when $t = T$ (the red curve in figure 3). This means that the beetle managed to crawl the entire length of the rubber band. On the other hand, when $y(t)$ emerges victorious, the integral is too small and will never reach 1 (the blue curve in figure 3). Failure awaits the beetle in this case.

Well, that's it. Almost.

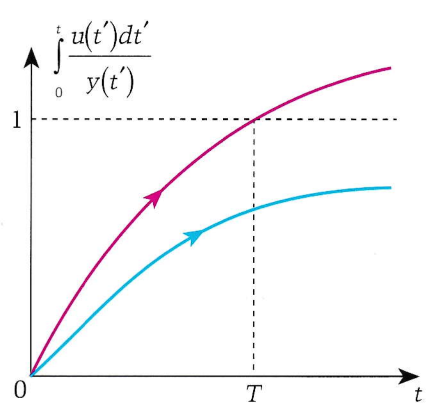


Figure 3

A few parting words

Let's go back to the very beginning of our inquiry, when we had just obtained equation (5) for the duration of the beetle's trip T. It depends exponentially on the relation of velocities v/u. This dependence is very strong. To demonstrate the power of an exponential function, let's assume u = 1 cm/s and v = 1 km/s, and let's plug them into equation (5). We get $T \sim \exp(10^5)$ s $\sim 10^{43400}$ s for the duration of the beetle's difficult journey, which far exceeds the age of the universe (10^{18} s). By journey's end, the rubber band will be longer than the universe is wide—it will measure a whopping 10^{28} cm.

Of course, a real bug would die quietly at the very outset of such a journey, and a real rubber band would become so thin that its molecules would be separated by vast distances of empty space. However, these considerations take nothing away from our detailed solution to the beetle problem.

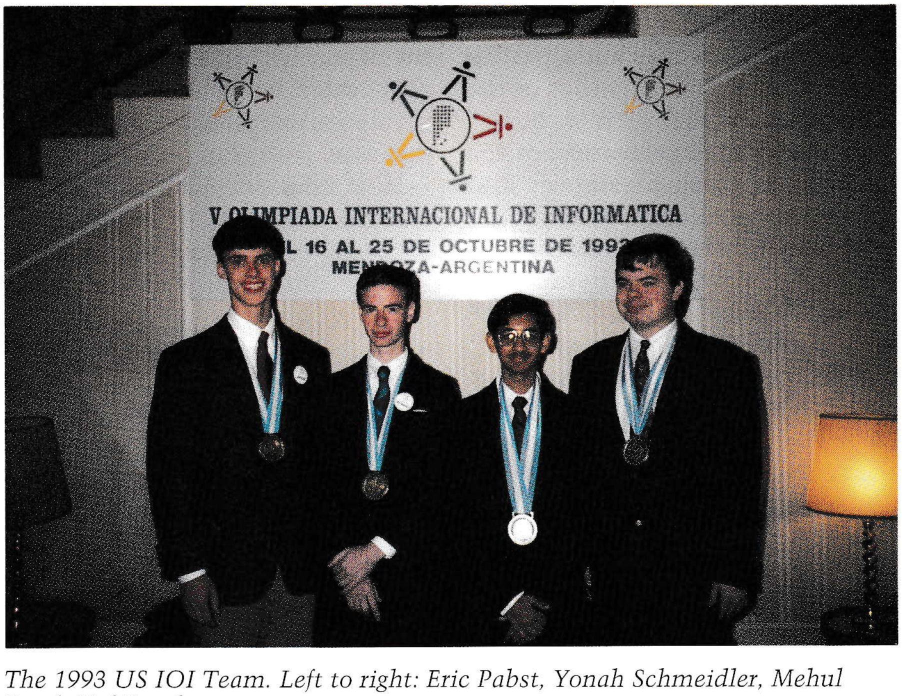
The Fifth International Olympiad in Informatics

Argentina serves up meaty problems to hungry young programmers

by Donald T. Piele

OUR ADVENTURE BEGAN Friday, October 15, at the Miami International Airport, where the USA Computing Olympiad team met for the first time since the summer training program at the University of Wisconsin–Parkside. Dr. Harold Reiter, the deputy team leader, flew back from London, where he was spending the year teaching mathematics at Kingston University. Team member Hal Burch, 18, flew in from Missouri, where he was a freshman at the University of Missouri at Rolla, having graduated in June from the Oklahoma School of Science and Mathematics in Oklahoma City. Eric Pabst, 17, came from Salt Lake City, Utah, where he was a senior at East High School, and Mehul Patel, 16, arrived from Houston, Texas, where he was a senior at Langham Creek High School. Yonah Schmeidler, 17, a graduate of Ramaz School in New York and now a freshman at MIT, had flown earlier to Buenos Aires and would meet up with us on Sunday. I flew in from Chicago's O'Hare airport after busing down from Wisconsin. Our next stop would be Santiago, Chile, with a connecting flight over the Andes mountain range to Mendoza, Argentina, the site of the Fifth International Olympiad in Informatics (IOI).

We left the United States at the peak of the fall colors and arrived in Mendoza in the full bloom of spring. We were met at the Mendoza airport by a contingent of college students from Mendoza University, whose job for the next ten days would be to guide the participants (273 students and team leaders from 45 countries) to various events within the city and one excursion to the Andes. Their enthusiasm and warmth were infectious. Eric savored the opportunity to try out the Spanish he had studied for five years, and he quickly established a special relationship with our hosts. His facility in the native language proved to be a big asset for him as well as all the members of the US team. Several times during our stay he would be called upon to give radio and TV interviews, talk with the press, help us translate stories that appeared in *Los Andes* (the local newspaper), and find the beef and chicken dishes on a restaurant menu.



The 1993 US IOI Team. Left to right: Eric Pabst, Yonah Schmeidler, Mehul Patel, Hal Burch.

In the shadow of Mt. Aconcaqua

IOI participants were housed in two hotels, and our team stayed at the Hotel Aconcaqua, named after the highest mountain in the Western Hemisphere, which is located near Mendoza. Fifty Compaq computers were set up on the hotel's second floor for students to use. This was a welcome sight for all teams and a place they often went to hash things over. Similar Compaq computers were housed in the Convention Center, approximately six blocks south of the hotel, where we had our meals and where the competition was held. The city of Mendoza had purchased 400 Compaq computers for the event, and they would be used by the city when the competition was over. There were enough computers around to completely outfit the team leaders' room with a networked system complete with e-mail and printing capabilities. This was a first for IOI and a very appreciated feature of this year's Olympiad. I used it to keep in touch with family and supporters back home.

On Sunday evening we all gathered at the Convention Center for the opening ceremonies. Argentinean officials, including the director of technology in education, the mayor of Mendoza, and the 1993 IOI organizer, Dr. Alicia Bañuelos, gave their addresses in Spanish, which were translated paragraph by paragraph into English, the official language at IOI. A festive mixer erupted soon after, with a Latin beat drowning out any attempt at conversation. The young hostesses, dressed in fashionable miniskirts, soon had the young group shaking and stomping to a fast Latin beat. It was nonstop aerobics and survival of the fittest, and I survived by watching and taking pictures from the sidelines.

Three problems to begin

Tuesday was the first day of competition. The team leaders and deputy team leaders were given a wake-up call at 3:30 A.M. so we would be ready for the early morning jury meeting at 5:00 A.M. in the Convention Center. The main order of business was the selection of three problems for the first day's competition. Those were selected from a set of nine problems submitted by the scientific committee from Argentina. Besides choosing the problems, all of the non-English-speaking countries needed to translate the problem statements into their native language and have copies made for each of their participants. There were approximately 35 different native languages represented, and everything needed to be ready at the appointed starting time.

At 11:00 A.M. the competition began, and 155 students went to their personal Compaq 386 machine, identified with a small flag of their country, in one of four different rooms, and "started their engines." They had five hours to solve three problems, using one of the officially installed languages: Turbo Pascal v. 6.0, Turbo C++ v. 2.0, QuickBasic v. 4.5, and LCN Logo v. 3.0. We had been working for six hours straight in a smoke-filled room, and it was time for a much-deserved rest. But before we could relax, we had to remain on call to translate any written questions the students might ask during the first hour. Then we were free to go and get some rest before the judging began.

At 4:00 P.M. the competition ended and the students filed out of their rooms with looks of confidence and relief. For the next several hours they would be called back, one at a time with their team leader, to have their programs checked by a local coordinator who had been trained to run the programs against a series of input data and evaluate the output file for the correct results. If all runs were perfect, the program was awarded 100 points. Hal and Mehul's programs were flawless, and Eric and Yonah's were close behind with 71 and 62. No scores are officially posted for the first day, but we quickly learned through word of mouth that a total of 16 students had perfect first-round scores.

Interlude: Argentinean delicacies

The next day was reserved for touring a local chocolate factory, followed by a barbecue at the country home of one of the organizers of the IOI. The feast began with trays filled with empanadas, a pastry filled with beef and spices and freshly baked in clay outdoor ovens.

One of the special treats in Argentina is to cook large hunks of fine beef very slowly over an open pit. The meat is then sliced off and placed on buns and topped with a special mustard sauce. This makes an excellent sandwich, and the lifetime of each platter full of meat could be measured in nanoseconds. Soon the hosts took to filling the platter and running through the crowd to reach those who weren't close enough to see the food before it vanished. As they traversed the lawn, scores of hungry participants reached out and snared their meal and quickly emptied the tray. It took quite a few runs to make it with anything left for the unlucky ones at the other end of the lawn.

After a delightful afternoon, we returned to our hotel to get ready for the final round.

Two hard problems (one unplanned)

Thursday morning began at 3:30 A.M. and was a repeat of Tuesday, except that this time one harder problem was selected from a set of three. One problem was eliminated because of its ambiguous wording and the difficulty of making it completely clear in 35 languages. Almost any problem can have different interpretations depending on how it's translated. For example, does the statement "all rectangles fall within the borders of an $a \times b$ sheet of paper" mean that rectangles can or cannot share a boundary with the sheet of paper? In English, the statement would imply that they could be on the boundary, but it all hinges on how you translate the word *within*. Explaining this in 35 languages can be difficult, so the jury overwhelmingly chose a problem that we believed had no ambiguities.

The problem was clear, but one thing we forgot to discuss was how the solutions to this problem would be graded. This, unfortunately, led to a major misunderstanding.

The trouble quickly became apparent when we walked into the computer room with the coordinator and saw for the first time the rules used to judge the eight sample runs. The first six data sets had a limit of two minutes and the last two a limit of five minutes. Everyone on the team had solutions that ran instantly for the first seven data points, but all ran over the five-minute limit for the last and most difficult data set. Since this run was worth 25 points, their hopes for a gold medal vanished as did the hopes for 12 other participants who had perfect scores the first day and also did not optimize for speed. They had fallen into the exponential time trap, which for many could have been avoided had they known that, for the first time at IOI, speed would be the deciding factor.

Last year, at the IOI in Germany, I was surprised to learn that the speed and efficiency of an algorithm wasn't considered a factor in grading. In fact, several programs were allowed to run for hours, even overnight, and others finished in seconds; yet no distinction was made between them. I thought this was rather odd, but everyone seemed to accept this as an unwritten rule of IOI. Students were aware of it, and we had told our team members to play it safe and go with any working algorithm and not to worry about speed unless it was explicitly stated in the problem.

It never occurred to the jury to ask how the problem would be graded, and when the time question surfaced after the competition was over, it was too late to correct. Many students were well aware that their programs could take years to complete if a large number of data points was used as test data, but since time had never been a factor before, they thought it wouldn't be a factor here. But this was not to be. The jury reacted to this situation by drafting additional competition rules to be considered for IOI '94, including: "When a time limit will be applied during evaluation, it should be explicitly stated in the problem description." Had this been done at IOI '93, it would have helped a great deal. Of course, speed of execution as a factor in grading solutions isn't a bad idea. Since this was the first Olympiad to broach the time barrier, it will now be on the minds of all team leaders as they prepare for IOI '94. In Mendoza, the omission affected everyone equally.

Into the windswept Andes

A long bus excursion into the mountains of the Andes was reserved for Saturday. Our final destination was Uspallata, a ski resort high in the mountains. Here we were treated once again to the famous open pit beef barbecue done on a grand scale. A wind barrier was set up around the pit to deflect the strong spring winds rolling off the mountains. The snow was completely gone, and so were all the people. But inside the dining hall, the participants were happy grazing on all the beef they could eat and toasting a local guide whose birthday had been discovered.

We soon had to leave to get back to an important jury meeting to decide the cutoff scores for the gold, silver, and bronze medals. It was a picturesque excursion into a dry, barren, and mountainous region. A few were able to drive to a spot where they could view Mt. Aconcaqua.

Back at the Convention Center the jury met to decide who would get the medals. According to the rules, only half of the students can receive a medal. This rule helps maintain the value of each award. Also, the gold, silver, and bronze awards must be given out in a ratio of 1 : 2 : 3, or at least as close as possible. Out of a possible 200 points, it worked out as follows: 13 gold medals were awarded for scores of 180–200; 27 silver medals for 160–179 points; and 39 bronze medals for 125–159 points.

Awards and plans

Hal Burch and Mehul Patel received silver medals, and Eric Pabst and Yonah Schmeidler got the bronze. Our team ranked seventh out of 45 in the total number of points, and for the first time two girls won silver medals, one from the Czech Republic and one from the Slovak Republic.

Seven teams won four medals:

	Pts	G	S	B
Slovak Republic	714	2	1	1
Romania	691	2	1	1
Russia	683	1	2	1
Iran	660	1	2	1
China	644	1	1	2
Korea	640	1	1	2
USA	633	–	2	2

Gold medals were also won by students from Sweden, the Czech Republic, Bulgaria, Belarus, and a United Nations team from Yugoslavia.

After the jury adjourned, the US delegation was invited to attend a meeting of the International Committee to see when we would be interested in hosting an Olympiad. Countries that had submitted proposals up to 1997 were Sweden, the Netherlands, Hungary, and South Africa. Several countries were invited to this meeting to announce tentative plans to submit proposals for years to come. They were Portugal (1998), Turkey (1999), China (2000), Thailand (2001), and Korea (2002). We were also interested in the year 2000, but since China has been a member of IOI longer, they were given precedence over any proposal from a newer member. Since 2003 is too far into the future to make any plans, Joann DiGennaro, president of the Center for Excellence in Education, our sponsor, didn't want to make a commitment at this time, and we decided to wait a year and see if the proposal for the year 2000 materializes.

The awards ceremony was held on Sunday and began at 9:30 A.M. in the Independence Theater. All medal winners were seated on the stage, and the delegates, other participants, and spectators were seated in the audience. After the opening ceremonies each team leader was invited to the stage to present the medals to their team members, starting with the bronze and ending with the silver. For the gold medal winners, the students received their award and prizes from local dignitaries. The top four students, who were tied at 200 points each, received computers and were awarded a new IFIP trophy that will go each year to the top student or students at the IOI.

Picture were taken as the trophy was hoisted into the air by four excited and deserving young men from the Czech Republic, Romania, Iran, and Sweden. The torch was passed to Sweden, whose team invited us all to the 1994 IOI in Stockholm, and the curtain rang down on another successful International Olympiad in Informatics. Thank you, Argentina, for a wonderful Olympiad. Our memories of your warm hospitality will always be with us.

On behalf of this year's US delegation to IOI, I would like to express our warm appreciation to the Center for Excellence in Education (CEE) and Joann DiGennaro, who funded the USA Computing Olympiad training program at the University of Wisconsin–Parkside and the IOI team's trip to Argentina. We are most grateful for the generous support we received from CEE.

I also want to thank USENIX for its financial contribution to the USA Computing Olympiad and the University of Wisconsin–Parkside for in-kind support of the USACO.

For more information about IOI, write or call

Donald T. Piele
USACO Director
University of Wisconsin–Parkside
Box 2000
Kenosha WI 53141-2000
piele@cs.uwp.edu
414 595-2231 (O)
414 634-0868 (H)

Problems from IOI V

First-round problems

Problem 1

You have a necklace of n beads ($n \le 100$) some of which are red, others blue and others white, arranged at random. Let's see two examples for $n = 29$:

```
       1 2                 1 2
     o x x o             x o o x
    o      x            x      x
   o        o          x        o
  o          o        @          o
 x            o      @            @
x              x    o              o
x              x    x              x
 x            x      o            x
  o          o        x          o
   x        o          o        o
    x      o            o      o
     o    o              o    x
      o x o               o o @
```

Figure 1 Figure 2

o = red bead
x = blue bead
@ = white bead

(The beads considered first and second in the text that follows have been marked in the figures.)

The configuration in figure 1 may be represented as a string of b's and r's, where "b" represents a blue bead and "r" a red one, as follows: brbrrrbbbrrrrrbrrbbrbbbbrrrrb.

Suppose you are to break the necklace, lay it out straight, and then collect beads of the *same* color from one end until you reach a bead of a different color, and do the same for the other end (which may not be of the same color as the beads collected before this). Determine the point where the necklace should be broken so that the *greatest* number of beads can be collected.

For example, for the necklace in figure 1, eight beads can be collected, with the breaking point either between beads 9 and 10 or between beads 24 and 25.

In some necklaces, white beads had been included as shown in figure 2. When collecting beads, a white bead that is encountered may be treated as either red or blue and painted with the desired color. The string that represents this configuration will include the symbols r, b, and w.

Write a program to do the following:

1. Read a configuration from an ASCII input file NECKLACE.DAT with each configuration in one line. Write these data into an ASCII output file NECKLACE.SOL. An example of an input file would be

```
NECKLACE.DAT
brbrrrbbbrrrrrbrrbbrbbbbrrrrb
bbwbrrrwbrbrrrrrb
```

2. For each configuration, determine the maximum number M of beads collectable, along with the breaking point.

3. Write to the outfile NECKLACE.SOL the number M and the breaking point. The solutions for different configurations should be separated with a blank record. Example of a possible solution:

```
NECKLACE.SOL
brbrrrbbbrrrrrbrrbbrbbbbrrrrb
8 between 9 and 10

bbwbrrrwbrbrrrrrb
10 between 16 and 17
```

Problem 2

Some companies are partial owners of other companies because they have acquired part of their total shares. For example, Ford owns 12% of Mazda. It is said that a company A controls company B if at least one of the following conditions is satisfied:

(a) A = B;
(b) A owns more than 50% of B;

(c) A controls k ($k \geq 1$) companies C(1), ..., C(k) so that C(i) owns $x(i)$% of B for $1 \leq i \leq k$ and $x(1) + \ldots + x(k) > 50$.

The problem to solve is:

Given a list of triples (i, j, p), which means that the company i owns p% of company j, calculate all the pairs (h, s) so that company h controls company s. There are at most 100 companies.

Write a program to do the following:

1. Read from an ASCII input file COMPANY.DAT the list of triples (i, j, p) to be considered for each case (that is, each data set), where i, j, and p are positive integers. Different cases (data sets) will be separated with a blank record.

2. Find all the pairs (h, s) so that company h controls company s.

3. Write to an ASCII output file COMPANY.SOL all the pairs (h, s) found, with h different from s. The pairs (h, s) must be written in consecutive records and in increasing order of h. The solutions for different cases must be separated with a blank record.

Example:

```
COMPANY.DAT     COMPANY.SOL
2  3  25        4  2
1  4  36        4  3
4  5  63        4  5
2  1  48
3  4  30
4  2  52
5  3  30

1  2  30        2  3
2  3  52        2  4
3  4  51        2  5
4  5  70        3  4
5  4  20        3  5
4  3  20        4  5
```

Problem 3

N rectangles of different colors are superposed on a white sheet of paper. The sheet's dimensions are a cm wide and b cm long. The rectangles are put with their sides parallel to the sheet's borders. All rectangles fall within the borders of the sheet. As a result, different figures of different colors will be seen. Two regions of the same color are considered to be part of the same figure if they have at least one point in common; otherwise they are considered different figures. The problem is to calculate the area of each of these figures. The numbers a and b are positive integers not greater than 30. The coordinate system considered has its origin at the sheet's center and the axes are parallel to the sheet's borders.

Different data sets are written in an ASCII input file RECTANG.DAT in which a, b and N are the first line of each data set, separated by a blank space. In each of the next N lines you will find

- The integer coordinates of the position where the left lower vertex of the rectangle was put;
- Followed by the integer coordinates of the position where the upper right vertex of the rectangle was put;
- And then the rectangle's color represented by an integer between 1 and 64, the color white represented by 1.

The order of the records corresponds to the order used to put the rectangles on the sheet. Different data sets will be separated with a blank record.

Write a program to

1. Read the next data set from RECTANG.DAT.

2. Calculate the area of each colored figure.

3. Write in an ASCII output file RECTANG.SOL the color and the area of each colored figure as shown in the example below. These records will be written in increasing order of color. The solutions to different data sets will be separated with a blank record.

Example:

```
RECTANG.DAT          RECTANG.SOL
20 12 5              1 172
-7 -5 -3 -1 4        2 47
-5 -3 5 3 2          4 12
-4 -2 -2 2 4         4 8
2 -2 3 -1 12         12 1
3 1 7 5 1

30 30 2              1 630
0 0 5 14 2           2 70
-10 -7 0 13 15       15 200
```

Second-round problem

You have won a contest sponsored by an airline. The prize is a ticket to travel around Canada, beginning in the westernmost point served by this airline, then traveling only from west to east until you reach the easternmost point served, and then coming back only from east to west until you reach the starting city. No city may be visited more than once, except for the starting city, which must be visited exactly twice (at the beginning and the end of the trip). You are not allowed to use any other airline or any other means of transportation. Given a list of cities served by the airline and a list of direct flights between pairs of cities, find an itinerary that visits as many cities as possible and satisfies the above conditions beginning with the first city and visiting the last city on the list and returning to the first city.

Different data sets are written in an ASCII input file ITIN.DAT. Each data set consists of the following:

- In the first line, the number N of cities served by the airline and the number V of direct flights that will be listed. N will be a positive integer not larger than 100. V is any positive integer.
- In each of the next N lines: a name of a city served by the airline. The names are ordered from west to east in the input file—that is, the ith city is west of the jth city if and only if $i < j$. (There are no two cities on the same meridian). The name of each city is a string of at most 15 digits and/or characters of the Latin alphabet—for example, AGR34 or BEL4. (There are no spaces in the name of a city.)

- In each of the next V lines: two names of cities, taken from the list of cities, separated by a blank space. If the pair *city1 city2* appears in a line, it indicates that there exists a direct flight from *city1* to *city2* and also a direct flight from *city2* to *city1*.

Different data sets will be separated by a empty record (that is, a line containing only the end-of-line character). There will be no empty record after the last data set. The following example is stored in the file ITIN.DAT:

```
ITIN.DAT
8 9                     5 5
Vancouver               C1
Yellowknife             C2
Edmonton                C3
Calgary                 C4
Winnipeg                C5
Toronto                 C5 C4
Montreal                C2 C3
Halifax                 C3 C1
Vancouver Edmonton      C4 C1
Vancouver Calgary       C5 C2
Calgary Winnipeg
Winnipeg Toronto
Toronto Halifax
Montreal Halifax
Edmonton Montreal
Edmonton Yellowknife
Edmonton Calgary
```

The input may be assumed correct. No checking is necessary.

The solution found for each data set must be written to an ASCII output file ITIN.SOL: in the first line, the total number of cities in the input data set; in the next line, the number M of different cities visited in the itinerary; and in the next $M + 1$ lines, the names of the cities, one per line, in the order in which they are visited. Note the first city visited must be the same as the last. Only one solution is required. If no solution is found for a data set, only two records for this data set must be written in ITIN.SOL: the first one giving the total number of cities, the second saying "NO SOLUTION."

A possible solution for the above example:

```
ITIN.SOL
8              5
7              NO SOLUTION
Vancouver
Edmonton
Montreal
Halifax
Toronto
Winnipeg
Calgary
Vancouver
```

Put your program solution into an ASCII file named DDD.xxx. The extension .xxx is .BAS for QBasic, .LCN for Logo, .C for C, and .PAS for Pascal.

Bulletin Board

American Regions Math League

If you like math, and you like playing on a team, the American Regions Math League (ARML) may be your cup of tea. This year the ARML competition will be held on June 4 at two sites: Pennsylvania State University and the University of Iowa. When the "power question" is posed, you will have a chance to work it through with your teammates, producing a single answer paper. Later, "quick problems" are given, and team members decide who will tackle what. And in the "relay round," the answer to each question forms part of the next, and only the final answer is scored.

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 W. 205th St., Bronx NY 10468.

"How's the weather up there?"

Next year, students at 25 US schools will have the chance to chat with an astronaut aboard the space shuttle via amateur radio. The Shuttle Amateur Radio EXperiment (SAREX) program is currently soliciting applications for a limited number of openings. Selections are made jointly by the American Radio Relay League (ARRL), the National Aeronautics and Space Administration (NASA), and the Radio Amateur Satellite Corporation (AMSAT).

Urban, suburban, and rural schools are encouraged to apply. To be eligible, school officials must complete an application and write a proposal that shows how they will integrate the program into the classroom. Applicants must also prove that they have the support of local radio amateurs (popularly known as "hams") to qualify for the program.

For information on how to apply to the SAREX program, write to Tracy Bedlack at the American Radio Relay League, 225 Main St., Newington CT 06111, or call 203 666-1541.

Wanted: a certain back issue

Occasionally readers call or write to *Quantum*, asking for copies of the September/October 1990 issue, which is out of print and unavailable. If any readers have a copy that they are willing to part with, or would like a copy, please contact us. We will maintain a list and try to link seller with purchaser.

Write to Tim Weber, Managing Editor, *Quantum*, 1840 Wilson Blvd., Arlington VA 22201-3000 (e-mail: 72030.3162@compuserve.com).

Readers write . . .

A. John Mallinckrodt, associate professor of physics at the California State Polytechnic University in Pomona, found that we came up short in our discussion of "telephoto shooting." He writes:

Contrary to a problem solution published in the January/February 1994 issue (Challenges in Physics and Math, P105), the apparent foreshortening of objects observed through a telephoto lens is quite definitely *not* a result of the ratio of transverse to longitudinal magnifications. To demonstrate this, it is enough simply to note that this ratio would seem to predict precisely the *opposite* effect.

The solution itself properly derives the ratio of transverse to longitudinal magnifications as $(L - F)/F$, where L is the distance to the object and F is the focal length of the lens. Unfortunately, it goes on to misinterpret this as a "flattening" factor in spite of the fact that this expression obviously *decreases* as F increases. (To see why this ratio is, in fact, irrelevant to the explanation of foreshortening, it is important to keep in mind that ordinary planar imaging systems—like the retina and film—have no way of conveying information about longitudinal magnification.)

The foreshortening effect is actually a relatively simple (and at least partially psychological) consequence of the fact that when everything in a view "looks" closer, then everything will necessarily also "look" foreshortened. Telescopes, binoculars, and telephoto lenses make things "look" closer by increasing the angular size of their images; when the subtended angle of an image is increased by a factor of m, the object it represents "looks" m times closer. Suppose, for example, that the front of a car is 100 meters away and the back is 104 meters away. Quite clearly if we make both "look" four times closer we have a car that "looks" like it is 1 meter in length.

The oscillating ring in physics challenge P100 in the November/December issue impelled Rouben Rostamian (Department of Mathematics and Statistics, University of Maryland–Baltimore County) to write:

The solution points to an interesting phenomenon: the period of oscillations of this system is identical to that of a simple pendulum of length L.

Both in the statement of the problem and its solution, it is assumed that $r = L/2$. The problem is actually more interesting than that. An inspection of the solution indicates that the assumption $r = L/2$ is a red herring—the period of oscillations is independent of the radius of the ring! The unfortunate wording of the problem may mislead some readers into believing that the equality of the oscillatory and flexural periods is a consequence of the $r = L/2$ assumption.

In that same issue, Mary E. Violett of Haymarket, Virginia, noticed something rotten in "Bushels of Pairs." On page 6, column 1, equation (4), "$x \neq 0$" should have been "$x \neq 1$," and "$x = 0$" should have been "$x = 1$."

Robert A. Rosenbaum (University Professor of Mathematics and the Sciences, emeritus, at Wesleyan University in Connecticut) found an alternative approach to brainteaser B99 that "may be of interest because it uses a physical principle to obtain a mathematical result of which the conclusion to B99 is a corollary." He writes:

Let A, B, C, D be any four points (not necessarily lying in one plane, not necessarily with no three collinear, not necessarily forming a convex quadrilateral). Let P be the midpoint of segment AB, Q of BC, R of CD, S of DA, T of AC, and U of BD. Then the segments PR, QS, TU are concurrent at point O, which is the midpoint of each of these three segments.

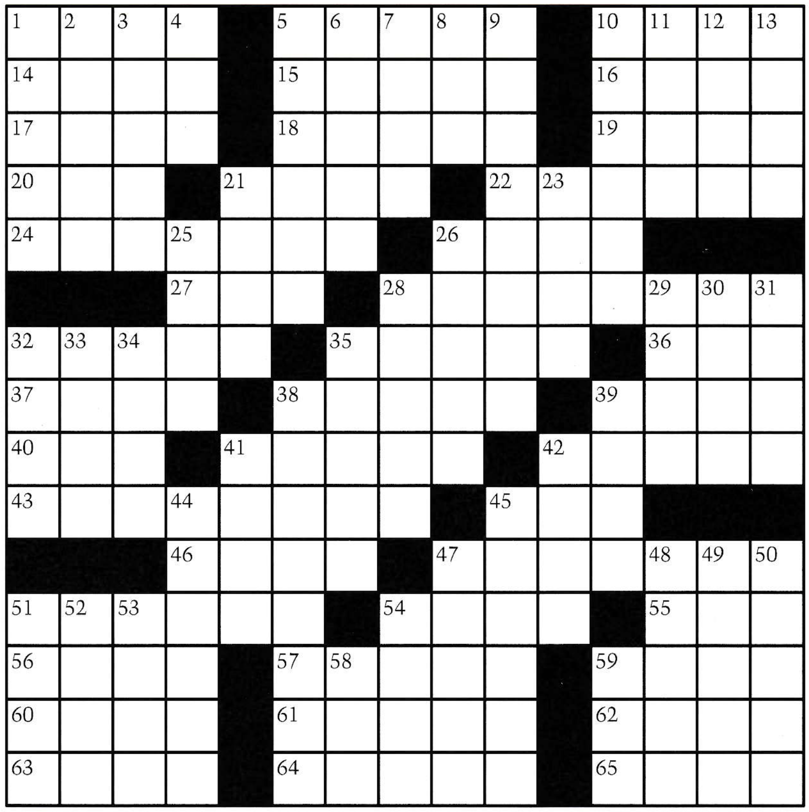
The result can be obtained by placing a mass m at each of A, B, C, D. To locate the center of mass of the system we can replace the masses at A and B by a mass $2m$ at P; and we can replace the masses at C and D by a mass $2m$ at R. Then we can replace the masses at P and R by a mass $4m$ at O, which is then the center of mass of the original system. But, starting again, . . ."

Jim Moskowitz at the Franklin Institute Science Museum noticed a flub in math challenge M93 (September/October 1993). The statement should say "positive integers," not simply "integers." As our correspondent noted, "If any integers are allowed, there are lots of pairs of integers whose sum is 30,030 and whose product is divisible by 30,030. For example, {60,060, –30,030} or the trivial {30,030, 0}."

We thank all our readers who have taken the time to send us their comments and corrections.

Criss × cross science

by David R. Martin



Across

1 Large plasma ball
5 Football's "Papa Bear"
10 Happy
14 Heraldic border
15 Abreast
16 Capital of Peru
17 Placed
18 Simple tool
19 Specified measure
20 Hydrogen and oxygen
21 Procreate
22 Liquid alkane
24 Mass/volume
26 Girl's given name
27 Single
28 ___ motion (of tiny particles)
32 Sesame seed
35 Rugged rocks
36 Jen___ Chou (Chinese essayist)
37 Whoopee! (in ancient Greece)
38 Roman writer
39 Dues
40 Distilled coal
41 Flat surface
42 Unpleasant
43 Cell division stage
45 Understand
46 Decays
47 Equilibrium mechanics
51 Noncorrodible alloy
54 Plant part
55 Transform mathematically
56 Anthropologist ___ Hrdlicka
57 French river
59 ___ oil (juniper tar oil)
60 Paper quantity
61 Jewish calendar month (var.)
62 Monster
63 Danson and Knight
64 Abounds
65 Plant starter

Down

1 State of matter
2 Sum of diagonal matrix elements
3 Strange
4 Ruby
5 Rock salt
6 Mimicry
7 Wash
8 Expert
9 Study of blood serum
10 Wheat protein
11 Soviet physiologist Stern
12 Idi ___
13 Appointment
21 Trigonometric function
23 Female bovines
25 Unit of loudness
26 Lifting device
28 H_2O and NaCl
29 Willow
30 Alloy of gold and silver
31 Inquisitive
32 ___ particle (emitted electron)
33 ___ Picone
34 Writer Ephron
35 Organism category
38 Blood particle
39 Units of length
41 Unit of illuminescence
42 Joint
44 Nicol and Wollaston, e.g.
45 Iron alloys
47 Play a guitar
48 Real or virtual follower
49 Group of leaders
50 Distance/time
51 Hand truck
52 Toward shelter
53 Element 82
54 ___ *et philosophus esto*
58 Spanish cheer
59 Trig function

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE JANUARY/FEBRUARY PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M106

The answer to the general question (b) is $n-2$ (for $n = 5$, the answer is 3). Clearly, the number in question can't be less than $n-2$, because an n-gon can be cut into $n-2$ triangles (by all diagonals drawn from one vertex, for example), and every triangle must contain a marked point. To see that the number $n-2$ is sufficient for a convex n-gon $A_0A_1 \ldots A_{n-1}$, let's mark a point B_k, $k = 1, 2, \ldots, n-2$, inside the triangle $A_0A_kA_{n-1}$ very close to the vertex A_k (see figure 1)—more exactly, in the triangle cut off from the triangle $A_0A_kA_{n-1}$ by the line $A_{k-1}A_{k+1}$. Any triangle formed by three vertices of the polygon can be written as $A_iA_kA_j$, $0 \le i < k < j \le n-1$. It contains point B_k, because angle $A_iA_kA_j$ contains angle $A_{k-1}A_kA_{k+1}$, and the diagonal $A_{k-1}A_{k+1}$ crosses the segments A_iA_k and A_jA_k. (N. Vasilyev)

M107

In the situation described, we can prove that $AM^2 + BN^2 + CP^2 = AP^2 + BM^2 + CN^2$. Substituting the given values yields $3^2 + 4^2 + 5^2 = 5^2 + 2^2 + x^2$, and $x = \sqrt{12}$.

To see that $AM^2 + BN^2 + CP^2 = AP^2 + BM^2 + CN^2$, we draw AO, BO, and CO (fig. 2). Then $AO^2 + BO^2 + CO^2 = (AM^2 + OM^2) + (BN^2 + ON^2) + (CP^2 + OP^2) = (AP^2 + OP^2) + (BM^2 + OM^2) + (CN^2 + ON^2)$. Equating the two expressions on the right and cancelling like terms gives the desired result.

M108

Somewhat unexpectedly, each term on the left side of the inequality can be estimated separately. We'll prove that

$$\frac{t}{1-t^2} \ge \frac{3\sqrt{3}}{2}t^2$$

for all t, $0 \le t < 1$. On this interval, the inequality is equivalent to

$$f(t) = \frac{3\sqrt{3}}{2}t\left(1-t^2\right) \le 1,$$

which can be proved by a routine analytical investigation of the function: find the derivative

$$f'(t) = \frac{3\sqrt{3}}{2}\left(1-3t^2\right).$$

It has only one zero in $[0, 1)$: $t_0 = 1/\sqrt{3}$. Since $f(0) = f(1) = 0 < f(1/\sqrt{3}) = 1$, t_0 is the maximum point of $f(t)$—that is, $f(t) \le f(1/\sqrt{3}) = 1$ for all t in $[0, 1)$.

It follows that the left side of the inequality in the problem is not less than

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

Equality is achieved only at the point $x = y = z = 1/\sqrt{3}$.

M109

Consider the 90° rotation R of the square about its center, taking B into A and A into D (in figure 3 it's a counterclockwise rotation). Let's show that this rotation takes the segment KL into MN—this will prove the statement of the problem.

Since $\angle KBA = 90° - \angle KAB = \angle KAD$ and, likewise, $\angle KAB = \angle ADM$, the ray BK is rotated into the ray AM (these rays form equal angles with rays BA and AD, respectively, and BA goes into AD by the definition of our rotation), and the ray AK is rotated into DM. Consequently, the intersection point K of rays BK and AK is taken into the intersection point of AM and DM—that is, into M: $R(K) = M$. In the same way we can prove $R(L) = N$, which means $R(KL) = MN$. This completes the proof. (V. Dubrovsky)

M110

If you've looked into "The Light at the End of the Tunnel" in this issue, you may have noticed that this problem is related to so-called *monovariants*—a useful device for

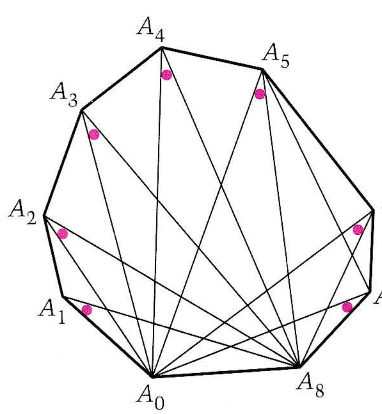


Figure 1

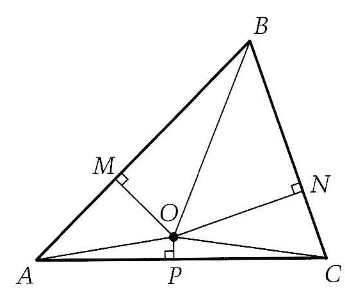


Figure 2

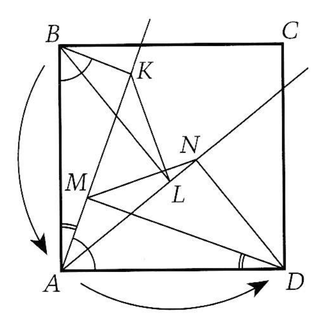


Figure 3

dealing with a certain class of problems. Refer to that article for details and compare the solution below to the method discussed there.

For every student, take the number of the room he or she lives in, square this number, and denote by s the sum of all these squares (the number of terms in this sum is equal to the number of students).

Every time two students living in rooms k and $k+1$ move apart, the sum s increases by $[(k-1)^2 + (k+2)^2] - [k^2 + (k+1)^2] = (2k^2 + 2k + 5) - (2k^2 + 2k + 1) = 4$.

Let's show that it can't increase forever.

Note that if someone were staying in any of three successive rooms $k-1$, k, and $k+1$ on some day, at least one of these rooms will be occupied on any subsequent day. Indeed, one can leave the block of these rooms only by moving from room $k+1$ to room $k+2$, or from $k-1$ to $k-2$. But either move can occur only if there is a student in room k who moves to room $k-1$ or $k+1$, respectively. In any case this student remains in the three-room block. It follows that a student from room k can't get to a room with a number greater than $k + 3N$, where N is the number of students; otherwise in each of the $N+1$ three-room blocks $(k, k+1, k+2)$, $(k+3, k+4, k+5)$, ..., $(k+3N, k+3N+1, k+3N+2)$ at least one of the N students must stay—which is, of course, impossible. Similarly, our student can't go in the opposite direction farther than room $k - 3N$. So if a and b $(a \le b)$ are the smallest and the greatest numbers of rooms occupied originally, then our young pianists will always stay in rooms with numbers between $a - 3N$ and $b + 3N$. Therefore, the sum s considered above cannot be greater than Nm, where m is the greatest of the numbers $(a - 3N)^2$ and $(b + 3N)^2$. So s will stop increasing some day, which means that students will stop moving.

The sum s in this solution could be replaced by other characteristics of the students' distribution in the rooms. Try, for instance, the sum of the "distances" between all pairs of students, where "distance" is understood as the absolute value of the difference between the numbers of the rooms where the two students live.

Physics

P106

There is no "most correct answer" to estimation problems like this one. What's important is the route one takes to the solution. You have to supply some of the data yourself, approximating the magnitudes based on your knowledge of familiar phenomena.

So, what is the approximate value for the velocity needed for an astronaut to leave the surface of the asteroid? We know that a person can jump to a height $h \cong 1$ m without extreme effort (one can do better without a spacesuit, but the point is, an astronaut isn't a grasshopper or an Olympic high-jumper!). This corresponds to a takeoff velocity equal to

$$v_0 = \sqrt{2gh} = 4.4 \text{ m/s}.$$

Let's assume that the density of the asteroid is equal to the average density of the Earth ρ. Then the acceleration due to gravity on the surface of an asteroid of radius r is

$$g_a = G\frac{M_a}{r^2} = G\rho\frac{4\pi r}{3}.$$

The corresponding value for the Earth, which has a radius $R = 6{,}400$ km, is

$$g = G\rho\frac{4\pi R}{3}.$$

The so-called escape velocity (the minimum velocity needed to leave a planet or other massive object) is obtained in a straightforward way by equating the kinetic energy at the surface to the change in the gravitational potential energy in going from the surface to infinity $\frac{1}{2}mv^2 = GM_a m/r$:

$$v_0 = \sqrt{2g_a r} = \sqrt{G\rho\frac{8\pi r^2}{3}} = \sqrt{\frac{2gr^2}{R}} = r\sqrt{\frac{2g}{R}}.$$

Equating this value to the initial velocity of the jump gives us the radius of the asteroid:

$$r = \sqrt{Rh} = 2.5 \text{ km}.$$

This is a reasonable value for an actual asteroid. It's clear that our answer wouldn't change drastically if we took another value for the height of the jump. You can show that our answer doesn't depend very much on the assumption that the densities of the asteroid and Earth are the same.

P107

This is another estimation problem. Let the balls of fat be the same size and uniformly distributed in the milk. We can determine the number of balls if we recall that the fat content of milk in stores ranges from negligible (skim milk) to about 5%. Let's choose 2%. Taking into account the approximate equality of the densities of fat and water, we'll assume that fat accounts for 0.02 of the milk's volume.

Let's designate the radius of the ball r and the thickness of the film of spilled milk h. Then the number of balls of fat N in a puddle with an area S can be found from the equation

$$N\frac{4\pi r^3}{3} = 0.02Sh.$$

The balls cover the area of the puddle completely with their cross sections. Thus,

$$\pi r^2 N = S.$$

Dividing the first equation by the second yields an estimate of the radius of a ball of fat:

$$r = 0.015h.$$

The thickness of the film can be estimated by using the known value of surface tension for water, but for our rough estimate we can use a value from our own milk-spilling experience—say, 1–2 mm. So the approximate radius of a ball of fat in milk is 0.01–0.02 mm.

P108

From the statement of the problem it's clear that the power of the heating element is equal to the power dissipated in the surroundings (the temperature of the water does not change with time). So if we turn the heater off, the dissipated power will be p = 100 W, and the time it takes the water to cool 1°C is equal to

$$t = \frac{cm\Delta T}{p} \cong 42 \text{ s},$$

where the specific heat c for water is 4.2 kJ/kg · C°.

P109

Because the sum of the voltages across the capacitors is always equal to the voltage of the battery V_0, when the voltage across one of them reaches the maximum, the voltage drop across the other falls to the minimum. Because there is no current flowing through either capacitor, there is no current in the coil either, and so the total energy of the system is equal to the sum of the energies of the capacitors. Notice that when the second capacitor is connected a certain amount of heat energy Q can be dissipated (it's also possible that no energy is dissipated if at the moment of connection the voltage drop across the first capacitor is equal to the battery's voltage, so that the second capacitor is connected to a source with zero voltage). The total work of the battery can be expressed in terms of the charge of the first capacitor. So we can write

$$\frac{q_1}{C} + \frac{q_2}{C} = V_0,$$

$$q_1 V_0 = \frac{q_1^2}{2C} + \frac{q_2^2}{2C} + Q.$$

Eliminating q_1 from these equations, we obtain the following equation for the second capacitor:

$$\frac{q_2^2}{C} = \frac{CV_0^2}{2} - Q.$$

It's clear that the maximum value of q_2 will be achieved when there is no heat dissipation in the system. So, finally,

$$q_2 = \frac{CV}{\sqrt{2}}.$$

Note that the value for the inductance L does not appear in the answer. The inductor determines the time scale for the problem, but does not affect the maximum values of the capacitors' energies.

P110

If we are close to the chain link fence, even a single piece of wire can block a substantial portion of our field of view. Moving away from the fence, we increase our field of vision, and the relative amount of light coming from the tennis players behind the fence and from the individual pieces of wire in the fence changes in favor of the former—approximately the ratio of the area of the holes in the fence to the overall area of the fence. (A player is now covered with many more holes.)

When we move briskly along the fence, we can see the players more clearly because, due to the persistence of our vision, several different momentary images now take part in forming the image of the players that we perceive. We see different parts of the players in the various images so that, by "summing" all the images, we get a single blurred but substantially more transparent fence, and the players are much more visible.

Brainteasers

B106

Answer: 6, 5, 4, 6, 5, 4, 6, 5, 4, 6, 5. Let a, b, c, d be any four consecutive numbers. Then $a + b + c = b + c + d$, and so $a = d$. Thus, the sequence on the blackboard must be periodic with a period of 3—that is, it can be written as $a, b, c, a, b, c, a, b, c, a, b$. By the statement of the problem, $a = 6$, $c = 4$. Therefore, $b = 15 - 6 - 4 = 5$.

B107

Let r and R be the radii of the inner and outer circular paths. If θ is the angle between the radial lines through the starting and ending points, then the value of θ (in radians) for which the two walks have the same length is given by

$$R\theta = r\theta + 2(R - r).$$

The answer, $\theta = 2$ radians (about one third of the way around the circle), is independent of the two radii. Thus, even if there is no inner circle ($r = 0$) but you are limited to only the circular and radial paths, it's better to stay on the circular path if you are going less than one third of the way around.

B108

Let's consider the motion of the particles of the surface of the river after the stone drops into it as the sum (superposition) of two motions: wave motion in the form of expanding circles and a translational motion with the velocity of the river. The waves will form expanding circles with a center moving with the velocity of the river. (We're assuming here that all portions of the river have a uniform velocity.)

B109

The answer is 942/1413. Since $NINE = (3/2)SIX < 1{,}500$, we know that $N = 1$. Now we can write the equation in the form $2{,}020 + 200I + 2E = 300S + 30I + 3X$, or

$$300S = 2{,}020 + 170I + 2E - 3X.$$

The right side of this equation is certainly larger than 1,800, so S cannot be less than 7. Now we examine three possibilities. For $S = 7$, we have $80 = 170I + 2E - 3X$, which is impossible. For $S = 8$, we have $380 = 170I + 2E - 3X$, or $40 - 2E + 3X = 170(I - 2)$, which also is impossible, because the left side of this equation is greater than zero and less than, say, 70, so it's not divisible by 170.

Finally, we put $S = 9$ and get $680 = 170I + 2E - 3X$, or $170(4 - I) = 2E - 3X$. It follows that $I = 4$, and $2E = 3X$. Taking into account that E and X must be different numbers not equal to $9 = S$ and $4 = I$, we arrive at the unique possibility $X = 2$, $E = 3$.

B110

Join the center O of the circle to all the five given points (fig. 4). Then the angles AOC, BOD, EOC are right angles, because each of them is a central angle that intercepts the same arc as the corresponding inscribed angle measuring 45°. It follows that the center O lies inside the

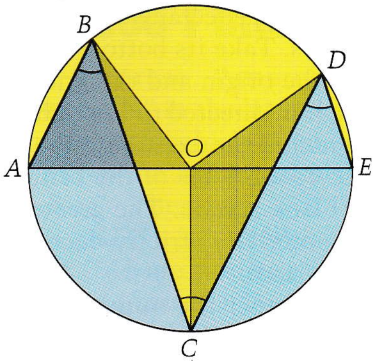


Figure 4

angle BCD, and the yellow area is divided by the radii drawn above into five pieces: two circular segments AB and DE, circular sector BOD, and two triangles OBC and ODC. Note that $\angle DOC = 180° - \angle AOB$, $\angle BOC = 180° - \angle DOE$. Therefore, the area of triangle ODC, which is equal to $\frac{1}{2}OC^2 \sin \angle DOC$, is equal to the area of triangle AOB ($\sin \angle DOC = \sin \angle AOB$), and area($\triangle OBC$ = area($\triangle ODE$). Now, if we replace the triangles ODC and OBC in the yellow figure by triangles OAB and ODE with the same areas, respectively, we'll turn this figure into the semicircle $ABCDE$ without changing its area, which, consequently, equals the blue area. (V. Dubrovsky)

Kaleidoscope

In the tic-tac-toe on the torus, the player who makes the first move can always force a win. The first move is irrelevant (because on a toroidal board all the squares are equivalent). The second move of the first player can be any but the silliest one (which marks the square in line with the first X and O). The remaining moves are practically uniquely determined.

When one torus is turned inside out through a hole in it, the second one, initially linked with the first, is absorbed by it, and will finally end up completely inside the first torus.

Toy store

1. It's clear that a peaceful arrangement of seven queens must leave one row and one column free of queens. If we cut the chessboard with a number of queens on it along the line between two files and exchange the two pieces (together with the queens), we'll get a peaceful arrangement whenever it was a peaceful arrangement initially, because a full diagonal becomes another full diagonal after this operation. (In fact, these two arrangements represent simply the *same* arrangement on the torus: they are obtained by cutting the torus along different circles—see figure 3 in the Kaleidoscope.) Using this operation, we can rework any arrangement of no more than seven queens so as to move free files to the extreme top and right positions. Below we consider only arrangements on the remaining 7×7 chessboard. Any symmetry of this reduced board (rotation about its center or reflection about a line through the center) takes any pair of its parallel diagonals at a distance of d—that is, two segments of a full diagonal—into another such pair. So these symmetries preserve "peacefulness" of arrangements. Number the vertical and horizontal files 1, 2, ..., 8 (fig. 5). Suppose there exists a peaceful arrangement of seven queens on the 7×7 subboard; let $(1, q_1)$, ..., $(7, q_7)$ be the squares with the queens. We can apply the reasoning in the article to see that the remainders r_k of the differences $q_k - k$, $k = 1, \ldots, 7$, when divided by 8 must all be different, while their sum must by divisible by 8 (because the seven differences themselves add up to zero). So the sum $r_1 + \ldots + r_7$ is obtained from $0 + 1 + \ldots + 7$ by crossing out one term to make the remaining sum divisible by 8. It's easy to see that the number we should cross out is 4. Thus, there is no queen with $r_k = 4$, and there is exactly one queen with any other r_k, $0 \le r_k \le 7$. The squares (i, j), with $j - i = 4$ modulo 8, are (1, 5), (2, 6), (3, 7) ($j - i = -4$) and (5, 1), (6, 2), (7, 3) ($j - i = 4$). By symmetry, there are 6 other squares that can't have queens on them (obtained from the first six squares, say, by 90° rotation about (4, 4)). All 12 forbidden squares are shaded in figure 5. We know that there is a queen with $r_k = 0$—it stands in one of the unshaded squares of the diagonal (1, 1)–(7, 7). Symmetrically, there must be a queen on one of the unshaded squares of the diagonal (1, 7)–(7, 1). We have only two essentially different possibilities (positions that are not symmetric to each other). Either there is a single queen on (4, 4) for both diagonals, or there is a queen on (1, 1) and a queen on (3, 5). In the first case, the queen does not attack any of the four squares (3, 2), (3, 6), (7, 2), (7, 6), and a second queen placed on any of these squares will attack the others. We have also ruled out all the other squares in columns

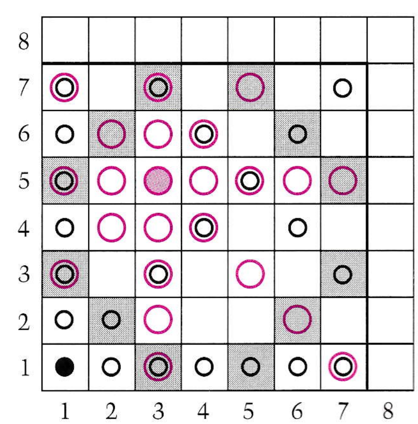


Figure 5

3 and 7, which means that there can only be one queen in those two columns, and thus cannot be seven on the board. A slightly longer argument, proceeding case by case, rules out the possibility of having queens on (1, 1) and (3, 5). So we cannot place seven queens peacefully on the toroidal chessboard.

2. One peaceful arrangement (in the usual chess notation) is a1, b3, c6, d2, e7, f5.

3. The reasoning based on divisibility can be applied to any even n, because the sum $1 + 2 + \ldots + (n - 1)$ is not divisible by n, and twice this sum is divisible by n. For odd n not divisible by 3 it's easy to verify that the squares (k, i_k), where i_k is the remainder of $2k$ upon division by n, $k = 1, \ldots, n$, are all different, and queens on them form a peaceful arrangement.

Chess problem. After 1. Qf5–h7! the black king has only two possible moves: (a) 1. ... Ke8–f8 2. Qh7–g6 Kf8–e7 3. Ke2–e1 Ke7–d7 4. Qg6–e8#!, or (b) 1. ... Ke8–d8 2. Qh7–c7+ Kd8–e8 3. Nb5–h6! (jumping over the edge) Ke8-f8 4. Qc7–e1#!. Check that all the moves of the black king are forced by white pieces attacking "over the edge."

Light at the End

1. Join the red points to the blue points with N arbitrarily drawn segments and apply the operation described in problem 3 in the article to any pair of intersecting segments, replacing them with a pair of disjoint segments with endpoints of different colors. Repeat so long as there remain intersecting segments. The total length of the segments is monotonically decreasing.

2. It's easy to see that the number of segments with endpoints of the same color (we'll call them *monochromatic*) is a monovariant.

3. Paint the points arbitrarily. Let m be the number of monochromatic segments (see the previous solution). We'll show m is a monovariant—more exactly, that m can be decreased by recoloring points whenever $m > N$.

Suppose $m > N$. First we show that there exists a point joined to at least three points of its own color. For suppose the contrary—that each point in our set is connected to at most two points of its own color. Then the N original points are, altogether, endpoints of at most $2N$ monochromatic segments. This counts the monochromatic segment twice (once for each endpoint), so there are at most N monochromatic segments, or $m \le N$, contradicting our assumption.

So let point A be joined to at least three points of its own color. Then there are at most 11 segments issuing from A, and one of the four colors—say, black—occurs on their endpoints fewer than three times. Painting A black decreases m. The rest of the proof is standard.

4. No, it's impossible. At each stage, consider the difference d between the maximum and minimum numbers. Each of these numbers is a quarter sum of the four numbers obtained previously on the adjacent faces, and these two sums must have exactly two common terms. When we subtract one sum from the other to find d, the common terms cancel out. So d can be written as the quarter sum of *two* differences between certain previous numbers. Each of these differences is not greater than the previous value of d, which means that every step reduces d by half or even more.

It can be shown that in fact the sequence of numbers on any face of the cube tends toward the arithmetic mean of all six numbers as a limit. This means that d is not a monovariant as it was defined in the article: it may take infinitely many values. And the process we consider isn't finite either, though d is decreasing and bounded from below. However, this isn't needed to solve the problem.

5. If $a < b$, then the greatest common divisor of a and b is not greater than a and strictly less than b. It follows that the sum of all the numbers around the circle is a monovariant.

6. The answer is 1/512. Consider the number $c = m/2^n$, where m is the smallest positive number written on the blackboard and n is the number of zeros. When two positive numbers are averaged, n doesn't change and m can only grow, so c can only grow, too. When a zero is averaged with a positive number a, n decreases by one and m either remains unchanged or becomes equal to $a/2 \ge m/2$, so the new value of c will be greater than or equal to $(m/2)/2^{n-1} = m/2^n$. In either case c can't decrease, so it's a monovariant. Originally, $c = 1/2^9 = 1/512$; therefore, we'll always have $m \ge 2^n/512 \ge 1/512$. To obtain exactly 1/512, we must successively average the number appearing in place of the original one with all the nine zeros.

7. Surround the set of black squares with a rectangle bounded by grid lines. Take its bottom left vertex as the origin and the grid lines through it (directed to the right and upwards) as the axes. Find the sum of the coordinates of the center of each black square. The greatest of these sums is a strictly decreasing monovariant.

8. Consider the number of candies that each player holds in a certain round of the game. Notice that the smallest of these numbers doesn't decrease during the game, and the maximum number either increases by one (if it's odd and occurs in two neighboring places) or doesn't increase at all. This means that the maximum never becomes greater than $M_0 + 1$, where M_0 is its original value. Thus, the total number of candies held by all the children never exceeds $n(M_0 + 1)$, where n is the number of players. Therefore, at some point the leader stops giving out candies, which means that all the numbers have become even by this time.

From that point on, we can show that the number $S = nM + k$ is a monovariant, where M is the maximum number of candies held by a player and k is the number of players with M candies. For $k < n$, consider the player with M candies whose neighbor on the left has fewer than M candies. After the operation, this player will have fewer than M candies; therefore, either M remains the same while k decreases or M decreases while k may increase by at most $n-1$. In any case, S decreases by one or more.

Once S stops decreasing, all the numbers will have become equal.

9. Let a_i be the number of the place occupied by the kth volume, $d_i = |a_i - i|$. Then the sum of all the numbers d_i is a decreasing monovariant. Indeed, when the kth volume is set in the kth place, the term d_k in the sum vanishes, and only the $d_k - 1$ terms d_i with the numbers i between k and a_k can change, but each at most by one. So the sum decreases by at least $d_k - (d_k - 1) = 1$.

Penrose Patterns

(See "Penrose Patterns and Quasi-crystals" in the last issue)

1. In figure 6 (which reproduces figure 7 from the article) AC bisects angle BAC. Therefore, $\angle BAC = \angle CAD = \angle ADB = 36°$. It follows that ABC and DAB are similar isosceles triangles, and since $DC = CA = AB = 1$ (because $\angle CAD = \angle CDA$),

$$x = DB = \frac{DB}{AB} = \frac{AB}{CB}$$

$$= \frac{1}{DB - DC} = \frac{1}{x-1}.$$

Now the quadratic equation $x^2 - x - 1 = 0$ yields $x = \tau = (1 + \sqrt{5})/2$.

Triangle FHI has the same angles as triangle ADC (36° – 36° – 108°).

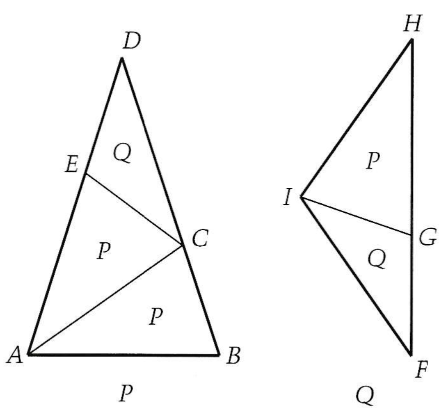


Figure 6

In addition, $FI = AC = 1$. So $FH = AD = \tau$.

2. Calculate angles and use the previous solution.

3. Let P_0 and Q_0 be the P- and Q-triangles of the initial size, and P_n the P-triangle emerging after the nth inflation. By construction, each triangle P_n is tiled with P_0 and Q_0 such that the tiling of P_{n+1} is an extension of the tiling of P_n and, in addition, contains one more congruent copy of P_n tiled in the same way as P_n. So P_{n+2} contains two copies of P_{n+1} and, therefore, four copies of P_n; in general, P_{n+k} contains 2^k copies of P_n tiled in the same way.

Now, if T_0 is any finite part of the entire tiling, we take n such that P_n covers T_0 (this is possible because $P_0 \subset P_1 \subset P_2 \subset \ldots$ and triangles P_n together cover the plane). Then each new copy of P_n (there are infinitely many of them) will contain a copy of T_0.

Proofs of the statements from the editor's postscript "Why does it work?"

1. Since F_1 and F_2 are separated only by a line of the ith set, they lie in adjacent strips of the ith set and in the same strip of each of the four other sets. So $n_i(F_1)$ differs by ± 1 from $n_i(F_2)$, while $n_j(F_1) - n_j(F_2) = 0$ for all $j \neq i$. By the construction of the tiling,

$$\begin{aligned}\overrightarrow{A_1A_2} &= [n_1(F_2) - n_1(F_1)]\mathbf{e}_1 + [n_2(F_2) \\ &\quad - n_2(F_1)]\mathbf{e}_2 + \ldots + [n_5(F_2) \\ &\quad - n_5(F_1)]\mathbf{e}_5 \\ &= [n_i(F_2) - n_i(F_1)]\mathbf{e}_i \\ &= \pm\mathbf{e}_i.\end{aligned}$$

2. The nodes N_i ($i = 1, 2, 3, 4$) were constructed corresponding to faces F_i of the grid G such that F_1 and F_2 are adjacent across the same line of the grid as F_3 and F_4, with F_1, F_4 on one side of this line and F_2, F_3 on the other. According to the previous proof, $\overrightarrow{N_1N_2} = \overrightarrow{N_4N_3}$ ($= \pm\mathbf{e}_i$). Similarly, $\overrightarrow{N_2N_3} = \overrightarrow{N_1N_4}$ ($= \pm\mathbf{e}_j$). Since the angle between $\pm\mathbf{e}_i$ and $\pm\mathbf{e}_j$ is a multiple of 36°, the quadrilateral is a rhombus with the unit side length and angles 36°, 144° or 72°, 108°.

3. Suppose the angle between $\overrightarrow{PP'}$ and $\mathbf{e}_i$ is acute. Then, moving along the line PP' from P to P', we'll cross every line $l_i(n)$ we meet in the "positive" direction—that is, we'll pass from the strip between $l_i(n-1)$ and $l_i(n)$ to the strip between $l_i(n)$ and $l_i(n+1)$ (because the lines are numbered in the direction of $\mathbf{e}_i$). If, crossing line $l_i(n)$, we pass from face F_k to F_{k+1}, then $n_i(F_{k+1}) - n_i(F_k) = n - (n-1) = 1$, and

$$\overrightarrow{A_kA_{k+1}} = [n_i(F_{k+1}) - n_i(F_k)]\mathbf{e}_i = \mathbf{e}_i$$

(compare this with the proof of statement 1).

Similarly, if the angle between $\overrightarrow{PP'}$ and $\mathbf{e}_i$ is obtuse, we'll cross lines $l_i(n)$ only in the "negative" direction, from larger numbers to smaller. So in this case $\overrightarrow{A_kA_{k+1}} = -\mathbf{e}_i$ whenever A_kA_{k+1} is parallel to $\mathbf{e}_i$.

If $\overrightarrow{PP'}$ is perpendicular to $\mathbf{e}_i$, we move parallel to the lines of the ith set and will never cross any of them. So none of the vectors in the sum will be equal to $\pm\mathbf{e}_i$.

4. We have

$$\begin{aligned}n_i(F') - n_i(F) &= n_i(F_n) - n_i(F_1) \\ &= [n_i(F_2) - n_i(F_1)] + [n_i(F_3) - n_i(F_2)] \\ &\quad + \ldots + [n_i(F_n) - n_i(F_{n-1})].\end{aligned}$$

The argument we used in the proofs of statements 1 and 3 shows that the nonzero terms in this sum correspond to all the terms $\pm\mathbf{e}_i$ in the sum $\overrightarrow{A_1A_2} + \ldots + \overrightarrow{A_{n-1}A_n}$ and add up to the coefficient of $\mathbf{e}_i$ after terms in the vector sum are collected—that is, to c_i.

5. By the definition of the tiling T, the given nodes correspond to two faces F and F'. Then the construction discussed in statements 3 and 4 applied to F and F' yields the required progressive path joining the given nodes.

To prove the second statement, join the given nodes with a progressive path $A_1 \ldots A_n$. All its edges make acute angles with a certain direction d. Draw all ten vectors $\pm\mathbf{e}_i$ from a point O and a perpendicular to d through O (fig. 7 on the next page). Choose the five vectors of the ten on the "positive" side of the perpendicular—the side at which the direction d points (if two vectors, $\mathbf{e}_i$ and $-\mathbf{e}_i$, happen to be perpendicular to d,

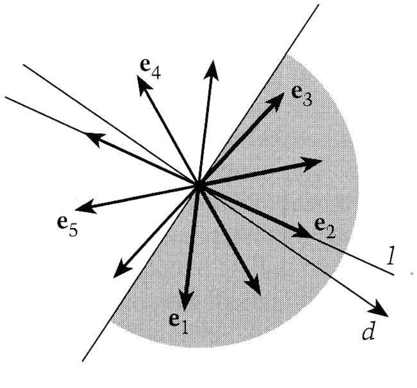


Figure 7

we take the four "positively directed" vectors and add either of these two). The angle between the two "extreme" vectors of these five is $4 \cdot 36° = 144°$; the other three vectors divide it into four equal parts equalling 36°. For definiteness, assume that one of the "extreme" vectors is $\mathbf{e}_1$ and the other one is $\mathbf{e}_3$. Then the remaining vectors are $-\mathbf{e}_4$, $\mathbf{e}_2$, and $-\mathbf{e}_5$. Every vector $\overrightarrow{A_kA_{k+1}}$ along our path is equal to one of these five vectors.

Let's draw a line l along $\mathbf{e}_2$ and consider the projection $A'_1 \ldots A'_n$ of the path onto this line. Each vector $\overrightarrow{A'_kA'_{k+1}}$ is the projection of one of our five vectors onto l, so it's aligned with $\mathbf{e}_2$ and its length is equal to cos 72° (if $\overrightarrow{A_kA_{k+1}} = \mathbf{e}_1$ or $\mathbf{e}_3$), cos 36° (if $\overrightarrow{A_kA_{k+1}} = -\mathbf{e}_4$ or $-\mathbf{e}_5$), or 1 (if $\overrightarrow{A_kA_{k+1}} = \mathbf{e}_2$). In any case, it's not shorter than cos 72° = 0.309..., so the distance r between the given nodes satisfies $r = A_1A_n \geq A'_1A'_n = A'_1A'_2 + A'_2A'_3 + \ldots + A'_{n-1}A'_n \geq (n-1)\cos 72°$.

If the number $n-1$ of edges in the path is greater than three, then $r \geq 4 \cos 74°$.

If this number is three and one of the edges is parallel to $\mathbf{e}_2$, $\mathbf{e}_4$, or $\mathbf{e}_5$, then $r \geq \cos 36° + 2\cos 72° > 0.8 + 2 \cdot 0.3 > 1$.

If $n - 1 = 3$ and each of the three vectors $\overrightarrow{A_1A_2}$, $\overrightarrow{A_2A_3}$, $\overrightarrow{A_3A_4}$ is equal to $\mathbf{e}_1$ or $\mathbf{e}_3$, then two of them are equal to, say, $\mathbf{e}_1$, and

$$r = |\overrightarrow{A_1A_4}| = |2\mathbf{e}_1 + \mathbf{e}_i| > |2\mathbf{e}_1| - |\mathbf{e}_i| = 1$$

by the Triangle Inequality (here i = 1 or 3).

Finally, if the path consists of two edges A_1A_2 and A_2A_3, the angle α between them is a multiple of 36°. From the triangle $A_1A_2A_3$ we have $r = A_1A_3 = 2\sin(\alpha/2)$, which is equal to $2 \sin 18° = 2 \cos 72°$ for $\alpha = 36°$, and is no less than $2 \sin 36° > 1$ for $\alpha > 36°$.

6. Suppose two different faces F and F' of grid G define the same node A of T. Applying the construction discussed in statements 3 and 4 to F and F' yields a progressive path $A_1 \ldots A_n$ that consists of at least one edge, and so, by the previous proof, $A_1 \neq A_n$. But the endpoints of this path are the nodes corresponding to F and F'—that is, $A_1 = A = A_n$. This contradiction completes the proof.

7. Let $F_1 = F$, F_2, ..., $F_n = F'$ be the faces of G associated with the nodes A_1, A_2, ..., A_n, respectively. By statement 6, they are defined uniquely, so F_k must necessarily border on F_{k+1} (because A_k is joined with an edge to A_{k+1}). Now the proof of statement 4 above applies without any change.

8. Let A_1A_2 be an edge of T and F_1, F_2 be the faces of grid G corresponding to the nodes A_1 and A_2. By the construction of T, A_1A_2 is a side of a rhombic tile if and only if the node of grid G corresponding to this tile is an endpoint of the common side of polygons F_1 and F_2 (see the proof of statement 2). But F_1 and F_2 have exactly one common side, which has two endpoints, each associated with one tile.

9. Move from A to X and label the edges we successively meet s_1, s_2, s_3, ... and their intersections with AX as X_1, X_2, X_3, ... Every two successive edges s_k and s_{k+1} are sides of the same tile, so if they are parallel, then X_kX_{k+1} is no less than the shortest height h of the narrow rhombus. If three successive edges s_{k-1}, s_k, and s_{k+1} form a zigzag ($s_{k-1} = BC$, $s_k = CD$, $s_{k+1} = DE$), then it's easy to see that $X_{k-1}X_{k+1} \geq h$ (because neither angle ABC nor angle BCD is less than 36°).

Now divide segment AX into a finite number of pieces of length less than h. Then no two successive edges intersecting one of these pieces are parallel, and no three such edges make a zigzag. So all the edges intersecting one piece issue from a common node. Since there are at most ten edges issuing from a node (they make angles not less than 36°), each of the pieces contains no more than ten points X_k (in fact, it can contain at most five of these points). So the total number of points X_k—and the total number of tiles in our chain—is finite.

10. Let A be a node of T. It is related to a certain face of grid G—a polygon $F = X_1X_2 \ldots X_n$. Then any rhombic tile with the vertex A is related to one of the vertices of this polygon. Label R_i the rhombus corresponding to X_i. By construction, the sides of R_i issuing from A are perpendicular to $X_{i-1}X_i$ and X_iX_{i+1}, so the angle α_i between them equals $180° - \angle X_{i-1}X_iX_{i+1}$—that is, it's equal to the exterior angle of the polygon F at X_i. Each rhombus R_i is wedged between R_{i-1} and R_{i+1}, which border on its sides issuing from A (R_1 is between R_n and R_2, and R_n is between R_{n-1} and R_1), because the node X_i is joined with edges to X_{i-1} and X_{i+1}. So in the sequence R_1, R_2, ..., R_n, R_1 the tiles border on the next and the previous ones without overlapping with them, and the go around A without gaps. Since $\alpha_1 + \alpha_2 + \ldots + \alpha_n = 360°$ (this sum is equal to the sum of exterior angles of F), the rhombi circumnavigate the node A exactly once. Thus, no two of them overlap.

11. The existence of the required edges is almost obvious, but the proof is full of petty details; so we'll omit it here.

If M is a single common point of edges AB and CD, then one of the distances MA and MB is not greater than 1/2 (because $AM + MB = AB = 1$). The same is true for CD. Suppose $AM \leq 1/2$ and $CM \leq 1/2$; then $AC < AM + CM \leq 1$.

12. We regret that a misprint marred the satement of the problem. It should have read: ". . . no two edges drawn from A and C" (not "from A to C").

Consider the tile $ABPQ$ lying on the same side of AB as edge BC. Since $\angle ABP \geq 36°$, the edge BC can't lie outside this tile. It can't lie strictly inside it either, because in this case a tile on BC would overlap with $ABPQ$, contradicting statement 10. Therefore, BC coincides with the side BP of tile $ABPQ$ ($C = P$).

For similar reasons, any two edges AD and CE must lie outside our tile $ABCD$. So if they intersect at M, the triangle AMC contains either B or Q. Then $AM + MC > AB + BC = AQ + QC = 2$. But this is impossible, because $AM + MC < AD + CE = 2$.

13. Take an arbitrary node B of the tiling T. The translation by the vector $\mathbf{p}$ takes it into another node B' of T. Connect these nodes with a path $b = B_0B_1 \ldots B_n$, where the B_i are all nodes of T, and $B_0 = B$, $B_n = B'$.

Replace the terms in the sum $\overrightarrow{B_0B_1} + \overrightarrow{B_1B_2} + \ldots + \overrightarrow{B_nB_{n-1}} = \overrightarrow{B_0B_n}$ with appropriate vectors $\pm\mathbf{e}_i$, and collect like terms. We get

$$\mathbf{p} = \overrightarrow{B_0B_n} = p_1\mathbf{e}_1 + \ldots + p_n\mathbf{e}_n.$$

This representation is the required one.

To prove this, consider an arbitrary face F of the grid G and the corresponding node A of T. Connect A to B with a path a. Since $\mathbf{p}$ is a period of T, our translation takes this path into a path a' from some node A' ($\overrightarrow{AA'} = \mathbf{p}$) to B'. Let F' be the face of G corresponding to A'. Now consider the path from A to A' that goes first along a to B, then along b to B', then along a' from B' to A'. Again write the sum of vectors corresponding to the path $ABB'A'$ and collect like terms. The terms corresponding to the edges in a and a' will cancel out, because these two paths consist of the same (up to translation) edges, but we pass them in opposite directions. So after simplifying we'll get the same sum $p_1\mathbf{e}_1 + \ldots p_n\mathbf{e}_n$. Finally, statement 7 applied to the path $ABB'A'$ from A to A' shows that $n_i(F') = n_i(F) + p_i$ for $i = 1, 2, \ldots, 5$.

14. We'll show that this statement is true even for triple intersections of strips—namely, for strips of the first, second, and fifth sets. To simplify notations, suppose that $n_1 = n_2 = n_5 = 0$ (otherwise we can renumber the lines of the grid G). The intersections $r_{12} = s_1(0) \cap s_2(0)$ and $r_{15} = s_1(0) \cap s_5(0)$ are congruent rhombuses (they are seen in the left corner in figure 8; r_{12} is shaded). If the two rhombuses have no common point, we are done. So assume they both contain point A. Draw vectors along their sides: $\mathbf{v}_1$ (red in figure 8) along the strip $s_1(0)$, $\mathbf{v}_2$ (blue) along the strip $s_2(0)$, $\mathbf{v}_5$ (black) along the strip $s_5(0)$. All three vectors are the same length v. We can see that the intersection of $s_1(kp_1)$ and $s_2(kp_2)$ is the rhombus r'_{12} obtained from r_{12} under translation by the vector $kp_2\mathbf{v}_1 + kp_1\mathbf{v}_2$, and the intersection r'_{15} of $s_1(kp_1)$ and $s_5(kp_5)$ is obtained from r_{15} under translation by $kp_5\mathbf{v}_1 + kp_1\mathbf{v}_5$ (these intersections are shown in the right corner in figure 8). Let B and C be the images of A under the translations specified above. Point B belongs to r'_{12}, C to r'_{15}, and $\overrightarrow{BC} = k(p_5\mathbf{v}_1 + p_1\mathbf{v}_5) - k(p_2\mathbf{v}_1 + p_1\mathbf{v}_2) = k[(p_5 - p_2)\mathbf{v}_1 + p_1(\mathbf{v}_5 - \mathbf{v}_2)] = k\mathbf{d}$. Notice that $\mathbf{v}_1$ and $\mathbf{v}_5 - \mathbf{v}_2$ are parallel, but the ratio of their lengths is irrational (draw $\mathbf{v}_2$ and $\mathbf{v}_5$ from one point and join their tips—you'll get a P-triangle, so the length of $\mathbf{v}_5 - \mathbf{v}_2$ is v/τ). Since p_1, p_2, and p_5 are integers, $\mathbf{d} \neq 0$. Therefore, the distance between B and C (equal to $k|\mathbf{d}|$) can be made arbitrarily large by choosing a large enough k. This means that we can slide the rhombi r'_{12} and r'_{15} arbitrarily far apart and thus make the intersection of $s_1(kp_1)$, $s_2(kp_2)$, and $s_3(kp_3)$ empty.

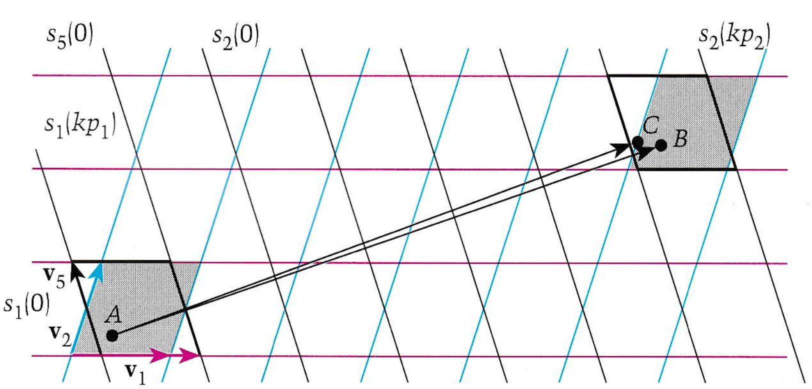


Figure 8

15. This statement follows directly from the definitions.

16. Represent the grid G as the union of two grids: G', comprising the lines of the first, second, and fifth sets; and G'', comprising the lines of the third and fourth sets. Consider a face F' of grid G' and a face F'' of G''. If we shift F'' so that its sides do not pass through the vertices of F' and its vertices do not hit the sides of F' as F'' is moving, then the intersection of F' and F'' will remain empty or nonempty, according to what it was initially. In addition, the boundary of the new intersection (if it's nonempty) will be formed by the lines of the same sets and with the same numbers as initially.

Denote by G_0 the part of G that corresponds to T_0 (the union of all the faces associated with nodes in T_0). For each node in G_0 measure the distance to the closest line of G (not passing through this node); let δ be the minimum of these distances (it's positive, because this set of nodes in G_0 is finite). If we shift the grid G'' by a vector whose length is smaller than δ, the aforementioned condition on the nodes and sides of any two faces F' (of G') and F'' (of G'') whose intersection is a face of G_0 will hold. This means that each face F of G_0, though slightly altered under the shift, will remain the intersection of the strips with the same numbers bounded by the grid lines with the same numbers. So the node corresponding to F will remain in its place together with all the edges issuing from it. Thus, the entire subtiling T_0—all its nodes and edges—is preserved.

Notice also that our operation can't create any new nodes and edges in the area covered by T_0, because that would mean the appearance of new tiles overlapping with the old tiles of T_0.

17. Lines of the second set make an angle of 72° with a line of the first set, and the interval between them is 1. Therefore, the length of vector $\mathbf{v}$ is $1/\sin 72°$. Similarly, the length of $\mathbf{v}'$ is $1/\sin 36°$. The ratio of these lengths is equal to $\sin 36°/\sin 72° = 1/(2\cos 36°)$. We can see from the Q-triangle FHI in figure 6 above and problem 1 that $2\cos 36° = FH/FI = \tau$, so $\mathbf{v} = (1/\tau)\mathbf{v}'$ and $|n\mathbf{v} - m\mathbf{v}'| = |n - m\tau| \cdot (1/\sin 72°)$. Now the statement follows immediately from the Fractional Parts Theorem, because $\tau = (1 + \sqrt{5})/2$ is irrational.

18. First, consider two grids G_1 and G_2 (constructed and numbered like G) such that G_1 can be translated into G_2. Using the same fixed point O, construct tilings T_1 and T_2 associated with G_1 and G_2 (see page 17 in the original article). In view of statement 1, it's quite clear that T_2 is obtained from T_1 under a certain translation as well. More exactly, suppose that the numbers of lines in the ith set of G_2 differ by t_i from the numbers of corresponding lines in G_1 (obviously the difference between the numbers of corresponding lines is the same in each set). Then, for any face F_1 in G_1 and its image F_2 under the translation that takes G_1 into G_2, we'll have $n_i(F_2) = n_i(F_1) + t_i$, which means that the nodes A_1 and A_2 of T_1 and T_2 associated with F_1 and F_2, respectively, will differ from each other by the vector $\overrightarrow{A_1A_2} = t_1\mathbf{e}_1 + t_2\mathbf{e}_2 + \ldots + t_5\mathbf{e}_5$ (provided the tilings were constructed using the same pole O, of course). Therefore, the translation by this vector takes T_1 into T_2.

Now take as G_1 the grid G' obtained from G by shifting the third and fourth sets of lines through the vector $m\mathbf{v}' - n\mathbf{v}$, and as G_2 the grid G itself. Note that the translation through vector $n\mathbf{v}$ takes G' into G. Indeed, for i = 1, 2, and 5 any line $l_i'(k)$ of G' simply coincides with $l_i(k)$ and, by statement 15, is taken into $l_1(k)$ (for $i = 1$), $l_2(k + n)$ (for $i = 2$), or $l_5(k - n)$ (for $i = 5$). Then, $l_3'(k)$ is taken into the image of $l_3(k)$ under the translation through $(m\mathbf{v}' - n\mathbf{v}) + n\mathbf{v} = m\mathbf{v}'$—that is, by analogy with statement 15, into $l_3(k + m)$, and, similarly, $l_4'(k)$ is taken into $l_4(k - m)$. As we saw above, this means that the tiling T' associated with G' is taken into T under translation by vector $0 \cdot \mathbf{e}_1 + n\mathbf{e}_2 + m\mathbf{e}_3 - m\mathbf{e}_4 - n\mathbf{e}_5 = \mathbf{t}$. But, by statement 16, T' contains T_0, and we're done.

Corrections

(See also "Readers Write . . ." on page 52.)

Vol. 4, No. 3:

p. 3, col. 1, ¶2, l. 7: *for* Dick Rutan *read* Burt Rutan. (Burt Rutan is the noted aircraft designer; Dick Rutan flew the *Voyager*, designed by his brother Burt, around the world nonstop with Jeana Yeager.)

p. 18, col. 3, l. 3: *for* from A to C *read* from A and C.

Vol. 4, No. 2:

p. 44, col. 1, ¶2, ll. 11–12: *for* Can a^x be greater than x? *read* Can a^x be less than x?

p. 44, col. 2, ¶2, l. 3: *for* $f'(x) = a^x \ln a - \ln x$ *read* $f'(x) = a^x \ln a - 1$.

p. 44, col. 2, ¶2, l. 5: *for* $-\ln x < 0$ *read* $-1 < 0$.

p. 56, col. 1: *for* $\overrightarrow{A_1B_2}$, $\overrightarrow{A_2B_3}$, etc., *read* $\overrightarrow{A_1A_2}$, $\overrightarrow{A_2A_3}$, etc.

Vol. 1, No. 4:

A credit line was inadvertently omitted from the piece by I. M. Gelfand on the Moscow Correspondence School in the March/April 1991 issue of *Quantum*. Dr. Tanya Alekseyevskaya-Gelfand translated the article and helped prepare it for publication.

E-mail your comments and questions to 72030.3162@compuserve.com

Torangles and torboards

What you can do with a little glue

by Vladimir Dubrovsky

IF YOU ARE AN ACTIVE READer of this magazine, by now you will have created a sizable collection of paper models of peculiar geometric objects based on our Toy Store articles. Here we add two new items to our *Quantum* Paper Model Collection. They aren't flexible like most of the models we've described. What makes them remarkable is that they're the simplest polyhedrons with a hole through them—in other words, topologically equivalent to the torus (see the Kaleidoscope in this issue). "The simplest" here doesn't mean "the easiest to create"—in your head or with your hands. The word is understood more formally as "having the smallest possible number of vertices." The models are shown in figure 1: the left one has $v = 7$ vertices, the right one $v = 8$.

Polyhedral tori

Before I explain how to make them, let's try to understand why these polyhedral tori are really minimal. Start with any polyhedral torus, and let's see what must be true if it is to be minimal. We'll make use of Euler's formula for the torus $f - e + v = 0$, where f, e, and v are the numbers of the faces, edges, and vertices of any toroidal (torus-shaped) polyhedron. Since any face can be subdivided into triangles (by its diagonals) without changing the number of vertices, we can assume that all the faces are triangles. If we count the sides of each face and add up the results, we get, on the one hand, $3f$ (all faces are triangles) and, on the other hand, $2e$ (every edge is a side of two faces), so $3f = 2e$, and consequently, $3v = 3e - 3f = e$. Every edge joins two vertices, so the number of edges is no greater than the number of pairs of vertices, which is $v(v-1)/2$. Thus, we get

$$3v = e \le \frac{v(v-1)}{2},$$

and, after simplifying, $v \ge 7$. But this is only a necessary condition. To construct a toroidal polyhedron with, say, seven vertices, we must first draw the network of its edges and faces—that is, a map on a torus consisting of triangular countries (faces) any two of which can have either a common side (edge), a common vertex, or no common points at all. Then the network must be realized as a polyhedron. In the case $v = 7$, the inequality above becomes an equality: $3v = v(v-1)/2$. This means that every two vertices are connected by an edge. But, of course, not every triple of vertices will be the vertices of a face.

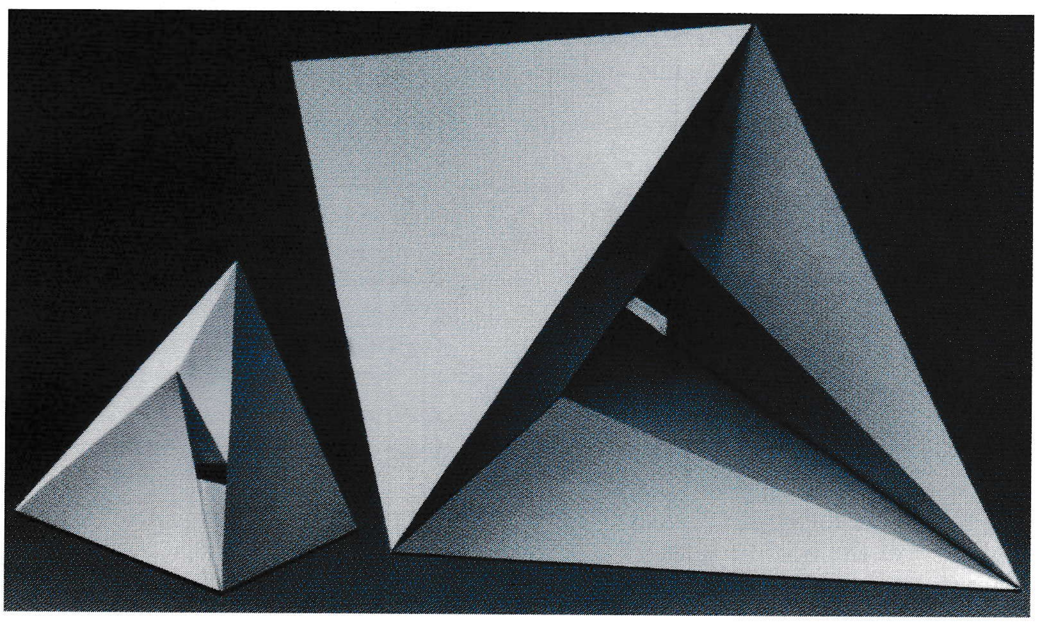


Figure 1

It's a good exercise to create the required network. Try to do it, then compare your result with figure 2a (on the next page), or with the triangular network in figure 4 in the Kaleidoscope. Note that these two networks are equivalent

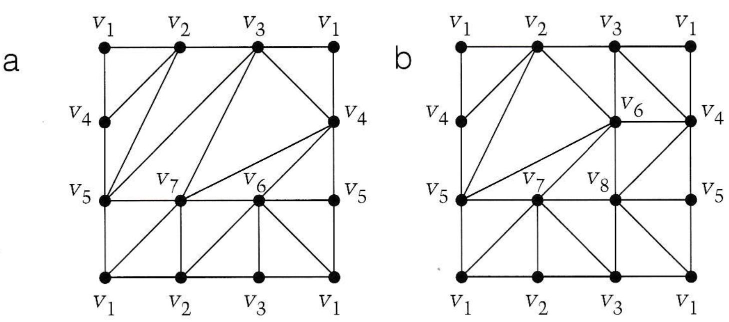


Figure 2
Networks of (a) 7-vertex, (b) 8-vertex toroidal polyhedrons.

even though they look different—don't forget that the opposite sides of the squares in these figures should be thought of as glued to each other! Can you establish a correspondence between the vertices of the two networks such that any three vertices of a face in one network correspond to the vertices of a face in the other network?

Now that the network is ready, try to imagine a polyhedron $v_1v_2\ldots v_7$ based on this network of edges and faces. This is a challenging but engrossing problem. One of the possible polyhedrons can be constructed using the following edge lengths (in centimeters):

$v_1v_4 = v_2v_3 \cong 10.3,$
$v_4v_7 = v_3v_7 \cong 7.8,$
$v_4v_3 = 10.0,$
$v_2v_4 = v_1v_3 \cong 8.7,$
$v_4v_5 = v_3v_6 \cong 7.0,$
$v_1v_6 = v_2v_6 = v_1v_5 = v_2v_5 \cong 7.0,$
$v_1v_2 \cong 6.0,$
$v_5v_6 \cong 4.1,$
$v_1v_7 = v_2v_7 \cong 3.4,$
$v_6v_7 = v_5v_7 \cong 4.9,$
$v_4v_6 = v_3v_5 \cong 3.8.$

Cut triangles with these side lengths out of thin cardboard, join them together according to figure 2a, and you'll get the model on the left in figure 1. The model was created by two undergraduate students at Moscow University, A. Bushmelev and S. Lavrenchenko. In fact, they worked on the more general and serious problem of the existence of a toroidal polyhedron with a given network of edges. Such a theorem for ordinary polyhedrons was discovered by E. Steinitz in 1917 (see the Kaleidoscope in the May/June 1993 issue). The young Moscow mathematicians proved that any triangular network on the torus represents a certain polyhedron. In their proof, they reduced an arbitrary network to one of the simplest ones (with no more than 10 vertices). One of the four possible 8-vertex networks is shown in figure 2b, and its realization as a polyhedron is shown in figure 1 at the right. The edge lengths for this model are

$v_iv_j \cong 15.0,$
$v_iv_k \cong 18.1$ $(i = 1, 6; j = 3, 4, 8;$
$k = 2, 5, 7),$
$v_7v_2 = v_2v_5 = v_5v_7 \cong 18.5,$
$v_8v_3 = v_3v_4 = v_4v_8 \cong 6.3,$
$v_7v_8 = v_2v_3 = v_4v_5 \cong 7.0,$
$v_7v_3 = v_2v_4 = v_5v_8 \cong 12.9.$

The Queen Problem on the torus

The Queen Problem is an old and well-known problem of recreational mathematics, often associated with the name of the great Carl Friedrich Gauss (1777–1855). One has to place eight queens on a chessboard so that they don't attack one another, and find the number of such arrangements. Obviously it's impossible to arrange more than eight queens so that they don't attack one another, because in any such arrangement two or more queens would appear on the same horizontal or vertical line.

What happens if we glue the opposite sides of the board together, turning it into a torus? The difference between the central, edge, and corner squares then disappears, because we wipe out the edges themselves, and a queen will always attack $7 \cdot 4 = 28$ squares regardless of its position. Horizontally and vertically, these are the same 14 squares as on the ordinary chessboard. But either diagonal attacked by a queen (if it's not one of the main 8-square diagonals a1–h8 or a8–h1) can be "extended" onto the parallel short diagonal at a fixed distance d, equal to four diagonal lengths of a square. Two such "extended diagonals"—we'll call them *full diagonals*—are illustrated in figure 3. You may imagine that a queen moving along a full diagonal jumps back to the opposite side of the board the very moment it reaches an edge of the board, and then it continues moving in the same direction: the torus has no edges!

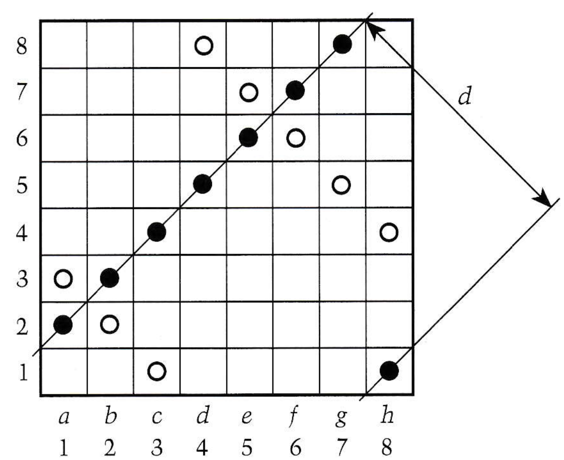


Figure 3

Thus, queens on the toroidal chessboard are more powerful, and "peaceful coexistence" of eight queens becomes impossible. The proof of this fact is based on *divisibility*.

Let's number both the horizontal and the vertical files 1, 2, ..., 8 (fig. 3). Then each square receives a pair of coordinates (i, j), $1 \le i, j \le 8$. Suppose we can put eight queens on the board so that they don't attack one another (we'll call such arrangements *peaceful*). Then there will be exactly one queen in each horizontal and vertical file, so we can write the coordinates of the queens as $(1, q_1)$, $(2, q_2)$, ..., $(8, q_8)$, where the numbers q_1, ...,

q_8 take each value from 1 to 8 once. Also, no two queens can stand on the same full diagonal parallel to the main diagonal a1–h8. This means that the remainders of the differences $q_k - k$ when divided by 8 must all be different. (Check that for all squares (i, j) on any of these full diagonals the remainder of $j - i$ modulo 8 is the same!) So the remainder of the sum $(q_1 - 1) + \ldots + (q_8 - 8)$ must equal that of $0 + 1 + \ldots + 7 = 28$—that is, it's equal to 4. But this is impossible, because $(q_1 - 1) + \ldots + (q_8 - 8) = (q_1 + \ldots + q_8) - (1 + \ldots + 8) = (1 + \ldots + 8) - (1 + \ldots + 8) = 0$.

The following problems elaborate on this theme.

Problems

1. Show that *seven* queens can't be peacefully arranged on the 8×8 toroidal chessboard. (Hint: you can reduce the number of search possibilities by proving that the top and far right files can be assumed to be free of queens, and all symmetries of the remaining 7×7 board turn a peaceful arrangement into another peaceful arrangement; also, use the above considerations of divisibility.)

2. Find a peaceful arrangement of six queens.

3. Prove that on the $n \times n$ toroidal chessboard a peaceful arrangement of n queens doesn't exist for any even n, and exists for any odd n not divisible by 3.

What is the maximum number of queens in a peaceful arrangement on the $n \times n$ toroidal chessboard for n divisible by 2 or 3? It seems that a plausible answer is $n - 2$. You may want to prove or disprove this conjecture.

Checkmate on the torus

In conclusion, here's a toroidal chess problem by E. Mach[1] (fig. 4): white must checkmate black in four moves. Don't forget that on the torus each piece attacks the same number of squares regardless of its position (like a queen), so they become stronger, but the black king also has more possibilities to escape from check.

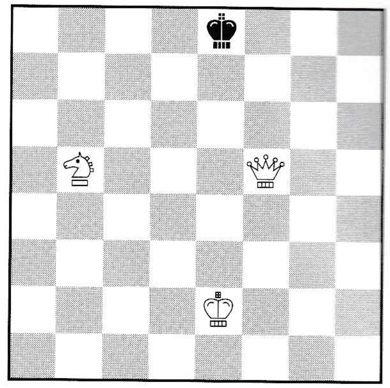


Figure 4

Problem by E. Mach: toroidal mate in four moves.

[1]This problem is borrowed from *Mathematics on the Chessboard* by Yevgeny Gik (in Russian).

ANSWERS, HINTS & SOLUTIONS ON PAGE 57