QUANTUM

JANUARY/FEBRUARY 1994

$5.00

SPRINGER INTERNATIONAL

GALLERY Q

Allegory (1556) by Lorenzo Lotto

IMAGINE YOU COULD REACH INTO THE PAGE AND turn this picture over like the cover of a book. Behind it you would see a portrait of Bernardo Rossi, Bishop of Treviso. That is, if you happened to be a close friend of his. The portrait and its cover panel have been separated for centuries—the bishop's celebrated visage now resides in a Naples museum.

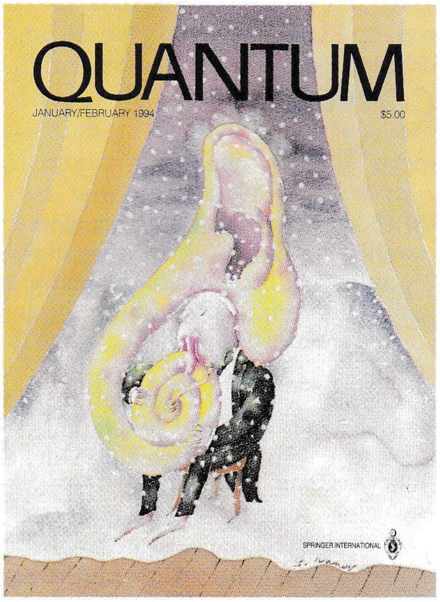
Those who found "Allegory" had little difficulty making the connection. Even if there hadn't been an explicit reference to the portrait on the back, art historians would have recognized the coat-of-arms propped against the tree from Rossi's signet ring in the Naples painting.

We can only speculate why Lorenzo Lotto (c. 1480–1556) painted this scene contrasting the life of virtue and the life of vice as a companion to the bishop's portrait. The Renaissance love of learning is evident in the child's "playthings": scroll, carpenter's square, flute, compass. The satyr in the gloomy woods seeks truth in wine. An empty amphora with a piglike snout lolls at the feet of the beastly sensualist. The background hints at the ultimate fates of these two souls: a shipwreck for one, a quick ascent to heaven for the other.

The tree in the foreground has much to say as well. Blasted by lightning or wind, it nonetheless has sprouted anew—on the "good" side. It will no doubt try to regain the stature of its peers on the right (whose virtue we need not impugn—Nature is not prone to human vice). But why stop there? What keeps it from growing taller? Our first feature article takes some data, makes some estimates, and draws some conclusions about the upward-striving life of trees.

QUANTUM

JANUARY/FEBRUARY 1994 VOLUME 4, NUMBER 3



Cover art by Sergey Ivanov

What's that we hear? Across the snowy stillness, in the enveloping murk of midwinter . . . the dulcet tones of the Magic Flute? the Fiddler on the Roof? seventy-six trombones? We're getting close . . .

We humans take great pleasure in sound. Before we saw the world we heard it, in the long prenatal darkness. And in our typically anthropocentric way, we tend to think that sound was made for us. If *we* can't hear it, it's not a "sound." We now know that the range of vibrations that human beings can hear is actually quite narrow, and that nonaudible sound can be very useful. Paradoxically, we build machines to help us see what we can't hear and, perhaps more importantly, see what we couldn't otherwise see.

The Kaleidoscope in this issue explores the world of sound. And as the oblivious tubist on our cover plays "Auld Lang Syne," we wish our readers health and happiness in 1994.

FEATURES

DEPARTMENTS

PUBLISHER'S PAGE

Swords into plowshares

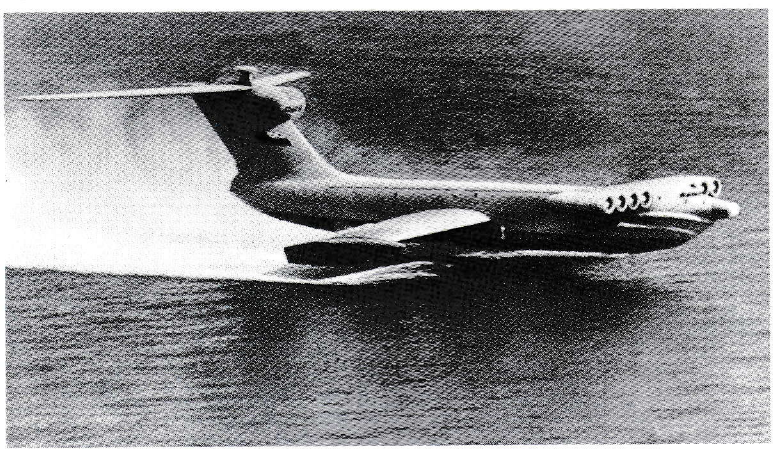
One example of military-to-civilian conversion

MY YOUNGEST SON, RICHard Aldridge, is an aerospace engineer, commercial pilot, and certified flight instructor. Much to my chagrin, he also flies aerobatics! Richard has had a lifelong love of flying and has learned more about airplanes than anyone I've ever met. He will help me describe an interesting and unique technology that only Russia has, one that could have a timely application in the United States.

Last January, Richard learned of a remarkable new kind of vehicle that had been a military secret for many years. The Russians decided to declassify it and try to market it for civilian use. This vehicle flies like an airplane but is neither an airplane nor a hovercraft. It is called a WIG (for "wing-in-ground effect") or, more commonly, a wingship.

The Russians began developing such aircraft in the early 1960s and had produced several versions. They are designed to take off and land on water, and fly most efficiently just above ground or water level. They were to be used as military cargo vehicles and troop transports, and also as weapon platforms that could fly undetected under enemy (that was us!) radar. More recently, it was proposed that the wingships be used for rescues at sea. The Russians currently have two models of wingship in use: the Orlenok (Eaglet) and the Lun (Hen-harrier).

The Lun is huge—some 74 m long, with a 44-m wingspan and a tail height of 16 m. It cruises at about 500 km/hr with a range of 3,000 km and a maximum altitude



of 3,000 m. But what is most significant is its huge payload—around 100,000 kg. The Lun can take off from water with a total mass of 400,000 kg.

Here is Richard's description of the wing-in-ground effect (perhaps a more quantitative version will appear in a subsequent issue of *Quantum*):

The wing-in-ground effect is the phenomenon of drag reduction when an airplane flies near the ground. "Drag" is the retarding force that the plane's engines must overcome. It is made up of two components. *Parasite drag* is the drag due to viscous effects and pressure differences on the fore and aft sections of a body. This is the drag most of us are familiar with. The second type of drag occurs when lift is produced, and so is called *lift-induced drag*. This is the drag that is reduced during flight close to the ground.

The easiest way to explain lift-induced drag is by vortex generation. According to Bernoulli's principle, when a wing produces lift, the total pressure on the upper surface of the wing is less than the total pressure on the bottom surface. Thus, a net pressure pushes the wing up—giving it lift. Because of these pressure differences, air from under the wing at higher pressure spills around and over the wingtips to the area above the wing, where the pressure is lower. When this happens vortices are produced—little tornadoes that are tilted sideways. These vortices can be a hazard (which is why air traffic controllers delay small airplanes from taking off behind large airplanes).

Why would these vortices cause drag? Because they are columns of spinning air that contain rotational energy. How then does an airplane operating near the ground create less lift-induced drag? These vortices at the wing tips interact with the ground, which hinders their production and reduces their size. Thus, they require less energy to produce, which means less drag. This reduced drag makes the wingship much more efficient than a regular airplane.

To help bring this technology to a North American market, Richard flew to Moscow in midwinter and took an overnight train to Nizhny Novgorod (formerly Gorky, where Andrey Sakharov had been exiled). He was brought onto the frozen surface of the Gorky Sea (a reservoir 100 km north of Nizhny) to observe one of the wingships in flight. On his next trip there, he brought with him Peter Garrison of *Flying* magazine and leaders of ARPA, a US government group interested in possible uses for the wingships.

Richard arranged for a full-scale demonstration flight on the Caspian Sea, where an ARPA team of engineers and scientists observed the spectacular performance of the Lun's smaller sibling, the Orlenok. (The team included Dick Rutan, who designed the plane that flew around the world nonstop—the long-winged *Voyager*.)

Readers in the US who watched the horrifying California fires in November may have guessed where this editorial is headed. When Richard saw the C-130 transport planes and helicopters try to carry enough water to put out those fires, he realized that this task was a perfect match for the Lun. With modifications, it could do the job far better than any other kind of aircraft. The Lun could take off and land on water, and it could be loaded with water with high-speed pumps built into its hull.

Since the density of water is about 1,000 kg/m^3 and the Lun has a payload of 100,000 kg, it could carry 100 m^3 of water, and more if modified with smaller fuel tanks. This is 26,400 gallons of water for each flight, allowing a blanket of water 1 foot deep, 20 feet wide, and 177 feet long to be dropped on a fire for each flight from the water to the fire. Two or three of these wingships might put out fires quickly and easily, well before they could do the terrible damage they now do.

What a great opportunity to fit Russian technology to the task of saving the homes and lives of people in the nation that this technology was originally designed to threaten.

—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 offices. **Postmaster:** send address changes to: *Quantum* Springer-Verlag New York, Inc., Journal Fulfillment Services Department, 44 Hartz Way, Secaucus, NJ 07096-2491. Copyright © 1994 NSTA. Printed in U.S.A.

Subscription Information: North America

Student rate: $15; Personal rate (nonstudent): $20; Institutional rate: $34; Single Issue Price: $5. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus, NJ 07096; 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.

Trees worthy of Paul Bunyan

Why do trees grow so tall—but no taller?

by Anatoly Mineyev

EVERYONE KNOWS THAT there are no trees as tall as one kilometer. What factors influence the height of a tree? What quantitative laws determine its thickness and height? Can one understand and explain them using elementary physical estimates?

Let's open a reference book to find data for the height h and trunk diameter d (measured near the ground) of some trees. Figure 1 shows the data with the ratio h/d plotted against $1/d$. Also shown are some lines delimiting our data: the horizontal line $h/d = 50$ and three slanted lines h = const.

It's evident that all varieties of trees here on Earth seem to follow certain regularities. To figure out what they are, let's formulate a few obvious questions.

Why is the ratio h/d limited to 50 for thick trees with a base diameter $d > 0.4$–0.5 m (to the left of point B in figure 1)? Why can the ratio h/d exceed 50 for thin ($d < 0.2$ m) trees, but only under a certain height (10–20 m)? Why is the maximum height of trees on Earth 100–150 m?

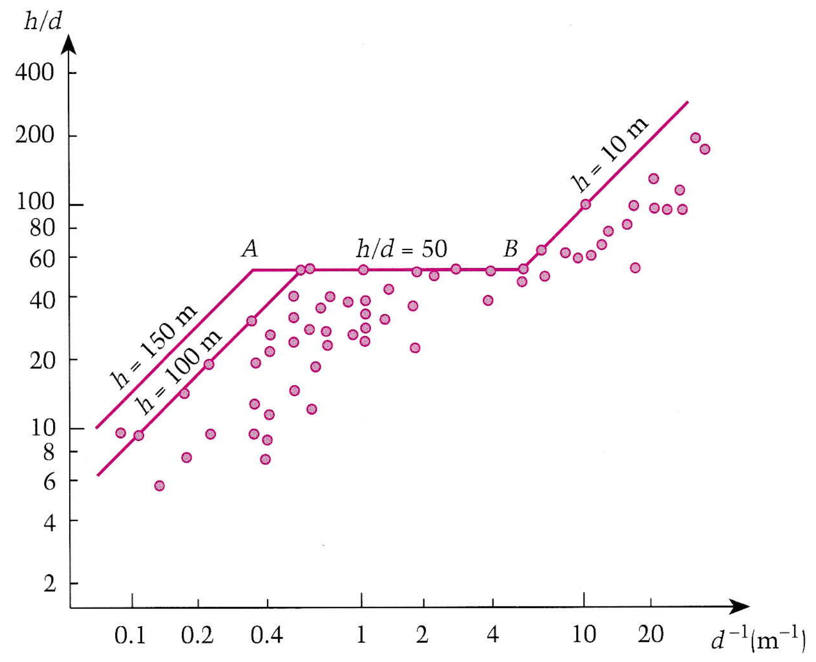


Figure 1

Dependence of a tree's relative height (h/d) on its diameter (d). The left-hand portion corresponds to the thickest trees, the right-hand portion represents thin trees.

Mechanics: strength and elasticity

Let's start at the center of figure 1 and consider the horizontal line AB ($h/d = 50$). What is the origin of this limit? It's natural to suppose that a slender tree with a high value of h/d can't bear the load and will get broken. Indeed, one of the basic parameters determining the maximum value of h/d is the strength of the trunk. Let's consider the limits connected with this strength (a) in the ideal case of a vertical tree and in two approximations intended to model Nature: (b) a tree at an angle with the vertical and (c) a tree influenced by the wind.

To simplify the estimates, a tree is assumed to be a solid cylinder of diameter d and height h. The density ρ and the maximum stress σ_{max} of the tree will be taken to be

$$\rho \cong 500 \text{ kg/m}^3$$

and

$$\sigma_{max} \sim 10^7 \text{ N/m}^2.$$

(a) A vertical tree. For such a tree the maximum stress is equal to the pressure at the ground $\sigma = \rho g h$, where $g = 10$ m/sec^2 is the acceleration due to gravity. Substitution of the numerical data for σ_{max} and ρ yields an estimate of the maximum height for a strictly vertical tree: $h_{max} = 2$ km. Too high! Apparently no tree in nature is strictly vertical.

At right: "A resourceful little guy."

Art by Leonid Tishkov

Находчивый карлик

(b) A tilted tree. In the more realistic situation of a tree inclined at an angle α, there is a torque acting on the trunk:

$$\tau_1 = mg\frac{h}{2}\sin\alpha,$$

where $m = \pi\rho hd^2/4$ is the mass of the trunk. At equilibrium τ_1 is compensated by the torque τ_2 of the elastic forces in the trunk, which is estimated to be $\tau_2 = l_2F_2$, where l_2 is the characteristic distance from the axis of rotation to the point of application of the force and F_2 is the force. The factor l_2 is equal to the radius of the tree at its base ($r = d/2$); the force F_2 is equal to the product of the average stress σ and the cross-sectional area of the base $F_2 \sim \sigma r^2$, from which we get $\tau_2 \sim \sigma r^3$. The exact expression for the torque τ_2 is similar to this estimate:

$$\tau_2 = \frac{\pi}{4}\sigma r^3 = \frac{\pi}{32}\sigma d^3.$$

Equating τ_1 and τ_2 results in

$$\frac{h}{d} = \frac{\sigma}{4\rho gh\sin\alpha}.$$

Inserting the values for the density and maximum stress yields

$$\left(\frac{h}{d}\right)_{\max} \cong \frac{500}{h\sin\alpha},$$

or

$$\left(\frac{h}{d}\right)_{\max} \cong \frac{22}{\sqrt{d\sin\alpha}},$$

with h and d in meters.

Apparently only low trees can tilt to a marked degree. Thus, when $h/d \sim 50$ and $h \sim 20$–30 m (a common tree), the possible inclination is less than 20° ($\sin\alpha < 0.3$), and it's less than 3° ($\sin\alpha < 0.1$) only for the tallest trees ($h \sim 100$ m). This means that a very tall tree must be nearly vertical or quite thick. If the tilt of the tree is near the limit, the tree either breaks or is "forced" to get thicker in order to decrease the ratio h/d. When the tilt is large (which is the case with a shrub or a tree with a well-developed crown), the limit of the ratio h/d is

$$\left(\frac{h}{d}\right)_{\max} \cong \frac{30}{\sqrt{d}} \text{ to } \frac{40}{\sqrt{d}},$$

which corresponds to $\sin\alpha \sim 0.3$–0.5.

(c) A tree and the wind. Wind is another cause of the torques that stress a tree. Let's assume a value of $v \sim 30$ m/sec (100 km/h) for the velocity of a strong wind. At such a high velocity the aerodynamic force acting on the trunk is determined by the inertia of the wind—that is, the resistance to drastic changes in the wind velocity blowing around the trunk. This force can be written as the product of a wind-blast pressure and a characteristic area:

$$F_w \sim P_{air}S_{char}.$$

When v is large enough, the pressure P_{air} depends only on the parameters of the air flow: its density ρ and velocity v. Dimensional analysis gives us

$$P_{air} \cong \rho_{air}v^2.$$

Now let's write the characteristic area of a trunk as $S_{char} = SK_f$, where $S = dh$ is the area of the section of the trunk perpendicular to the air flow and K_f is the coefficient of the shape and characterizes the deviation of S_{char} from S. For trees with a small crown, $0.5 < K_f < 1$; for trees with a large crown, $K_f > 1$.

Now we can write the following expression for the force of the wind:

$$F_w \cong \rho_{air}v^2dhK_f.$$

The corresponding torque can be written

$$\tau_w = F_w\frac{h}{2}.$$

Using the relationship between the torque and the stress in the tree, we get the limit:

$$\left(\frac{h}{d}\right)_{\max} \cong \sqrt{\frac{\pi}{16K_f}}\sqrt{\frac{\sigma_{\max}}{\rho_{air}v^2}}.$$

Finally, substituting the values for the parameters—$\sigma_{\max} \sim 10^7$ N/m^2, $\rho_{air} \sim 1.3$ kg/m^3, $v \sim 30$ m/sec—we obtain

$$\left(\frac{h}{d}\right)_{\max} \cong \frac{40}{\sqrt{K_f}}.$$

Thus, a strong wind can explain the nature of the plateau AB in figure 1. It should be noted that among the different parameters of a tree it is only the ratio $\sigma_{\max}/K_f$ that enters into the formula for $(h/d)_{\max}$, and the influence of this ratio is not very critical.

Now let's look at the more "regular" dependence on the right side of figure 1, corresponding to young trees with small base diameters. Why is it possible for them to increase the ratio h/d to 100 and even more? Let's turn again to the formula

$$\left(\frac{h}{d}\right)_{\max} \sim \sqrt{\frac{\sigma_{\max}}{K_fv^2}}.$$

The first thing that comes to mind is that these trees are partially screened from the wind by taller trees, remembering the dependence $(h/d) \sim v^{-1}$. Also, the younger trees usually have smaller crowns ($K_f \sim 0.5$–1). These reasons might explain the increase in the ratio h/d for slender trees.

But why is there a drastic change in the dependence $h/d = f(d)$ (point B in figure 1) at $h \sim 10$–20 m? Apparently this is the characteristic height at which a tree can bend to a great extent or deviate from the vertical. Assume that the wind has tilted a tree such that $\sin\alpha \sim 0.3$–0.5. Then we obtain a ratio h/d equal to 50 when the tree's height h is at that 10–20-m mark.

The stability of the trunk's shape

Now let's consider some estimates of the stability of the trunk's shape when it is firmly anchored in the ground. The trunk is assumed to bend a little under the action of external forces. This results in (a) storage of energy U_d related to elastic deformation and (b) a decrease in gravitational energy U_g due to a lowering of the center of gravity.

Clearly, the condition $\Delta U_g \leq U_d$ must be satisfied—otherwise the bend would increase and the trunk will break. Now let's formulate the limitations resulting from this condition. We'll bend a trunk with a constant cross section along its length. Intuition tells us that the trunk assumes the shape of an arc. The corresponding change in the gravitational energy is

$$\Delta U_g = mg(h - h_1),$$

where mg is the weight of the trunk, h is the height of the unbent tree, and h_1 is the height of the tree when bent. The values of h and h_1 are shown in detail in figure 2, which explains as well the way to calculate them:

$$h = R\beta,$$
$$h_1 = R \sin \beta,$$

where β is in radians. Thus,

$$\Delta U_g = mgR(\beta - \sin \beta).$$

When the bend is small ($\beta << 1$), we can use the approximation $\sin \beta \cong \beta - \beta^3/6$, which yields

$$\Delta U_g \cong \frac{mgR\beta^3}{6} = \frac{mgh^3}{6R^2}.$$

Now let's evaluate the potential energy stored in a rod when it is bent slightly. Note that the bending of a thin rod ($d << h$) reduces to simple deformations of the cross section: *compression and tension*. The related stress σ in the cross section can be obtained by using Hooke's law

$$\sigma = \varepsilon E,$$

where E is Young's modulus and ε is the relative deformation (compression or tension). In our case, $\varepsilon = Z/R$, where Z is the absolute deformation. Layers to the right of point O in figure 3 will undergo compression, and those to the left will undergo tension. The energy of deformation per unit volume can be estimated as the product of the force per unit area and the relative displacement:

$$F\frac{\Delta l}{V} = \frac{F}{S}\frac{\Delta l}{l} = \sigma\varepsilon.$$

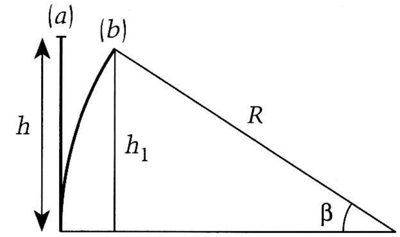


Figure 2

(a) An upright tree and (b) the same tree bent by the wind.

In the entire trunk, the energy of elastic deformation will be $U_d \cong \sigma_m \varepsilon_m V$, where σ_m and ε_m are the level of stresses and relative deformation at the middle of the radius. Since $\sigma_m = \varepsilon_m E$, $\varepsilon_m = r/2R$, $V = \pi r^2 h$, and $r = d/2$, we obtain

$$U_d \cong \frac{d^2EV}{16R^2} = \frac{\pi}{64}\frac{d^4hE}{R^2}.$$

Thus we have arrived at the resulting limit:

$$\frac{h}{d} \le \sqrt{\frac{3}{8}\frac{E}{\rho gh}},$$

or

$$\frac{h}{d} \le \sqrt[3]{\frac{3}{8}\frac{E}{\rho gd}}.$$

Inserting the values of $E \cong 5 \cdot 10^9$ Pa and $\rho \sim 500$ kg/m^3, which are characteristic of wood, we obtain

$$\frac{h}{d} \le \frac{600}{\sqrt{h}},$$

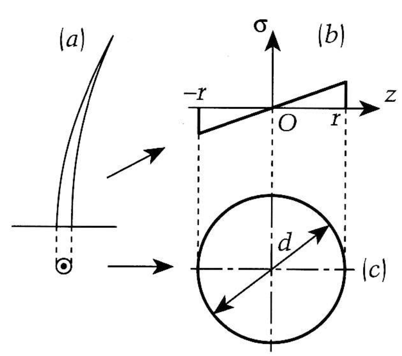


Figure 3

(a) Parameters of a bent tree; (b) dependence of stress on trunk radius r; (c) top view of the trunk's cross section.

or

$$\frac{h}{d} \le \frac{70}{\sqrt[3]{d}}.$$

Preliminary results

Now, what have we come to understood from the "mechanical" parameters of wood that describe its elasticity and density? Quite a bit, actually: the middle and right-hand portions of figure 1. The red line in figure 4 (on the next page) marks the initial curve; the thick lines show the limits associated with the strength of a vertical trunk, stability under the action of its own weight, and the effect of strong winds.

As you can see, the ideal case of a vertical tree doesn't occur in nature. Strong winds account for a plateau of h/d = const. Stability considerations for the trunk under the influence of gravity on very thick trees (several meters in diameter) results in even more severe limits for the ratio h/d than the influence of a strong wind. However, purely mechanical considerations do not clarify the nature of the maximum height for trees. As one would expect, a tree proves to be much more complicated than a dried-up stick driven into the ground.

Water runs quietly

Now let's turn our attention to the *movement of water inside a tree.* Water unites the roots, trunk, and leaves into a single living system. Let's estimate the characteristic velocities of water in the three most important cases (fig. 5): (a) during evaporation from the foliage—v_{fol}; (b) during flow in the trunk—v_{tr}; and (c) during absorption by the roots from the soil—v_{root}. We'll move from the top to the bottom of a tree—and from simple to complicated estimates.

(a) The velocity of water in the foliage. The crown of a tree needs to be cooled continuously as it is heated by the Sun. Otherwise the overheated foliage would cease to function. When there is a gentle breeze, the solar heat can basically be carried

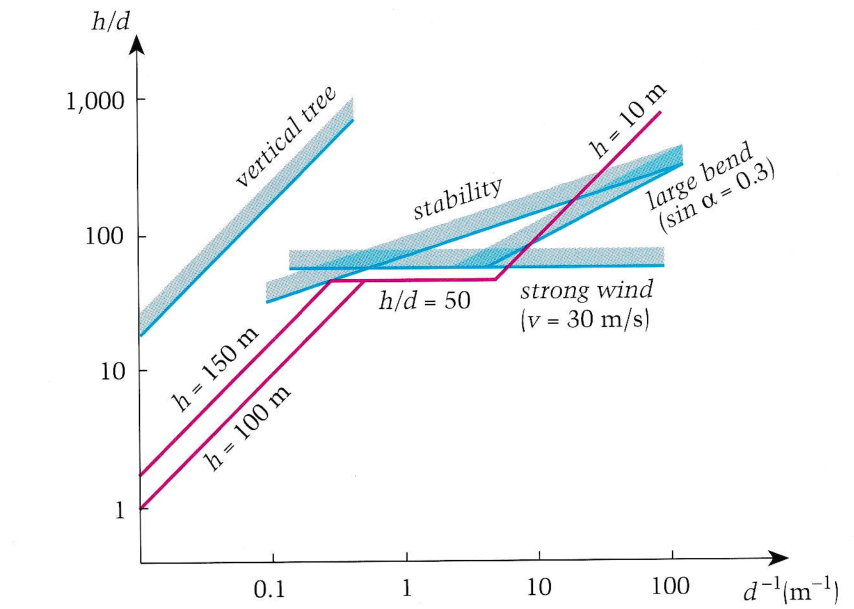


Figure 4

Limitations related to a tree's physical parameters.

off by the evaporation of water from the surfaces of leaves or needles. This primes the "pump" that extracts water from the soil. The heat balance for a leaf of area S_1 consuming solar radiation is

$$\alpha q_{\odot} S_1 = Q_{evap} \rho_w v_{fol} S_1.$$

On the left-hand side we have the input of solar energy, and on the right-hand side we have the output of energy due to evaporation from the

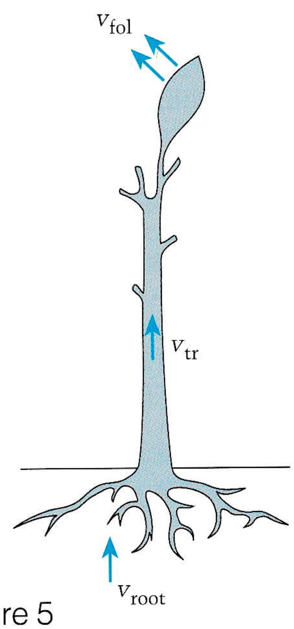


Figure 5

Water flow in a tree.

leaf's surface. Here $q_{\odot}$ denotes the average flux of solar energy at the Earth's surface ($q_{\odot} \sim 10^3$ W/m^2); α is the portion of energy consumed by a leaf ($\alpha \sim 0.2$–0.3 accounts for the reflection of light from the leaf's surface, its orientation, and the screening effect of the crown); ρ_w is the density of water ($\rho_w = 10^3$ kg/m^3); and Q_{evap} is the latent heat of vaporization for water ($Q_{evap} \sim 2 \cdot 10^6$ J/kg). From this it follows that the velocity of water supplied to the crown's surface for subsequent evaporation is

$$v_{fol} = \frac{\alpha q_{\odot}}{Q_{evap} \rho_w}.$$

With the values of the parameters given above, $v_{fol} \sim 10^{-7}$ m/sec $\sim 4 \cdot 10^{-4}$ m/h. This formula shows that v_{fol} depends only on the illumination and is proportional to it. When the wind is blowing, the rate of heat loss rises, which results in an increase in v_{fol}.

(b) The velocity of water in the trunk. Within the trunk water flows through very thin channels that are the "skeletons" of once-living cells. The diameter of these channels is extremely small—of the order of 10^{-5} m in conifers and as large as 1–$2 \cdot 10^{-4}$ m in deciduous trees. When water moves slowly through the thin channels, the resistive force F_{res} is determined by the viscosity of water.

If the layers of a liquid move with different velocities, a kind of "friction" between them arises—the viscous force. If we want to quantitatively describe a viscous liquid, we can consider the elementary example of a liquid flowing between two parallel plates set close to one another. If one plate is fixed and the other moves with a velocity v, then a viscous force arises between them:

$$F_{visc} = \eta S \frac{v}{h},$$

where S is the area of each plate, h is the distance between them, and the factor η is the coefficient of viscosity ($\eta \sim 10^{-3}$ kg/m · sec for water).

When water flows through a tube with a diameter d and a length l, the change in the velocity from the wall to the center of the tube is of the order of v/d; the area of the wall is $S \sim ld$. This gives us

$$F_{visc} \sim \eta v l.$$

In the case of uniform flow, the viscous force is balanced by the differential pressure:

$$F_{visc} = \Delta p \frac{\pi d^2}{4}.$$

This leads to an estimate of the differential pressure along the length l:

$$\Delta p \sim \frac{\eta l v}{d^2}.$$

The precise formula (Poiseuille's law) gives the value

$$\Delta p \cong 32 \frac{\eta l v}{d^2}.$$

Lifting a liquid to a height h with a velocity v_{tr} in the Earth's gravitational field through a channel of diameter d_{tr} requires a differential pressure

$$\Delta p_{tr} = \rho_w g h + 32 \frac{\eta h v_{tr}}{d_{tr}^2} = \rho_w h \left(g + \frac{32 \nu v_{tr}}{d_{tr}^2} \right),$$

where $\nu \cong 10^{-6}$ m²/s is the so-called kinematic viscosity of water. This formula shows that when v_{tr} is small, the value of Δp does not depend on velocity, and in the opposite extreme it is proportional to v_{tr}. The characteristic value of the velocity at which the dependence of Δp on v_{tr} drastically changes (it becomes much more difficult for a tree to pump the water) is $v_{char} = gd_{tr}^2/32\nu$. After substituting the values for g and ν, $v_{char} \sim (3 \cdot 10^{-7})d_{tr}^2$ (here v_{char} is measured in m/sec and d_{tr} in microns). This estimate yields $v_{char} \sim 10$ cm/h for conifers with vessels 10^{-5} m in diameter; for deciduous trees the corresponding values are $d_{tr} \sim 10^{-4}$ m and $v_{char} \sim 10$ m/h.

Despite the approximate nature of the formula for v_{char}, the experimental values for the velocity of water in the vessels of actual trees that can be found in the literature are quite close to our estimates. In a way, a tree "feels" the region where the dependence of Δp on velocity changes drastically.

The differential pressure along the trunk that corresponds to the characteristic velocity is

$$\Delta p_{tr} \sim 2\rho_w gh.$$

For high trees with $h \sim 100$–150 m, the value of Δp is about 20–30 atm.

(c) The velocity of water in the roots. To penetrate into the roots of a tree, water must cross the boundary of the root cell. This is not a simple task, since the water molecular must cross a barrier—the cell membrane—that is a double-layered, semipermeable boundary ~ 10^{-8} m wide. Water diffusion across the membrane is very slow. Simple though rather cumbersome estimates result in $v_{root} \sim 10^{-8}$ m/sec $\cong 4 \cdot 10^{-5}$ m/h.

Now all three characteristic velocities for water in a tree have been obtained. A comparison of these values is quite interesting. The amount of water flowing through the roots, the trunk, and the foliage must be the same. Thus, the area of the roots (or, strictly speaking, of the root-hairs) S_{root}, of the water-carrying system of the trunk S_{tr}, and that of the crown S_{fol} must vary inversely with the corresponding velocities. Taking S_{tr} as the unit area, for conifers we get (at a velocity $v_{tr} \sim 0.1$ m/h under natural illumination and in rather wet soil) $S_{fol} : S_{tr} : S_{root} = 200\text{–}300 : 1 : 2{,}500\text{–}3{,}000$. From this it follows that the area of the bottom is greater than that of the top, and each of them is far greater than the total cross-sectional area of the vessels in the trunk. When it forms the crown, a tree must increase its surface to a great extent. How does it do this while expending the minimum amount of building materials?

Similar problems often appear in both Nature and engineering. A lightning bolt is "compelled" to branch out into smaller and smaller channels. The frost on your windowpane must make the same "decisions." So does a snowflake.

The method of solving the problem is the same, as a rule: repeated division into smaller and smaller structures, and the last structure is the one engaged in the fundamental process (absorption of water by the root hairs, evaporation of water by the leaves, the dissipation of electric energy by lightning).

So now we know how water travels inside a tree. This knowledge has helped us estimate the velocity of water transport and understand the reason for the vast branching structures of roots and twigs. But the mist that hangs over the left-hand curve in figure 1 has not yet lifted.

The growth of trees

It's quite possible that the crux of the problem lies not in strict physical limitations on a tree's height. Perhaps trees just don't have enough time to grow as tall as we might wish. Let's look at the evolution of the size of an individual tree. When the tree is short, the relation h/d assumes approximately a steady-state value. In this case the growth of the tree slows down greatly as time goes on.

Let's take apart one simple example and show how it can be realized in nature. Let the increase in trunk mass per year be Δm = const. Then the mass increases linearly with time: $m(t) \sim t$. But the height of the tree, when the ratio h/d is constant, will increase at a substantially lower rate: the estimate

$$m \sim d^2h \sim \frac{h^3}{(h/d)^2}$$

results in

$$h \sim \sqrt[3]{m} \sim \sqrt[3]{t}.$$

Thus, the dependence $h \sim \sqrt[3]{t}$ results from the postulate m = const. Is it really valid? The increment of mass per year is determined by the amount of foliage formed in the crown (where the photosynthesis takes place), which is then distributed throughout the tree.

For all trees the ratio of the mass of the water pumped to the top to the mass of the organic substances synthesized in the crown varies within narrow limits—on average it is $k \sim 300$. This means that the increase in trunk mass is proportional to the amount of water pumped to the treetop: $\Delta m = m_w/k$. During one year this yields $m_w = \rho_w v_{tr} T_1 S(t)$, where ρ_w is the density of water, v_{tr} is the velocity of water flowing upward, T_1 is the duration of pumping during a year, and $S(t)$ is the effective area in the trunk's cross section that is involved in water transport. The first three factors on the right-hand side can be assumed to be constant ($\rho_w = 10^3$ kg/m³, $v_{tr} = v_{char}$—the characteristic velocity of water transport through the trunk, and T_1 is the length of the spring–summer growing season). The effective area taking part in pumping is

$$S(t) \sim S_{ring}(t) \cdot N_r,$$

where S_{ring} is the area of an individual ring in the tree and N_r is the number of rings. As a rule, the thickness of a yearly ring decreases as the years go on (you can examine the rings on a tree stump to verify this).

Thus, the area of the rings remains virtually constant.

To simplify our reasoning we can assume $S_{\text{ring}}(t)$ = const. As for the number of rings in a tree that engage in water transport, we must keep in mind that eventually the cells in the inner layers become obstructed with the substances being transported and water ceases to flow in them. Thus, only a small layer of wood near the bark actually participates in lifting water. So the number of rings N_r feeding the crown is rather small. It's not more than a few dozen in conifers and even fewer (less than 10) in deciduous trees.

So, where have we gotten? All the factors in the expression for the increase in the trunk's mass are virtually constant in time. So it seems probable that tree growth gradually slows: $h \sim \sqrt[3]{t}$.

What's so interesting about this dependence? Well, it makes it clear that the growth rate decreases drastically with time. If a tree grew to a height of 50 m in 50 years, it would pass the 100-meter mark at the age of 400, and it would reach a height of 200 m in 3,200 years. It looks like this is the end of the story. Of course, we must remember the numerous not-too-rigorous notions and suppositions we allowed ourselves in our long journey in the mist.

To add the finishing touch to our investigation, let's list some evolutionary considerations that work against gigantism in trees.

1. The vegetation around a tree competes with it, if not for airspace above, then certainly for nutrients in the soil. The low-lying plants adapt to changes in nutrient levels more easily than the big plants.

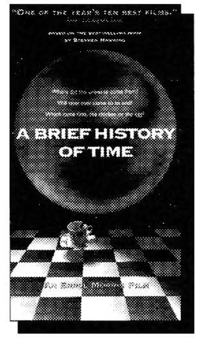
2. Over a period of hundreds and thousands of years, insects and microorganisms have plenty of time to undermine a tree's strength and vitality. Destruction begins from the core of the trunk, where the cells have practically ceased functioning.

3. Every 10^4 years or so the Earth's climate changes drastically. The last Great Ice Age, for example, ended about 10^4 years ago.

4. There is no competition among 100-m trees for sunlight. So there are no evolutionary reasons for growing to 200–300 m: the greater effort needed to lift water that high would not be compensated by any conceivable gain.

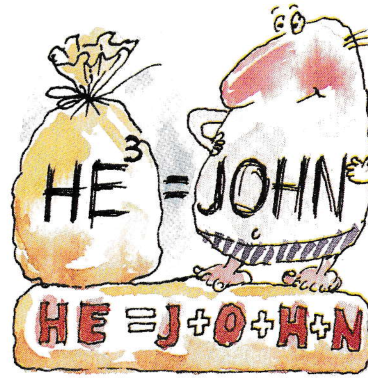
5. For a tree to be a candidate for record height, it must have an extremely vertical and very thick trunk, and it must have a comparatively small crown as well. We might say that giants work for themselves and not for posterity. Such trees have a lower reproductive rate in comparison with species that are smaller but have a well-developed crown of foliage.

All this seems to explain why there are no "leafy mastodons" more than 150 m tall anywhere on Earth.



BRAINTEASERS

Just for the fun of it!

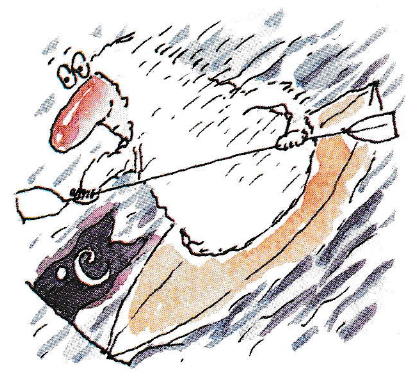


B101

Alphanumeric exchange. Replace the letters in the equations given in the figure with digits (the same letters with the same digits, different letters with different digits) so as to make the equations true. (A. Savin)

B102

Broken calculator. Each digit on the display of a standard hand calculator is represented as a certain combination of seven bars. Something has happened to my calculator: some of the bars in the rightmost place fail to turn on. However, I learned how to tell all ten digits in the broken place from one another, but if one more bar were broken, I'd be unable to always figure out the correct digit. How many bars are inoperative? (V. Dubrovsky)

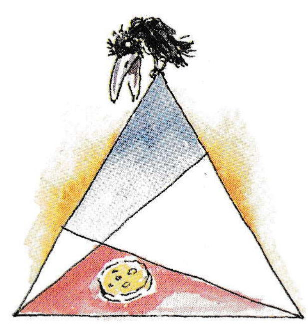


B103

Staying ahead of the iceman. Why don't fast-flowing rivers freeze at temperatures far below zero (Celsius)? (S. Krotov)

B104

Arithmetic of the hunt. For years now the famous Baron Münchhausen has gone to a lake every morning to hunt ducks. Since August 1, 1993, he's been saying to his cook every day, "Today I shot more ducks than two days ago, but fewer than a week ago." What is the greatest number of days he can make this statement, taking into account that the baron never lies? (A. Kimartsev)

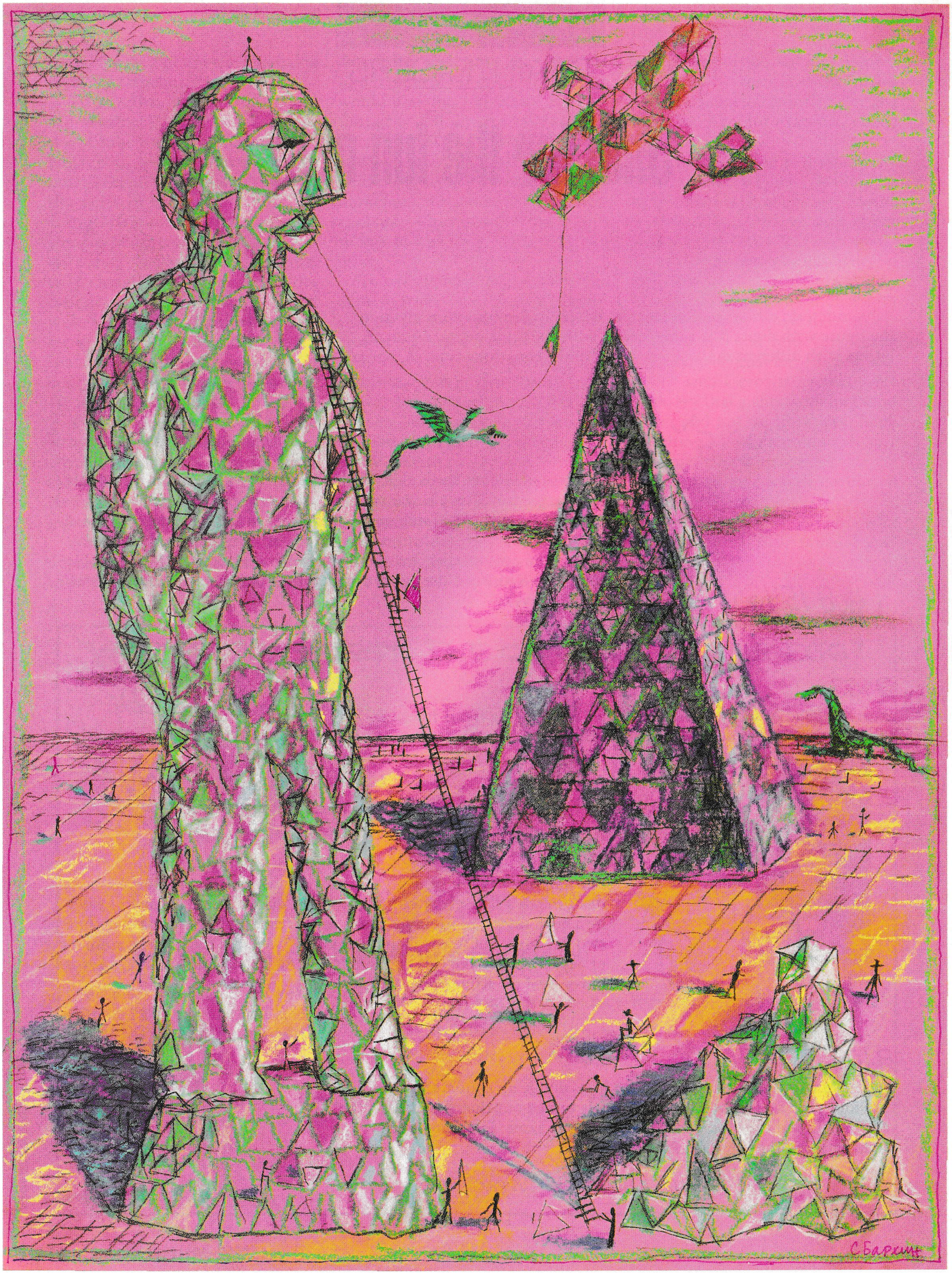


B105

Cutting an equilateral triangle. Two lines divide an equilateral triangle into four pieces as shown in the figure. It turned out that the areas of the red triangle and the blue quadrilateral are the same. Find the obtuse angle between the lines. (V. Proizvolov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 59

Art by Pavel Chernusky


С.Бархин

Penrose patterns and quasi-crystals

What does tiling have to do with a high-tech alloy?

by V. Koryepin

THIS ARTICLE IS DEVOTED to the mathematics of quasi-crystals—a new kind of material that was discovered in 1984. Actually, their "physical" discovery was preceded by the creation of intriguing (and absolutely elementary!) two-dimensional mathematical models of quasi-crystalline structures. At first these constructions, called Penrose patterns, were viewed merely as elegant little trinkets, but at the present time hundreds of research articles on the physics and mathematics of quasi-crystals have been published. So this topic offers an opportunity that is all too rare nowadays: to get a glimpse of a new, rapidly developing area of modern science by studying very elementary, almost recreational material.

We'll be considering a certain kind of tiling of the plane. A *tiling* is a covering of the entire plane with nonoverlapping figures. It's likely that an interest in tiling first arose in connection with the manufacture of mosaics, ornaments, and other decorative designs. Many ornaments are composed of repeated motifs.

One of the simplest tilings is shown in figure 1. The plane here is covered with identical parallelograms. Any parallelogram in this til-

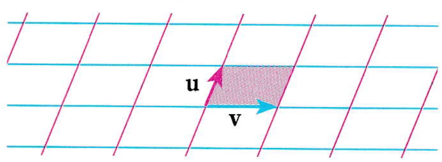


Figure 1

ing is obtained from the shaded one by translation by a vector $n\mathbf{u} + m\mathbf{v}$ (where $\mathbf{u}$ and $\mathbf{v}$ are the vectors along the sides of the shaded parallelogram, n and m are integers). It should be noted that the entire tiling is taken into itself when it is translated by either of the vectors $\mathbf{u}$ or $\mathbf{v}$. This important property can be taken as a definition: a tiling is called *periodic with periods* $\mathbf{u}$ *and* $\mathbf{v}$ if the translations by these vectors take it into itself. Periodic tilings can be quite intricate, and some of them are very beautiful.[1]

Quasi-periodic tilings of the plane

There exist interesting *nonperiodic* tilings of the plane, too. In 1974 the English mathematician Roger Penrose discovered *quasi-periodic* plane tilings. The properties of these tilings naturally generalize the properties of periodic tilings.

[1]See the Gallery Q and Kaleidoscope departments in the November/December 1991 issue of *Quantum* for a number of charming periodic tilings by M. C. Escher.—*Ed.*

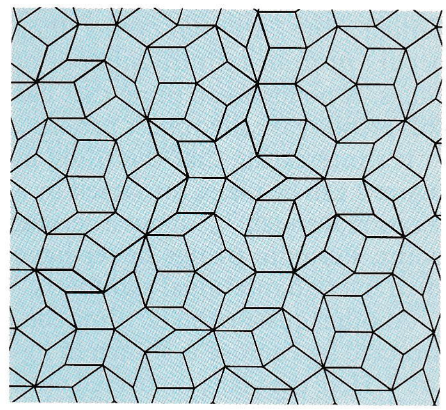


Figure 2
Penrose's example of a quasi-periodic tiling of the plane with rhombi of two types.

An example of such a tiling is shown in figure 2. The plane is completely covered without gaps and overlaps by rhombi of two kinds. These are the *wide rhombus*, with angles of 72° and 108°, and the *narrow rhombus*, with angles of 36° and 144° (fig. 3). Of course, we assume that the side lengths of both rhombi are the same. This tiling is non-

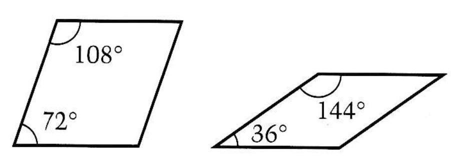


Figure 3
Wide and narrow rhombi.

Art by Sergey Barkhin

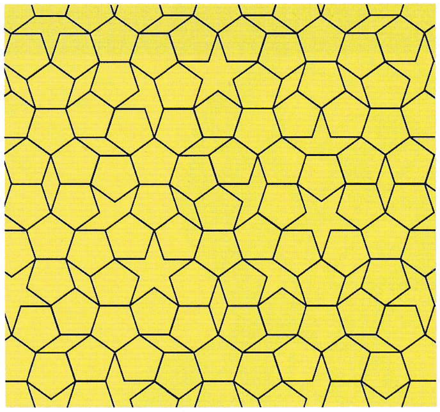


Figure 4

A quasi-periodic tiling with four types of polygons.

periodic—it can be shown that no translation can take it into itself. But it has another important property that makes it similar to periodic tilings and justifies the name quasi-periodic: *any finite part of this tiling occurs in it an infinite number of times.* A nonperiodic tiling that has this property is called *quasi-periodic.*

It's interesting that the tiling in figure 2 can be taken into itself by a rotation through 72° about a certain point. Periodic tilings never have such fivefold rotational symmetry.[2]

Another quasi-periodic tiling constructed by Penrose is shown in figure 4. Here the plane is covered with four kinds of polygons: star, rhombus, regular pentagon, and 7-gon in the shape of a paper boat. The best-known example of a quasi-periodic tiling is shown in figure 5; Penrose managed to tile the entire plane with two kinds of chickens.

Inflation and deflation

Each of the three quasi-periodic tilings above is a covering of the plane with translations and rotations of a finite number of tiles. It can't be fit onto itself by any translation, but any finite part of it is found in the entire covering infinitely many times and "equally often" throughout the plane.

[2]This fact was proved in the article "Diamond Latticework" in the January/February 1991 issue of *Quantum*; see also problem M100 in the last issue.—*Ed.*

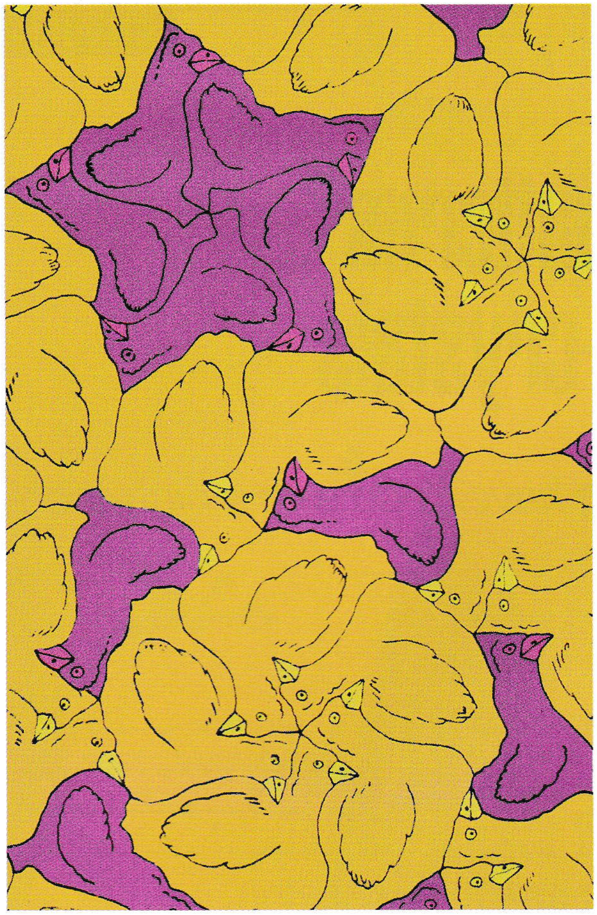


Figure 5

Penrose's two-chicken tiling.

These tilings can be subjected to some special operations. Penrose called them inflation and deflation. They'll help us understand the structure of the coverings described above. They also can be used to create Penrose patterns.

The illustration of inflation that is easiest to grasp is provided by the so-called Robinson triangles. These are two isosceles triangles P and Q with angles 36°, 72°, 72° and 108°, 36°, 36°, respectively, such that the equal sides of Q and the base of P are the same length—let's say they are 1 unit long (fig. 6).

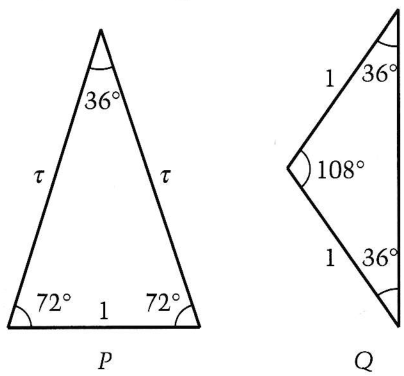


Figure 6

Problem 1. Using figure 7, prove that the equal sides of P and the base of Q are also the same length—namely, τ, where τ is the famous "golden section"[3] $(1 + \sqrt{5})/2$ (the positive root of the equation $\tau^2 = \tau + 1$).

The triangles P and Q can be cut into smaller triangles similar to the initial ones with the ratio of similarity $1/\tau$, as illustrated in figure 7: the line AC in triangle P is drawn to bisect the angle DAB, and the line CE so as to cut off an isosceles triangle CED ($\angle ECD = \angle EDC$). The "Q-triangle" FHI can be cut into two "Robinson pieces" by the line IG such that $HG = HI$.

Problem 2. In figure 7, prove that if $AB = IF = 1$, then ABC, ACE, and IGH are congruent triangles similar to triangle P, and ECD and GIF are congruent triangles similar to Q.

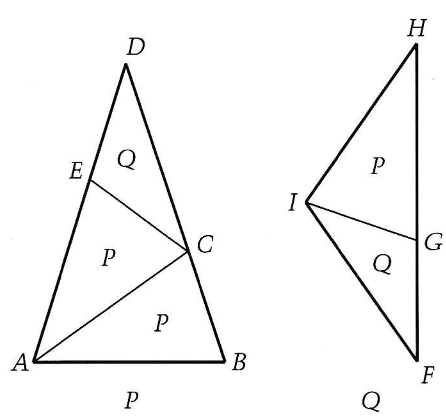


Figure 7

[3]See "The Ancient Numbers π and τ" in the Kaleidoscope of the January/February 1991 issue of *Quantum*.—*Ed.*

Thus, we've cut up the two original triangles into three equal triangles of type P and two of type Q. Comparing figures 6 and 7, we see that the dimensions of the new triangles are smaller than those of the initial ones by a factor of $1/\tau$. This special cutting is called *deflation.* The inverse operation of gluing P- and Q-triangles together into bigger triangles of the same shape is called *inflation.* As figure 7 shows, two P-triangles and one Q-triangle can be combined to make one P-triangle, and a pair of triangles, one of either kind, make up a bigger Q-triangle. The measurements of the new triangles are τ times greater than those of the initial triangles.

It's clear that we can repeat the operation of inflation to get a pair of triangles whose dimensions are τ^2 times greater than the initial triangles. If we repeat the inflations a sufficient number of times, we can get arbitrarily large Robinson triangles.

Now imagine that we take a P-triangle, inflate it by adding a congruent P-triangle and a corresponding Q-triangle, then complete the inflated triangle to a still bigger triangle, and so on (for instance, as in figure 8). Clearly, we can think of the triangles that are added to the one obtained in the previous step of this process as partitioned into P- and Q-triangles of the initial size. Thus we'll get an expanding portion of the plane tiled with the copies of the original Robinson triangles. Sooner or later, any point in the plane will be covered with this growing tiled area, so the process actually defines a tiling of the entire plane with Robinson triangles. This can be compared to the process of gradual crystallization in two dimensions except that, as we'll soon see, it's more like "quasi-crystallization."

I'll give a sketch of the proof that the chain of iterated inflations can be organized so as to obtain a nonperiodic tiling.

Let's consider the limiting ratio r of the number of triangles to that of Q-triangles in an arbitrary tiling of the entire plane with triangles of these two sorts. That is, we draw a large circle of radius R and count the numbers p_R and q_R of P- and Q-triangles inside this circle. Then r is the limit of the ratio p_R/q_R as R approaches infinity:

$$r = \lim_{R\to\infty} \frac{p_R}{q_R}.$$

I'll show first that for any *periodic* tiling using the triangles P and Q, the number r is always rational. Suppose a tiling has periods $\mathbf{u}$ and $\mathbf{v}$. Draw a grid of parallelograms with sides $\mathbf{u}$, $\mathbf{v}$ (fig. 9), choose (and shade) one of them, and denote by p the number of triangles with their bottom vertices (with respect to the grid) lying in the shaded parallelogram. (If a triangle has a horizontal base, we look at its left bottom vertex.) Do the same for Q-triangles to define the corresponding number q. The union of all these $p + q$ triangles is a so-called *fundamental region* of our periodic tiling—a polygon made of tiles whose translations by vectors $n\mathbf{u} + m\mathbf{v}$ with integer m and n tile the entire plane without overlaps. Then

$$r = \frac{p}{q}.$$

Indeed, let N_R be the number of parallelograms in our grid that lie completely inside the circle. Then p_R is approximately equal to the number of triangles with their (left) bottom vertices in these parallelograms $p_R \cong pN_R$. The error of this approximation accounts for the triangles in the circle whose respective parallelograms stick out of the circle, and for the triangles that don't fit in the circle themselves but their parallelograms do. All such triangles lie within a ring between circles of radii $R - d$ and $R + d$ with some fixed d. So their number is not greater than the area $\pi[(R + d)^2 - (R - d)^2] = 2\pi dR$ of the ring divided by the area of the triangle, which is proportional to R, while N_R is proportional to R^2. Similarly, $q_R \cong qN_R$, where the error of the approximation is negligible with respect to N_R. It follows that in the limit the ratio of the numbers of P- and Q-triangles is equal to this ratio counted for the fundamental region—that is, to p/q. This shows

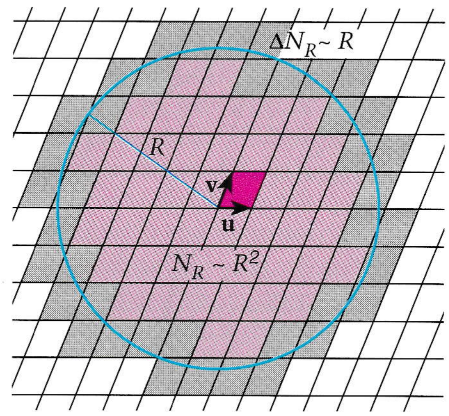


Figure 9

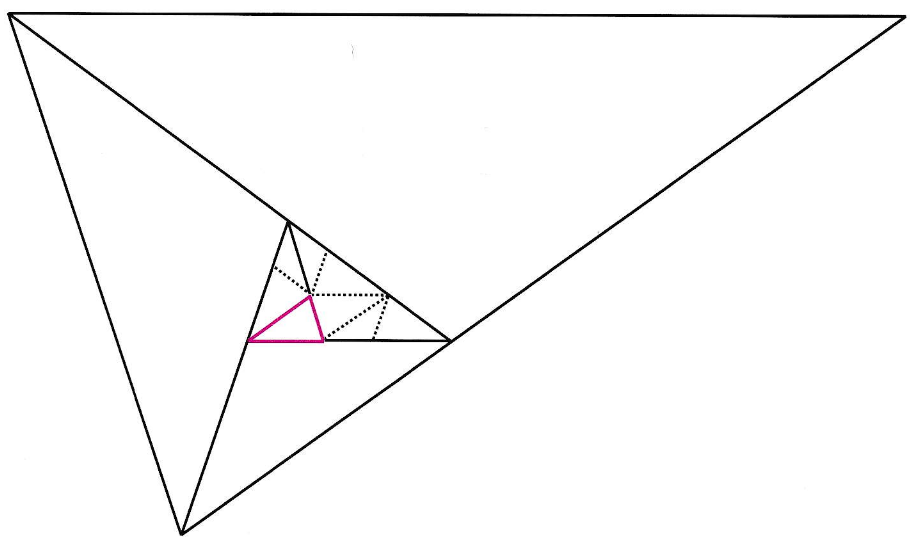


Figure 8
Inflating P-triangle devouring the plane.

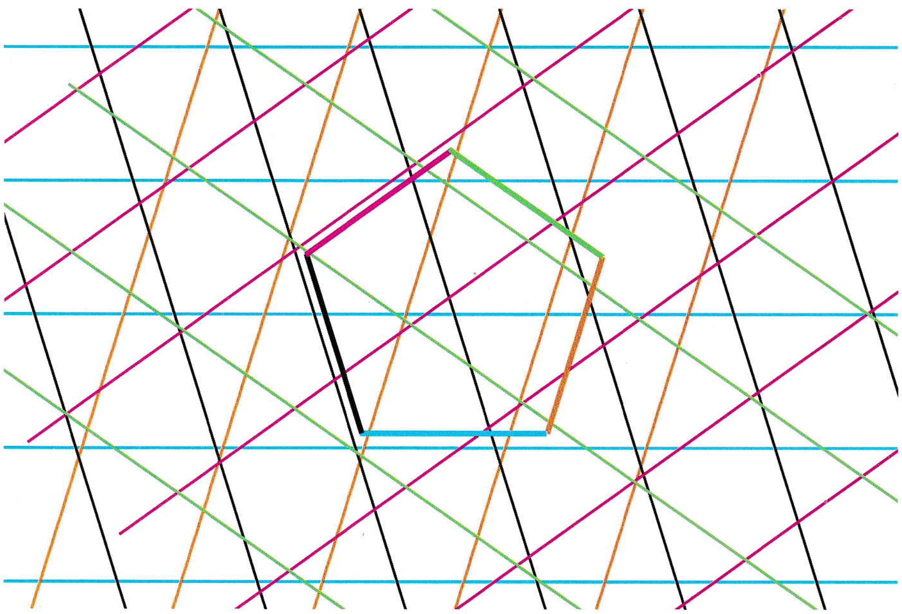


Figure 10

that for any periodic tiling, the value of R must be rational.

We complete the proof that the tiling is nonperiodic by showing that its R is irrational. Imagine that we deflate our tiling. Then it is not hard to see that the initial fundamental region will contain $p' = 2p + q$ smaller triangles and $q' = p + q$ smaller Q-triangles. A little thought will show that we can assume, without loss of generality, that in the sequence of inflating triangles used to construct our tiling the position of each triangle (except the first) with respect to the previous one is the same. This means that each triangle in this sequence, just like the very first triangle, generates the tiling of the plane by its copies and the copies of the corresponding Q-triangle, and this tiling repeats the original one except that it's scaled up by a fixed factor. So the limit of the ratio of the numbers of P- and Q-tiles in the inflated tiling is the same as in the original tiling. And, of course, the same as in the deflated tiling as well. This can be written as

$$\frac{p}{q} = \frac{p'}{q'} = \frac{2p+q}{p+q} = 1 + \frac{p/q}{p/q+1}.$$

Solving this equation for p/q, we get $p/q = \tau = (\sqrt{5} + 1)/2$, which is a contradiction, because this number is irrational. Thus, our tiling cannot be periodic.

Problem 3. Show that a tiling produced by iterated inflations contains infinitely many copies of any of its finite parts. Along with the nonperiodicity that we have already proved, this will show that our tiling is quasi-periodic.

The tiling of the plane with Robinson triangles isn't unique—there are infinitely many such tilings. Roughly speaking, this is because the deflation of the triangle ABD in figure 7 could be done by drawing the bisector from vertex B rather than A. Using this freedom of choice, we can turn the triangular tiling into the rhombic tiling in figure 2. The chicken tiling is also generated by Robinson triangles.

Duality

The method of creating quasi-periodic tilings given above looks like a clever guess. But there is a regular way of constructing quasi-periodic coverings. It's based on the so-called *duality transformation*, introduced by the Dutch mathematician de Bruijn.

I'll illustrate his idea by way of the example of tiling the plane with rhombi (see figure 2). To begin with, we draw a grid (denoted by G) that consists of five sets of lines. The lines of each set are parallel to one of the sides of a fixed regular pentagon. The distance between two neighboring lines of each set is one unit (see figure 10). The only restriction on the relative positions of the sets is that no three lines can meet at one point (it can be proved that such a grid can indeed be constructed). The lines of grid G cut the plane into an infinite number of polygons called the *faces* of the grid; the sides and vertices of these polygons are called the *edges* and *nodes* of the grid, respectively. (A similar terminology will be used for the quasi-periodic tiling T we're going to construct: its faces, edges, and nodes are the tiles (rhombi), their sides, and their vertices, respectively.)

Now let's perform the duality transformation. It relates the faces, edges, and nodes of the grid G to the nodes, edges, and faces of the tiling T in accordance with the rule described below.

Let's assign the numbers 1, 2, 3, 4, 5 to the sides of the regular pentagon used in the construction of the grid (fig. 11) and the corresponding sets of parallel lines. Denote the vectors drawn from the center of the pentagon to the midpoints of the sides with corresponding numbers as $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, $\mathbf{e}_4$, $\mathbf{e}_5$; we can assume that these are unit vectors. Assign the numbers 0, ±1, ±2, ... to the lines of each set so that the numbers increase along the directions of vectors $\mathbf{e}_i$ (starting

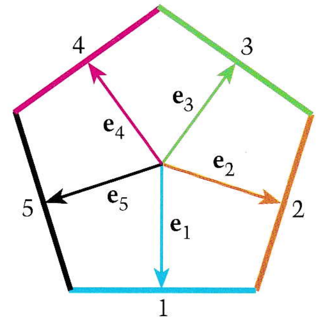


Figure 11

from the initial pentagon). Thus, every line can be denoted as $l_i(n)$, where i is the number of its set and n is its number in this set.

Now we associate with every face F of the grid G a set of five numbers $n_1(F), n_2(F), \ldots, n_5(F)$, where $n_i(F) = n$ of the face F lies between the lines $l_i(n)$ and $l_i(n+1)$ of the ith set (i = 1, 2, 3, 4, 5). These numbers are uniquely defined for any face, because for any of the five sets of lines, the face F belongs to exactly one strip between the lines of this set. Now we can locate the node $N = N(F)$ of our tiling that is "dual" to the face F of the grid G. This node is the point whose position vector $\mathbf{v}(F)$ is given by

$$\mathbf{v}(F) = n_1(F)\mathbf{e}_1 + n_2(F)\mathbf{e}_2 + n_3(F)\mathbf{e}_3 + n_4(F)\mathbf{e}_4 + n_5(F)\mathbf{e}_5.$$

To construct the nodes of the new tiling, we choose a fixed point O and use it as the endpoint of each of these position vectors. Thus, every face of G is assigned to a node of the tiling T. To form the tiling, some of the nodes must be joined with edges—these will be the sides of the rhombic tiles. The rule of joining is simple: two nodes, $N_1 = N(F_1)$ and $N_2 = N(F_2)$, are joined with an edge of the tiling if and only if the corresponding faces F_1 and F_2 of the grid G have a common edge. So each edge of G is associated with an edge of T.

Finally, consider an arbitrary node A of the grid G. It is always the intersection of exactly two straight lines of the grid, so it's a vertex of exactly four faces of the grid F_1, F_2, F_3, F_4 around A. Then the corresponding nodes of the tiling N_1, N_2, N_3, N_4 will be joined in just this order to form the quadrilateral $N_1N_2N_3N_4$. It is not difficult to show (see problem 1 of the editor's postscript) that every edge of our tiling T is of unit length and is parallel to one of the vectors $\mathbf{e}_i$. This means that $N_1N_2N_3N_4$ is a rhombus similar to one of the two rhombi in figure 3. Thus, the construction of the required rhombic tiling is completed.

It remains to show that this tiling is indeed quasi-periodic; for that, see the last section of this article.

The duality transformation illustrated above is a general method for constructing quasi-periodic tilings. The regular pentagon in our construction could be replaced by any regular polygon with not less than seven sides.

Quasi-periodic tiling of three-dimensional space

Another way to generalize is to apply duality to quasi-periodic tilings in space. Indeed, there exists a three-dimensional generalization of Penrose patterns. Space can be tiled with parallelepipeds of a special kind without gaps and overlaps. Each of these "tiles" is a copy of one of two parallelepipeds—the so-called Ammann–Mackay parallelepipeds. A parallelepiped is determined by the three vectors joining one of its vertices to the endpoints of the edges issuing from this vertex. For the first Ammann–Mackey parallelepiped these three vectors are

$$\mathbf{e}_1 = (0, 1, \tau),$$
$$\mathbf{e}_2 = (-\tau, 0, -1),$$
$$\mathbf{e}_3 = (\tau, 0, -1);$$

for the second they are

$$\mathbf{e}_4 = (0, -1, \tau),$$
$$\mathbf{e}_5 = (\tau, 0, 1),$$
$$\mathbf{e}_6 = (0, 1, \tau).$$

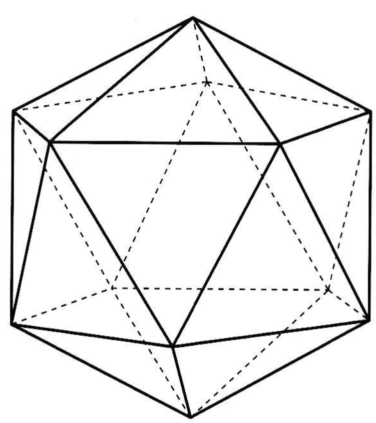


Figure 12
The icosahedron—a regular polyhedron with 20 triangular faces, 30 edges, and 12 vertices.

The tiling of space with translations and rotations of these parallelepipeds can't be taken into itself by translation, but any of its finite parts is repeated infinitely many times in the entire tiling, so it's quasi-periodic. This tiling is connected with the symmetries of the *icosahedron* (fig. 12), which never occur in periodic tilings.

It turned out that this very symmetry is characteristic of an aluminum and manganese alloy $Al_{0.86}Mn_{0.14}$ (discovered in 1984) when it is cooled quickly. Thus, Penrose tilings shed light on the structure of this new substance. And not just this one—other quasi-crystals have been found recently. Scientists are currently studying these novel substances both experimentally and theoretically.

Editor's postscript: why does it work?

The article you've just read deals mostly with the "whats" and "hows" of quasi-periodic tilings. It explains what the term means and how these tilings can be constructed. However, many of the "whys" remain unexplained, especially those concerning the duality transformation. Actually, it's not very easy to prove that the construction based on duality described above really leads to a quasi-periodic tiling. On the other hand, we believe that many of our readers would be interested in the proof and would even like to do it themselves. So we've decided to divide it into a number of simpler steps and present these in this postscript as a series of statements to be proved. We'll freely use notations introduced in the article.

The first group of statements will bring about a better understanding of the structure of T and the connection between the tiling and the grid G that generates it.

1. Let A_1 and A_2 be the nodes of T corresponding to adjacent faces F_1 and F_2 of grid G. If the common side of F_1 and F_2 lies on a line of the ith set, then the vector $\overrightarrow{A_1A_2} = \pm\mathbf{e}_i$.

2. The quadrilateral $N_1N_2N_3N_4$ formed from the edges of T as de-

scribed in the article for any node of G is always a "wide" or "narrow" rhombus (fig. 3). These rhombi will be called "tiles."

Now let's take any two faces F and F' of G, choose arbitrary points P and P', respectively, in these faces, and move along a straight line from P to P'. As we move, we'll pass through a series of faces. Denote them by $F_1 = F, F_2, \ldots, F_n = F'$ and label A_k the node of T corresponding to F_k $(k = 1, 2, \ldots, n)$. Consider the vector sum $\overrightarrow{A_1A_2} + \overrightarrow{A_2A_3} + \ldots + \overrightarrow{A_{n-1}A_n}$.

3. All the vectors in this sum that are parallel to $\mathbf{e}_i$ have the same direction—that is, they are all equal either to $\mathbf{e}_i$ or to $-\mathbf{e}_i$. The first possibility occurs if the angle between vectors $\mathbf{e}_i$ and $\overrightarrow{PP'}$ is acute, the second if the angle is obtuse. What happens if these vectors are perpendicular?

4. Suppose that after replacing each vector in the sum $\overrightarrow{A_1A_2} + \overrightarrow{A_2A_3} + \ldots + \overrightarrow{A_{n-1}A_n}$ with the appropriate vector $\pm\mathbf{e}_i$ and collecting like terms, the sum $c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3 + c_4\mathbf{e}_4 + c_5\mathbf{e}_5$ emerges. Then the following relation holds:

$$n_i(F') = n_i(F) + c_i$$

(the numbers $n_i(F)$ were defined above, in the section about duality).

This statement isn't quite trivial, because the same vector (here, $\overrightarrow{A_1A_n}$) can be represented as a linear combination of the vectors $\mathbf{e}_i$ in many different ways. So the specific way in which we defined the coefficients c_i is important.

In the sequence of nodes $A_1 \ldots A_n$ each node (except A_n) is joined with an edge to the next (because faces F_k and F_{k+1} are adjacent). Any such sequence of nodes with the edges connecting them will be called a "path." If all the edges in a path make acute angles with a certain direction (as in statements 3 and 4), we'll say it's a "progressive" path.

5. Any two nodes in T can be connected with a progressive path. The distance r between any two nodes is greater than 1 except when they are the endpoints of an edge ($r = 1$) or of two edges issuing at an angle of 36° from one node ($r = 2 \cos 72°$).

6. The duality correspondence associates different faces of G with different nodes of T—that is, it's a one-to-one correspondence.

7. Statement 4 holds for any path from $A_1A_2 \ldots A_n$ with F and F' defined as the faces of G associated with its endpoints.

Now it can be shown that the rhombi resulting from the duality transformation really tile the plane without gaps and overlaps. (This fact by itself is not used, though, in the subsequent proof that these rhombi form a quasi-periodic structure.)

8. Every edge of T is a common side of exactly two rhombi.

9. Take any rhombus from T and fix a point A in it. Move from A to an arbitrary given point X along a straight line and consider the emerging chain of rhombi, each of which (except the first) borders on the side of the previous one that is crossed by the segment AX. This chain is finite.

Since the last rhombus in the above chain covers X, it follows that our rhombic tiles leave no gaps on the plane.

10. Two overlapping tiles (that is, different but having common interior points) cannot have a common vertex.

11. If tiling T contains overlapping tiles, then there are two edges with a single common point (that is not their common vertex). The shortest distance between their endpoints is less than 1.

12. If two edges AB and BC make an angle of 36°, then no two edges drawn from A to C can have a common point other than their common endpoint.

Comparing statement 11 with statements 5 and 12, we conclude that our tiles do not overlap, so T is an actual tiling. The next two statements show that it is nonperiodic. We'll use the notation $s_i(n)$ for the strip between the lines $l_i(n)$ and $l_i(n+1)$ of grid G.

13. If T has a period $\mathbf{p}$, then $\mathbf{p}$ is representable as $p_1\mathbf{e}_1 + \ldots + p_5\mathbf{e}_5$ with integer coefficients p_i such that for any face F there exists a face F' that satisfies the equation $n_i(F') = n_i(F) + p_i$.

14. Given two arbitrary sets of five integers each—$(n_1, \ldots, n_5)$ and $(p_1, \ldots, p_5)$—there is a positive integer k such that the intersection of five strips $s_i(n_i + kp_i)$, $i = 1, \ldots, 5$, is empty.

The nonperiodicity of T follows from statements 13 and 14, because if $p = p_1\mathbf{e}_1 + \ldots + p_5\mathbf{e}_5$ is a period of T, then, by statement 13, together with any face F the grid G must contain the face F_1 such that $n_i(F_1) = n_i + p_i$, where $n_i = n_i(F)$, and the face F_2 such that $n_i(F_2) = n_i(F_1) + p_i = n_i + 2p_i$, and, in general, the face F_k such that $n_i(F_k) = n_i + kp_i$ for any k. But that would mean that the intersection of the strips $s_i(n_i + kp_i)$ is a face of G for any k and so is a nonempty set, which contradicts statement 14.

To make sure that T is quasi-periodic, we have to show that any finite part T_0 of T is repeated in T infinitely many times. The concluding series of statements demonstrates that there

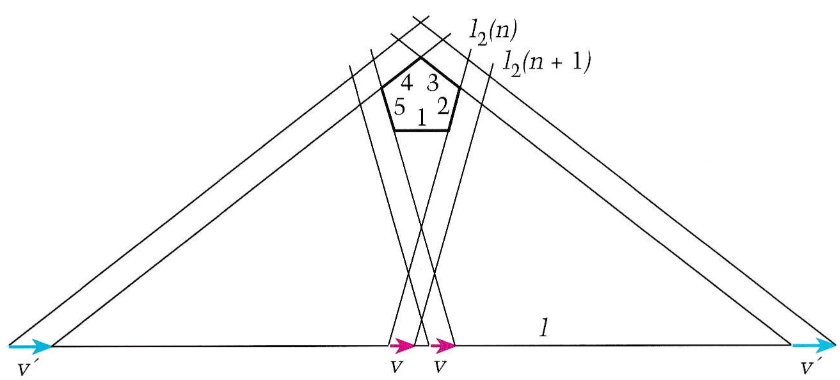


Figure 13

are arbitrarily many translations that carry all the nodes in T_0 into other portions of T, and, therefore, T_0 has arbitrarily many copies in T.

Choose an arbitrary line l from, say, the first set of lines that form grid G, and denote by $\mathbf{v}$ the vector joining the points where this line meets two successive lines of the second set, $l_2(n)$ and $l_2(n+1)$. Similarly, let $\mathbf{v}'$ be the vector intercepted on l by two lines of the third set. Then, because of our order of numbering lines in a set, the corresponding vectors for the fourth and fifth sets will be $-\mathbf{v}'$ and $-\mathbf{v}$ (fig. 13).

15. For any integers i, j, and k the translation along the vector $\mathbf{v}$ takes the strips $s_1(i)$, $s_2(j)$, and $s_5(k)$ into $s_1(i)$, $s_2(j+1)$, and $s_5(k-1)$, respectively.

16. If we shift all the lines of the third and fourth sets by a vector with a small enough length (less than a certain δ) without moving the three other sets of lines, the part T_0 of T remains unchanged.

17. For any $\varepsilon > 0$ one can find integers n and m such that $|n\mathbf{v} - m\mathbf{v}'| < \varepsilon$. (Hint: show that the ratio of the lengths of $\mathbf{v}$ and $\mathbf{v}'$ is irrational—in fact, it equals $1/\tau$—and use the Fractional Parts Theorem from the article "Ones Up Front in Powers of Two" in the last issue of *Quantum*.)

18. Let's choose n and m using statement 17 for $\varepsilon = \delta$, where δ is the number from statement 16. Then the translation by the vector $\mathbf{t} = n\mathbf{e}_2 + m\mathbf{e}_3 - m\mathbf{e}_4 - n\mathbf{e}_5$ will take T_0 into another finite part of T.

The last statement ensures the existence of at least one copy of T_0. The more pairs of integers m, n satisfying $|n\mathbf{v} - m\mathbf{v}'| < \varepsilon = \delta$ we can find, the more copies of T_0 we get. Choosing n and m with a sufficient margin—say, for $\varepsilon = \delta/100$—we'll get a hundred pairs (n, m), $(2n, 2m)$, ..., $(100n, 100m)$ satisfying this inequality, and so, a hundred reproductions of T_0. We can show in the same way that the number of reproductions of T_0 is arbitrarily large.

The proofs will be published in the next issue.

Backtracking to Faraday's law

Electrolysis and energy conservation

by Alexey Byalko

THE MOLECULES OF ANY chemical compound consist of atoms. Electrons are responsible for the interaction between atoms. However, the electrons in an atom are not all on an equal footing. The fundamental role in chemistry is played by the electrons in the outer shell. Stable molecules are formed when the energy of the coupled system of atoms is lower than the sum of the energies of these atoms when they're on their own.

In a number of cases the forces holding the individual parts of a molecule together are those of electrostatic (Coulomb) attraction. This happens when one part of a molecule has an excess number of electrons—this is a negative ion. Another part of the molecule lacks the same number of electrons—this is a positive ion.

When the ions of a substance are in a medium with a high permittivity ε, the Coulomb forces between them are weakened by the same factor ε. As a result some ions become freer and begin to move relative to one another. Such solutions are called electrolytes.

Electrolytes are good conductors of electric current. The carriers of charge in electrolytes are ions, and so the conductivity of electrolytes is called ionic. The passage of current through an electrolyte is accompanied by chemical reactions at the electrodes that lead to the separation of the elements constituting the electrolyte. This process is called electrolysis. And it is with electrolysis that the electronic nature of a chemical interaction manifests itself most vividly.

The great English physicist Michael Faraday (1791–1867) was the first to investigate electrolysis. He proved experimentally that the mass of the substance separated during electrolysis is proportional to the charge and the chemical equivalent of the substance. The discovery of the laws of electrolysis played an essential role in the formation of modern concepts of the structure of matter. But now we can do the opposite: we can derive Faraday's laws from these concepts.

It will be easier to do this using a concrete example. Let's look at the electrolysis of water (H_2O). Pure water dissociates rather easily into the ions H^+ and OH^-. To increase the ionic conductivity we add a salt, acid, or base that strongly dissociates into ions—for instance, NaOH—to the water.

Let's place electrodes—that is, chemically inert conductors—into this electrolyte and connect them to a battery. Positively charged Na^+ ions now move to the negatively charged cathode, and the negative OH^- ions move to the anode. This is why positively charged ions are called *cations* and negatively charged ions *anions*. When they reach the electrodes, the ions are neutralized: anions give their extra electrons to the anode, and cations gain the electrons they lack.

We can write down the chemical reactions taking place at the electrodes:

$$4OH^- \rightarrow 4e^- + O_2 + 2H_2O \text{ (anode)},$$
$$4H_2O + 4e^- \rightarrow 2H_2 + 4OH^- \text{ (cathode)}.$$

These reactions eventually lead to the dissociation of the water:

$$2H_2O \rightarrow 2H_2 + O_2.$$

For every two hydrogen atoms and one oxygen atom formed, two electrons must pass through the circuit. Measuring the current I and the time t of its passage, we find the total charge: $q = It$. Dividing this charge by the electronic charge e, we find the total number of hydrogen atoms formed:

$$n_H = \frac{It}{e}.$$

The number of oxygen atoms formed during this time is

$$n_O = \frac{It}{2e}.$$

In general, the number of atoms of

Art by Sergey Ivanov

the substance formed on the electrode is equal to

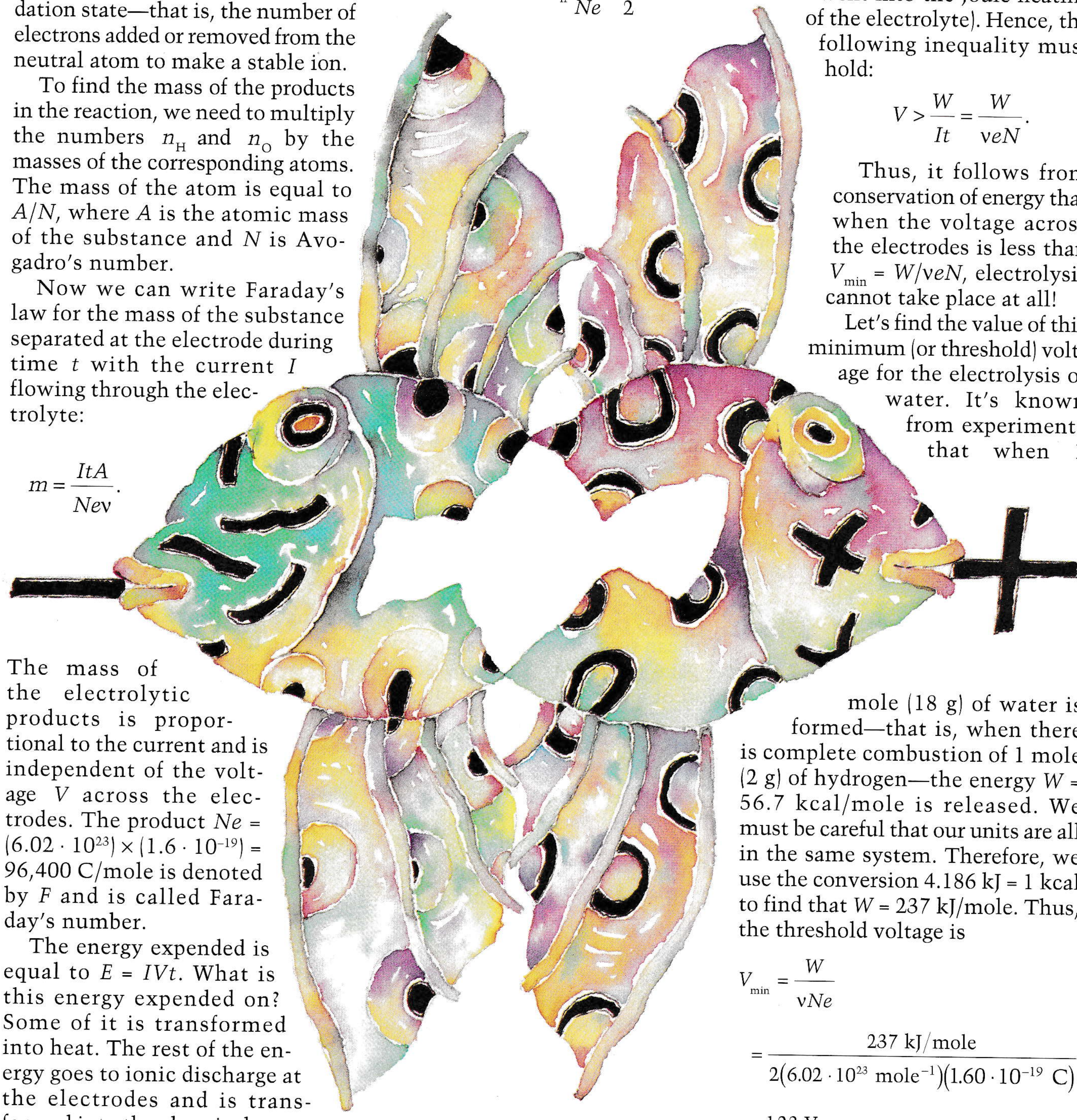
$$n = \frac{It}{\nu e},$$

where ν is the atom's valence or oxidation state—that is, the number of electrons added or removed from the neutral atom to make a stable ion.

To find the mass of the products in the reaction, we need to multiply the numbers n_H and n_O by the masses of the corresponding atoms. The mass of the atom is equal to A/N, where A is the atomic mass of the substance and N is Avogadro's number.

Now we can write Faraday's law for the mass of the substance separated at the electrode during time t with the current I flowing through the electrolyte:

$$m = \frac{ItA}{Ne\nu}.$$



The mass of the electrolytic products is proportional to the current and is independent of the voltage V across the electrodes. The product $Ne = (6.02 \cdot 10^{23}) \times (1.6 \cdot 10^{-19}) =$ 96,400 C/mole is denoted by F and is called Faraday's number.

The energy expended is equal to $E = IVt$. What is this energy expended on? Some of it is transformed into heat. The rest of the energy goes to ionic discharge at the electrodes and is transformed into the chemical energy of the products obtained during electrolysis. Let's return to the electrolysis of water and determine this energy.

Using a catalyst, let's make the hydrogen and oxygen formed during the electrolysis of water react. This reaction releases energy in the form of heat:

$$W = Q_h\left(m_{H_2} + m_{O_2}\right) = Q_h \frac{It}{Ne}\frac{\mu_{H_2O}}{2},$$

where μ_{H_2O} is the molar mass of water and Q_h is the heat of combustion for hydrogen.[1] This means that the energy transformed into the chemical energy of the electrolytic products is equal to W. From the conservation of energy it follows that $W < E = IVt$ (the difference $E - W$ went into the Joule heating of the electrolyte). Hence, the following inequality must hold:

$$V > \frac{W}{It} = \frac{W}{\nu eN}.$$

Thus, it follows from conservation of energy that when the voltage across the electrodes is less than $V_{min} = W/\nu eN$, electrolysis cannot take place at all!

Let's find the value of this minimum (or threshold) voltage for the electrolysis of water. It's known from experiments that when 1 mole (18 g) of water is formed—that is, when there is complete combustion of 1 mole (2 g) of hydrogen—the energy $W =$ 56.7 kcal/mole is released. We must be careful that our units are all in the same system. Therefore, we use the conversion 4.186 kJ = 1 kcal to find that $W =$ 237 kJ/mole. Thus, the threshold voltage is

$$V_{min} = \frac{W}{\nu Ne} = \frac{237\ \text{kJ/mole}}{2\left(6.02 \cdot 10^{23}\ \text{mole}^{-1}\right)\left(1.60 \cdot 10^{-19}\ \text{C}\right)} = 1.23\ \text{V}.$$

A potential difference is established between the electrolyte and the electrode placed in it. This occurs because metals are capable of

[1]Remember that the heat of combustion for a fuel is the amount of heat released with complete combustion of 1 g of fuel. (Don't mix up the notations: q is the total charge, Q_h is the heat of combustion.)

passing into a solution in the form of ions. Here the electrode is negatively charged and the electrolyte is positively charged. Along the surface of the electrode a layer of positive ions is formed. Notice that this potential difference doesn't cause a current to flow through the half cell (that is, one electrode with its corresponding electrolyte). The ions are in a state of dynamic equilibrium with the electrode.

Another variant is possible, whereby the metal is positively charged and the electrolyte is negatively charged. This occurs when the electrode placed in the electrolyte doesn't "dissolve," but rather negative ions of the electrolyte precipitate onto its surface.

It's impossible to measure the voltage in a half cell directly. However, we can choose a specific half cell as a standard and measure the potential difference when it is combined with various other half cells. If we assign a zero potential difference to the standard half cell, we can attribute the measured potential difference to the other cell. This potential difference is known as the *oxidation potential* ϕ.

The table above presents results obtained by this method. Platinum was chosen as the standard electrode and the electrolyte is a one molar[2] solution of hydrogen ions. The energy needed to reduce 1 mole of the given substance from ions is determined by the oxidation potential.

Since we are usually interested in the voltage between a pair of electrodes rather than the potential of an individual electrode, it's clear that the choice of standard electrode can be arbitrary. What's important is that a certain minimum potential difference is necessary for the chosen pair of electrodes if electrolysis of each substance in the electrolyte is to occur.

The oxidation potentials for the metals in the table are listed in descending order. Notice that the sequence of metals (including hydrogen)—K, Na, Mg, Al, Zn, Fe, Pb, H, Cu, Ag, Au—is the electrochemical series for the metals, so well known from chemistry. In this series each metal can replace any metal to its right in a salt solution. For example, the following reactions are possible:

$$Zn + Cu^{++} \rightarrow Cu + Zn^{++},$$
$$2Na + 2H^{+} \rightarrow 2Na^{+} + H_2.$$

You can see from this example how physics gives a quantitative description of a law noted empirically in chemistry. The reactions occur in the direction in which energy is released, and the threshold voltage is directly proportional to the amount of energy released.

Now, using the table as a guide, we can say why hydrogen, and not sodium, is liberated at the cathode during electrolysis of an aqueous solution of NaOH. The threshold voltage for electrolysis of the NaOH solution is equal to 3.94 V (2.71 V + 1.23 V). If the potential difference across the electrodes V < 1.23 V, electrolysis cannot take place; if 1.23 V < V < 3.94 V, then metallic sodium cannot form. If V > 3.94 V, sodium can be formed, but it will rapidly react with the water.

The fact that there are different electrolytic threshold voltages for different metals provides a way of separating one metal from another. Imagine you have copper with small impurities of silver and gold. How can you extract the noble metals? Electrolyze the copper salt $CuSO_4$ by using this unpurified copper as the anode. The following reactions are possible at the anode: liberation of oxygen, dissolution of copper (that is, ion formation), and dissolution of silver and gold. Metallic copper precipitates on the cathode. If the potential difference of the electrodes is taken to be rather small (smaller than 0.80 V – 0.34 V = 0.46 V) to prevent the dissolution of silver and gold and the electrolysis of water, the anode is destroyed. Ions of copper Cu^{++} pass into solution and precipitate on the cathode, while the silver and gold—which are insoluble at this potential difference—remain at the anode.

Now let's look at the voltage–current characteristics of the electrolytes. The situation here is different for different kinds of reactions—for instance, for reactions occurring in batteries and the electrolysis of water.

In a battery the chemical reaction is chosen so that the substances formed when the battery is discharged during use either remain in solution or precipitate on the electrodes. When the battery is charged, the reaction will be reversed, which will restore its initial state.

Reaction	ϕ (V)	Reaction	ϕ (V)
$K^+ + e^- \rightarrow K$	2.92	$H^+ + e^- \rightarrow \frac{1}{2}H_2$	0
$Na^+ + e^- \rightarrow Na$	2.71	$\frac{1}{2}Cu^{++} + e^- \rightarrow \frac{1}{2}Cu$	–0.34
$\frac{1}{2}Mg^{++} + e^- \rightarrow \frac{1}{2}Mg$	2.37	$Ag^+ + e^- \rightarrow Ag$	–0.80
$\frac{1}{3}Al^{+++} + e^- \rightarrow \frac{1}{3}Al$	1.66	$\frac{1}{3}Au^{+++} + e^- \rightarrow \frac{1}{3}Au$	–1.50
$\frac{1}{2}Zn^{++} + e^- \rightarrow \frac{1}{2}Zn$	0.76	$Br^- \rightarrow \frac{1}{2}Br_2 + e^-$	1.06
$\frac{1}{2}Fe^{++} + e^- \rightarrow \frac{1}{2}Fe$	0.44	$OH^- \rightarrow \frac{1}{2}H_2O + \frac{1}{4}O_2 + e^-$	1.23
$\frac{1}{2}Pb^{++} + e^- \rightarrow \frac{1}{2}Pb$	0.13	$Cl^- \rightarrow \frac{1}{2}Cl_2 + e^-$	1.36

[2]A one molar solution contains one gram-atom of hydrogen per liter of solution.

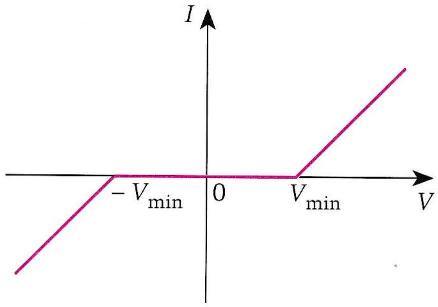


Figure 1

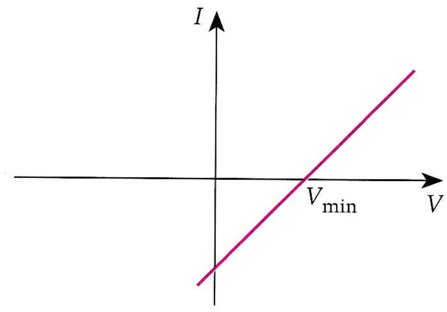


Figure 2

During electrolysis hydrogen or other gases might be liberated; a metal in the solution might be reduced and precipitate. Then the reaction cannot be reversed.

In the second case, at a voltage lower than the threshold V_{min} there is no current, and the resistance of the electrolyte is practically infinite. When the potential difference becomes negative, the anode and cathode change places and, starting at the voltage $-V_{min}$, the current reappears. So the voltage–current graph looks like figure 1.

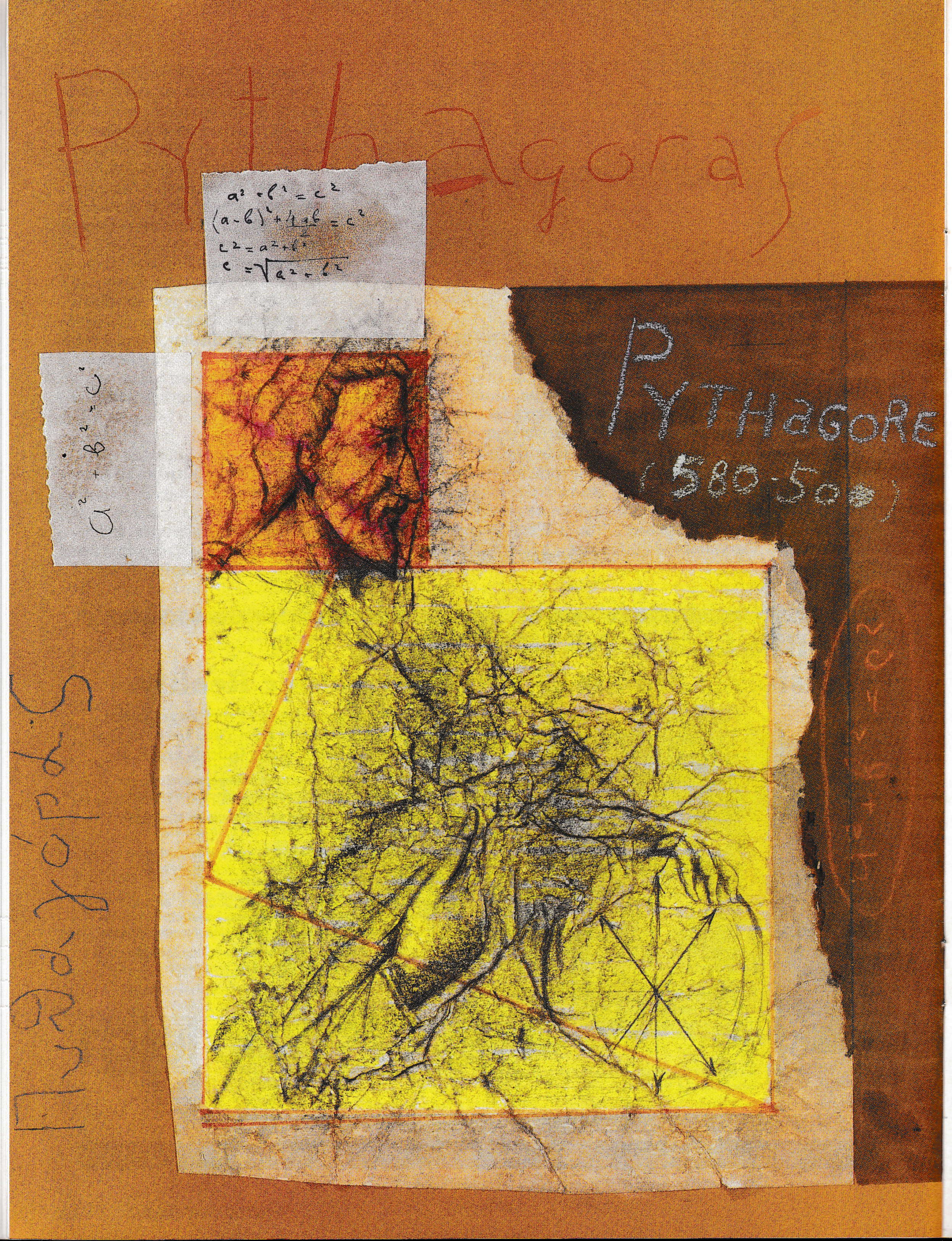
If, during a reaction of the first type, a substance has already been liberated at the electrodes, the electrolyte, along with the electrodes, becomes a source of current with an electromotive force exactly equal to the threshold voltage. Therefore, the voltage–current relationship obeys Ohm's law but is shifted by the threshold voltage (fig. 2). When the applied voltage is zero, the battery acts like the usual current source while the chemicals last.

The slope of the lines in figures 1 and 2 are characteristic of the electrolyte's conductivity and in the case of flat electrodes is directly proportional to the electrode cross section and ion concentration. It is inversely proportional to the distance between the electrodes. (The graphs are valid for low current densities.)

To conclude, I'd like to underscore the fact that the existence of a threshold voltage in electrolysis is a direct consequence of Faraday's law and the law of conservation of energy. This phenomenon demonstrates the interrelationship of chemical and electrical forms of energy.

Exercise

Electrolysis of 1 liter of an aqueous solution of $AgNO_3$ is performed at a voltage $V = 1$ V. One gram of silver precipitates. By how many degrees does the temperature of the solution rise? (Neglect heat losses.)



Pythagoras
$a^2 + b^2 = c^2$
$c^2 = a^2 + b^2$
$c = \sqrt{a^2 + b^2}$
$a^2 + b^2 = c^2$
PYTHAGORE
(580-500)
Πυθαγόρας

The good old Pythagorean Theorem

We never tire of using it—and "reproving" it!

by V. N. Beryozin

SOME THEOREMS AND problems have unusual fates. For instance, why have professional and amateur mathematicians always been so interested in the Pythagorean theorem? Why were so many of them not satisfied with the existing proofs, but kept seeking for their own, which brought the total number of proofs—during the twenty-five comparatively observable centuries—to several hundred?

When we talk about the Pythagorean theorem, peculiarities immediately start cropping up, starting with its name. It's widely thought that it was *not* Pythagoras who gave the first wording of the theorem. And it's equally doubtful that he gave any proof at all. If Pythagoras was a real person (and some people doubt even that!), then he most probably lived in fifth and fourth centuries B.C. He never wrote anything himself. He called himself a philosopher—which meant, to his way of thinking, a "lover of wisdom." He founded the Pythagorean union, whose members occupied themselves with music, gymnastics, mathematics, physics, and astronomy. He must have been an outstanding orator, too, judging from a legendary account of his arrival at Croton. "The first appearance of Pythagoras before the people of Croton began with a speech addressed to the youth, in which he gave such a rigorous and, at the same time, fascinating presentation of the duties of youth that even the elders of the town asked him not to leave them without instructing them as well. In his second speech he pointed to the law and moral purity as the bases of the family. In the following two he addressed children and women. As a consequence of his last speech, in which he particularly censured luxury, thousands of valuable women's dresses were brought to Hera's temple, since the women dared not appear in them in public." However, even as late as the second century A.D.—that is, 700 years after Pythagoras—there lived and worked very real people, prominent scientists, who were under the clear influence of the Pythagorean union and held in great respect what was allegedly created by Pythagoras.

Art by Yury Vashchenko

It's also beyond doubt that the interest in the theorem is generated both by the important place it occupies in mathematics and by the satisfaction experienced by the authors of new proofs, who have overcome a problem mentioned by the great Roman poet Horace (65–8 B.C.): "It is difficult to express well matters of common knowledge." Originally, the theorem established the relation between the areas of the squares constructed on the hypotenuse and legs of a right triangle: *the square constructed on the hypotenuse is equal in area to the sum of the squares on the legs.*

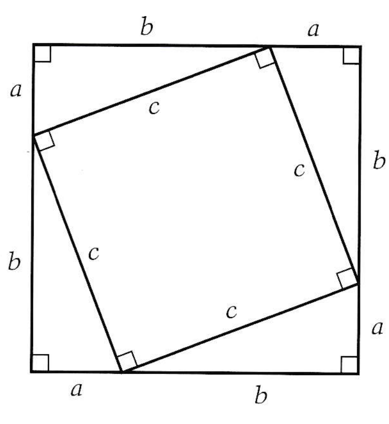


Figure 1

Several relatively simple proofs of the theorem are based on this geometric version of its statement. One of these is based on figure 1. In it, a square of side $(a + b)$ is divided into four congruent right triangles with sides a, b, and c, and a smaller square with side c. Equating the areas of the large square with the sum of the areas of the four triangles and the small square, we find $(a + b)^2 = c^2 + 4(ab/2)$, which leads algebraically to the result.

A second relatively simple proof was devised by a young general in the American army named James A. Garfield, who would later serve as President of his country. In figure 2, two copies of triangle ABC are arranged to fit into a trapezoid. It follows that triangle ABE is an isosceles right triangle. Equating the area of the trapezoid with the sum of the areas of the triangles gives the desired result.

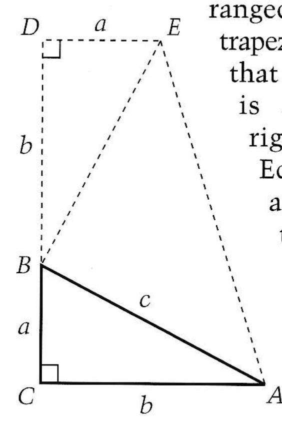


Figure 2

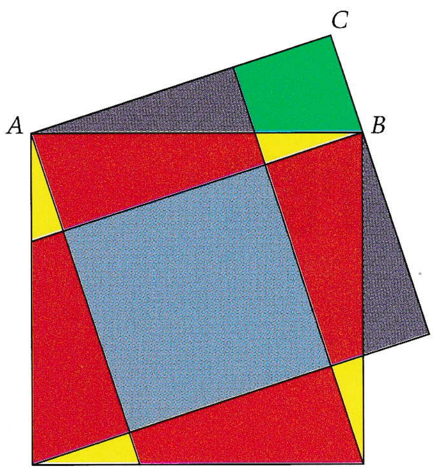


Figure 3

Actually, this is not a new proof: figure 2 is half of figure 1 (can you see how?). A genuinely new proof is illustrated in figure 3. The original triangle is denoted ABC, with its right angle at C; the square on the hypotenuse AB is built externally, two other squares—internally. The sides of the smaller squares are extended wherever the squares overlap. We see that the extension of the side of the square on BC parallel to BC passes through the vertex of the square on AB opposite A, and a similar fact is true for the square on the other leg AC.

Problem[1]

1. Devise a proof of this observation. Also, prove that it would remain true if the squares on the legs were constructed externally, and on the hypotenuse—internally.

This is the crucial point for the proof in question, as well as for a handful of other proofs. Now, it's easy to understand that the figures of the same color in figure 3 are congruent. The big square on the hypotenuse is cut into four red trapezoids, four yellow triangles, and a blue square; we can write this as $4r + 4y + b$. The square on the bigger leg, AC, is cut into two red trapezoids, one yellow triangle, one blue square, two purple triangles, and one green trapezoid: $2r + y + b + 2p + g$; and the third square can be represented as $y + g$. But a red trapezoid and a yellow triangle clearly come together to

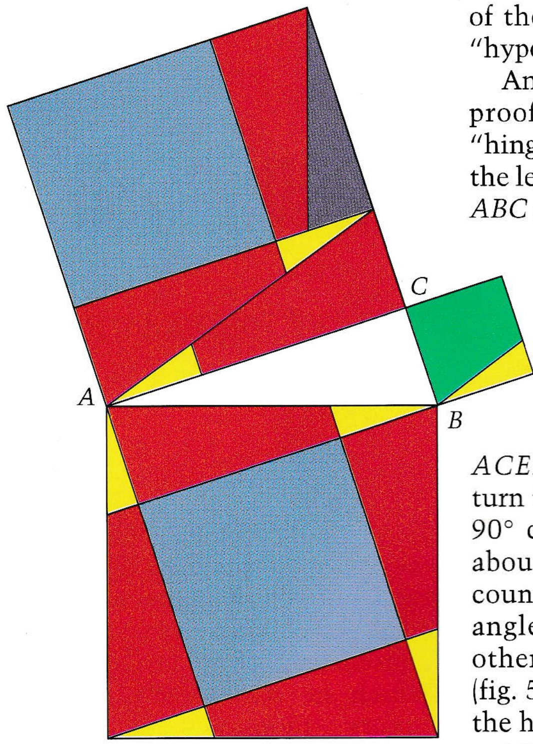


Figure 4

make the original right triangle, which is cut into $p + g$; so $p + g = r + y$. Finally,

$$2r + 2y + b + 2p + 2g = 4r + 4y + b,$$

which completes the proof. A graphic illustration of this cut-and-paste argument is given in figure 4, where all the squares are constructed outside the triangle, and one purple–green triangle in the square on AC is replaced with a congruent red–yellow triangle. (The remaining purple and green pieces in the "leg squares" correspond to one of the red–yellow triangles in the "hypotenuse square.")

And here's another cut-and-paste proof—it could also be called a "hinge" proof (fig. 5) The squares on the legs of the original right triangle ABC are constructed internally, and the smaller one—on BC—is shifted along BC so as to abut on the big one from the outside. The triangle BDE thus formed in figure 5a is congruent to ABC. Now we cut the two-square figure $ACEDFG$ along BA and BD, and turn triangle ABC about A through 90° clockwise and triangle BDE about D through the same angle counterclockwise. Bingo! The triangles neatly fit together on the other side of the remaining piece (fig. 5b) and form . . . the square on the hypotenuse!

Modern geometry prefers an algebraic formulation of the Pythagorean theorem: if the sides of a right triangle are measured with the same unit of length, then the square of the number expressing its hypotenuse equals the sum of the squares of the numbers expressing the legs. To put it more succinctly, the square of the hypotenuse is equal to the sum of the squares of the legs. I'll give two proofs that use this algebraic formulation.

The square in figure 6 is divided into four congruent right triangles and a smaller square. This figure accompanied the famous proof of the Pythagorean theorem in a treatise by the great Indian mathemati-

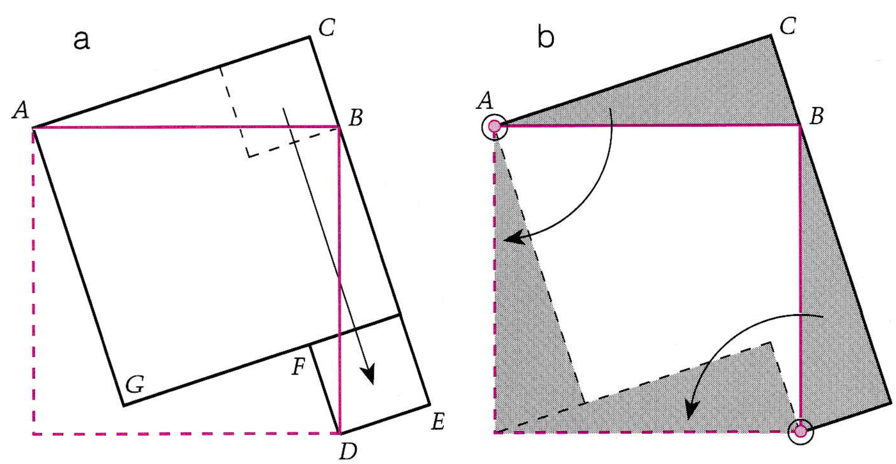


Figure 5

[1]Problems supplied by the editor.

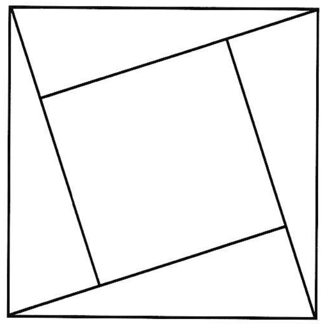


Figure 6

cian of the 12th century, Bhaskara Acharia. (The text of the proof consisted of one word: "Behold!")

If the side length of the big square (which is the hypotenuse of the given right triangle) is c, and the legs of the triangle are a and b, then the side length of the smaller square is $|a - b|$, and we have

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

or

$$c^2 = a^2 + b^2.$$

In figure 7 the given right triangle is divided into two smaller triangles by the altitude from the right angle. All three triangles are similar to one another. And this is the clue to the next proof. The areas of any similar figures constructed on the sides of the given triangle are in the same ratio as the areas of the squares on the respective sides. So the areas of the three triangles in figure 7 can be written as ka^2, kb^2, and kc^2 with a certain constant factor k. But the area of the big triangle equals the sum of the other two—that is, $kc^2 = ka^2 + kb^2$, or $c^2 = a^2 + b^2$. This is a very instructive proof, and it has the simplest possible diagram.

In the centuries that have passed since the discovery of the Pythagorean theorem, a lot of students must have gotten bad grades for mistakes they made in its proof. But undoubtedly the converse statement is even more treacherous in this sense, because students often mix the two theorems and, instead of the latter, refer to the direct statement. Here's what the converse theorem says: if the sides a, b, and c of a triangle satisfy the relation $a^2 + b^2 = c^2$, then the triangle is a right triangle, and its right angle is opposite the side c. The proof doesn't require a drawing and is quite simple. Suppose the equality $a^2 + b^2 = c^2$ is true for a given triangle. Construct a right triangle with legs a and b. Then, in accordance with the direct Pythagorean theorem, its hypotenuse is $c^2 = \sqrt{a^2 + b^2}$. It follows that the side lengths of this triangle are equal to those of the given triangle, so the two triangles are congruent, and the given triangle is also a right triangle.

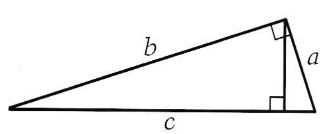


Figure 7

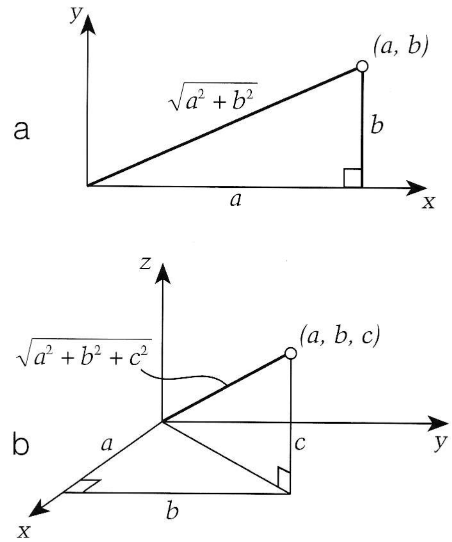


Figure 8

The Pythagorean theorem can be generalized in many ways. First of all, it can be rewritten in the following coordinate form, which is clearly equivalent to the original statement: *the square of the distance from a point in the coordinate plane to the origin is equal to the sum of the squares of its coordinates* (fig. 8a). In this form it remains valid for space of three (fig. 8b) as well as any other number of dimensions.

Problem

2. Prove the three-dimensional variant of the coordinate version of the Pythagorean theorem.

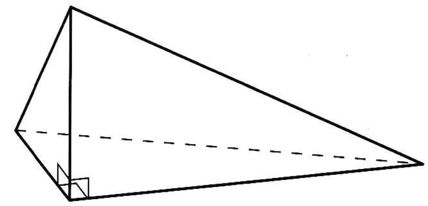


Figure 9

Another three-dimensional generalization was probably established in the 17th century and is quite often used in applied mathematics. It turns out that if three faces of a tetrahedron are right triangles with right angles at their common vertex, then the sum of the squares of their areas equals the square of the area of the fourth face (fig. 9).

Problem

3. Prove this "Pythagorean theorem for a tetrahedron."

In closing, I should mention that the Pythagorean theorem has a plane generalization, too. It belongs to Pappus of Alexandria (3rd century A.D.) and states that *if three parallelograms are constructed on the sides of an arbitrary triangle, two of them externally and one internally, in such a way that the sides of the first two parallelograms parallel to sides of the triangle pass through the vertices of the third parallelogram* (fig. 10), *then the area of the latter is equal to the sum of the areas of the first two.* The Pythagorean theorem follows from this by the observation made in the third proof above (see problem 1).

To prove Pappus's theorem, shift the sides of the external parallelo-

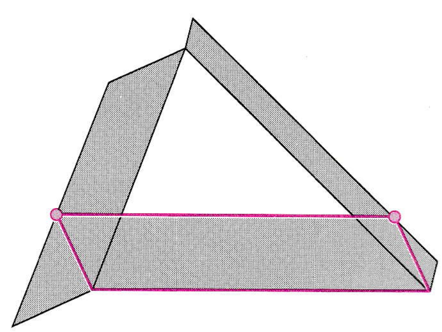


Figure 10

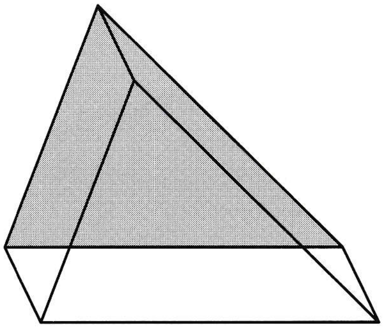


Figure 11

grams parallel to the triangle's sides so that they issue from the vertices of the third parallelogram as shown in figure 11. Of course, the areas of the parallelograms don't change under this operation. Now if we cut the original triangle off from the pentagon thus obtained (see figure 11), the remaining area will be equal to the sum of the areas of the external parallelograms. And if we cut off the shaded triangle (which is clearly congruent to the original one) from the same pentagon, we'll leave the third parallelogram. This immediately proves the theorem.

Problems

4. Figures 12–15 illustrate four other proofs of the Pythagorean theorem. Restore these proofs.

5. *Lunes of Hippocrates.* Three semicircles are constructed on the sides of a right triangle as shown in figure 16. Prove that the total area of the two shaded "lunes" thus obtained is equal to that of the triangle.

6. A right triangle is divided into two triangles by the height from the right angle (fig. 7). The radii of the incirles of these small triangles are r_1 and r_2. Find the inradius of the big triangle.

7. In a right triangle, a and b are the legs, c is the hypotenuse, and h is the height on the hypotenuse. Prove that a triangle with side lengths $a + b$, h, and $c + h$ is also a right triangle. ◘

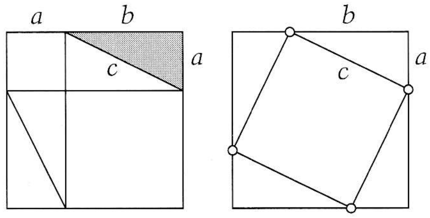


Figure 12

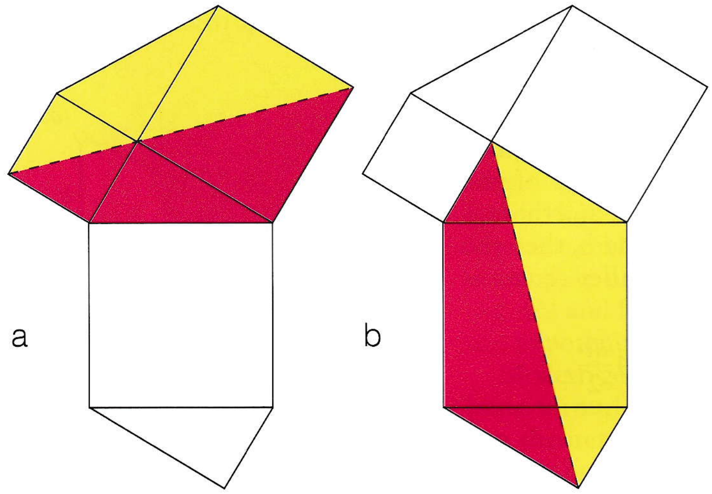


Figure 13

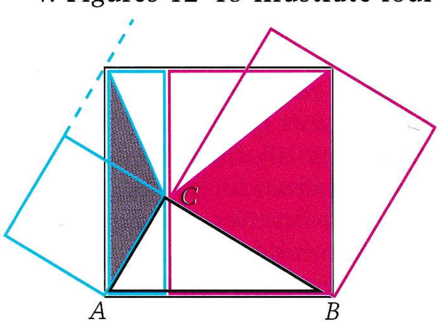


Figure 14

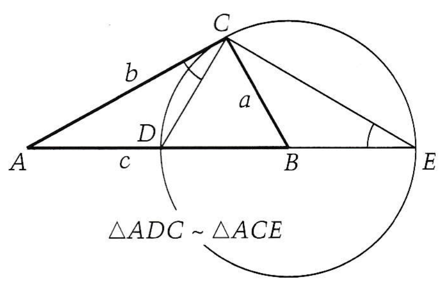


Figure 15

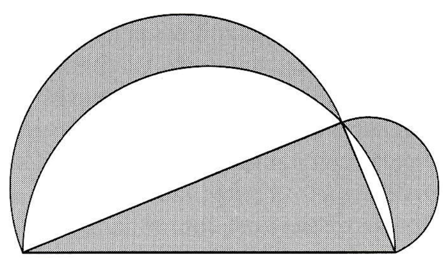


Figure 16

MATH INVESTIGATIONS

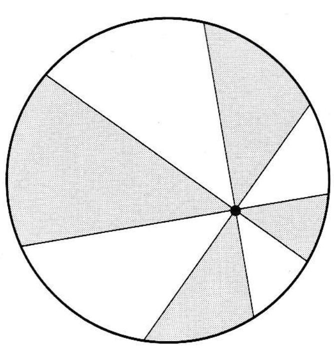
The Pizza Theorem—Part I

A tribute to Joe Konhauser's memory

by George Berzsenyi

IN THE FIRST PART OF THIS two-part column my purpose is to introduce my readers to a wonderful problem area, which was recently called to my attention by Professor Stan Wagon of Macalester College in St. Paul, Minnesota. Stan Wagon learned about it from the late Professor Joseph D. E. Konhauser, who featured it in his well-known "Problem of the Week" (P of W) program. In Stan's words, this problem has "vastly improved my mathematical life and eaten up a lot of my free time." I trust my readers will not escape unscathed either.

In its simplest form, the Pizza Theorem is as follows: *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*



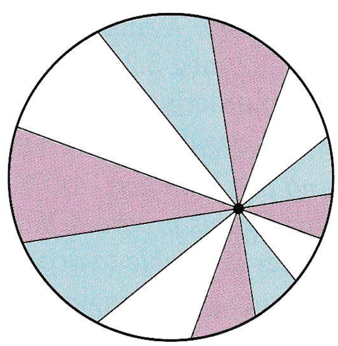
resulting "pseudosectors" are colored alternately black and white, then the sum of the black areas is equal to the sum of the white areas. My first challenge is: **Verify this result**. There are at least three different ways to accomplish this task; two of them don't even require calculus.

> The late Professor Joseph D. E. Konhauser was a superb problemist, an insightful geometer, an inspiring teacher, and a great friend. In particular, he has made many memorable contributions as a member of the Putnam Problems Committee, as a member of the USA Mathematical Olympiad (USAMO) Committee, and as the editor of the *Pi Mu Epsilon Journal* and its "Puzzle Section." From 1968 till his untimely death (in 1992), he was also well known for his "Problem of the Week" program at Macalester College, never repeating a problem and keeping meticulous files on all the wonderful challenges he shared with his students and colleagues.

My second challenge is: **Generalize and extend the above theorem.** Clearly, one may be able to vary the number of chords, one may attempt to use more than two colors, one can search for extensions to three dimensions, and may consider ellipses in place of circles. One may also define the pseudosectors via equidistant points on the circle rather than equal angles between the pseudoradii. One can abandon even the common interior point in favor of sectors drawn in succession, forming specified angles with one another; a problem of this type was recently posed and solved in *Kvant*, the Russian-language sister journal of *Quantum*.

In part II of this column, I'll provide precise references, but for the present, it may be best for you to investigate your own ideas.

In conclusion, I want to thank Stan for letting me use his material. After Joe's death, he not only took over his P of W program, but he is making the problems available via e-mail to many other mathematicians around the world. Presently, with Prof. Daniel Velleman of Amherst College, he is busily editing the best of the 700+ problems that appeared in Joe's program, for a volume of the Dolciani Series of the Mathematical Association of America. Perhaps my readers are familiar with Stan's earlier books and will look forward to the appearance of this one. I also wish to thank Prof. Hung Dinh of Macalester College for sharing with me the P of W materials while he was running the program during Stan's sabbatical,



and Prof. John Duncan of the University of Arkansas and Prof. Thomas Banchoff of Brown University for their speedy responses to my inquiries about various extensions to the problem.

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M101

Thirty plus. The six most active students in a class formed 30 different committees, every two of which intersected with each other—that is, had at least one member in common. Prove that it's possible to form one more committee intersecting with each of these 30 committees. (S. Fomin)

M102

In terms of increasing polynomials. Prove that (a) the polynomial $y = x^2$, (b) any polynomial, can be represented as the difference of two polynomials each of which is a monotonic increasing function. (V. Pikulin)

M103

A test for congruence? Equal sides of two acute isosceles triangles are the same length, and the radii of their incircles are the same, too. Are these triangles necessarily congruent? (A. Yegorov)

M104

Integer roots of integer quadratics. Find (a) at least one pair, (b) all pairs of nonzero integers (p, q), such that the equations $x^2 + px + q = 0$ and $x^2 + qx + p = 0$ both have integer roots. (E. Turkevich)

M105

Swapping apartments. In a certain city only paired exchanges of apartments are permitted: if two families swap their apartments, they can't take part in any other exchanges on the same day. Prove that any complex exchange of several apartments can be performed in two days. (N. Konstantinov, A. Shnirelman)

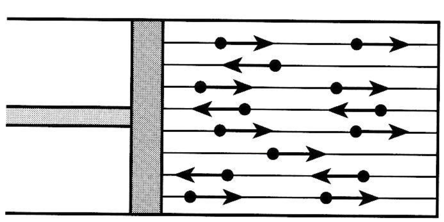
Physics

P101

Float and sinker. A thin homogeneous cylindrical float is made out of a light substance with a density ρ_1. A lead sinker of density ρ_2 is tied with fishing line to the bottom of the float. What conditions must the ratio of the masses of the sinker and the float satisfy for the float to rest vertically in the water? (Neglect the forces of surface tension. The density of water is ρ_0.) (M. Semyonov)

P102

One-dimensional ideal gas. Behind a piston in a cylindrical vessel there is a "one-dimensional" ideal gas of "molecules" that are small elastic balls moving only in the direction of the cylinder's axis (see the figure). The mass of each molecule is m, the initial velocity is v_0, the initial concentration is n_0, and the initial volume behind the piston is V_0. Find the expression relating the pressure and volume for such a "gas." (M. Semyonov)



P103

Removing stains. Some people, when they get a greasy stain on their clothes, use a hot iron to remove it. What is the physics underlying this procedure? (S. Krotov)

P104

Curious meter reader. Once when I was looking at my electric meter I found a curious thing. When I switched on a lamp of power P_1 for 1 minute, the disk in the meter made N_1 revolutions; when another lamp of power P_2 was switched on, the disk made N_2 revolutions. How many revolutions N_3 did the disk make when I switched on both lamps? (A. Deshkovsky)

P105

Telephoto shooting. It's known that when you take a picture of long objects from a distance with a telephoto lens, these objects appear flattened along the line of sight. Determine the change in the ratio of transverse to longitudinal dimensions in an image obtained with a telephoto lens having a focal length $F = 1$ m in comparison with the object itself—an automobile moving toward the photographer at a distance $L = 200$ m.

ANSWERS, HINTS & SOLUTIONS ON PAGE 55

Songs that shatter and winds that howl

How sound is your reasoning about these sonic phenomena?

THE PHYSICS OF SOUND IS one of the most "lively" areas of science. Suffice it to say that humans have long felt that they possess the most perfect "device" in acoustics (and in optics as well). However, step by step people discovered that the world is filled with sounds imperceptible to our ear. Only by keen observation of animals, who have far better acoustical "equipment" than we do, and by inventing new, artificial senors, have we greatly enlarged the palette of sounds and put them to use on our behalf. Architecture, music, medicine, engineering—these are but a few of the fields where our modern understanding of sound is applied. This installment of the Kaleidoscope is full of examples.

Questions and problems

1. The sound of an artillery shell exploding reached one observer in 3 s and another observer in 4.5 s. Use graphical techniques to find the position of the explosion if the distance between the observers was 1 km.

2. It's known that if a source of sound and a person are at approximately the same height, the sound is heard at a greater distance in the direction of the wind than otherwise. Why?

3. Why does the wind howl?

Art by Dmitry Krymov

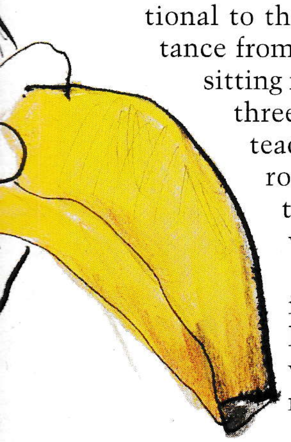
"The theory of sound, as it is normally understood, is part of the same field of study as the theory of oscillations in general. . . . As a rule we shall consider only those classes of oscillatory motion for which our ears proved to be a good sensing device. Without hearing it is doubtful whether we would be as much interested in oscillations as the eyes are in light."—Sir John William Strutt Rayleigh

4. What kind of wave is produced by a violin bow—transverse or longitudinal?

5. The air pressure in an automobile tire can be determined by the sound it makes when you strike the tire with a metal object. How?

6. Why does the banjo have a ringing sound, while the harp has a soft, singing tone?

7. The loudness of a sound is inversely proportional to the square of the distance from the source. A student sitting in the fifth row is about three times as far from the teacher as one in the first row, yet they hear the teacher almost equally well. Why?



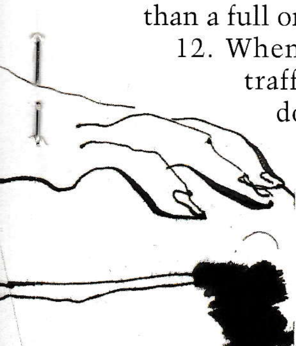
8. Which of two tuning fork resounds longer—one held in a vise or one placed on a resonating box?

9. The sound absorption of glass is much lower than that of air. Yet we can greatly decrease street noise by closing the window, and if there is a storm window, we can almost completely prevent any sound from coming into the room. How can we explain this?

10. Why is it so quiet after a snowfall?

11. Why does a half-full kettle make more noise just before boiling than a full one?

12. When there is automobile traffic beneath your windows, sometimes the glass hums loudly. This annoying sound can be significantly damped by sticking a small piece of modeling clay in the center of the windowpane. Can you explain how this works?



13. An opera singer can break a large wine glass by singing a certain note loudly for a few seconds. Why?

14. Why does a whip "crack"?

15. Why does a bullet whistle when you shoot it out of a rifle but fly silently when you throw it?

16. What is the shape of the shock wave when an airplane flies by at supersonic speeds?

17. Why does a door opened slightly from a noisy corridor hardly decrease the sound level at all?

Microexperiment

If you blow near one end of a narrow pipe, sound of a certain frequency will be produced. Try to estimate the frequency.

It's interesting that . . .

. . . in tenth-century Russia the interiors of churches and temples were "acoustically enhanced." Special clay vessels were placed in the walls and domes, serving as resonators.

. . . the system of sound signals used by some African tribes was so highly developed that they could be considered as having a "telegraph" more sophisticated than the optical telegraphs in Europe that preceded the electrical telegraph. This sonic telegraph was used to report the sinking of the *Lusitania*: "The Great Ship of White People Has Sunk, Many People Perished" resounded in drum language all over the continent, from Cairo to Ibadan.

. . . the echo in the Castle of Woodstock in Great Britain clearly repeats no fewer than 17 syllables. And there's a castle near Milan, Italy, where a loudly spoken word reverberates as an echo 30 times!

. . . the frequency range of the human voice is far less than that of human hearing (20–20,000 Hz). The highest notes produced by a modern female singer correspond to frequencies of about 2,350 Hz; the record at the lower end (held by a man, we assume) is 44 Hz.

. . . the energy carried by an ordinary sound wave is very small. If a thermally insulated glass filled with water absorbs all incident sound energy corresponding to a value of 70 decibels (the volume of loud talking), it would take about 30,000 years to warm it to the boiling point from room temperature.

. . . the secret of dolphins being able to "see" distant objects with ultrasound lies in the narrow directionality of their acoustic signals. In this way bottle-nosed dolphins in the Black Sea can accurately swim up to a piece of buckshot (4 mm in diameter) thrown into the water at a distance of 20–30 m from the animal.

. . . one of the many practical applications of ultrasound in medicine arises because it's possible to concentrate it in a very small volume of tissue without affecting the rest of the body. ◘

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

Stop on red, go on green . . .

"When you walk on a path going north, you will only meet people coming from the north. At the crossroads, you'll meet people coming from the east, from the west . . ."—Nouk Bassomb

by Arthur Eisenkraft and Larry D. Kirkpatrick

When you're driving down a road and you see a yellow light, don't you wonder when the light will turn red? Maybe there should be an additional light—say, a blue one—that tells you that the yellow light will be changing to red any moment now. But then again, maybe there should be an orange light that tells you that the blue light will be ending soon and that the red light is imminent. But then again . . .

Who needs yellow lights at intersections? Who decides whether the yellow light should be one second, two seconds, or four seconds? Are yellow lights always set to encourage safe driving? Let's analyze what happens when you approach a yellow light.

As you drive down the road at a certain speed, you may see the light turn from green to yellow. You must make a decision to keep going or to step on the brakes and come to a stop. If you're relatively close to the intersection, you know that you can continue at the same speed and make it through while the light is still yellow. If your distance to the intersection is larger, you may decide to stop.

Let's assume that you want to keep going. To calculate your safe distance from the intersection, we simply calculate the distance you must go to get through the intersection while the light is yellow. This may be easier to follow if we use some real numbers as an example. Let's assume that the speed limit is 50 km/h, which is equivalent to 30 mph or 23 m/s. Let's also assume that the yellow light is on for 3.0 s before the light turns red. Therefore, you must travel a distance of 69 m during the time the light is yellow. If the width of the intersection is 15 m, you can safely proceed through the intersection if you are closer than 54 m. We'll call this the "go zone."

If you decide to stop when you see the light turn yellow, you must know the distance you will travel as you move your foot from the gas pedal to the brake (the coasting distance) and the distance it takes your car to stop (the braking distance). Once again, let's look at some real numbers and perform a calculation. The car is once again traveling at 23 m/s. If your response time is 1.0 s, the car will travel a distance of 23 m. If the deceleration of the car is 5 m/s^2, the car will travel an additional 53 m while braking. This distance is calculated according to the following equation:

$$v_f^2 - v_0^2 = 2as,$$

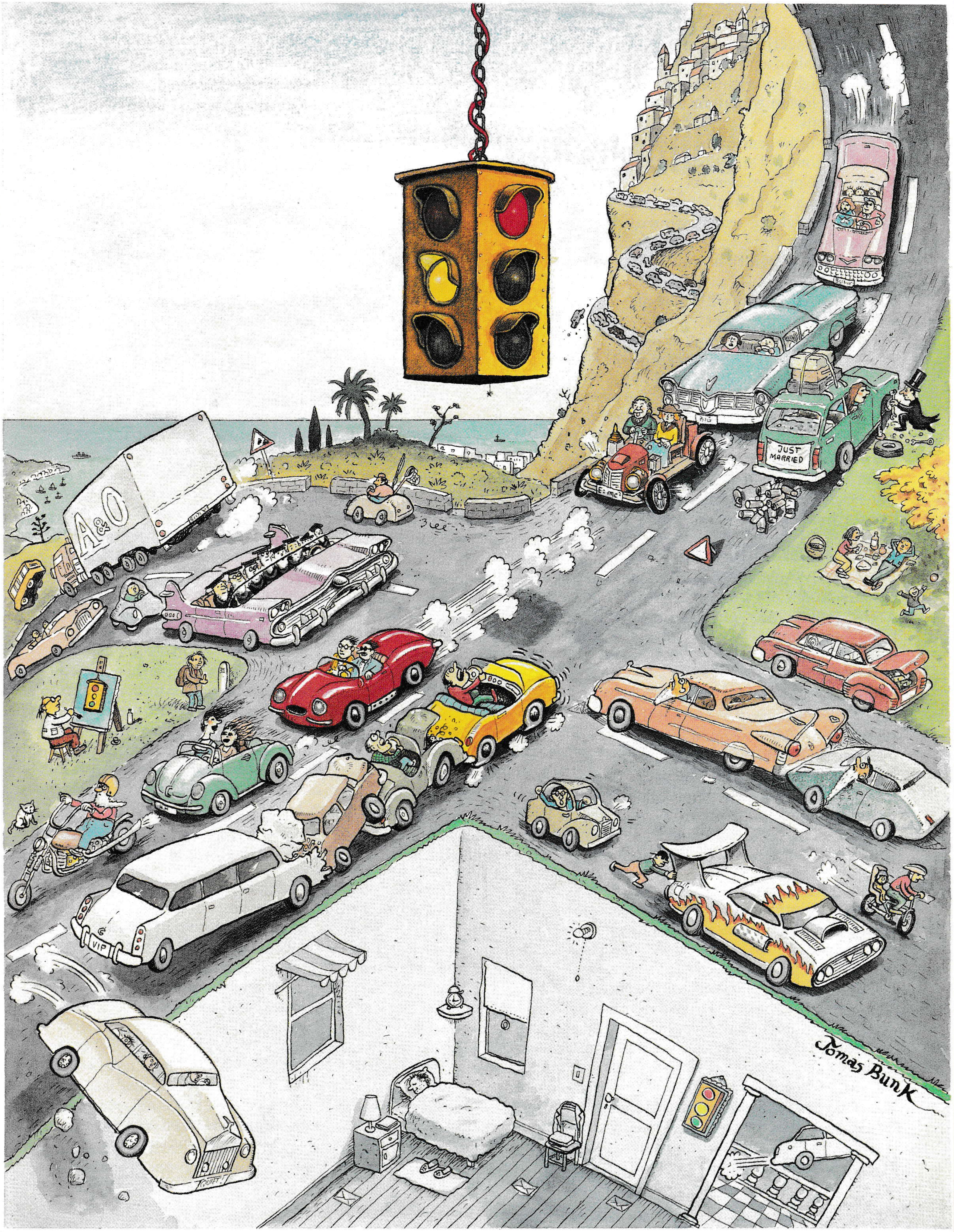
where v_f is the final velocity, v_0 is the initial velocity, a is the acceleration, and s is the distance traveled. The car can be safely stopped if it is at least 76 m from the intersection. We'll call this the "stop zone."

But wait—what happens if you're 65 m from the intersection? If you try to stop, you'll find yourself in the intersection. If you try to continue, you'll find yourself going through a red light. You're in trouble! We'll call this the "dilemma zone."

A safer intersection would not have a dilemma zone. If the yellow light time were 4.0 s, the go zone would be 77 m. The stop zone would still be 76 m. If you are closer than 77 m, you can safely proceed. If you are farther than 76 m, you can safely stop. If you are between 76 and 77 m, you can safely go or stop. This "overlap zone" provides for a safe intersection.

Rather than using data from a

Art by Tomas Bunk



A&O
JUST MARRIED
E=mc²
VIP
Tomas Bunk

single intersection, this month's contest problem asks you to do the work of a highway engineer and provide the relevant equations for safe intersections.

A. What is the general equation for the (a) go zone, (b) stop zone, (c) dilemma zone or overlap zone?

Assume a response time t_r, a maximum braking acceleration a, a yellow light time t_y, a speed v_0, an intersection width w, and a car length l.

B. For what speeds will there always be a dilemma zone?

C. Rewrite the equations in part A assuming that the car is going downhill when you see the yellow light.

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

Atwood's marvelous machines

In the July/August issue we asked readers to solve two versions of Atwood's machine. Correct solutions were submitted by Jeff Dodson of Vista, California, and Scott Wiley of Weslaco, Texas.

Figure 1 shows the situation for part A. Let's choose a coordinate system in which down and left are positive. This means that both masses will have positive displacements when the system is released. Using the notation in figure 1 we can write down Newton's second law for each mass:

$$Mg - T = MA, \qquad (1)$$
$$T = ma. \qquad (2)$$

With our choice of coordinates, we can write the connection between the accelerations as

$$A = a. \qquad (3)$$

Solving equations (1–3) yields

$$A = \left(\frac{M}{M+m}\right)g = \frac{2}{3}g,$$

$$T = \left(\frac{Mm}{M+m}\right)g = \frac{1}{3}Mg.$$

Notice that the acceleration of the system is larger than we calculated for the full Atwood's machine. This makes sense because the sack is no longer being retarded by gravity.

If we place the system in part A on a moving cart, we have the situation depicted in figure 2. Using the notation in figure 2 and our earlier choice of coordinates, we can once again write down Newton's second law for each mass:

$$Mg - T = Ma_1, \qquad (4)$$
$$T = ma_2, \qquad (5)$$
$$-T = m_3a_3. \qquad (6)$$

We have used the observation that the tension in the rope exerts a force T to the left on mass m and therefore by Newton's third law must exert a force T to the right on the beam (and hence on the cart).

We now need a relationship between the accelerations. To do this we look at very small displacements of the three masses. Mass M will fall a distance d_1 that is equal to the distance that mass m moves to the left and the distance the cart moves to the *right*. Therefore,

$$d_1 = d_2 - d_3$$

and

$$a_1 = a_2 - a_3. \qquad (7)$$

The easiest way to solve equations (4–7) is to use equation (7) to replace a_1 in equation (4) and use equations (5) and (6) to replace a_2 and a_3. Solving for T we get

$$T = \frac{mMm_3}{Mm + Mm_3 + mm_3}g.$$

Plugging this value for T back into equations (4–6) gives us the values for the accelerations:

$$a_1 = \frac{M(m+m_3)}{Mm + Mm_3 + mm_3}g,$$

$$a_2 = \frac{Mm_3}{Mm + Mm_3 + mm_3}g,$$

$$a_3 = \frac{-Mm}{Mm + Mm_3 + mm_3}g.$$

Notice that in the limit $m_3 \to \infty$ these equations reduce to the answers we obtained in part A.

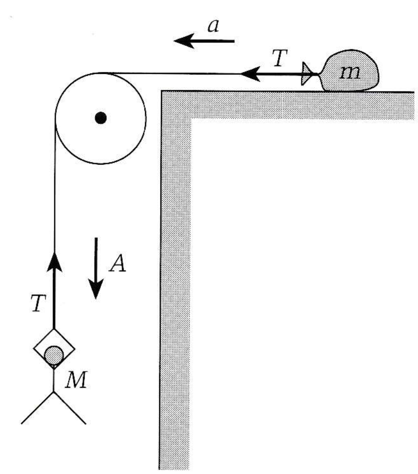


Figure 1

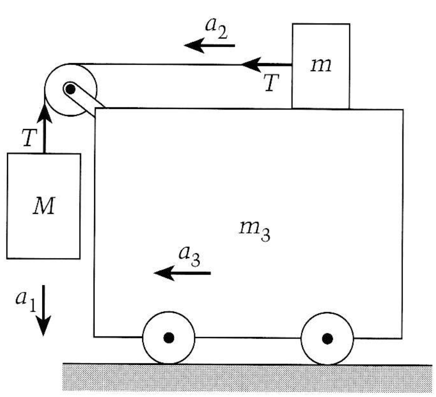


Figure 2

A princess of mathematics

Sofya Kovalevskaya proves a thing or two

by Yuly Danilov

A REMARKABLE EVENT IN THE LIFE OF Sweden's capital city of Stockholm took place in November 1883 and caught the attention of many Swedish newspapers. One democratic newspaper wrote: "Today we shall report not about the arrival of some banal prince of royal blood or some other person of no significance. No, it is a princess of mathematics, Mrs. Kovalevskaya, who honored our town with a visit and who will become the first female privat-dozent[1] in all of Sweden." It would not be an affront to truth if the newspaper had added: "And not only in Sweden, but also Germany, England, France, Russia, and every other country."

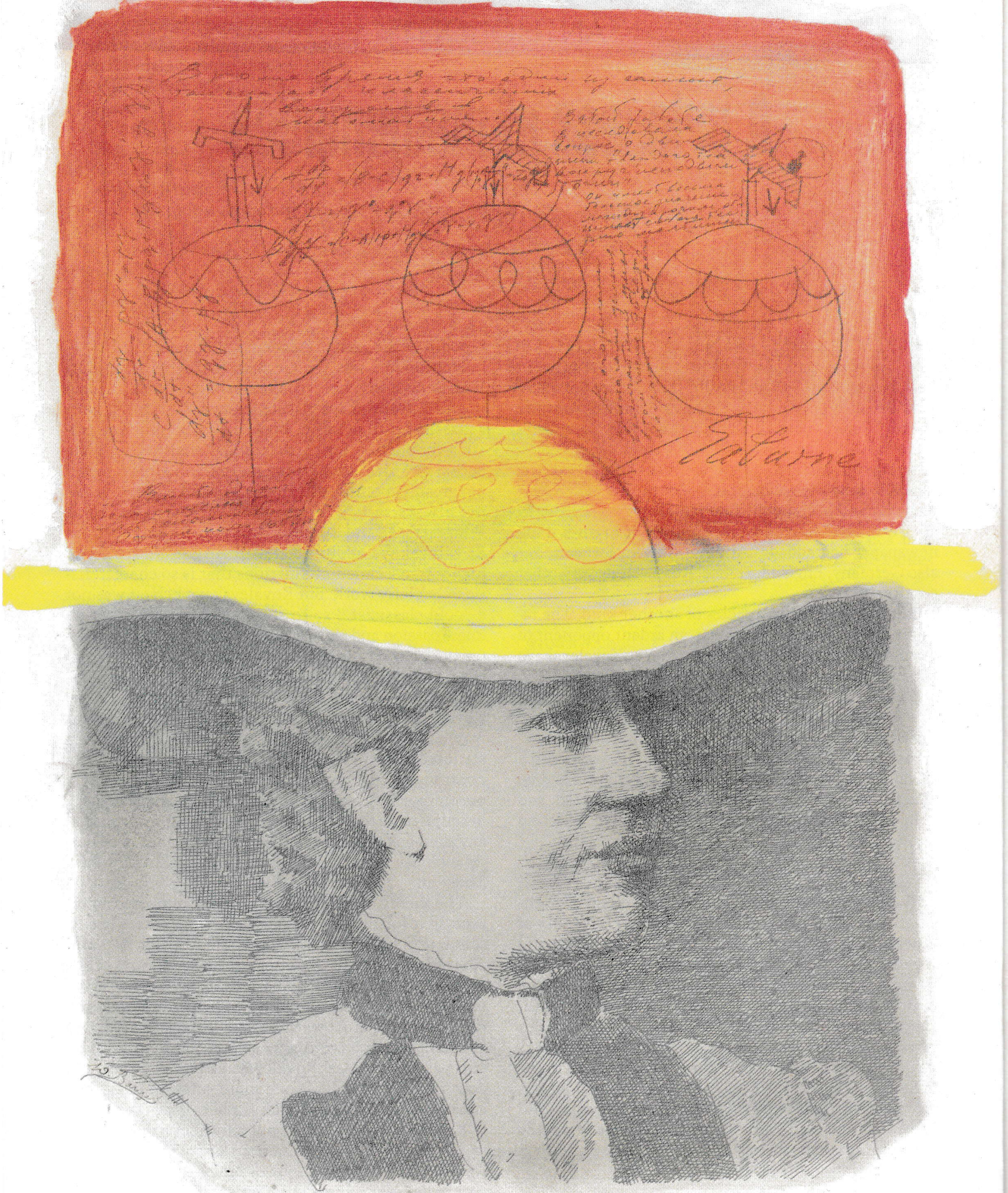
But there were also articles of another kind. The famous writer August Strindberg, an ardent opponent of women's emancipation, tried to prove, as Kovalevskaya herself had jokingly remarked, "how harmful, useless, and uncomfortable is such a monstrous phenomenon as a female professor of mathematics." To the credit of the Scandinavian people we can say that the sympathies of the overwhelming majority were with Kovalevskaya, and Strindberg's angry philippics had no effect. Sofya Kovalevskaya was heartily accepted, and "our professor Sonya" acquired in Sweden a second homeland.

And it was there that her extraordinary mathematical talent and literary gifts were revealed with a special brilliance. She delivered lectures on selected mathematical topics with great success at the University of Stockholm (twelve courses from 1884 to 1890) and wrote *Reminiscences of Childhood* (1890), the novel *Woman-nihilist* (1891), and (together with the Swedish writer Loeffler-Edgren, who was the sister of the mathematician and rector of the University of Stockholm Mittag-Loeffler) the drama *Struggle for Happiness* (1887), to say nothing of her shorter pieces. It was there, in Sweden, that S. V. Kovalevskaya created her principal mathematical work, "On the Rotation of a Solid Body about a Fixed Point," which became a bona fide mathematical sensation and was awarded the Prix Borodin of the French Academy of Sciences in 1888.

[1]A kind of assistant professor.

Her mentor was the German mathematician Karl Weierstrass, who not without hesitation kindly agreed to give Kovalevskaya private lessons. A scientist with a classical turn of mind, Weierstrass did much to create a foundation for mathematical analysis after the period of *Sturm und Drang* (storm and stress), when new results poured as if from a horn of plenty but did not always satisfy the requirements of rigor. It was Weierstrass (and some of his pupils) from whom Kovalevskaya acquired that brilliant control over higher transcendental functions that is so characteristic of her later work. But Kovalevskaya herself was not a "cold classicist," so to speak. Her romantic nature knew passion and fantasy. One of her very first works—on the analytical dependence of partial differential equations (brought to a so-called normal form) on initial values—carries the clear imprint of her genius: Kovalevskaya not only proved, in a simple and clear way, a result achieved earlier by Cauchy, she also constructed a quite unexpected example showing the nontriviality of the result. Weierstrass experienced a joy uncommon for a teacher: he saw that he was able to teach his pupil more than he himself knew! From that time on, any textbook on partial differential equations necessarily includes the very important Cauchy–Kovalevskaya Theorem.

The theory of the spinning top long remained an inaccessible fortress and did not yield to the efforts of outstanding mathematicians. One could solve (or, as mathematicians prefer to say, integrate) the equations of the top for only a few particular cases. Before Kovalevskaya only two such cases were known (if we don't count a few insignificant variants): "Euler's top" and "Lagrange's top." Noting what was common to these two particular cases, Kovalevskaya posed a question: does there exist at least one more solution of the top equations that possesses this same feature? To answer it, Kovalevskaya had to use not only her virtuosic skills in the theory of higher functions but also the original way of thinking and flight of fancy that was such a part of her mental makeup. Taking an extraordinarily



daring step, Kovalevskaya posited that time is not a real magnitude (as everybody thought then and many think now) but a complex magnitude. This "crazy" assumption enabled her to analyze the top equations more completely than her forerunners did and to find the *only* remaining unknown case of integrability: the famous "Kovalevskaya top."

After it was proved that there are no other *general* solutions of the top equations other than those found by Euler, Lagrange, and Kovalevskaya, a search began for *particular* solutions of a *given form*. Many other questions were resolved regarding the stability of the solutions, their connection with the symmetry of the problem, and so on. New life was breathed into the theory of the top, and it continues to live today.

I won't describe the delight and admiration with which the mathematical world received Kovalevskaya's remarkable result and how ashamed were those who stubbornly refused to make university chairs accessible to women. You'll learn more about this in Kovalevskaya's reminiscences, presented below.

Reports of Kovalevskaya's remarkable success were followed with special attention in Russia. Russian mathematicians, with the great P. L. Chebyshev at their head, were unanimous in their desire to bring their famous compatriot back to her homeland, but—alas!—such decisions were made (and continue to be made) by persons who aren't scientists. General Kosich appealed to the president of the Russian Academy of Sciences, Grand Prince Konstantin Konstantinovich Romanov—a poet who published his verses under the pseudonym "K. R." In his letter general Kosich reminded the prince of Napoleon's words: that "any state must value the return of its outstanding citizens much more than conquering a rich city."

The answer given by the secretary of the academy, K. S. Veselovsky, read: "Sofya Vasilyevna Kovalevskaya, who has acquired abroad so high a reputation with her scientific works, is no less known among our own mathematicians. The brilliant success of our fellow citizen abroad is the more flattering for us, in that it must be ascribed completely to her own high qualities, insofar as patriotic feelings could not have worked on her behalf there. It is especially flattering for us that Mrs. Kovalevskaya was appointed to the position of professor of mathematics at the University of Stockholm. The appointment of a woman to a university chair could only be due to an especially high and absolutely extraordinary opinion of her talents and knowledge, and Mrs. Kovalevskaya had completely justified such an opinion by her truly remarkable lectures. . . ."

But "because there is no access to a professorship in our universities for ladies, whatever talents and knowledge they may possess, there is no position for Mr. Kovalevskaya in our fatherland so honorable and so well remunerated as that to which she has been appointed in Stockholm. The position of teacher of mathematics in Women's Higher Courses is much lower than a university chair. And in other educational institutions where women are allowed to be teachers, the course of mathematics is restricted to only the elementary parts."

Translated from the bureaucratic into common language, this meant that in Russia there was no place for Kovalevskaya. But "our professor Sonya" took a place all the more secure in the hearts and minds of those who were fortunate to know her personally and those who made her acquaintance many years later, and will continue to meet her, in her scientific works, which haven't lost their significance. Now "our Sonya" belongs to the whole world.

KOVALEVSKAYA

Self-portrait

(Excerpt)

My love of mathematics was manifested for the first time, as far I can recall, in the following way. I had an uncle, Pyotr Vasilyevich Korvin-Krukovsky (the brother of my father), who lived 20 versts[1] from our estate in his village of Ryzhakovo. A man well on in years, he had handed over his farm and household to his only son, and because he now had a lot of time on his hands, he often visited us and lived with us for months at a time. My uncle was an idealist in the full sense of the word and in many respects was a man not of this world, as they say. He was educated at home, but he nevertheless had vast and various but (as is common with self-taught persons) far from solid ideas about things, which he acquired exclusively as a result of his

[1]The verst is an old Russian unit of distance, roughly equal to one kilometer.—*Ed.*

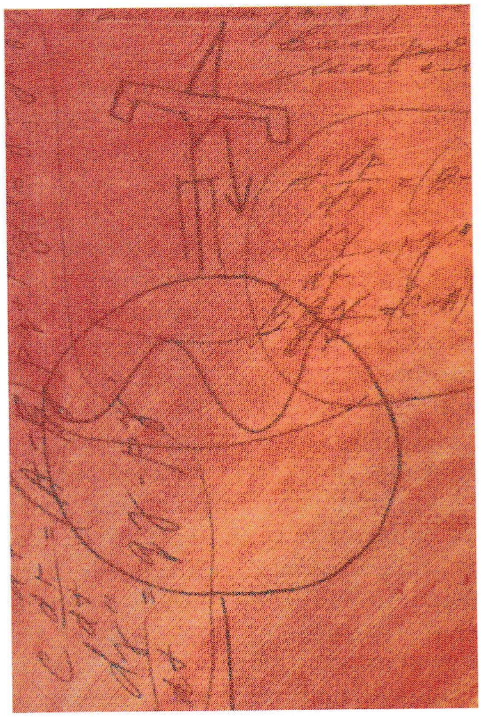
This excerpt is taken from the stenographic record of a conversation with S. V. Kovalevskaya in May 1890 in the editorial offices of the journal *Russkaya starina*. It was reviewed and prepared for publication by Kovalevskaya's brother, F. V. Korvin-Krukovsky. From S. V. Kovalevskaya, *Reminiscences. Novels.* Moscow: Pravda Publishing House, 1986, pp. 388–401.

Art by Yury Vashchenko

own curiosity, without any help from anyone and with the most inadequate preparation.

His favorite occupation and the sole pleasure remaining in his life was reading. In this respect the library in our country house held a strong attraction for him.

He read indiscriminately and with equal pleasure everything that he came across—novels, historical essays, works of popular science, and scholarly treatises. Kind and sweet-tempered by nature, he loved children immensely. Although at that time he was a 60-year-old man, he had the soul of a child. So despite the difference in our ages, I developed an extremely close, almost schoolmatish kind of friendship with my uncle. I was drawn to him by his stories; and he, soaring in the realm of fantasy, often forgot that there was a child present and, feeling the need to share his thoughts with someone, poured out his soul to me. I vividly recall the many long hours we spent together in the corner room of our large country house—in the so-called "tower" (which was in fact the library). My uncle told me fairy tales, taught me how to play chess; then, suddenly carried away by his ideas, he would initiate me into the mysteries of various economic and social projects that he dreamed of bequeathing to mankind. But mainly he liked to retell what he had learned and read over the course of his long life. And it was during these conversations that I happened to hear about some mathematical notions, and these left an especially deep impression on me. I learned from my uncle about the squaring of the circle, about asymptotes (lines to which a curve continually approaches but never reaches), and about many other things that were quite incomprehensible at the time, but which nevertheless seemed mysterious and at the same time particularly attractive to me.



To all this, fate decreed that another, purely accidental event would be added that intensified the impression that these mathematical expressions had made on me.

Before our arrival from Kaluga, our country house was being completely renovated. As part of this work, wallpaper had been ordered from St. Petersburg. But the necessary quantity was not calculated exactly, and there was not enough wallpaper for one of the rooms. At first it was decided that more would be ordered from St. Petersburg, but as often happens in such cases, due to the casual nature of country life and the laziness inherent in all Russian people, everything got put on the back burner. Time passed, and while plans were made and alternatives were considered and reconsidered, the renovation came to an end. Finally it was decided that one piece of wallpaper was not worth sending to the capital for—it was some 500 versts away. It was decided that all the other rooms were in good order, let the nursery stay the way it was—without wallpaper. The walls could simply be covered with plain paper. A lot of newspapers had collected over many years in the attic and had sat there utterly useless.

By a happy coincidence, in one pile of old newspapers and other junk there were printed notes of lectures on differential and integral calculus by the academician Ostrogradsky that my father had attended as a very young officer. The walls of my nursery were papered with those very pages. I was eleven years old at the time.

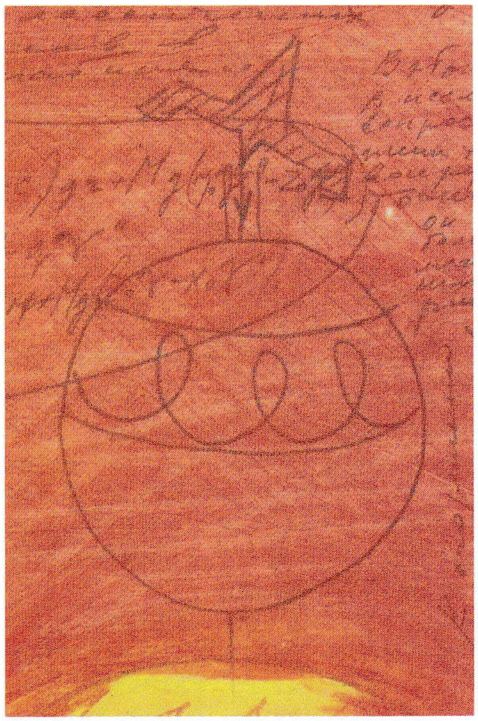
One day, looking at the walls, I noticed some things depicted there that I had heard about from my uncle. Since I was electrified by his stories in general, I began to scrutinize the walls with added attention. It amused me greatly to look at these sheets, yellowed with time, all peppered with strange hieroglyphs whose significance completely escaped me, but which I felt certain must mean something very clever and interesting. For hours on end I would stand in front of that wall, reading and rereading what was written on it. I must confess that at that time I did not have the foggiest idea what it meant, but I was strangely drawn to this activity. As a result of long examination, I learned many places by heart, and certain formulas, just by their appearance, became imprinted in my memory and left a deep mark there. I remember especially well that in the most conspicuous place on the wall there was a page with an explanation of infinitesimals and limits. How deeply these notions impressed me can be seen from something that occurred years later. I was taking lessons from A. N. Strannolyubsky, and as he explained these very concepts to me, he was surprised at how quickly I understood them. "You understood them as if you had already known them." And indeed, from a purely formal point of view, I had known many of them for a very long time.

I received my first systematic instruction in mathematics from I. I. Malevich. It was so long ago that now I do not remember his lessons at all. They remain only as dim memories. Nevertheless, they influenced me greatly and were very significant in my development.

Malevich taught arithmetic especially well and in his own unique way. But I must confess that when I began my studies, arithmetic did not interest me much. It is likely that, due to the influence of my uncle Pyotr Vasilyevich, I was more partial to various abstract discussions—for example, about infinity. In general, during my entire lifetime, the philosophical aspect of mathematics appealed to me more. I saw mathematics as a science that opens completely new vistas.

In addition to arithmetic, Malevich also taught me elementary geometry and algebra. Only after I had become acquainted to some extent with the latter did I feel such a strong attraction to mathematics that I began to neglect other subjects.

Seeing my preference, my father, who in general had a strong prejudice against learned women, decided to stop my mathematics lessons with Malevich. But somehow I managed to wheedle a copy of Bourdon's *Course in Algebra* from Malevich, which I began to study assiduously. Because all day long I was under the vigilant supervision of my governess, I had to engage in a bit of subterfuge. When I went to bed, I would put the book under my pillow and then, when everybody else was asleep, I would read all night long by the dim light of an icon lamp or night light.



Given the situation I could not even dream of continuing regular studies in my favorite subject, and it appeared that my mathematical knowledge was doomed to remain for a long time within the realm of Bourdon's *Algebra*, but something happened that caused my father to change his mind to some extent regarding my education.

One day Professor Tyrtov, the owner of a neighboring estate, brought us his elementary textbook on physics. I tried to read this book, but to my distress in the section on optics I found trigonometric formulas, sines, cosines, tangents. What is a sine? I was at a loss to answer this question, and I was forced to turn to Malevich for a solution to the riddle. But because his program of instruction did not include this topic, he told me that he did not know what a sine is. So, working with the formulas given in the book, I tried to explain it myself. By a strange coincidence, I chose the same path that was used historically—that is, I took a chord instead of a sine. For small angles these magnitudes almost coincide. And because in Tyrtov's textbook all the formulas involved only infinitesimal angles, all these formulas squared perfectly with my basic definition. But this set my mind at ease.

A short time later, when in conversation with Tyrtov the subject of his textbook arose, he expressed doubt at first about my ability to understand it, and when I said I had read it through with great interest, he said, "You're bragging!" But when I told him the route I took to explain the trigonometric formulas, his tone changed completely. He immediately went to see my father and urgently began trying to convince him of the need for me to be taught in the most serious way. While making his case, he compared me with Pascal. After some hesitation my father agreed to hire A. N. Strannolyubsky as my teacher. We began working together and over the winter we covered analytic geometry and differential and integral calculus.

The next year I married V. O. Kovalevsky, and soon we went abroad, but there we soon separated. I went to Heidelberg to continue my studies in mathematics, and he went to another university to study his discipline—geology. . . .

From Heidelberg I went to Berlin, but at first I was disappointed there . . . The capital of Prussia turned out to be . . . backward. Despite all my petitions and efforts, I failed to receive permission to attend the university in Berlin.

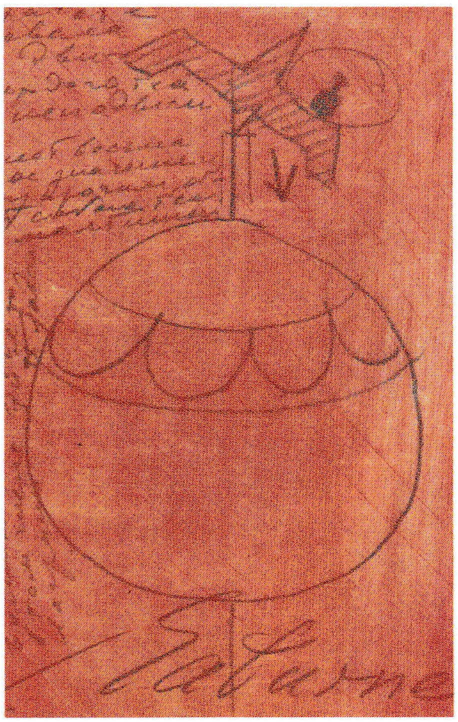
Then . . . professor Weierstrass took an interest in me. Noting the references from my Heidelberg professors and seeing that I was well prepared and eager to learn, and not just because it was the fashionable thing to do, he proposed that I study with him privately. These studies were the most important influence on my whole mathematical career. They determined irreversibly, once and for all, the direction I took in my further scholarly activity, and all my works were done in the spirit of Weierstrassian ideas.

Weierstrass himself I consider one of the greatest mathematicians of all time and without doubt the most remarkable among living mathematicians. He gave to all of mathematics a completely new direction and created not only in Germany but also in other countries a whole school of young scholars who travel the path he indicated, developing his ideas.

While attending Weierstrass's lectures, I also began to prepare myself for the doctoral degree. But because at that time the doors of the University of Berlin were closed to me as a woman, I decided to turn to Göttingen. According to the rules of German universities, in addition to passing an examination, one also was required to present a scholarly work—the so-called "inaugural dissertation"—to receive a doctoral degree. Weierstrass suggested several topics to me for development, and during the two years I spent in Berlin I produced not one work, as required, but three. Two of the treatises were in pure mathematics ("On Partial Differential Equations" and "On Reducing a Class of Abelian Functions to Elliptical Functions") and the third was astronomical in nature ("On the Shape of Saturn's Rings").

I submitted all three works to the University of Göttingen. They were acceptable to such a degree that the university, contrary to the established rules, found it possible to free me from the examination and public defense of my dissertation, which in essence is a pure formality, and conferred on me the degree of doctor of philosophy *summa cum laude*.

At that same time the first of the aforementioned works was published in Crelle's Journal (*Crelles Journal für die reine und angewandte Mathematik*) under the title "Zur Theorie der partiellen Differenzialgleichungen" ("On the Theory of Partial Differential Equations"). This was an honor that few mathematicians received and was even greater for a beginning mathematician, since this journal was then considered the most important mathematical publication in Germany. The best mathematical minds worked on it, and previously such mathematicians as Abel and Jacobi published their works there. My astronomical work "On the Shape of Saturn's Rings" was not published until many years later (1885) in the journal *Astronomische Nachrichten*.



In 1874 I returned to Russia. Here I studied with far less intensity, and the conditions of daily life were much less conducive to scholarly endeavors than in Germany. I worked with long and frequent interruptions, and I barely had the time to keep current in the mathematical field. During the entire time I was in Russia, I did not produce a single independent work. The only thing that provided some mathematical support was my correspondence and exchange of ideas with my dear teacher Weierstrass.

Various circumstances distracted me from serious scholarly activities in Russia. They had to do both with Russian society at large and the conditions under which I had to live. At that time all of Russian society was seized by the spirit of moneymaking by various commercial enterprises. This current swept up my husband and—I must confess my sins—it partially took me with it as well. We went in for constructing grandiose stone houses with commercial bathhouses attached. But in the end the market crashed, and we were ruined completely.

Soon after I returned to Russia the newspaper *New Times* was created. My husband was a close acquaintance of the publisher, and so we fell into the *New Times* circle. I tested my literary skills on this paper as a theater critic.

In 1882 I again went abroad, and since then I have lived there almost permanently, only now and then visiting Russia for short periods to settle some business matters. During my life I have had an opportunity to see many cities and countries, and so can justifiably say that except for Italy and Spain, I am rather well acquainted with Europe. And, with the exception of Sweden, I know Paris best of all. I was there on many occasions, and even now I generally spend my vacations in France.

Returning abroad, I again energetically took up the science from which I rested for so many years in Russia. First off I went to Paris and made the acquaintance of the outstanding mathematicians working there, including the renowned Hermite, and also met the younger mathematicians Poincaré and Picard. These two are, in my opinion, the most gifted among the new generation of mathematicians in all of Europe.

At that time I was engaged in writing an extended new work, "On the Refraction of Light in Crystals." Generally in mathematics one stumbles on topics for independent research by reading the treatises of other scientists. And so I was brought to this theme by studying the works of the French physicist Lame.

I finished work on my treatise in 1883 and created a bit of a sensation in the mathematical world, because the question of the refraction of light had not yet been sufficiently clarified, and I considered it from a completely new angle.

I submitted this treatise in 1884 to the new journal *Acta Mathematica* (it was founded in 1882). Although the *Acta* are published in Sweden, it is a true international publication, because it is subsidized not only by the Swedish king but also by foreign states, including France, Germany, Denmark, and Finland. Now [in 1890] it has become one of the biggest and most influential mathematics journals. Leading scholars of all nations work together on it, and it touches on the most burning questions, so to speak, that attract the attention of modern mathematicians. It often happens that several people are engaged with one and the same problem simultaneously. In general the conditions under which a serious mathematics journal is published are quite different from those of other periodicals. That is why the *Acta Mathematica* are issued not at a predetermined time but according to how material is accumulated, new problems ripen, and solutions are found. Usually two volumes are published in a year.

In addition to my treatise on the refraction of light, several other papers of mine have appeared in the *Acta Mathematica*. In 1883 the second of the works I presented as a dissertation in 1874 at the University of

Göttingen was published under the title "On the Reduction of a Class of Abelian Functions to Elliptical Functions."

All of my scientific papers are written in German or French. I know them on a level with my native Russian. But in mathematical works language plays an insignificant role. The main thing in this case is content, ideas, notions; and to express them mathematicians have their own language: formulas.

In the early 1880s the recently founded University of Stockholm began to develop. At that time I was well-enough known in the mathematical world both because of my writings and because of personal acquaintance with almost all the European mathematicians of note. Especially often I had a chance to meet in Berlin and Paris with the chief mathematician (now the rector) of the University of Stockholm, Professor Mittag-Loeffler, who was one of the best pupils of a teacher we had in common, Karl Weierstrass.

So in 1883 I was invited to Stockholm to deliver lectures on mathematics. . . .

When I first arrived in Sweden I proposed to deliver my lectures either in German or in French. Most of my students preferred that I lecture in German. But a year later I was able to deliver lectures in Swedish. This did not present any great difficulties for me, because immediately after my arrival I was accepted by Swedish society and began taking lessons in the Swedish language.

At first I was invited as a privat-dozent. But in less than a year I was appointed professor and continue to hold this position. In addition to delivering lectures, my duties include participation in meetings of the university council, where I have the right to vote on a level with other professors. A professor's salary in Sweden is 6,000 kronas a year (a Swedish krona is a little more than a deutsche mark; 700 kronas are equal to 1,000 francs). I deliver four lectures a week—that is, I speak for two hours straight on two days each week. Because my lectures are devoted to very special topics, I do not have very many students—about 17–18.

During my first year in Sweden I worked with great diligence and seriousness. It was there that I wrote the most important of all my mathematical works, for which I received a prize from the Academy of Sciences in Paris. In this paper I investigated the problem "On the Motion of a Rigid Body About a Fixed Point under the Influence of Gravity." It was a problem of great significance that includes, among other things, the theory of a pendulum. At the same time it was one of the most classical problems, so to speak, in mathematics. The greatest minds directed their efforts at a solution—Euler, Lagrange, Poisson, and others. But in spite of that, it was far from being completely solved, and we knew only a few cases for which a completely rigorous mathematical solution had been found. In the history of mathematics one can point to only a few questions that, like this one, made one wish so strongly for a solution and to which so much talented effort and diligent work had been brought to bear with, in the majority of cases, so few results of any substance. Not without cause did a German mathematician call this problem *die mathematische Nixe* ("the mathematical mermaid").

This problem had always interested me greatly, and many years earlier—almost from the time I was a student—I started to test my strength on it. But for a long time all my efforts remained fruitless, and only in 1888 was my work crowned with success. You can imagine how happy I was when at last I was able to achieve a really important result and take an important step toward solving so difficult a problem.

In the same year the Academy of Sciences in Paris announced a competition for the best paper on the following topic: "On the Motion of a Rigid Body." One of the conditions set by the Academy was that the paper must bring to completion or elaborate in some substantive way the knowledge attained to date in this area of mechanics.

At that time I had already achieved the main results of my work. But thus far they had remained in my own head. Because the problem I had solved was perfectly appropriate to the topic proposed by the Academy, I began to work with even greater zeal in order to impose order on all my material, work out the details, and write the paper before the deadline.

When all of this was finished, I sent my manuscript to Paris. According to the rules of the competition, it had to be sent anonymously—that is, I wrote a motto on my work and attached a sealed envelope containing a piece of paper with my name and the same motto written on it. So when the works were evaluated, their authors remained unknown.

The results exceeded my expectations. About fifteen papers were submitted, but only mine was considered worthy of a prize. But that was not all. Because the same topic had been proposed three times in a row and each time had gone without an answer, and taking into account the importance of the results I had achieved, the Academy decided to increase the prize from 3,000 to 5,000 francs. Then the envelope was opened, and all learned that I was the author of that work. I was immediately informed and went to Paris to participate in a meeting of the Academy that had been called in conjunction with the competition. I was received with extraordinary solemnity. They asked me to take a seat near the president, who delivered a flattering speech, and in general I was heaped with honors.

As I had mentioned earlier, I have been living in Sweden since 1883 and in that time I have assimilated Swedish life to such an extent that I feel at home there. Stockholm is a very beautiful city, and the climate is not bad at all—only the spring is unpleasant.

I have a wide circle of acquaintances, and I spend quite a bit of time in society. I even pay visits to the royal court. . . .

A polarizer in the shadows

One physicist's contribution to the wave theory of light

by Andrey Andreyev

IN THE HISTORY OF SCIENCE the work of Étienne Malus is overshadowed by the more weighty achievements of his contemporaries—the great French physicists of the early 19th century. Yet his talents were highly regarded by such scientists as Laplace and Lagrange, Arago and Young. He was a remarkable experimenter and geometer, well versed in chemistry and engineering. He was also a good conversationalist, a staunch soldier who participated in a long war, and a man of enormous persistence and ambition that allowed him to advance rapidly in his scientific career. Malus was destined to live a short life (and he spent only seven years working in physics), yet his scientific legacy is great and might have been far greater.

Malus did his most important work in the area of optics. He began his experiments in the years when a firm basis for the wave theory of light was gradually being formed. You may recall that two theories of light vied for prominence by the turn of the 19th century. The first considered light a flow of particles—"corpuscles"—that travel along a straight line in accordance with the laws of mechanics. Reflection of these corpuscles is similar to that of an elastic ball, and refraction of particles passing from one medium to another is explained by the attraction of one of the media. It was Sir Isaac Newton himself who elaborated the so-called corpuscular theory most rigorously. The competing theory considered light a wave moving in an elastic medium—"ether"—and explained both reflection and refraction by a change in the direction of the wavefront. The creator of the wave theory of light was a contemporary of Newton's, the Dutch physicist Christiaan Huygens.

The corpuscular theory dominated science for the entire 18th century, but signs of a conversion to the wave theory appeared as the century drew to a close. In particular, physicists had long noted the wonderful optical properties of Iceland spar, a clear, colorless form of calcite. A thin light beam passing through a crystal of this mineral splits into two rays: one that follows the well-known refraction law—the ordinary ray—and one that doesn't obey the law—the extraordinary ray. Surprisingly, the light beam is split even at normal incidence on a natural facet of the crystal, and the extraordinary ray is deflected to form a nonzero angle of refraction (fig. 1). It was noted also that the rays passing through Iceland spar differed from the original beam and did not divide if they passed through another crystal of Iceland spar. This phenomenon came to be called double refraction.

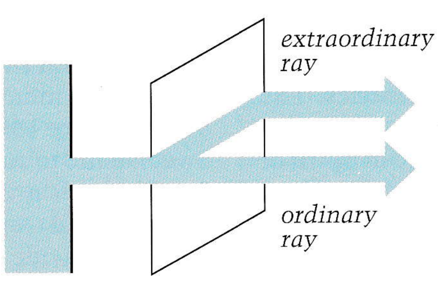


Figure 1

Malus was the first to see similar features in any reflection of light, thus demonstrating the general nature of a phenomenon resulting from some intrinsic natural property of light. He coined the term "polarization" for this phenomenon, and since then the branch of optics bearing that name has become one of the most interesting and important.

Although Malus adhered to the corpuscular theory all through his life, his discovery stimulated much research in support of the wave theory. His experiments were simple and can be repeated in a school lab. But before we turn to them, let's skim the pages of his biography. We'll see that his wonderful discovery was preceded by no less wonderful and difficult times filled with danger, great deeds, love, and glory.

Youth

Étienne Louis Malus was born in Paris on July 23, 1775. Little is known about his parents. His father, a member of the gentry, held the position of treasurer. This was a profitable position with a comfortable income to support his family. Little Étienne was educated at home and proved to be an able pupil. His mother introduced him to the world of Greek and Roman poetry. In his later years Malus loved the classical writers and cited by heart passages from Homer, Anacreon, Horace, and Virgil. In his leisure time he often composed verses in Latin. However, his friends did not consider versification his strong suit.

The boy's other favorite subjects were algebra and geometry, and it's no wonder he entered the Paris Engineering School in 1793. But the French Revolution had been going on for four years, and as fate willed it, the school closed just before the final examinations. The young engineer enlisted as a volunteer in a battalion of the Paris militia. And this was the start of his military career.

At this time a campaign was being conducted in the north of France. At the Dunkirk fortress Malus was assigned to field fortification work. One of the engineers supervising the work noticed the unexpected skill with which it was done. A brief conversation with Malus showed him that here was a young man of promise. Just at that time the Polytechnical School (École Polytechnique) opened in Paris, and at this engineer's urgent request the youth with a bright future was enrolled as one of its first students.

The Polytechnical School immediately became one of the most significant educational institutions in France, where outstanding physicists and mathematicians were hired to teach. The curriculum spanned two years and was very intensive. As time went on it graduated many French physicists of world stature.

After graduating from the Polytechnical School, second lieutenant-engineer Malus left Paris to join the army somewhere in central Europe. Military roads brought him to the right bank of the Rhine, where he stayed for eleven months in the garrison in the old town of Gissen. These months were probably the happiest for the passionate young man. He was on the verge of marrying the eldest daughter of the chancellor of the university, a Professor Koch, but suddenly he received orders to present himself at Toulon and join the expeditionary army, whose destination nobody knew.

Campaign

The name of Napoleon Bonaparte now intrudes powerfully into the leisurely course of our narrative. Across Europe wars were raging—what historians would later call the Napoleonic wars. And one of their most dramatic chapters was about to be written: the Egyptian campaign.

By 1797 Great Britain was the last undefeated foe of the French Republic. The English fleet under Admiral Horatio Nelson was too strong and could ruin any direct attack on the British Isles. To achieve victory over England, Bonaparte needed to attack from another direction—in Egypt.

Equipping the expedition, Bonaparte planned to carry out a comprehensive investigation of the "cradle of civilization," and he decided to bring with him to Egypt almost the entire French Academy of Sciences. This scientific contingent was led by a devoted friend of Napoleon's and one of the founders of the Polytechnical School, the renowned mathematician Gaspard Monge. Preparations for the risky gambit were top secret. In undertaking his first naval campaign, Bonaparte relied more on good luck than on his own forces.

On February 27, 1798, Malus embarked on a ship at Toulon that was part of the advance guard of the "Egyptian squadron." On June 10 he took part in the attack on Malta. Admiral Nelson was delayed by a storm at Gibraltar and unwittingly passed the French near the island of Crete. So Malus disembarked successfully with the rest of the expeditionary force on the shores of Egypt.

The campaign began successfully—the poorly equipped Egyptian troops were defeated in a few fierce battles. Malus took part in the famous battle at the Pyramids, where more than once he stared death in the face. But Malus's days in Egypt weren't filled entirely with war. Gaspard Monge remembered his pupil and invited him to collaborate in preparing his *Description of Egypt*. Once, while he was engaged in cartographic work, Malus discovered the ruins of the remarkable town of San. Soon Bonaparte founded the Egyptian Institute in Cairo. Monge was elected its president, and Malus became one of the fellows.

For some time after the defeat of the Mamelukes, Malus lived in Cairo. From there he went to Upper Egypt, where he led a geographical and archaeological expedition at the Nile Delta and visited the ancient pyramids at Giza with the celebrated general Cleber. The general liked the young officer, and he decided to take him on the Syrian campaign as a member of his division.

As you may know from reading history, this campaign was a disaster for Bonaparte. Malus shared all the calamities suffered by the French army in Syria. However, as a brave officer he distinguished himself during the siege of Jaffa, where he constructed trenches and other fortifications and took part in bitter fighting in the streets of the town.

Plague

When Bonaparte's army left to begin its siege of the Fortress of Saint Jean d'Arc, once a major stronghold of the Crusaders and now in the hands of the sultan's army, Malus received orders to remain in Jaffa. From the depths of the Syrian desert, or perhaps from the heights of the Lebanese mountains, one of the most terrible diseases known to mankind, the constant companion of wars and destruction, stole toward the uninvited guests: the plague (or Black Death). It followed in the wake of the French army. In the garrison left at Jaffa there were 300 wounded, 400 suffering from the plague, and only 50 healthy soldiers. Around them the town was dead, the traces of the recent defeat were still fresh in the streets. On the eleventh day Malus took ill with the same horrible disease that raged among the troops. Malus's notes give only a pale idea of what he lived through during those days when the signs of the plague became evident, surrounded by dying comrades.

The siege of Akkra dragged on. The sick kept arriving at Jaffa and increased the number of infected persons. The plague was now in every home. The only servant who helped Malus faithfully during his illness died along with the last of his friends. He was alone and helpless.

But Malus was lucky: after a month of illness he embarked at last on a ship sailing for Egypt. He wrote that on his return voyage "the sea air had a strong effect on me; it seemed to cure me of asthma; head winds delayed the ship at sea, and that greatly promoted my recovery." The ship dropped anchor near the Delta, but the ordeals of the young captain-engineer were not over yet. He was to spend a month in a quarantine surrounded by the diseased, which delayed his recovery. But all ended well. Malus fully recovered and could now leave the hospital, though he had to part with all his personal belongings.

The epidemic of the Black Death brought Bonaparte's army to the brink of disaster. Under these conditions he couldn't get favorable terms for an armistice, and in addition the shortage of provisions was felt more and more. The scientific activity of the Egyptian Institute gradually decreased and finally stopped. At the same time as Malus was battling his sickness, Gaspard Monge fell dangerously ill (though fortunately he soon recovered). Meanwhile the victorious troops of Russian Field Marshal Suvorov routed the French in Italy. Fortune, which had smiled on Bonaparte for so long, now seemed to betray him.

Return

Soon after his recovery Malus was given a new assignment, one that looked more like leave than duty. The soldiers of the garrison where he was sent lived in huts made of palm leaves just as the Arabs did. Near his hut horses and camels peacefully grazed, the fenced yard was full of hens, geese, and ducks. There was utter calm. The peaceful amusements had a restorative effect on Malus. Here he decided to study physics, and he wrote an essay on the nature of light, in which he discusses in detail the basic principles of the corpuscular theory.

Meanwhile the French army continued its campaigns. In August 1799 Bonaparte decided to leave his troops and return to Paris. He was forced to do this by the obvious failure of the Syrian campaign, which complicated the situation in Europe and threatened to deprive him of all the power he had attained. Command of the army was assumed by Malus's comrade, General Cleber. Despite some partial successes, the condition of the expeditionary army got worse. Malus resumed his military duties and again took part in battles. However, the brave general was soon assassinated by a Turk, which led to a weakening in control over the Army. In the fall of 1801 it lay down its arms.

Malus returned to France, still cherishing in his heart an old affection. Throughout the three years he was away he received letters from Gissen. At last he could marry Louise Koch. It was a happy union.

Another ten years of military life sent Malus all over Europe. By 1810 he had reached the rank of major-engineer, but his scientific interests drew him to Paris. In August 1810 the brave soldier of Napoleon's army was elected a member of the French Academy of Sciences. He took his honored place among the leading scientists of Europe, and he had a right to it: by that time Malus had discovered and described the phenomenon that would immortalize his name in science—the polarization of light.

Discovery

On January 4, 1808, the Academy of Sciences announced a competition for the answer to the following problem: "Devise an experimentally verified mathematical theory for double refraction of light in different crystals." The prize was awarded to Malus. By this time he was known in scientific circles as the author of a treatise on analytical optics in which he generalized several partial geometrical methods of constructing characteristic optical surfaces. He had also presented the academy with his experimental work on measuring the refractive index of transparent and opaque wax by the Wollastone method. So his success in the contest wasn't wholly unexpected. His new work, like his previous works, was favorably received by the committee, composed of such persons as Lagrange, Gaiyu, Gay-Lussac, and Biot. However, it should be mentioned that Malus,

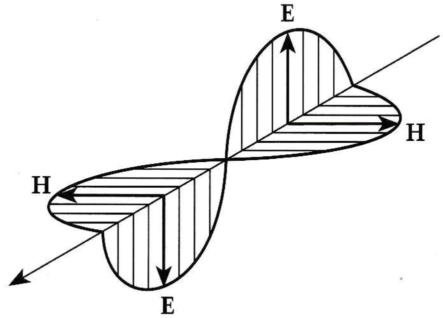


Figure 2

fearful that one of his rivals might block him by discovering the new properties of light that he had just observed, presented a substantial portion of his work to the academy on December 12, 1808, not waiting for the deadline for the competition. So his discovery dates from the end of 1808.

Let's look briefly at the sequence of observations Malus made. As mentioned above, a narrow light beam passing through a crystal of Iceland spar is split to produce two rays: an ordinary ray and an extraordinary ray. If we direct one of these rays through another sample of Iceland spar, double refraction does not always occur. It's interesting that, if we rotate the second crystal about the direction of the light beam, we can find a position for which the beam refracts normally and another position for which the same beam refracts abnormally. In intermediate positions two beams of different intensity emerge from the second crystal.

These properties were discovered by Huygens at the end of the 17th century. One day, at home in Paris, Malus looked through a double-refracting crystal at sunbeams reflected from the windowpanes of the Luxembourg Palace, which was opposite his apartment. Rotating the crystal, he suddenly noticed the very same change in the refraction of the beam passing through the crystal, just as if it had passed through Iceland spar. Instead of the two equally strong images that he expected, Malus observed only a single image—now ordinary, now extraordinary. This strange phenomenon startled him: he tried to explain it by variations of light in the atmosphere. But as night fell the light from a wax candle confirmed the daytime results. This time, though, Malus observed the reflection from a water surface. This was how he came to understand that light changes its properties—becomes polarized—not only after passing through Iceland spar but after being reflected by any surface. Thus, polarization is one of the fundamental properties of light.

In his article Malus wrote: "Under certain conditions the effect of some substances on light causes the reflected and refracted light to acquire certain properties that make them essentially different from the

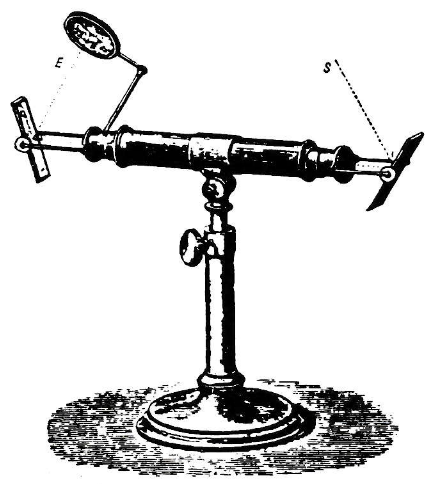


Figure 3

emitted source light. The property of light I shall describe is connected with this kind of change. The particular property was observed previously under certain conditions in the form of the appearance of a double image after light passed through calcite. It was thought, however, that this phenomenon was a feature of that particular crystalline substance, and it was not supposed that this phenomenon might arise not only in all doubly refractive crystals but also in all solid and fluid transparent or even opaque bodies."

Let's examine polarization from the modern viewpoint. A direct corollary of the electromagnetic theory of light is that light waves are transverse waves. This means that the vectors representing the oscillating electric **E** and magnetic **H** fields are perpendicular to the direction of wave propagation, and not only that, these vectors are perpendicular to each other (fig. 2). Given the direction of the beam **n**, we can choose the direction of **E** (and, correspondingly, **H**) arbitrarily in the plane perpendicular to **n**. In natural sunlight all directions of **E** are equally probable. However, if the light is completely polarized, all of the **E** vectors point in the same direction. If this direction happens to be vertical, the light is said to be vertically polarized.

When double refraction occurs, the ordinary and extraordinary rays are completely polarized—the polarization of one ray is perpendicular to that of the other ray. This effect is due to the asymmetrical arrangement of the atoms within the crystal. One can imagine a crystal that strongly absorbs one of these rays, in which case a single polarized light beam will be emitted. Crystals of this kind are called polarizers, and physicists nowadays use tourmaline to polarize light.

Experiments

Now we'll look at Malus's experiment in some detail. Figure 3 is an engraving of his actual experimental setup, while figure 4 is a schematic rendering. Let a narrow beam of natural light strike a glass mirror M such that the angle of incidence is ϕ. The reflected ray passes through a sheet of tourmaline T that can be rotated about the reflected ray. As the sheet is rotated an observer sees the light intensity increase and decrease. The orientations for minimal and maximal intensities differ by 90°.

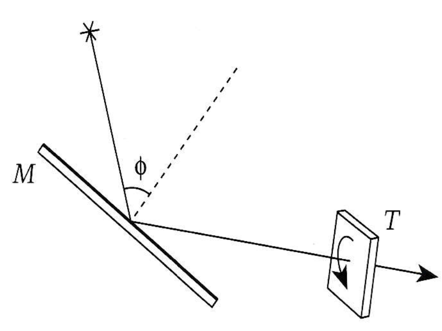


Figure 4

Now, if the reflected light retained the properties of natural light, we wouldn't see any change in the light intensity when we rotate the tourmaline crystal. On the other hand, if the reflected ray is completely polarized, certain orientations of the tourmaline crystal would not allow any light to pass through. Since we do not observe such an extreme case in our experiment, we conclude that the polarization of light reflected by glass is only partial—that is, the reflected ray is a mixture of natural and polarized light.

If we increase the angle of incidence ϕ, we find that the reflected light becomes more and more polarized until the light is completely polarized at a special angle ϕ_0, now known as Brewster's angle. As we increase the angle of incidence even more, the polarization decreases. In 1815 Brewster derived the relationship $\tan \phi_0 = n$, where n is the index of refraction of the glass, by requiring that the reflected and refracted rays be perpendicular. (You should try obtaining this relationship yourself.) Malus measured Brewster's angle for water to be 53°, as would be expected for a refractive index of 1.33.

Having discovered polarization, Malus wanted to obtain wide beams of polarized light, but he did not have any tourmaline crystals. Once again, the Luxembourg glass came to the rescue. Malus noted that when light is reflected at an angle ϕ_0, the polarization of the refracted ray is also maximal—at about 15%. So if one combines several plates of glass and directs a light beam through all of them at an angle ϕ_0, the light ray that emerges will be almost completely polarized.

Malus also determined the intensity of the completely polarized light I after passing through a second polarizer. If the second polarizer is rotated through an angle α from its position of maximal intensity, then $I = I_0 \cos^2 \alpha$. In the modern view, the electric field vector **E** is imagined to consist of components perpendicular and parallel to the transmission axis of the polarizer. Only the

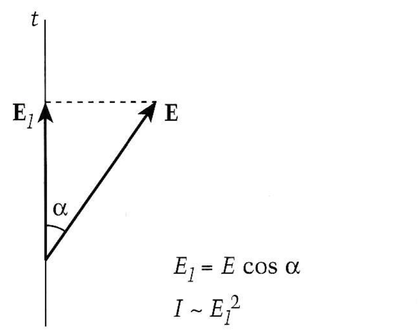


Figure 5

component parallel to the transmission axis passes through the polarizer; the other component is absorbed. This means that the electric field vector is reduced by a factor of $\cos \alpha$, where α is the angle between an incident electric field and the transmission axis (fig. 5). Since the intensity of the light is proportional to E^2, we obtain a factor of $\cos^2 \alpha$. (Remember that Malus didn't espouse the wave theory of light!)

Malus's ideas were developed further in a famous series of experiments by two French physicists, his contemporaries and friends Fresnel and Arago (1816), in which the interference of polarized beams was investigated. Fresnel also quantified the relative degree of polarization of light reflected from and passing through a dielectric. Unfortunately, Malus did not live to see these new achievements.

Glory

The treasure trove Malus discovered brought him well-earned glory for several years. He jealously guarded his reputation as the first to make the discoveries in this field, and was deeply offended when an academician questioned Malus's priority in one particular experiment. Malus was recognized around the world. The Scientific Secretary of the Royal Society in London, Thomas Young, notified Malus most graciously that he had been awarded the Rumford Medal and acknowledged that Malus's work caused him to doubt the theory of interference he had been developing for a number of years!

The pinnacle of Malus's scientific career was his election to the fellowship of the "Forty Immortals," as the members of the Academy of Sciences (Institute de France) were called in France. His friend Arago remembered that on the day of his election he promised to tell Malus the results, but for some reason the voting was delayed, and the unforeseen postponement caused the great physicist to think he had been rejected. This thought drove him to despair, and even his loving wife couldn't console him. A fearless warrior lost courage due to an imagined rebuff by the Academy. Arago saw this as a clear indication of the profound importance of academies in the life of science.

In 1810 Malus became an examiner at the Polytechnical School. This position was somewhat higher than that of an ordinary teacher: the examiner checked both the degree of preparation of the students and the level of teaching of the professors. A year later Malus took on the duties of principal after the aged Monge retired, and only a few formalities remained before he could take the post permanently. After the wartime deprivations of his youth, he now had everything necessary for the full enjoyment of life. His friends expected him to make even more discoveries. Then a sudden and acute case of consumption carried him off. It's possible that the traces of plague in his weak body hastened his end. Madame Malus never left his side in his last hours. Her husband's death was too much for her, and she died a few months later. Malus was 36 years old.

The discoveries of Malus have played a key role in the development of physics. In characterizing the polarization of light as a fundamental phenomenon, he capped the preceding work of Bartholine, Newton, and Huygens. The newly discovered properties of polarized light helped provide an experimental basis for the revolution in optics that occurred just a few years after Malus died: the transition to the wave theory of light.

Chopping up Pick's Theorem

But watch out for the frogs!

by Nikolay Vasilyev

TO EVALUATE THE AREA OF a polygon, we can rule the plane into a grid of equal squares, lay the polygon over the grid, and count the number N_1 of squares lying completely inside the polygon and the number N_2 of squares that have at least one common point with the interior of the polygon. Then, if the area of one grid square is 1, the area A of the polygon satisfies the relation

$$N_1 \le A \le N_2.$$

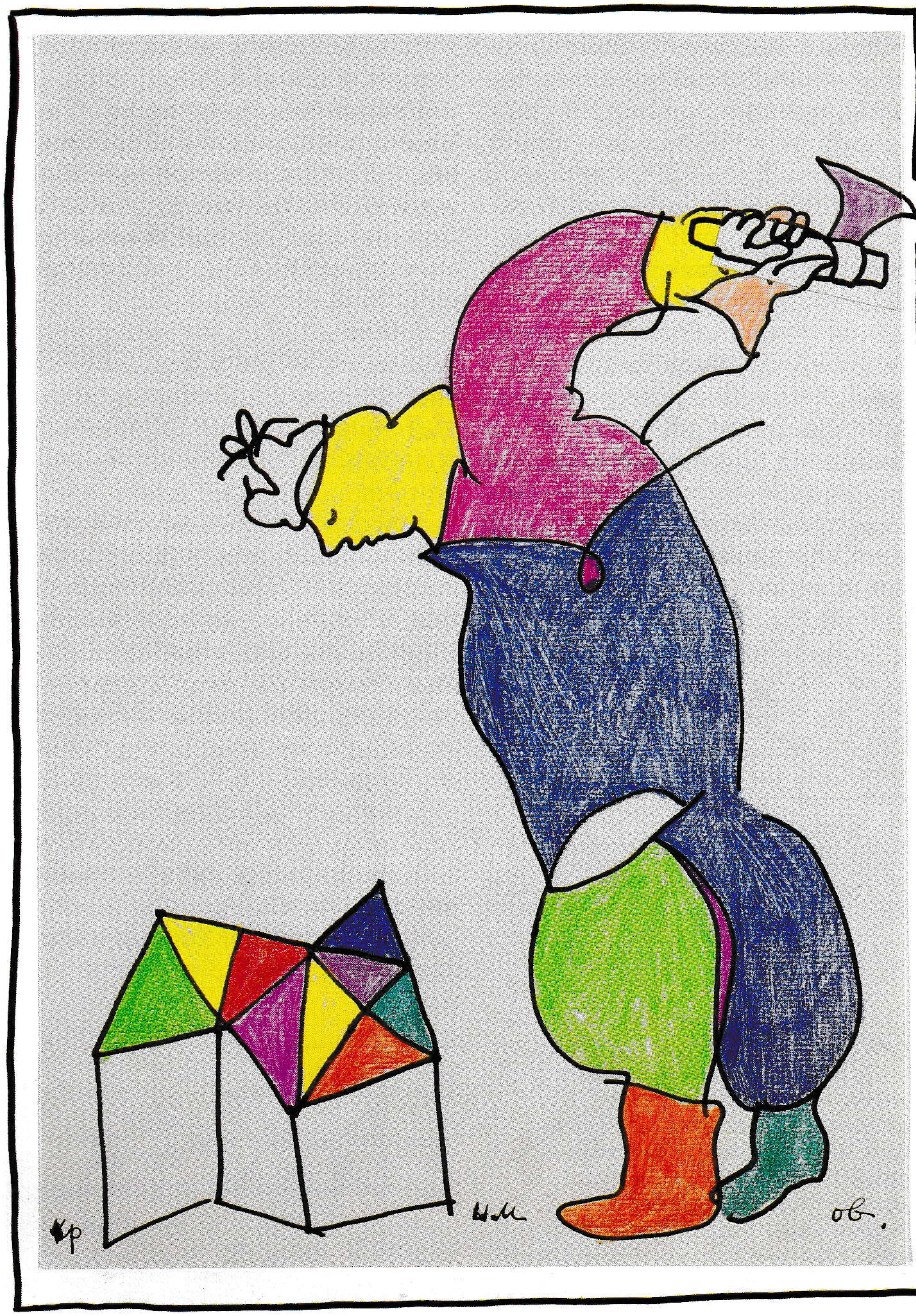
(This fact can be used for an accurate definition of the area of a polygon and other figures.)

Below we'll look only at polygons whose vertices are *nodes* of the grid—that is, intersection points of the lines of the grid. We'll call them *grid polygons*. It turns out that the area A of such a polygon can be expressed by this simple formula:

$$A = i + \frac{b}{2} - 1,$$

where i is the number of the nodes strictly inside and b is the number of the nodes on the border of the polygon (that is, on its sides and vertices).

This formula is usually called Pick's Theorem after the mathematician who discovered it in 1899. (However, we cannot be sure that this natural formula, which allows for a number of different proofs,



Art by Dmitry Krymov

never occurred to anyone before that.[1]) The proof of Pick's Theorem in this article, and some of its applications presented here, will have linkages with some of the math challenges in *Quantum*'s How Do You Figure? department.

Simple triangles

Let me remind you that all the polygons—in particular, the triangles—that we consider are grid polygons: they have their vertices on the nodes of the square grid. We assume that the grid is infinite in all directions and the side length of a grid square is 1.

The area of any grid triangle is easy to compute by representing it as the sum or difference of the areas of right triangles and grid rectangles whose sides lie on the grid lines through the vertices of the triangle in question. If you do this for the triangles in figure 1, you'll see that the area is always a "half-integer" number—that is, a number of the form $m/2$ with an integer m.

Let's say that a triangle is *simple* if there are no nodes inside it or on its sides, its vertices excepted. This name was chosen because any other triangle can be composed of simple ones (see statement 10 below). Notice that all the simple triangles in figure 1 have an area of 1/2. We'll see that this is no coincidence.

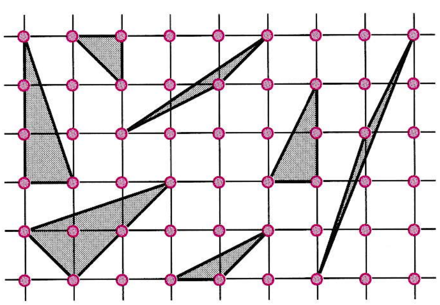


Figure 1

In math challenge M14 (in the September/October 1990 issue of *Quantum*) we examined the following situation: three frogs that sit initially at three vertices of a unit grid square start to play—what else?—leapfrog. Any one of them—say A—can jump over any other—say, B—to land at point C symmetric to A about B (fig. 2—it's clear that the frogs will always hit the nodes of our grid). *What are the triples of points where the frogs can find themselves after a series of leaps?* (This question was posed in the *solution* to M14—it generalized the question of the problem proper.)

We'll call a triangle *accessible* if there exists a series of jumps after which the frogs, starting at three vertices of one grid square, arrive at the vertices of this triangle. The transformation of a triangle consisting of a replacement of one of its vertices with the landing point of a frog that jumped from this vertex over any of the other two vertices will be called simply a *jump*.

THEOREM 1. *The following three properties of grid triangles are equivalent to each other: (1) a triangle has an area of 1/2; (2) a triangle is simple; (3) a triangle is accessible.*

You can convince yourself that this theorem is true by proving the subsequent 12 statements. Those that are supplied with hints in the solution section are marked with a star. The rest will be not very difficult if you prove them in this order:

$$\begin{array}{ccccccccccccc} 3 & \Rightarrow & 4 & & & & 1 & \Rightarrow & 2 & & 10 & & \\ & & \Downarrow & & & & & & \Downarrow & & \Downarrow & & \\ 5 & \Rightarrow & 6 & \Rightarrow & 7 & \Rightarrow & 8 & \Rightarrow & 9 & \Rightarrow & 11 & \Rightarrow & 12 \end{array}$$

1. A jump does not change the area of a triangle.

2. Any accessible triangle has an area of 1/2.

3*. If a line is added to a simple triangle ABC to form a parallelogram $ABCD$, then there will be no nodes of the grid inside or on the border of $ABCD$ (except for its vertices).

4*. A simple triangle remains simple after any jump.

5*. One of the angles of a simple triangle is always obtuse or right. The second case is possible only when all the vertices of the triangle belong to the same square of the grid. Such a simple triangle—with side lengths 1, 1, $\sqrt{2}$—will be called *minimal*.

6. With a single jump it is possible to turn any simple nonminimal triangle into a triangle whose longest side is shorter than the longest side of the original triangle.

7*. Any simple triangle can be turned into a minimal one in a finite number of jumps.

8. Any simple triangle is accessible.

9. Any simple triangle has the area 1/2.

10*. Any triangle can be cut into simple triangles.

11. The area of any grid triangle is equal to $m/2$, where m is an integer, and the integer m is also the number of simple triangles into which it can be cut (that is, the number is constant for any given triangle).

12. Any grid triangle of area 1/2 is simple.

Our Theorem 1 immediately follows from statements 2, 8, and 12.

Prove these additional properties of simple triangles:

13. For any segment AB joining two nodes of the grid there is a node C such that the triangle ABC is simple.

14. The node C in the previous statement can be chosen so that the angle ACB is obtuse or right.

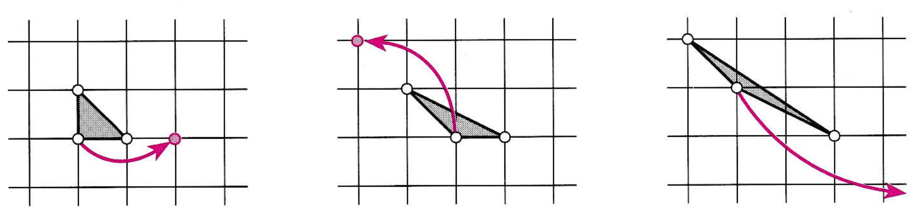


Figure 2

[1]This seems to be the only trace left in the history of mathematics by its author, G. Pick. At least, our mathematical encyclopedia has no entry for him. For another proof of Pick's Theorem see, for instance, *Introduction to Geometry* by H. S. M. Coxeter (New York: John Wiley & Sons, 1961).—*Ed.*

If the plane is tiled with congruent copies of a parallelogram such that any two of them either have an entire common side, a common vertex, or no common points at all, then the vertices of these parallelograms form a grid. We say that this grid is *generated* by the initial parallelogram. For instance, the grid considered throughout this article is generated by a unit square. But we can also think of it as generated by certain other grid parallelograms (like the red parallelograms in figure 3).

15. A parallelogram generates our square grid if and only if its diagonal divides it into two simple triangles.

This property can also be reformulated as follows.

16. A grid triangle is simple if and only if all the triangles obtained from it by translations that take one of its vertices into each of the nodes of the grid do not overlap (fig. 3).

Let's return to the leapfrog problem. Assume the frogs start at the vertices of some *fixed* minimal triangle rather than an arbitrary one (which was assumed in the original formulation of the problem). After any jump a frog always lands on a node that is, both horizontally and vertically, an even number of squares from where it started. Therefore, it always hops in its own grid of bigger squares (measuring 2×2). In figure 4 each of four such "subgrids" that together constitute the entire unit-square grid has its own color.

By Theorem 1, the frogs are always at the vertices of a simple triangle. This implies the following interesting property:

17. The vertices of any simple triangle always belong to three *different* subgrids described above (and are colored differently).

Now it's not difficult to answer the question about the possible positions of our three frogs (in the case of a fixed initial triangle).

18. Three frogs can get to three given nodes of the grid if and only if these nodes are the vertices of a simple triangle and are the same colors (fig. 4) as the vertices of the initial triangle. (The "only if" part of this statement has already been proved.)

This statement, together with statement 13, provides an answer to the second question posed in the solution to M14 (about the possible positions of two frogs):

19. If three frogs are jumping according to our rules, then any two of them can hit those and only those pairs of nodes that do not have other nodes on the segment joining them and have the same color as the initial nodes occupied by these two frogs.

Triangulations of a polygon

We've studied in detail a particular sort of grid polygon (triangle) that corresponds to the values $i = 0$, $b = 3$, $A = 1/2$ in Pick's Theorem. But this particular case allows us to pass directly to the most general one by using the theorem about cutting an arbitrary polygon into triangles (the grid is not needed any longer).

Consider a polygon (not necessarily convex) on the plane and a finite set K of points inside it or on its border such that all the vertices of the polygon belong to K. *A triangulation with vertices in the set K* is, by definition, a partition of the polygon into triangles with vertices in K such that every point of K is a vertex of all the triangles that contain this point. (That is, the points of K are not allowed to lie in the interior or on the sides of the triangles—see figure 5).

Theorem 2. *(a) Any n-gon can be cut into triangles by its diagonals; the number of these triangles always equals $n - 2$.* (Such a partition is a triangulation with the vertices at the polygon's vertices.)

(b) Mark b points on the border of a polygon (including all its vertices) and i points in its interior. Then there exists a triangulation with its vertices at these points. The number of triangles in such a triangulation always equals $b + 2i - 2$.

Part (a) is a particular case of (b) for $b = n$, $i = 0$. The proof of (b) will again be divided into a series of simple statements.

20*. From the vertex of the largest angle of an n-gon ($n > 3$), a diagonal that entirely lies inside the polygon can always be drawn.

21. If a diagonal of an n-gon cuts it into a p-gon and a q-gon, then $n = p + q - 2$.

22. Any n-gon can be cut into $n - 2$ triangles.

23. The sum of the angles of an n-gon (not necessarily convex) equals $(n - 2)\pi$.

24.* Any triangle with a number of points given inside it and on its border that include all its vertices has a triangulation with the vertices at these points.

25. The same is true for any n-gon.

26. The number of triangles in a triangulation is equal to $b + 2i - 2$, where i and b are the numbers of the vertices of the triangulation in the interior and on the border of the polygon, respectively.

This yields Theorem 2.

27. Derive Pick's Theorem $A = i + b/2 - 1$ from theorems 1 and 2.

A good way to prove statement 26 is to find the sum of angles of all the triangles in the triangulation in two

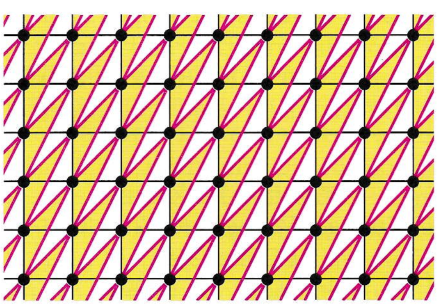


Figure 3

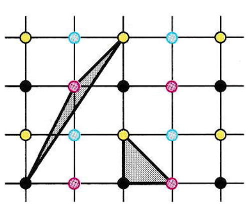


Figure 4

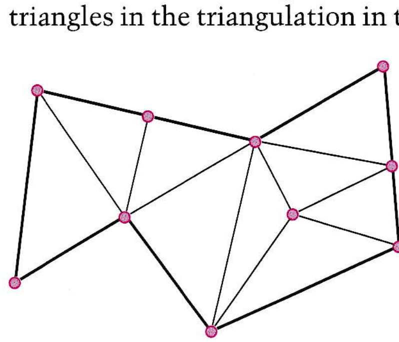


Figure 5

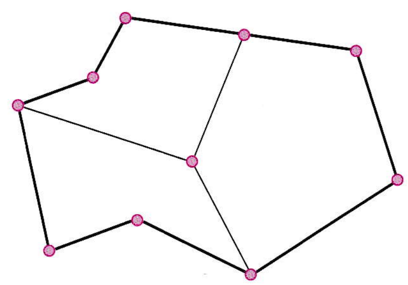


Figure 6

different ways. This method proves to be useful in many other problems of combinatorial geometry, including problems dealing with partitions of polygons. I'll give two more examples.

A partition of an n-gon into several polygons will be called *regular* if no vertex of any polygon lies on a side of another, unless it is also a vertex of the other polygon.

28. If an n-gon is divided regularly into k-gons (for a fixed integer k) so that i is the number of vertices inside and b is the number of vertices on the border of the n-gon, then the number m of k-gons is given by the formula (fig. 6)

$$m = \frac{b + 2i - 2}{k - 2}.$$

29. (This is a version of Euler's Theorem—see also "Topology and the Lay of the Land" in the September/October 1992 issue.) If N_0 points of the plane and N_1 segments joining some of them form a polygon divided regularly into N_2 polygons, then $N_2 - N_1 + N_0 = 1$.

Exercises

These applications of Pick's Theorem are connected mostly with various problems and theorems about polygonal curves on a grid, rather than with computing the areas of given polygons.

30. Suppose that the ratio of the area of a polygon to the square of one of its side lengths is irrational (as is the case for, say, an equilateral triangle). Prove that it is impossible to draw a similar polygon such that all its vertices are nodes of the grid.

31. Let A and B be two nodes such that B lies p squares to the right of A and q squares above it (so the distance between them is $\sqrt{p^2 + q^2}$. Find the distance from the line AB to the node closest to this line but not lying on it.

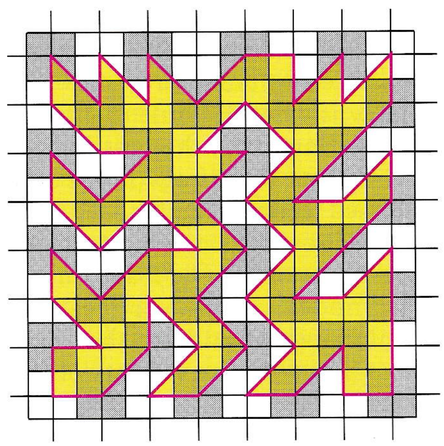


Figure 7

32. A chess king made the rounds of all the squares on the chessboard, visiting every square once. (A king can move from the center of any square to the center of any of the 8 neighboring squares.) The center of each square was joined to the center of the next square on the king's route (the last square was joined to the first one). It turned out that this path formed a closed polygonal path with no self-intersections. (a) What is the greatest possible length of such a path? (b) What is the greatest area that this path can bound? (The side of a square is 1 unit long.)

Part (a) has already been examined in *Quantum* (March/April 1991, math challenge M30). Here I'll give another solution. But let's begin with part (b). By Pick's Theorem the area bounded by the path is equal to $64/2 - 1 = 31$. The grid here consists of all 64 centers of squares, and all these nodes lie on the border of the polygon. Now let's look at part (a). Figure 7 shows a path in which 36 of 64 moves are diagonal and have a length of $\sqrt{2}$. Let's prove that it's impossible to increase this number. Since the largest path will have the largest number of such segments, this will show that the path in figure 7 has the largest length.

Look at the 1×1 squares that have their sides parallel to lines of our grid and whose diagonals are the diagonal line segments of the path.

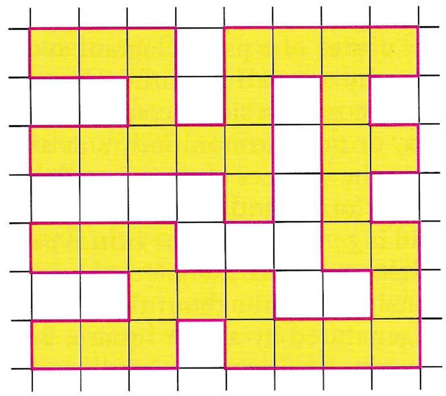


Figure 8

One half of this square lies outside the polygon bounded by the king's path. The total area of all these halves is not greater than $7^2 - 31 = 18$, because they all go into the 7×7 square on our grid. So the number of diagonal moves does not exceed 36.

Thus, the answers to this problem are (a) $28 + 36\sqrt{2}$, (b) 31.

33. You have to draw a closed non-self-intersecting polygonal path along the lines of a square grid that passes through all the nodes inside a rectangle measuring $p \times q$ squares (see figure 8, where $p = 10$, $q = 9$). (a) For what p and q is this possible? (b) What will the length of this path be? (c) What area will it bound?

HINTS FOR STARRED ITEMS ON PAGE 59

Bulletin Board

Boston University's PROMYS

"Have any plans for the summer?" It's a question you usually don't hear until May. But if you're a high school student with a passion for math, you might want to start thinking about the Program in Mathematics for Young Scientists (PROMYS) at Boston University.

PROMYS is a residential program designed for 60 ambitious students entering grades 10 through 12. From July 3 to August 13, participants will explore the creative world of mathematics. In addition to tackling interesting problems in number theory, more experienced participants may also study abstract algebra, dynamical systems, and the Riemann zeta function.

Problem sets are accompanied by daily lectures given by research mathematicians with extensive experience in Prof. Arnold Ross's longstanding Summer Mathematics Program at Ohio State University. A staff of 18 college-age counselors lives in the dormitories and is available to discuss mathematics with the students. Each participant belongs to a problem-solving group that meets with a professional mathematician three times a week. Special lectures by outside speakers offer a broad view of mathematics and its role in the sciences.

Admission is based applicants' solutions to a set of challenging problems, teacher recommendations, high school transcripts, and essays explaining their interest in the program. The estimated cost is $1,300 for room and board; books may cost an additional $100. Financial aid is available, and no student should be deterred from applying due to financial considerations.

For application materials, write to PROMYS, Department of Mathematics, Boston University, 111 Cunningham St., Boston, MA 02215, or call 617 353-2563. Applications will be accepted from March 1 until June 1, 1994.

Designing students

High school students from across the country have a chance to design their ideal youth community center and win college scholarships by entering the 1994 National Architectural Design Competition, sponsored by the New Jersey Institute of Technology School of Architecture.

Why a community center? "Because of its relevance to high school students in both urban and suburban areas," explains Prof. Mark Hewitt, the competition's coordinator, "and to emphasize the need for architects, new and practicing, to get involved in these types of community revitalization projects." Students will design both the interior and exterior of a small drop-in center that includes recreational facilities for teenagers.

The top prize is a five-year full-tuition scholarship to NJIT's School of Architecture, the eighth largest architecture school in the US. Second prize is a five-year half-tuition scholarship, and cash awards of $250 will be presented to four third-prize winners.

Entry forms must be submitted by January 28, 1994. Competition information packets will be mailed to entrants on February 11. The deadline for project submission is March 28, and winners will be announced on May 2.

To receive an entry form or for more information, call 201 596-3080 (in New Jersey, 1 800 222-NJIT), or write to Mark Hewitt, School of Architecture, New Jersey Institute of Technology, University Heights, Newark, NJ 07102-1982.

When blood goes bad

Blood—so accessible, renewable, and vital to life—has always been a primary focus of medical research. A new, lavishly illustrated report from the Howard Hughes Medical Institute takes a close look at recent scientific advances and how they apply to the treatment of blood disorders.

Blood: Bearer of Life and Death describes the progress made in understanding how blood cells develop and function, and how faults in the bloodstream can cause disease in humans. The report touches on the AIDS epidemic; cancer patients who cannot make new blood cells after chemotherapy; sickle cell disease, hemophilia, and other inherited disorders; and the high death toll from blood clots. The publication also describes gene therapy research for blood disorders, and how scientists are closing in on the "mother" cell of the bloodstream: the stem cell. The ability to manipulate stem cells could have far-reaching therapeutic value.

This report is the fourth in a continuing series and is available free of charge to teachers and students. Write to the Howard Hughes Medical Institute, Office of Communications, 4000 Jones Bridge Rd., Chevy Chase, MD 20815-6789.

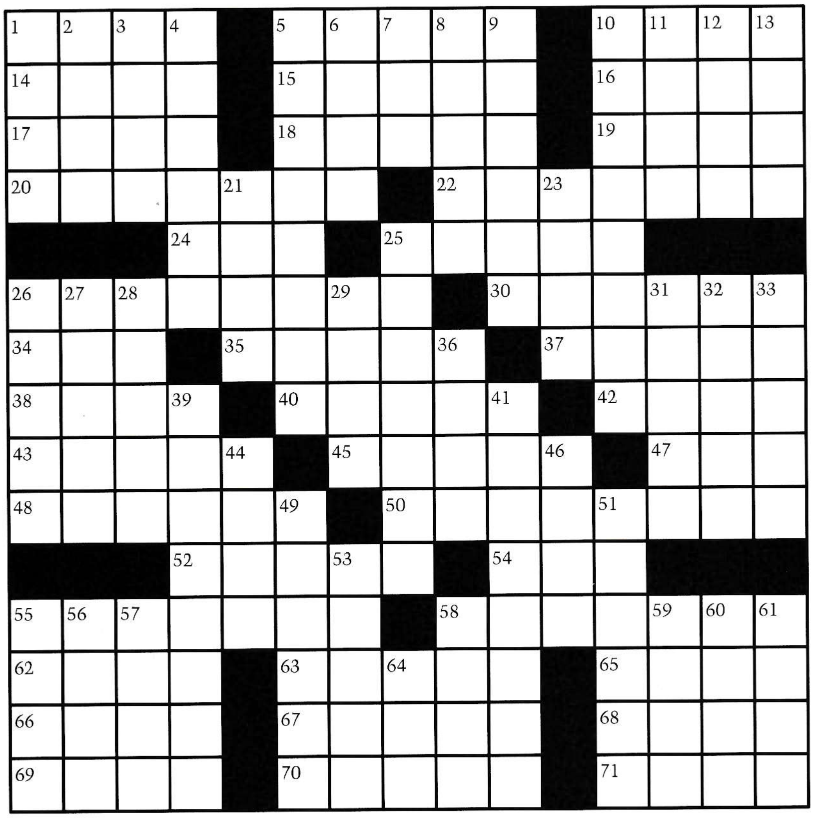
A place to publish

The Russian journal *Prologue* is seeking original student papers in mathematics, the sciences, and the humanities. Published quarterly, *Prologue* serves as a kind of training

CONTINUED ON PAGE 61

Criss × cross science

by David R. Martin



Across

1 Unit of life
5 Brazilian novelist Jorge ___
10 German river
14 Polynesian spirit
15 Fright
16 Engine speed
17 Goes for the gold
18 Ger. Pres. Friedrich ___ (1871–1925)
19 Knight's wife
20 Math course
22 Type of relativity
24 Betray
25 Unit of length
26 Force over area
30 "___ and stones ..."
34 Assist
35 Wash away
37 Booth
38 Remitted
40 Musical composition
42 Australian bird (var.)
43 Nypa palm
45 Mud
47 Brooch
48 Endured
50 Removing soluble constituents
52 Fragrance
54 Former Fed. agcy.
55 Jet part
58 Strongly alkaline
62 Carbamide
63 ___ a time
65 Donuts
66 Prohibits
67 Increase
68 Alter (text)
69 Toward shelter
70 Submerges
71 Sensory organ

Down

1 Bullfighter's cloak
2 Lat. abbr.
3 Respiratory organ
4 Monochromatic light sources
5 Camera part
6 Former people of Chad
7 Sue ___ Langdon
8 Funeral hymn
9 Stable electron groups
10 Iron ore
11 Jewish month
12 ___ mater
13 Film spool
21 Acid's opposite
23 Gains
25 ___ oblongata
26 Of John Paul
27 Lariat
28 Corrects
29 Decays
31 Artistic Italian family name
32 Kaluza–___ theory
33 Tossed
36 Singer Adams
39 Computer information
41 Emits
44 Persian fairy
46 Beige
49 Arsenic atoms in silicon
51 Hurry
53 Strait near Wales
55 Orchestra member
56 Russian mountain range
57 ___ Descartes
58 Barrel
59 Commotion
60 Eye part
61 Quote
64 One: German

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE NOVEMBER/DECEMBER PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M101

It's not difficult to show that a set of n elements has exactly 2^n subsets. In particular, six students, theoretically, could form $2^6 = 64$ committees (including an empty committee).

Divide all the possible 64 subsets of the set of our students into pairs $(A, \overline{A})$, where A is an arbitrary subset and $\overline{A}$ is its complement (which consists of all the students that don't enter the set A). Since there are 32 pairs and 30 committees, both subsets of at least one pair are nonempty and aren't committees. Let A be one of these subsets. If it intersects with all the 30 committees, it's the subset we need. If it's disjoint with a certain subset B, then the second set of the chosen pair, $\overline{A}$, contains the committee B. But B and, therefore, $\overline{A}$ intersect with all the other committees, so $\overline{A}$ is the desired subset.

You can prove that it's impossible to organize more than 32 committees satisfying the condition.

M102

(a) The required representation can be written out explicitly:

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

(b) First let us note that if $A(x)$ and $B(x)$ are polynomials that can be written in the required form, then so are $(A + B)(x)$ and $(A - B)(x)$. The proof of this result uses the fact that the sum of two monotonic increasing functions is monotonic increasing, and some simple algebra.

Let's prove the statement by induction over the degree of a polynomial. For degree zero (constant functions) the statement is trivial. Suppose it's true for any degree less than n. Write an arbitrary polynomial $P_n(x)$ of degree n in the form

$$P_n(x) = ax^n + P_{n-1}(x),$$

where $P_{n-1}(x)$ is a polynomial of degree not greater than $n - 1$ that satisfies our statement by the induction hypothesis. By the note at the beginning of this paragraph, it will suffice to prove our statement for the polynomial ax^n. For odd n, we can write $ax^n = (a + 1)x^n - 1x^n$ if $a > -1$, and $ax^n = 1x^n - (1 - a)x^n$ if $a < -1$. Since kx^n is monotonic increasing for $k > 0$, these representations satisfy the requirements of the problem.

If n is even we must go a bit further. Note that

$$(n+1)x^n = (x+1)^{n+1} - x^{n+1} - Q_{n-1}(x),$$

where $Q_{n-1}(x)$ is of degree $n - 1$, and hence is covered by the induction hypothesis. Now we can write

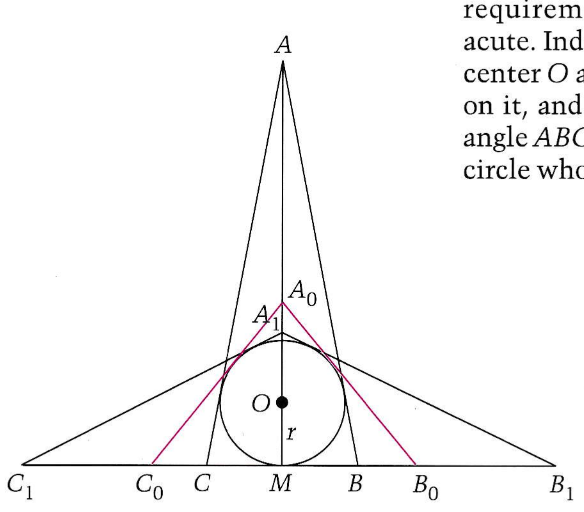


Figure 1

$$\begin{aligned} ax^n &= \frac{a}{n+1}(n+1)x^n \\ &= \frac{a}{n+1}(x+1)^{n+1} - \frac{a}{n+1}x^{n+1} \\ &\quad - \frac{a}{n+1}Q_{n-1}(x), \end{aligned}$$

which leads to the desired representation.

For an even $n = 2k$ we can give another argument. By the inductive conjecture, $x^k = P(x) - Q(x)$ with increasing polynomials P and Q. It follows that

$$\begin{aligned} x^{2k} &= [P(x) - Q(x)]^2 \\ &= 2P^2(x) + 2Q^2(x) - [P(x) + Q(x)]^2. \end{aligned}$$

Substituting $P(x)$ for x in the formula in part (a), we'll get a representation of $P^2(x)$ as the difference of increasing polynomials. The same can be done for $Q^2(x)$ and $[P(x) + Q(x)]^2$, yielding the required representation for x^{2k}. (A. Vaintrob, V. Dubrovsky)

M103

The answer is no (in general). This would be easy if we ignored the requirement that the triangle is acute. Indeed, let's fix a circle (with center O and radius r) and a point M on it, and consider an isosceles triangle ABC circumscribed about this circle whose base touches the circle

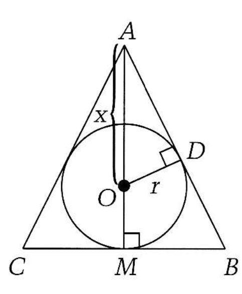


Figure 2

at point M. As we change the height AM of this triangle, we clearly see (fig. 1 on the preceding page) that its side AB can be made arbitrarily long by stretching the triangle either vertically ($AM \to \infty$) or horizontally ($AM \to 2r$). Since the side length changes continuously, any value of AB greater than its minimal value A_0B_0 is taken at least twice: once for the height greater than A_0M, once for the height smaller than A_0M.

However, it might happen that one of these two triangles with the same side lengths is always obtuse, so we have to undertake a more extended investigation.

Put $x = OA$ (fig. 2). Then, from similar triangles AOD and ABM, we have $AB/AM = OA/AD$. Letting $y = AB$, $x = AO$, we have $y = AM \cdot OA/AD = (x + r)x/\sqrt{x^2 - r^2}$. Letting $t = x/r$, we can rewrite the latter expression as

$$y = AB = rt\sqrt{\frac{t+1}{t-1}}. \qquad (1)$$

Since we are thinking of r as fixed, the expression gives $y = y(t)$ as a function of t. We are interested in the behavior of the function on the interval $\sqrt{2}, t, \infty$. Direct calculation shows that when $t = \sqrt{2}$, the side length of the right isosceles triangle is $y = r(2 + \sqrt{2}) \cong 3.41r$, while, say, for $t = 1.5$, which is greater than $\sqrt{2}$, $y(t) = r \cdot 3\sqrt{5}/2 \cong 3.35r$. Since $y(t) \to \infty$ as $t \to \infty$, a value such as $3.4r$ is taken at least twice: in the interval $(\sqrt{2}, 1.5)$ and in the interval $(1.5, \infty)$.

This concludes the solution of our particular problem, but an interested reader may go further and find, using derivatives, that the function $y(t)$ decreases for $1 < t \le \tau$ and increases for $\tau \le t < \infty$, where the point of minimum τ is the famous "golden section" $(\sqrt{5} + 1)/2 \cong 1.618$. This ubiquitous number, which you may have met many time in the pages of this magazine, crops up yet again! (V. Dubrovsky)

M104

We'll solve part (b) first. The answer to this problem consists of the following pairs (p, q): (4, 4), (5, 6), (6, 5), $(n, -n - 1)$ for all integer n ($n \ne 0, -1$). Let x_1, x_2 and y_1, y_2 be the integer roots of the trinomials $x^2 + px + q$ and $x^2 + qx + p$, respectively. Using the formulas for the sum and product of the roots of a quadratic,

$$x_1 + x_2 = -p,\ y_1 + y_2 = -q,$$
$$x_1x_2 = q,\ y_1y_2 = p.$$

Dividing the product of the first two equalities by the product of the other two ($pq \ne 0$) yields

$$\left(\frac{1}{x_1} + \frac{1}{x_2}\right)\left(\frac{1}{y_1} + \frac{1}{y_2}\right) = 1.$$

At least one of these factors must be greater than or equal to 1. Without loss of generality, we can assume that this is the first, so that

$$\left|\frac{1}{x_1} + \frac{1}{x_2}\right| \ge 1.$$

Similarly, one of the addends in the absolute value sign above must be greater than or equal to 1/2, and we can assume that this will be x_1 (if it is not the first, we can simply relabel them). Then

$$\left|\frac{1}{x_1}\right| \ge \frac{1}{2},$$

or

$$|x_1| \le 2.$$

Consider the two possibilities.

1. $|x_1| = 1$. Then x_2 must be a nonzero integer of the same sign as x_1. If $|x_2| = n$, then $p = -(x_1 + x_2) = \pm(n + 1)$, $q = n$. For natural numbers n, the polynomial $x^2 + nx - (n + 1)$ has integer roots ($-n - 1$ and 1).

The polynomial $x^2 + nx + (n + 1)$ demands closer examination. We can write its discriminant as $(n - 2)^2 - 8$. Since the roots are integers, this must be a perfect square of some integer b. Then we have $(n - 2)^2 - b^2 = 8$, or $(n - 2 - b)(n - 2 + b) = 8$, where n and b are positive integers. This equation is true only if $n - 2 - b = 2$ and $n - 2 + b = 4$—that is, $n = 5$ (and $b = 1$). Thus, we get the pairs $(-n - 1, n)$ (or $(n, -n - 1)$, which is the same) for all integer n other than 0 and -1, and the pair $(n + 1, n)$ for $n = 5$ only—that is, the pair (6, 5) and, symmetrically, (6, 5).

2. $|x_1| = 2$. Then $|1/x_1 + 1/x_2| \ge 1$ yields $|x_2| \le 2$. The case $|x_2| = 1$ was, with a change in notation, considered above, and the case $|x_2| = 2$ leads to the pairs (4, 4) and (−4, 4), the second of which is no good (the equation $x^2 + 4x - 4 = 0$ has no integer roots).

M105

Let's depict each apartment as a point on the plane, and a move from one apartment to another as an arrow joining the corresponding points. Clearly, the set of arrows that represent an arbitrary complex exchange is always such that in each point one arrow begins and one ends. Therefore, if we start at any point and move along the arrows, we'll always be arriving at new points until we come back to the beginning of our route (this is inevitable, because there is a finite number of points). Similar closed circuits are formed by other arrows, so all the arrows fall into a number of closed chains—*cycles* (fig. 3). Thus, it will suffice to prove the statement for a cyclic exchange of apartments.

Suppose we have to perform the exchange $A_1 \to A_2 \to \ldots \to A_n \to A_1$—that is, move the residents of apartment A_1 into apartment A_2, from A_2 into A_3, and so on. On the first day, let's perform the swaps $A_1 \leftrightarrow A_{n-1}$, $A_2 \leftrightarrow A_{n-2}$, ...—in general, $A_k \leftrightarrow A_{n-k}$ (if we make the convention that $A_0 = A_n$); the residents of the apartments A_n and, in the case of even n, $A_{n/2}$ stay where they are for

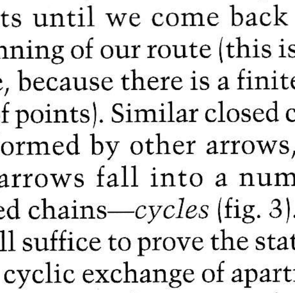


Figure 3

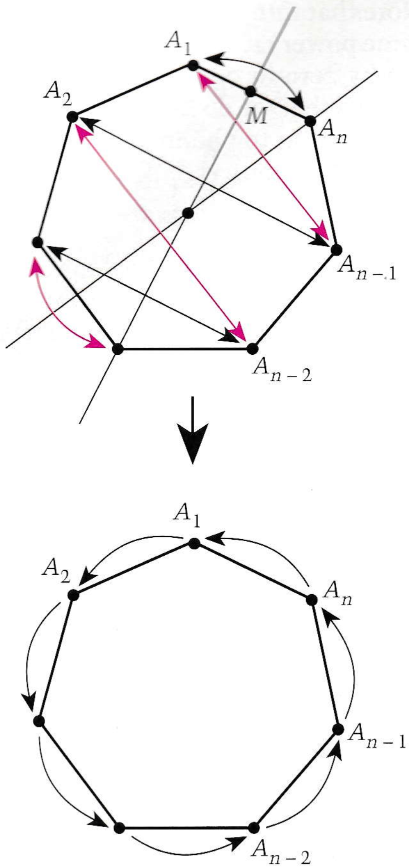


Figure 4

the time being. On the second day, the swaps $A_1 \leftrightarrow A_n$, $A_2 \leftrightarrow A_{n-1}$, ...—that is, $A_k \leftrightarrow A_{n+1-k}$—are carried out. As the result of both exchanges, the residents of apartment A_k move to $A_{n+1-(n-k)} = A_{k+1}$ for $k = 1, \ldots, n-1$, and from A_n to $A_{n+1-n} = A_1$, which is what we wanted.

This sequence of exchanges is geometrically illustrated in figure 4. If $A_1, \ldots, A_n$ are the vertices of a regular n-gon with center O, then the first series of exchanges corresponds to the reflection about line OA_n, and the second series results in the reflection about line OM, where M is the midpoint of A_nA_1. We can prove that two reflections about lines intersecting at point O at an angle α performed one after another result in a rotation about O through the angle 2α. In our case we get a rotation through $2\pi/n$. This rotation rearranges the vertices of the polygon cyclically.

Physics

P101

The buoyancy force (see figure 5) is equal to the weight of the displaced water and is given by $F_b = \rho_0 Sl'g$, where S is the cross-sectional area of the float and l' is the length of the submerged portion of the float. From the mass of the float we find that $S = m/\rho_1 l$, where l is the total length of the float. Combining these two relationships we obtain

$$F_b = mg\frac{\rho_0}{\rho_1}\frac{l'}{l}.$$

The float will float if the buoyant force balances the weight of the float and the force F exerted by the sinker. Since the sinker is also in the water, this force is given by $F + mg = F_b$, where

$$F = Mg - \rho_0\frac{M}{\rho_2}g = Mg\left(1 - \frac{\rho_0}{\rho_2}\right).$$

Therefore,

$$F_b = mg + Mg\left(1 - \frac{\rho_0}{\rho_2}\right).$$

Comparing these two expressions for F_b yields

$$\frac{l'}{l} = \frac{\rho_1}{\rho_0}\left[1 + \frac{M}{m}\left(1 - \frac{\rho_0}{\rho_2}\right)\right].$$

The inequality $l' \le l$ results in the constraint

$$\frac{M}{m} \le \frac{\frac{\rho_0}{\rho_1} - 1}{1 - \frac{\rho_0}{\rho_2}}.$$

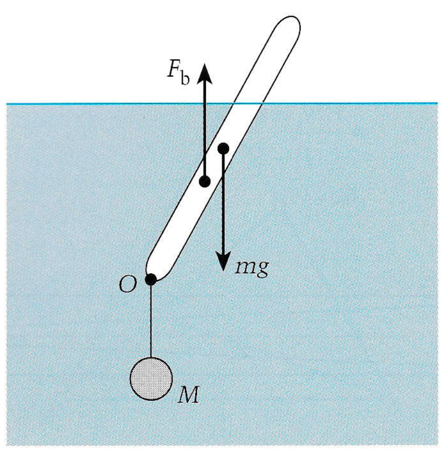


Figure 5

In order for the float to sit in a vertical position, the torque produced by the buoyant force about point O in figure 5 must be at least as large as the torque due to the float's weight:

$$\frac{F_b l' \sin\theta}{2} \ge \frac{mgl\sin\theta}{2},$$

where θ is the angle the float makes with the vertical. Substituting in our first formula for F_b, we get

$$mg\frac{\rho_0}{\rho_1}\frac{l'}{l}l' \ge mgl,$$

or

$$\frac{\rho_0}{\rho_1} \ge \frac{l^2}{l'^2}, \quad \frac{l'}{l} \ge \sqrt{\frac{\rho_1}{\rho_0}}.$$

Taking into account the relation for l'/l, we obtain

$$\frac{\rho_1}{\rho_0}\left[1 + \frac{M}{m}\left(1 - \frac{\rho_0}{\rho_2}\right)\right] \ge \sqrt{\frac{\rho_1}{\rho_0}},$$

which results in a second constraint:

$$\frac{M}{m} \ge \frac{\sqrt{\frac{\rho_0}{\rho_1}} - 1}{1 - \frac{\rho_0}{\rho_2}}.$$

Thus, the conditions on the mass ratio are

$$\frac{\sqrt{\frac{\rho_0}{\rho_1}} - 1}{1 - \frac{\rho_0}{\rho_2}} \le \frac{M}{m} \le \frac{\frac{\rho_0}{\rho_1} - 1}{1 - \frac{\rho_0}{\rho_2}}.$$

P102

Let's consider the collision of one molecule with the piston. The time between collisions is $\Delta t = 2l/v$, and the momentum transferred to the piston is $\Delta p_1 = 2mv$. The average force acting on the piston is $f_1 = \Delta p_1/\Delta t = mv^2/l$. When the piston moves with a speed v_p the molecules colliding with it gain an additional speed $\Delta v = 2v_p$. During the period between collisions the length of the gas cylinder changes by

$$\Delta l = -v_p\Delta t = -\frac{\Delta v}{2}\frac{2l}{v}$$

—that is,

$$\frac{\Delta v}{v} = -\frac{\Delta l}{l},$$

from which we get

$$v = \frac{l_0 v_0}{l}.$$

(We assume that the piston is much more massive than a molecule, and that $v_p \cong$ const.) Therefore,

$$f_1 = \frac{m l_0^2 v_0^2}{l^3}.$$

The total force acting on the piston due to all the molecules is $F = Nf_1$, where $N = n_0 l_0 S$ (S is the area of the piston), and so the pressure is

$$P = \frac{Nf_1}{S} = \frac{n_0 l_0 S}{S} \frac{m l_0^2 v_0^2}{l^3}$$

$$= \frac{n_0 V_0 \cdot m l_0^2 \cdot S^2 \cdot v_0^2}{S^3 l^3}$$

$$= \frac{n_0 m v_0^2 V_0^3}{V^3}.$$

From this we get

$$PV^3 = n_0 m v_0^2 V_0^3 = \text{const}$$

—that is, a one-dimensional adiabatic:

$$PV^\gamma = \text{const}, \quad \gamma = \frac{i+2}{i} = 3, \quad i = 1.$$

P103

The removal of grease stains from clothes by ironing them is based on the fact that surface tension decreases as the temperature increases. So if a hot iron is applied to one side of the stained fabric and a piece of ordinary paper is pressed against the other side, the grease is transferred to the paper (or another piece of fabric that absorbs fatty substances).

P104

The reading on any meter is proportional to the expenditure of the thing being measured. In our case, the reading on an electric meter is proportional to the amount of energy consumed, or (since the time t = 1 minute is fixed) to the power. It doesn't matter what units are used to measure the power—turns of a counter, numerals, and so on. Thus, $N = KP$. If $P_1/P_2 = N_1/N$, then it's clear that

$$N_3 = N_1 \frac{P_3}{P_1} = N_2 \frac{P_3}{P_2},$$

where $P_3 = P_1 + P_2$.

But what if $P_1/P_2 \neq N_1/LN_2$? It is reasonable to suppose that someone else in the apartment is consuming electricity (for instance, watching TV). Then

$$N_1 = K(P_1 + P_0), \qquad (1)$$
$$N_2 = K(P_2 + P_0), \qquad (2)$$
$$N_3 = K(P_3 + P_0). \qquad (3)$$

Solving the first and second equations simultaneously we obtain K and P_0:

$$K = \frac{N_1 - N_2}{P_1 - P_2} \qquad (4)$$

and

$$P_0 = \frac{P_1 N_2 - P_2 N_1}{N_1 - N_2}. \qquad (5)$$

Then

$$N_3 = \frac{P_1 N_2 - P_2 N_1 + (N_1 - N_2) P_3}{P_1 - P_2}$$

$$= \frac{P_1 N_1 - P_2 N_2}{P_1 - P_2}.$$

Note that if the two lamps have the same power rating, we will be dividing by zero in our equations. This tells us that we do not have enough information to obtain an answer in these situations.

P105

To find the longitudinal magnification we look at the similar triangles in figure 6 formed by the rays passing through the center of the lens:

$$K_1 = \frac{L_1'}{L_1} = \frac{b}{L}.$$

We now use the lens formula

$$\frac{1}{L} + \frac{1}{b} = \frac{1}{F},$$

$$\frac{1}{L - L_2} + \frac{1}{b + L_2'} = \frac{1}{F}, \qquad (1)$$

to find the ratio b/L and obtain

$$K_1 = \frac{F}{L - F}. \qquad (2)$$

To find the longitudinal magnification $K_2 = L_2'/L_2$ we let the car more forward one car length and again use the lens formula:

$$\frac{1}{L}\left(1 + \frac{L_2}{L}\right) + \frac{1}{b}\left(1 - \frac{L_2'}{b}\right) \cong \frac{1}{F}. \qquad (3)$$

Using the approximations $L_2 << L$ and $L_2' << b$, we obtain

$$\left(\frac{1}{L} + \frac{1}{b}\right) + \frac{L_2}{L^2} - \frac{L_2'}{b^2} \cong \frac{1}{F}. \qquad (4)$$

Using equation (1) to substitute for $1/F$, we have

$$K_2 = \frac{L_2'}{L_2} = \left(\frac{b}{L}\right)^2 = \left(\frac{F}{L - F}\right)^2.$$

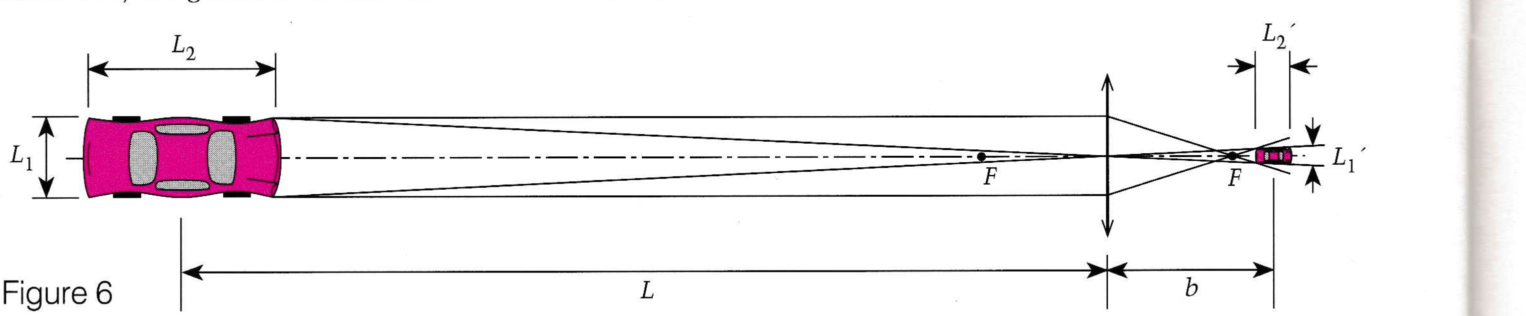


Figure 6

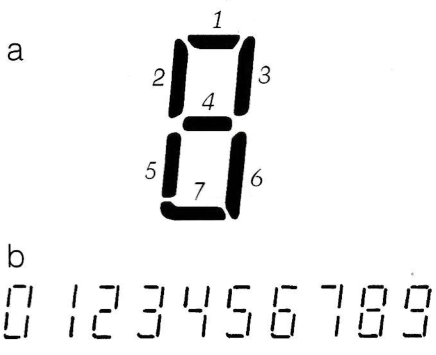


Figure 7

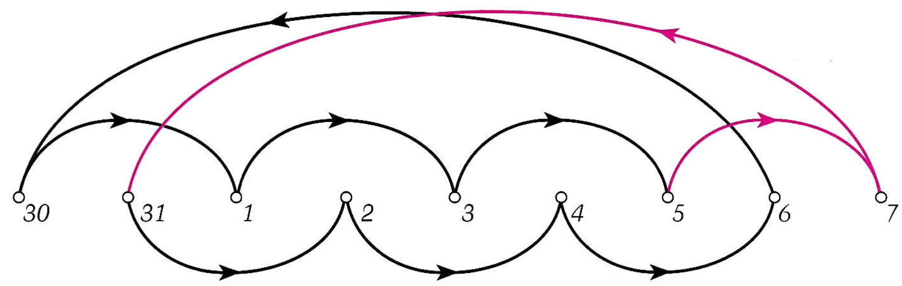


Figure 8

The ratio of transverse to longitudinal dimensions changes as follows:

$$\frac{L_2'}{L_1'} = \frac{K_2 L_2}{K_1 L_1} = \frac{L_2}{L_1} \frac{\left(\frac{F}{L-F}\right)^2}{\left(\frac{F}{L-F}\right)},$$

which reduces to

$$\frac{L_1'}{L_2'} = \frac{L_1}{L_2}\left(\frac{L-F}{F}\right)$$

$$= \frac{L_1}{L_2}\frac{200-1}{1} = 199\frac{L_1}{L_2}$$

—that is, the car in the image is "flattened" by factor of about 200!

Brainteasers

B101

Since $HE = J + O + H + N$, the numbers HE and $JOHN = (HE)^3$ yield equal remainders when divided by 9. Searching through all possible remainders, we find that the remainders of n and n^3 when divided by 9 are the same only for $n = 9k$, $n = 9k - 1$, or $n = 9k + 1$. It remains to check three possible values for HE—17, 18, and 19—since even 26^3 is a four-digit number. The answer is $HE = 17$, $JOHN = 4913$.

B102

Number the seven bars making a digit as shown in figure 7a. Comparing the calculator digit 8 with 0, 6, and 9 (fig. 7b), we see that the bars numbered 4, 3, 5 must work properly. Comparing 3 and 9, we find that the second bar must work, too. These four bars (2, 3, 4, 5) allow us to distinguish between any two digits except 4 and 9, which require either of the bars 1 and 7. So two bars were inoperative (6 and 1, or 6 and 7).

B103

The "delay" in the freezing of rivers at temperatures well below 0°C results from water motion. In a fast-flowing river the layers of water are continuously intermixed from the surface down to the bottom of the river. After being cooled to 0°C, the upper layers mix with the lower layers that are warmer, so their temperature increases above 0°C. Freezing begins only when all the water down to the very bottom is cooled to 0°C, and the deeper a river is, the longer it takes.

B104

The answer is 6 days. The numbered points in figure 8 represent days (point 1 corresponds to August 1, 1993). If, according to what he claims, Münchhausen shot fewer ducks on a certain day than on some other day, we draw an arrow from the point corresponding to the first day to the point of the second day. The figure shows that the baron couldn't repeat his claim for seven days (or more), because that would create a closed chain of arrows, which is impossible: moving along the arrows, we always get a greater number of ducks shot, so we can't return to the start. For six days, the red arrows in figure 8 should be erased, which destroys the circuit. It's easy to give an example satisfying Münchhausen's conditions: he could shoot one duck on July 31, two ducks on August 2, three ducks on August 4, four on August 6, five on July 30, and so on along the arrows, ending up with 8 ducks on August 5, while before July 30 he could shoot, say, 1994 ducks a day.

B105

The triangles ABN and CAM in figure 9 have the same area. A 120° clockwise rotation of the triangle ABC about its center takes it into itself and takes the triangle CAM into the triangle ABM_1 with the same area having point M_1 on BC—that is, into ABN. Since this rotation takes CM into AN, the angle between these lines is 120°, or 180° – 120° = 60° (which is the same).

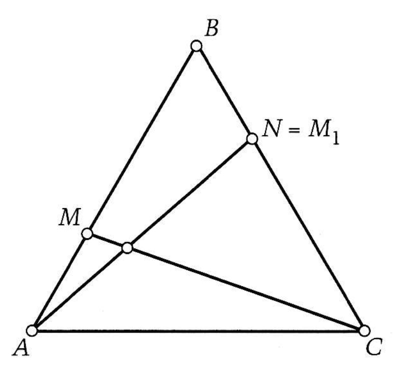


Figure 9

Pick's Theorem

(Hints for starred items)

3. Let M be the midpoint of a segment joining two nodes. Then the point symmetric to an arbitrary node with respect to point M is also a node.

4. Let the vertex A of a simple triangle ABC jump over vertex B onto a node A'. Complete parallelograms $ABCD$ and $BA'EC$. Then $ABCD$ is

obtained from $BA'EC$ by a translation that takes all nodes into nodes, so there cannot be any nodes inside and on the perimeter of $BA'EC$.

5. Draw grid lines through all vertices of the given simple triangle ABC and consider the rectangle bounded by four of these lines that contains the triangle. One of the triangle's vertices must be a vertex of the rectangle; let this vertex be A. Label the rectangle $ADEF$. If neither B nor C coincides with E—say, B lies on side DE, C on EF—then the perpendiculars to DE and EF drawn from B and C, respectively, meet in a node of the grid that lies in the triangle ABC (perhaps, on its border). This is impossible, since the triangle is simple. Otherwise, one of the vertices B and C—say, B—coincides with E. Then the angle ACB is obtuse or right.

7. The distances between nodes can only take values such that their squares are (positive) integers. Therefore, a decreasing sequence of these distances necessarily terminates.

10. Perform these operations in any order: any node lying in the interior or on the border of a triangle (one of the triangles of the partition already obtained) is joined to the vertices of this triangle.

20. This is obvious for a convex polygon. In a nonconvex polygon, choose its greatest angle. This angle is greater than 180°. Draw its bisector up to the border. Then slide the point thus obtained along the side until the segment joining it to the vertex of the chosen angle meets another vertex of the polygon.

24. Proceed by induction. Suppose that a triangulation exists for any n points. Take a set of $n + 1$ points and remove one of them. Draw the triangulation for the remaining points, then construct a new triangulation by connecting the "new" point to some of the others.

Kaleidoscope

1. See figure 10. Here $t_1 = 3$ s, $t_2 = 4.5$ s, $l = 1$ km, $v = 340$ m/s (the speed of sound).

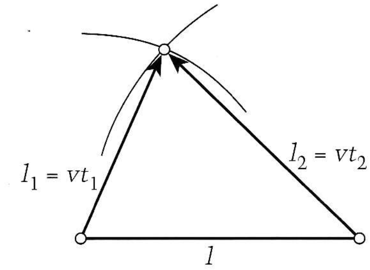


Figure 10

2. Usually the wind speed is greater at high places than near the ground, so the wavefronts become asymmetrical (fig. 11), which results in higher values for the speed of sound in the direction of the wind than in the opposite direction. Thus, the sound wave propagating against the wind will be deflected upward (the curve AB in figure 11), while the sound wave traveling with the wind will be lowered (the curve AC).

3. Air becomes unstable when it flows around wires or twigs, which results in vortices coming from the obstacles. These vortices create oscillations in the air pressure that we perceive as sounds.

4. In the string the waves are transverse; in the air they are longitudinal.

5. The greater the air pressure in the tire, the higher the pitch of the sound induced.

6. When a string is plucked by a hard fingernail or pick, higher harmonics are produced than when a soft finger is used. The higher harmonic oscillations add a ringing quality to the sound of a banjo.

7. Because the sound is reflected many times in a closed space, the room is more or less homogeneously filled with the energy of the sonic oscillations.

8. A tuning fork gripped in a vise rings longer because it emits a sound wave of lower intensity.

9. When a sound wave makes the transition from air to glass and then from glass to air, it is partially reflected, which results in a decrease in the amount of energy that penetrates the window.

10. There are small empty spaces between flakes of newly fallen snow that absorb sound (just as modern sound-absorbing panels do).

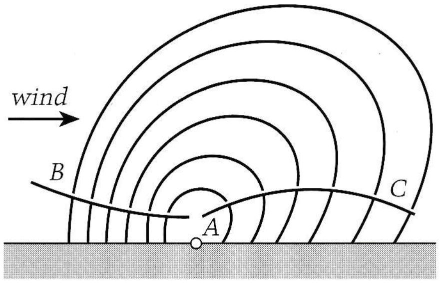


Figure 11

11. The air cavity in the kettle serves as a resonator for the sound.

12. A piece of modeling clay increases the mass of the windowpane, which leads to a decrease in the natural oscillation frequency of the glass. This destroys the resonance between the windowpane and the traffic noise.

13. A glass has certain resonance frequencies, and if one of them coincides with the frequency of the sound produced by the singer, the glass might break.

14. Either the tip of the whip strikes another part of the whip, producing a snapping sound, or it travels faster than the speed of sound and emits a shock wave.

15. A bullet shot from a rifle moves with a velocity greater than the speed of sound in air, so a shock wave is produced.

16. The shock wave takes the form of a conic surface moving at the same speed as the airplane.

17. The sound wave undergoes diffraction in the narrow opening, and when the sound passes through the opening it propagates throughout the room.

Microexperiment. The frequency can be determined from the wavelength, which is approximately equal to four times the length of the pipe.

Pythagorean Theorem

1. In figure 12, the counterclockwise rotation about the vertex B of the given right triangle ABC takes A into a vertex D of the external square on the hypotenuse, and line

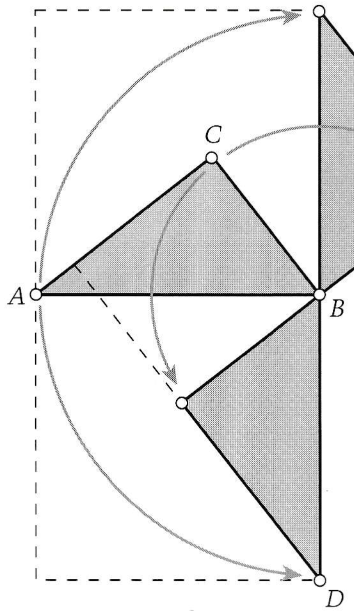


Figure 12

CA into the extension of the side of the internal square on BC parallel to BC, thus making the statement in question obvious. Similarly, the clockwise rotation proves the second statement—for the external square on BC and internal square on AB.

2. If $B(x, y, O)$ is the projection of $A(x, y, z)$ onto the xy-plane, then the triangle OBA, where O is the origin, is a right triangle. Therefore, $OB^2 = x^2 + y^2$, and hence $OA^2 = OB^2 + BA^2 = (x^2 + y^2) + z^2$.

3. Let $ABCD$ be the given tetrahedron with right angles at the vertex D, and let P be the projection of D onto ABC (fig. 13). Then the area of ABC equals the sum of the areas of ABP, BCP, and CAP. If α is the angle between the planes ABC and DBC, then area(BCP)/area(BCD) = cos α. The triangles BCP and BCD have a common side BC, so the ratio of their areas equals the ratio of their heights dropped on this side. But these heights have a common base H, since P is the projection of D, and form a right triangle PHD with angle $PHD = \alpha$ (see figure 13). Similarly, point D is the projection of A onto BCD, so area(BCD)/area(ABC) = cos α as well. It follows that

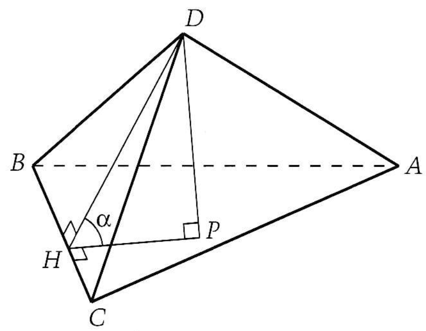


Figure 13

area(BCP) = area$(BCD)^2$/area(ABC).

Writing out similar expressions for the triangles CAP and ABP, and summing all the three of them, we get the formula in the box below.

As we noted above, the sum on the left side is equal to the area of ABC, and this completes the proof.

4. In figure 12 (in the article), the left side consists of four copies of the shaded triangle and the two squares on its legs. The right side consists of the same four copies and the square on the triangle's hypotenuse. Since both figures have an area of $(a + b)^2$, it follows that the sum of the areas of the squares on the left equals the area of the square on the right.

In figure 13, it's not hard to show that all the colored quadrilaterals are congruent to one another. Two of them are combined in two different ways (figures 13a and 13b) to make two hexagons, the first of which consists of two copies of a given triangle and two squares $(a^2 + b^2)$, and the second consists of the same two triangles and one square (c^2).

In figure 14, both the areas of the square on BC and of the right piece of the square on AB are equal to twice that of the red triangle; a similar relation is true for the blue triangle. So the sum of the areas of the two colored triangles is equal to half the area of the square on the hypotenuse, and at the same time, half the total area of the squares on the legs.

In figure 15, triangles ACD and AEC are similar (they have a common angle A and equal angles ACD and AEC), so $AD : AC = AC : AE$, or $b^2 = AC^2 = AD \cdot AE = (BA - BD)(BA + BE) = (c - a)(c + a) = c^2 - a^2$.

5. The Pythagorean Theorem shows quickly that the semicircle on the hypotenuse is equal in area to the sum of the semicircles on the legs.[1] Subtracting the areas of the white circular segments (see figure 16 in the article) from both sides of this equality, we get the desired relation.

6. The radii in question are proportional to the corresponding sides of their respective triangles. Therefore, $r^2 = r_1^2 + r_2^2$.

7. The area of the triangle can be expressed as $ab/2$ or $ch/2$, so $(a + b)^2 + h^2 = a^2 + b^2 + 2ab + h^2 = c^2 + 2ch + h^2 = (c + h)^2$. Now use the converse of the Pythagorean theorem.

[1]Hippocrates of Chios (c. 460 B.C.) used this diagram in his attempts to "square the circle." While he was able to draw a triangle equal in area to the lunes, which are bounded by arcs, this does not lead to a construction of a square equal in area to the entire circle. Much later, this construction was proved impossible. (D.E. Smith, *History of Mathematics*, Vol. II, New York: Dover, 1958, pp. 303–304)—*Ed.*

"BULLETIN BOARD" CONTINUED FROM PAGE 53

ground in scholarly exposition for high school students, giving them a feel for what is required in preparing and publishing academic research. A recent issue contained a paper on asymptotes (in English) by an American high school student and a paper on Edgar Allan Poe's "The Raven" (in Russian). Each issue contains English and Russian abstracts of the papers.

Prologue's editor in chief, Prof. Alexander Yurin, invites papers from high school students in the US and other countries. For more information, write to A. V. Yurin, *Prologue*, Box 124, St. Petersburg, 191065 Russia (e-mail: prologue@nit.spb.su). For a copy of *Prologue*'s editorial guidelines (in English), write to Tim Weber, *Quantum*, NSTA, 1840 Wilson Blvd., Arlington, VA 22201.

$$\boxed{\text{area}(BCP) + \text{area}(CAP) + \text{area}(ABP) = \frac{1}{\text{area}(ABC)}\left[\text{area}(BCD)^2 + \text{area}(CAD)^2 + \text{area}(ABD)^2\right]}$$

TOY STORE

A permutator's bag of tricks

Hints for solving a boundless class of puzzles

by Vladimir Dubrovsky

IN THE LAST INSTALLMENT of our Toy Store I described two puzzles with colored chips on a triangular grid (as you may recall, one of them was in fact a simplified version of a puzzle with "cannonball pyramids," each made of four balls, that are supposed to be rearranged by rolling). In both puzzles your task was to transform a given arrangement of chips into another given arrangement by moving them in "triads"—triangles of three chips touching each other—according to certain rules. Now it's time to explain how to do these and similar puzzles. However, only a few of the questions posed in the last issue will receive a complete and exhaustive answer below. A complete answer, I think, kills a puzzle—something must always be left for further improvement and investigation. Besides, I simply don't know the answers to a couple of questions about these puzzles. Write to *Quantum* if you have better luck with them.

Triads

I'll begin with elegant solutions to the three "specific" problems about the *triads* that were found by the author of this puzzle himself. (As you may recall, you have to move a triangle of three chips at at time—a "triad"—in attempting to transform the initial setup into the target arrangement.) These solutions are shown in figures 1–3; the arrows indicate what triad is moved and where it is placed in each particular move. Figure 1 also shows the original locations of the chips, so you can view the resulting displacement of the entire triangle of six chips: it's shifted "down" by a vector equal to its height. These solutions are probably the shortest possible (which is difficult to prove, though), and I don't know any regular way to discover them. My guess would rather be that the author of the puzzle started from the end: first he might have found several series of moves that restored the initial triangular shape, and then fit a beautiful coloration. However, using these operations as building blocks, we can construct new operations that generate many other rearrangements of chips. In fact, not just *many*, but *any!* And not only that: any rearrangement (or *permutation*) of chips can be composed of these "blocks" so as to bring the big triangle of chips back to its initial place. And this part of the job can be done in a quite regular way—that is, by using some technique applicable to virtually all "permutational puzzles."

How can we prove that every conceivable permutation in this puzzle is really possible? If we consider each of the six chips as distinct from the others, there are $6! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720$ of them, and it wouldn't be very

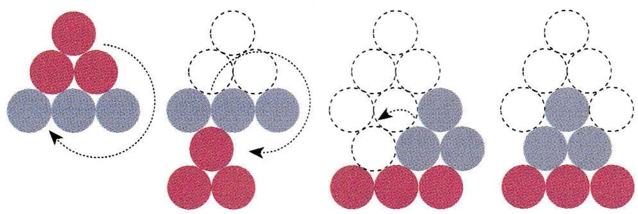


Figure 1

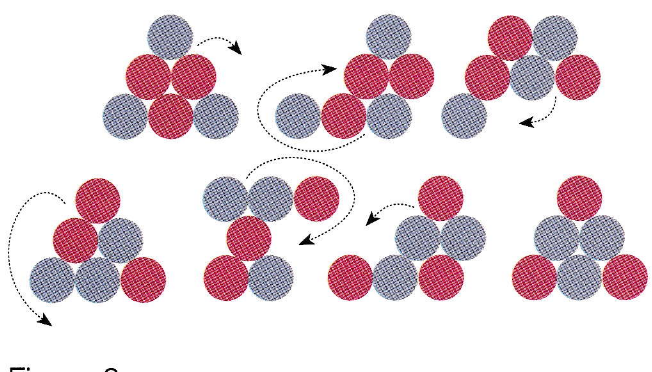


Figure 2

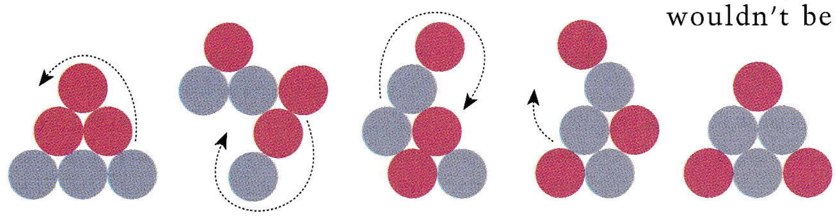


Figure 3

elegant to consider each one separately. But it will suffice to obtain a fairly small number of permutations that are known to generate any other. For instance, all *exchanges of pairs* of chips (*transpositions*) constitute such a set. Given a permutation that is to be performed, we take any chip not yet in its prescribed location and swap it with the chip in this location; then we repeat this operation until the required order is established.[1] A transposition of the chips in places a and b is denoted (a, b). In fact, we don't need *all* transpositions. The swaps (a, x) of a chip in a fixed location a with any other one will do: you can readily verify that the transposition (x, y) can be obtained as the result of three successive exchanges: (a, x) followed by (a, y) followed by (a, x) again, which is written as $(x, y) = (a, x) \circ (x, y) \circ (a, x)$. The sign $\circ$, which means "followed by," is often dropped. From place a the chip jumps to place x in the first swap, stays at x in the second swap, and jumps back to a in the third: $a \to x \to x \to a$. Similarly, $x \to a \to y \to y$ and $y \to y \to a \to x$. Or, if the locations of chips are numbered 1, 2, ..., 6, the "neighbor exchanges" $(i, i + 1)$, $i = 1, 2, \ldots, 5$, suffice to generate any transposition (you can easily prove this yourself or look up the proof in the solution to exercise 4 in "Some Things Never Change" in the September/October 1993 issue of *Quantum*). Of course, these remarks about representing permutations by means of transpositions are true for any finite number of chips or other objects and are useful in solving a lot of permutational puzzles.

So let's take a closer look at the operation in figure 1. Number the chips (fig. 4a). Then after this operation the distribution of numbers will look like figure 4b (check it!). The corresponding permutation can be depicted graphically by arrows joining the initial and final position of each chip (fig. 4c): we see that it consists of three disjoint pair exchanges (at this point we'll ignore the shift of

[1]See also math challenge M105 in this issue.

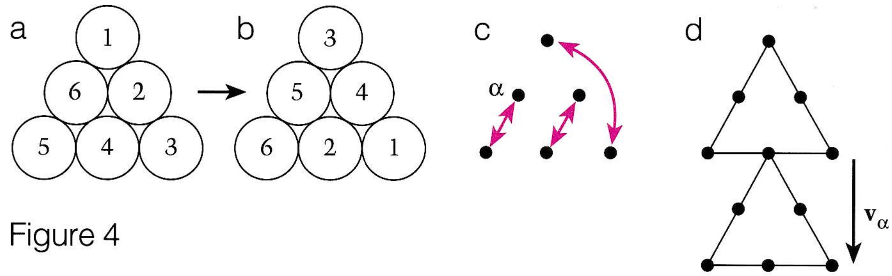


Figure 4

the triangle of chips as a whole, shown in figure 4d). This operation by itself, which will be denoted by α, is of little help, because all we can do with it alone is repeat it several times, and every second repetition restores the initial arrangement. But using the *symmetries* of the equilateral triangle, we can produce a family of "clones" of α (fig. 5). Two of them—β and γ—are the "rotations," so to speak, of α by ±120°, and the other three—$\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$—are the "reflections" of α, β, and γ about the corresponding symmetry axes of the triangle. Now we have enough material to undertake an item-by-item search through various combinations of our six "clones" in order to find simpler and easier-to-handle operations—those that don't move all the chips at one time. If you try to do this small investigation yourself, sooner or later—depending on how proficient you are with permutations—you'll come up with combinations like $\alpha\beta\gamma$ or $\bar{\alpha}\,\bar{\gamma}\,\alpha$ that leave two of the chips in their places and move the other four in cyclic order. We could get along even with these operations alone and complete the solution (fans of Rubik's cube will understand me—they can do wonders with four-cycles), but the operation in figure 2 provides us with a shortcut. The first three moves of this operation restore the triangular shape of the set of our six ships, but create a cyclic move of five of them: in our numeration, $1 \to 2 \to 3 \to 4 \to 6 \to 1$. This permutation, which will be labeled δ, is graphed in figure 6 (on the next page). Notice that none of the three moves that produce δ involves chip 5, so this chip—and the entire triangle, for that matter—remains in its initial place. So if we perform δ after the operation $\bar{\alpha}\,\bar{\gamma}\,\alpha$ (which moves four of the same five chips in the opposite direction—$6 \to 4 \to 3 \to 2 \to 6$) we get the desired result: the transposition $(1, 2)$ (fig. 6)! It remains to add to this operation its "symmetric clones" to obtain all neighbor exchanges $(1, 2)$, $(2, 3)$, $(3, 4)$, $(4, 5)$, $(5, 6)$, $(6, 1)$ and, therefore, any transposition—and any permutation at all.

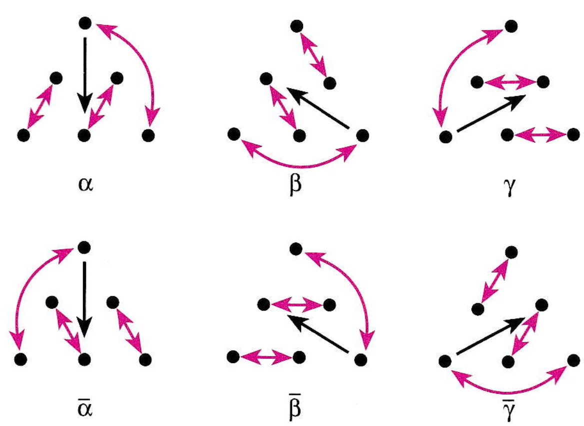


Figure 5

Now let's see whether we can perform the above permutations so as to eventually bring the big triangle to its starting position. As is shown in figures 4d and 5, the operations α and $\bar{\alpha}$ shift the triangle along the same vector $\mathbf{v}_\alpha$ equal and parallel to the altitude to the "bottom" side of the triangle. Similarly, the vector $\mathbf{v}_\beta$ of translations produced by β and $\bar{\beta}$ is equal and parallel to the altitude dropped to the left side of the triangle—that is, it's obtained from $\mathbf{v}_\alpha$ by a clockwise 120° rotation; and $\mathbf{v}_\gamma$, corresponding to γ and

$\bar{\gamma}$, is obtained from $\mathbf{v}_\alpha$ by a counterclockwise 120° rotation. The sum of the three vectors is zero, so the sum of any two of them is the negative of the third. Notice also that the inverse α^{-1} of the operation α creates *the same* permutation, but the *opposite* vector $-\mathbf{v}_\alpha$ (to perform α^{-1} just read figure 1 from the right to the left). These observations allow us to control the shift of the triangle without changing the permutation. For instance, the swap operation in figure 6, $\bar{\alpha}\,\bar{\gamma}\alpha\delta$, translates the triangle by $2\mathbf{v}_\alpha + \mathbf{v}_\gamma$ (δ yields a null shift), while $\bar{\alpha}\,\bar{\gamma}\alpha^{-1}\delta$ gives the same transposition and a shift by only $\mathbf{v}_\gamma$ ($= \mathbf{v}_\alpha + \mathbf{v}_\gamma - \mathbf{v}_\alpha$). We'd like to eliminate this shift, so let's try to create operations that only translate the triangle of chips without rearranging them. Such operations can be composed by using yet another trick: *iteration.* Any permutation, when repeated (iterated) a sufficient number of times, vanishes—all the permutated elements return to their initial places. (This is obvious for a cyclic permutation, and every permutation can be represented as several disjoint cycles—see the solution to problem M105 in this issue.) If a permutation is accompanied by a shift, then such an iteration yields a pure shift (by a multiple vector). For instance, $\gamma^2 = \gamma \circ \gamma$ results in the translation of all six chips by $2\mathbf{v}_\gamma$. Again, the permutation $\alpha \circ \beta$ is a combination of two disjoint three-cycles (fig. 7). Its shift is $\mathbf{v}_\alpha + \mathbf{v}_\beta = -\mathbf{v}_\gamma$. Therefore, $(\alpha\beta)^3$ puts every coin back where it was and shifts the whole triangle by $-3\mathbf{v}_\gamma$. So the product of these two iteration $\gamma^2(\alpha\beta)^3$ is the "pure" translation by $-\mathbf{v}_\gamma$. And, finally, adding this translation to the transposing operation created above, we get "pure" transposition of chips 1 and 2: $\tau = \bar{\alpha}\,\bar{\gamma}\alpha^{-1}\delta\gamma^2(\alpha\beta)^3$. Thus, *we can rearrange—*at least theoretically—*our six chips in any way, and in the same six places that they occupied initially.* There could be little doubt that the vast majority of solutions to particular triad problems based on the operation τ and its "clones" are

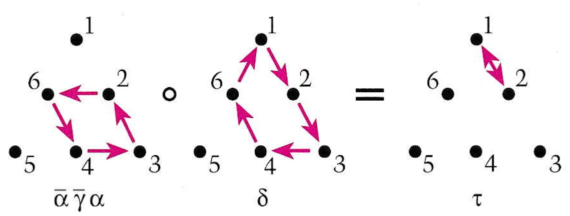


Figure 6

very uneconomical, but the search for a *short* solution is another, much more difficult problem.

One more question about the triads still remains open. We can get all *permutations*, but what about the *translations* of the triangle of chips? It follows from what was said above that it can be shifted by any vector of the form $n\mathbf{v}_\alpha + m\mathbf{v}_\beta$ with integer m and n: $\mathbf{v}_\gamma = -(\mathbf{v}_\alpha + \mathbf{v}_\beta)$ and adds nothing new. But these shifts constitute only some of all possible positions of the triangle on the grid over which our chips travel. It's not difficult to show that if there is at least *one* attainable position of the triangle other than those given by the above translations, then *all* possible positions are attainable. The question is *whether we can shift the triangle by a vector different from $n\mathbf{v}_\alpha + m\mathbf{v}_\beta$.* Let us know if you prove or disprove this statement.

Cannonball pyramids

In the last issue we also examined the "cannonball pyramids," a puzzle in which colored pyramids rolled around a board like the one in figure 8. To apply permutations to this puzzle, we first have to find sequences of moves (rolls) that rearrange the pyramids in such a way that the places in the box occupied by the whole set of pyramids before and after these moves are the same. One such sequence is illustrated in figure 8. If the pyramids are numbered as in this figure and rolled in the order 7, 2, 2, 1, 1, 4, 4, 6, 6, 10, 11, 11, 8, 8, 2, 2, 1, 4, 6, 10, 10, 11, 11, 8, 2, 7 about the hole marked with a star, they will form the following cycle of length seven: $1 \to 2 \to 8 \to 11 \to 10 \to 6 \to 4 \to 1$ (that is, pyramid 1 will arrive at the place of pyramid 2, which in turn moves to place 8, and so on). Using the

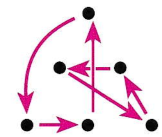


Figure 7

symmetry of the box, we can obtain a similar cycle in the right half of the box: $2 \to 3 \to 5 \to 9 \to 12 \to 11 \to 7 \to 2$. These two cycles—$l$ and r, respectively—are shown schematically in figure 9a. Such cycles appear in many permutational puzzles (Magic Rings, a puzzle consisting simply of two intersecting rings of colored beads, was popular several years ago), and in most cases they can be handled in the manner explained below.

Consider the operation $C = l^2r^2l^{-2}r^{-2}$, where l^2, r^2 are, of course, double iterations of l and r, and l^{-2} and r^{-2} are inverses of l^2 and r^2 (they move each element of their respective rings to the second place counting counterclockwise). The expression $ABA^{-1}B^{-1}$ is called the *commutator* of A and B, so C is the commutator of l^2 and r^2. (In the general case we would have to consider the commutator of l^n and r^m, where n and m are the numbers of segments of the left and right rings between their points of intersection, and use it in the way we'll use C.) You can verify that C exchanges two pairs of elements, as is seen in figure 9b. Next we modify C using "conjugation." The *conjugate* of operation A by B is the operation BAB^{-1}. Generally speaking, it creates the same effect as A, but in a different place. In our case, the permutation produced by the conjugate of C

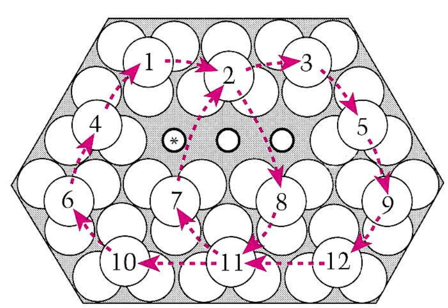
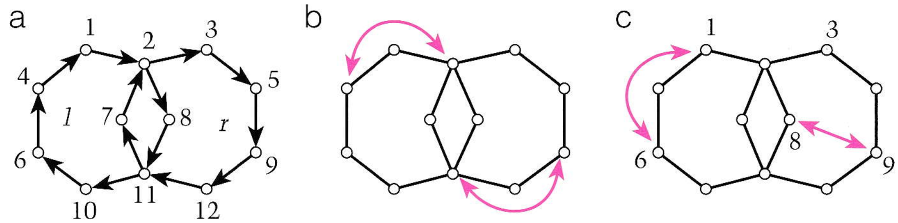


Figure 8

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION (Required by 39 U.S.C. 3685). (1) Title of publication: Quantum. A. Publication No.: 008544. (2) Date of filing: 10/1/93. (3) Frequency of issue: Bimonthly. A. No. of issues published annually, 6. B. Annual subscription price $34.00 (4) Location of known office of publication: Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010. (5) Location of the headquarters of general business offices of the publishers: 175 Fifth Avenue, New York, NY 10010. (6) Names and addresses of publisher, editor, and managing editor: Publisher: Bill G. Aldridge, National Science Teachers Association, 3140 North Washington Boulevard, Arlington, VA 22201, in cooperation with Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010; Editor(s): Larry D. Kirkpatrick / Mark Saul / Constantine Bogdanov / Vladimir Dubrovsky, National Science Teachers Association, 3140 North Washington Boulevard, Arlington, VA 22201; Managing Editor: Timothy Weber, National Science Teachers Association, 3140 North Washington Boulevard, Arlington, Va 22201. (7) Owner: National Science Teachers Association, 3140 North Washington Boulevard, Arlington VA 22201. (8) Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total of bonds, mortgages or other securities; none. (9) The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes: has not changed during the preceding 12 months. (10) Extent and nature of circulation. A. Total no. copies printed (net press run): Average no. copies each issue during the preceding 12 months, 11,200; no. copies single issue nearest to filing date, 12,200. B. Paid circulation: 1. Sales through dealers and carriers, street vendors, and counter sales: Average no. copies each issue during preceding 12 months, 1,009; no. copies single issue nearest to filing date, 1,459. 2. Mail subscriptions: Average no. copies each issue during preceding 12 months, 5,896; no. copies single issue nearest to filing date, 7,349. C. Total paid circulation: Average no. copies each issue during preceding 12 months, 6,904; no. copies single issue nearest to filing date, 8,808. D. Free distribution by mail, carrier, or other means. Samples, complimentary, and other free copies: Average no. copies each issue during preceding 12 months, 794; no. copies single issue nearest to filing date, 735. E. Total distribution: Average no. copies each issue during preceding 12 months, 7,698; no. copies single issue nearest to filing date, 9,543. F. Copies not distributed. 1. Office use, left-over, spoiled after printing: Average no. copies each issue during preceding 12 months, 3,098; no. copies single issue nearest to filing date, 2,657. 2. Return from news agents: Average no. copies each issue during preceding 12 months, 404; no copies single issue nearest to filing date, 0. G. Total. Average no. copies each issue during preceding 12 months, 11,200; no copies single issue nearest to filing date, 12,200. I certify that the statements made by me above are correct and complete.

Craig Van Dyck

Craig Van Dyck
Vice-President, Production



$C = l^2r^2l^{-2}r^{-2}$ $\qquad$ lCl^{-1}

Figure 9

by l is lCl^{-1} (fig. 9c). It can be written as two transpositions (1, 6)(8, 9), where only element 9 lies on the right ring. A simple further modification allows us to replace 9 here by any element on the right ring. For example, (1, 6)(8, 3) (fig. 9c) is produced by the conjugate of C by r^2l (that is, by $r^2lCl^{-1}r^{-2}$). Since we can exchange element 8 with any element of the right ring, we can arrange any permutation of the seven elements of this ring and element 8, though such a permutation may be accompanied by the transposition (1, 6). Symmetrically, we can arrange an arbitrary permutation of the elements of the left ring and element 7, possibly, accompanied by the transposition (3, 9). It's clear enough that using both kinds of permutations we can obtain any required arrangement except that two elements may be transposed. But with the given coloration of our cannonball pyramids, this doesn't create a problem: some of them are the same color and their transposition doesn't change the overall color design.

I've outlined the solution for the case when both in the initial and target states of the puzzle the pyramids are arranged according to figure 8—with the three empty spaces in the middle of the box. If this is not the case, we must find and remember a sequence of rolls R_1 that turns the initial arrangement into any standard one (with the "holes" in the center, as in figure 8), and another sequence, R_2, that turns a standard arrangement into the target one (ignoring the colors). Between R_1 and R_2 we can insert a permutation P of the standard arrangement of pyramids such that R_1PR_2 turns the initial color design into the target one.

I'd like to finish this discussion of the cannonball pyramids with an additional question: *if all the pyramids were colored differently, what permutations could be obtained from a given one*? Now we can't ignore the extra transposition that may occur when we use the method described above. This method allows us to produce any *even* permutation—that is, representable as an even number of successive transpositions—and nothing else. (The article "Some Things Never Change" cited above will help you understand why this is true.) But this doesn't mean that other (odd) permutations can't be obtained in a different way.

So, the next time you come across a new permutational puzzle, don't forget this magic unscrambling kit: *commutators*, *conjugates*, *iterations*, and *symmetries*. In my experience, it has proved to be of universal use.

Corrections

Vol. 4, No. 1:

p. 37, exercise 14, l.1: *for* 1993 *read* 1994; l. 6: *for* 1993, 1992, … *read* 1994, 1993, … .

p. 46, col. 3, ¶6, l. 2: *for* $m > n = 3$ *read* $m \geq n \geq 3$.

p. 60, col. 3, exercise 2, ll. 10–11: *for* method of problem 3 *read* method of problem 2.

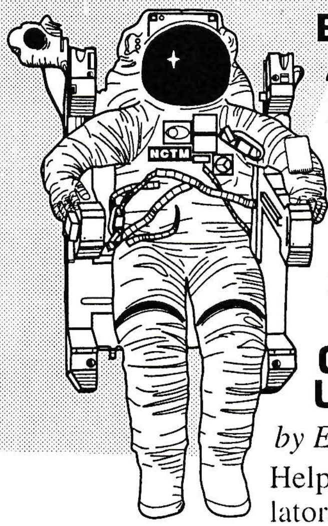
p. 62, col. 1, exercise 14, l. 10: *for* 1993 *read* 1994; *for* 1992 *read* 1993.

p. 64, col. 1, ¶2, ll. 10–15: *for* Pairs of these classes are marked *read* Two classes forming one such pair are marked; *for* belongs permanently to one of these eight classes *read* can have only eight orientations—one of each class.

Cover 3, col. 1, ¶2, l. 19: *for* figure 5 *read* figure 5b.

Cover 3, figure 5b: the circle above the one marked *c* should be considered empty; all the other nine circles have chips in them.

Explore the Universe of Math
with NCTM Books



Exploratory Problems in Mathematics

by Frederick W. Stevenson

Presents open-ended problems that introduce students at all levels to the creative side of mathematics. Offers you specific problems that are strictly for intellectual enjoyment and excitement yet have ample room for proper mathematical exploration. Guidelines are given on how to begin, but the solutions are up to you. 1992, 169 pp., ISBN 0-87353-338-0, #495B2, $16.00*.

Calculator Logic Systems and Mathematical Understandings

by Enid R. Burrows

Helps junior and senior high school students explore the ways different calculators can be used effectively and how to construct appropriate algorithms for calculator computation. Includes many exercises as well as a review and a clarification of binary operations and functions. 1990, 133 pp., ISBN 0-87353-295-3, $436B2, $22.50*.

Fractals for the Classroom

by Heinz-Otto Peitgen et al.

Part One: Introduction to Fractals and Chaos

Part Two: Complex Systems and the Mandelbrot Set

An exciting, dynamic view of chaos theory and fractal geometry and how they relate to each other, to other aspects of mathematics, and to natural phenomena. Both contain beautiful color photographs. Part 1, 1992, 450 pp., ISBN 0-387-97041-X, #440B2, list price $29.00*. Part 2, 1992, 500 pp., ISBN 0-387-97722-8, #507B2, list price $29.00*.

Fractals for the Classroom: Strategic Activities, Vols. 1 & 2

by Hartmut Jürgens et al.

These hands-on activities will show you the underlying mathematical principles, characteristics, and beauty of fractals. Vol. 1 includes nine color slides, 1991, 128 pp., ISBN 0-387-97346-X, #482B2, list price $19.95*. Vol. 2, 1992, 187 pp., ISBN 0-387-97554-3, #483B2, list price $19.95*.

SPECIAL PACKAGE PRICE for all four fractals books . . .
NOW ONLY $80* ($97.90 if bought separately), ISBN 0-87353-370-4, #589B2.

NCTM has more materials for math enthusiasts like you. Ask for a free catalog today!

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS
1906 Association Drive, Reston, VA 22091-1593 • Tel. (703) 620-9840, fax (703) 476-2970. For orders only, call (800) 235-7566

*20% discount to NCTM individual members