QUANTUM

JULY/AUGUST 2000

US $7.50/CAN $10.50

Springer

NSTA

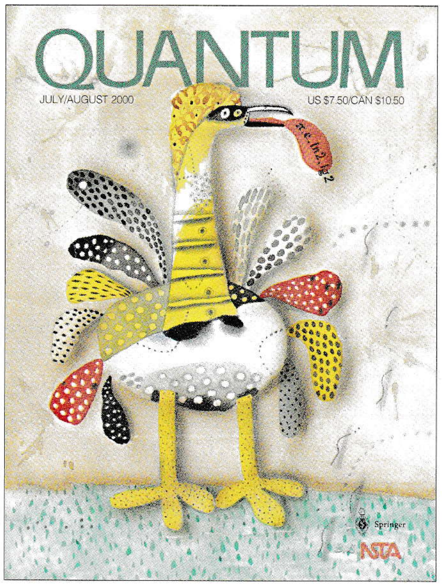
Oil on wood, 15 7/8 × 18 3/4, Gift of Edgar William and Bernice Chrysler Garbisch, © 2000 Board of Trustees, National Gallery of Art, Washington, D.C.

The Mounted Acrobats (1825 or after) American

THE THREE-RING CIRCUS HAS DELIGHTED audiences for over a century, but not everyone enjoys watching animals perform under the Big Top. There are two diametrically opposed points of view as to whether or not the use of animals in the circus is a cruel practice. Coincidentally, there are also two diametrically opposed points on the ring that the above horse is circling where a continuous function takes equal values. We're sure you'll never look at certain circus acts the same way again after you have explored this theorem in greater detail. Turn to page 16 to learn more about the "Borsuk—Ulam theorem."

QUANTUM

JULY/AUGUST 2000 VOLUME 10, NUMBER 6



Cover art by Vasily Vlasov

"What came first, the chicken or the egg?" is a question that has transcended time. But our fowl-feathered friend on the cover seems more concerned with certain transcendental numbers than debating his origin. If you'd like to find out more about the topic that has just passed this chicken's lips, turn to Algebraic and Transcendental Numbers on page 22.

Indexed in *Magazine Article Summaries, Academic Abstracts, Academic Search, Vocational Search, MasterFILE*, and *General Science Source*. Available in microform, electronic, or paper format from University Microfilms International.

FEATURES

DEPARTMENTS

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 11 (6 issues) will be published in 2000–2001. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau (Moscow, Russia), as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000; telephone: (703) 243-7100; electronic mail: quantum@nsta.org. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York NY 10010-7858.

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

Subscription Information:
North America: Student rate: $18; Personal rate (nonstudent): $25. This rate is available to individual subscribers for personal use only from Springer-Verlag New York, Inc., when paid by personal check or charge. Subscriptions are entered with prepayment only. Institutional rate: $56. Single Issue Price: $7.50. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Mail order and payment to: Springer-Verlag New York, Inc., Journal Fulfillment Services Department, PO Box 2485, Secaucus, NJ 07094-2485, USA. Telephone: 1(800) SPRINGER; fax: (201) 348-4505; e-mail: custserv@springer-ny.com.

Outside North America: Personal rate: Please contact Springer-Verlag Berlin at subscriptions@springer.de. Institutional rate is US$57; airmail delivery is US$18 additional (all rates calculated in DM at the exchange rate current at the time of purchase). SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, Germany.

Advertising:
Representatives: (Washington) Paul Kuntzler (703) 243-7100; (New York) Joe Antino (212) 460-1575; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, Germany, telephone 49 (0) 30-827 87-0, telex 185 411.

Printed on acid-free paper.

Visit us on the internet at www.nsta.org/quantum/.

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

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

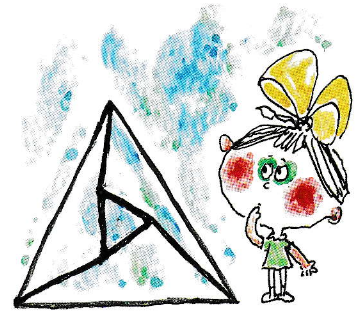
The mission of the National Science Teachers Association is to promote excellence and innovation in science teaching and learning for all.

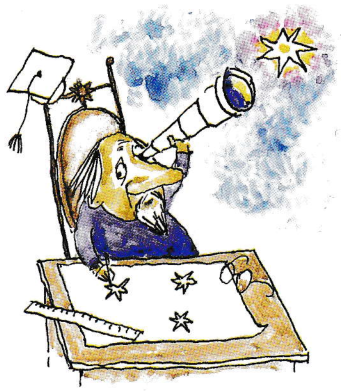
BRAINTEASERS

Just for the fun of it!

B296

Picture perfect. A teacher put a pentahedron (a polyhedron with five faces) on the table. Two of the faces were triangles, and the other three were quadrilaterals. Jean drew a top view of this polyhedron (see the figure). Is this drawing correct?



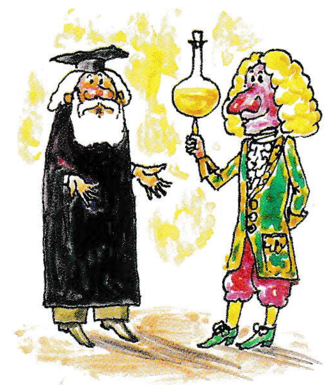


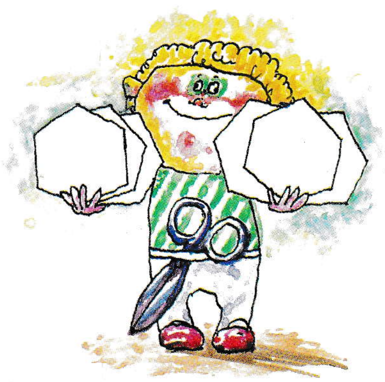
B297

Connect the dots. Is it possible to mark six points in the plane, and connect some pairs of them with nonintersecting segments, so that every point is connected to four others?

B298

Polls apart. One hundred persons—chemists and alchemists—attended a conference. They were asked a question: "Which group is more numerous here, chemists or alchemists (not including yourself)?" The first 50 persons said that alchemists were more numerous. Now, it's known that alchemists always lie and chemists always tell the truth. How many chemists attended the conference?





B299

Vexing hexagons. Six congruent regular hexagons are given. Cut three of them into two parts such that the nine parts obtained (three hexagons and six "halves") can be used to compose an equilateral triangle.

B300

Wrong-way mercury. A thermometer is quickly removed from molten tin. At first the mercury goes up, not down! Explain this strange phenomenon.

Art by Pavel Chernusky

ANSWERS, HINTS & SOLUTIONS ON PAGE 54

Self-similar mosaics

When the whole is the sum of its parts

by N. Dolbilin

MOSAICS MADE OF PIECES that are similar to the whole are called self-similar. One of the first famous examples of a self-similar mosaic was constructed by the English physicist Roger Penrose. After quasicrystals were discovered in 1984, Penrose's patterns became a generally recognized model used for analyzing their geometrical properties. The mathematicians John Conway and William Thurston found new and unexpected relations of these mosaics to other fields of mathematics. Self-similarity plays an important role in modern fields of mathematics such as dynamic systems, fractals, and quasicrystals.

Self-similar figures

It is well known that the medial lines (the lines connecting the midpoints) of a triangle split the triangle into four equal triangles (figure 1).

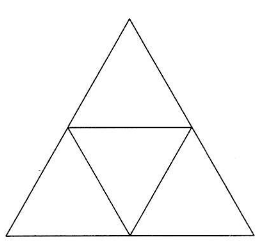


Figure 1

Each of these small triangles is similar to the original one. In this sense, the triangle is a self-similar figure.

A figure F is *self-similar* if it can be cut into several figures, $F_1, F_2, \ldots, F_m$, each of which is similar to the initial figure. Since every constituent F_i is similar to F, there exists a similarity transformation h_i that takes F onto $F_i : h_i(F) = F_i$. The coefficients of these similarity transformations are not necessarily equal, but all are less than 1.

Similarity transformations

Recall that a *similarity transformation* is a transformation, h, of a plane (or space) such that the distance $d(x, y)$ between any two points x and y changes by the same number k:

$$d(x, y) = kd(h(x), h(y)), k > 0.$$

In the case $k = 1$, the similarity transformation is a *rigid motion*. Some rigid motions—for example, *translation* by a nonzero vector—move any point to a different point. Others—for example, a *rotation* g of the plane about a point O by a certain angle—leave certain points where they were. Such a point (for which $g(O) = O$) is called a *fixed* point of the transformation g. Thus, a rotation about a point has a *single fixed point*. Motions of a third type, reflections in a line l, have infinitely many fixed points: the points of the line l itself. Thus, some rigid motions have no fixed points, others have a single fixed point, and still others have infinitely many.

This situation changes if we consider a similarity transformation h that is not a rigid motion. This happens if the similarity coefficient, k, is not equal to 1. In this case, the transformation h has a unique fixed point.[1]

By virtue of this remarkable fixed point theorem, any similarity transformation of the plane with a coefficient $k \neq 1$ has a unique fixed point, say, a point O. For this reason, any similarity transformation can be represented as a dilation[2] (centered at O with some coefficient k), fol-

[1]This remarkable theorem holds not just for similarity transformation, but for a more general type of mapping: a *contraction mapping*. This is a mapping $f(x)$ for which $d(x, y) \leq kd(h(x), h(y))$ for a certain $0 < k < 1$.

[2]A dilation (or homothecy, or homothety) is a transformation with a center O and a coefficient k, which takes any point P onto a point P' such that O, P, and P' are collinear, and $OP' = k \cdot OP$. If k is positive, O is outside line segment PP'. If k is negative, O is outside this segment.

lowed by a rotation about O through a certain angle, and perhaps reflection through a line that passes through point O. In particular, such a transformation can be a simple dilation without any rotational component.

Examples of self-similar figures

The triangle. We have seen that any triangle can be decomposed into four similar triangles by its medial lines. What similarity transformations map the original triangle ABC into these similar parts? Three of these transformations are dilations centered at the vertices of the triangle, with coefficient 1/2. The fourth is the dilation centered at the intersection of the triangle's medians, with coefficient –1/2. This last transformation can also be described as a dilation[3] centered at the intersection of the medians with positive coefficient 1/2, followed by a rotation about the same point by 180°.

The right triangle. A right triangle can be decomposed into two similar triangles (figure 2). Let us

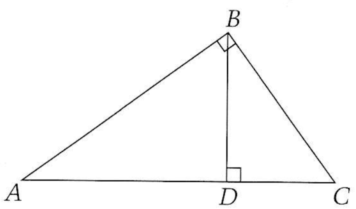


Figure 2

map $\triangle ABC$ onto $\triangle ABD$. Consider the similarity transformation g_1 that is the product of the dilation h_1 centered at A and coefficient $k_1 = AB/AC$, followed by a reflection about the bisector of $\angle BAC$. It is not hard to see that

$$g_1(\triangle ABC) = \triangle ABD.$$

In the same way, we can see that the transformation g_2 that maps $\triangle ABC$ into $\triangle BDC$ is the product of the dilation centered at C with coefficient $k_2 = BC/AC$ and a reflection about the bisector of $\angle ACB$.

Note one difference between this example and the preceding one. In the case of the right triangle, transformation coefficients differ in absolute value.

Domino. The *domino* figure consists of two equal squares. It can be easily cut into four pieces similar to itself (figure 3).

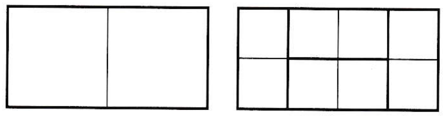


Figure 3

Problem 1. Find the fixed point for each of the corresponding similarity transformations.

Problem 2. Find another way to cut the domino into four pieces, each similar to the original.

Chair. The *chair* figure (tromino) consists of three equal squares. It can be cut into four similar copies of itself: F_1, F_2, F_3, and F_4 (figure 4). Let

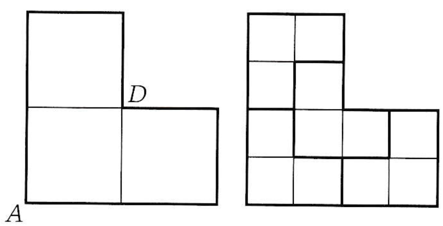


Figure 4

h_1, h_2, h_3, and h_4 be the similarity transformations that transfer the "big" chair into the corresponding parts. They have a coefficient of 1/2. The dilations h_1 and h_2 have their centers at the points A and D, respectively.

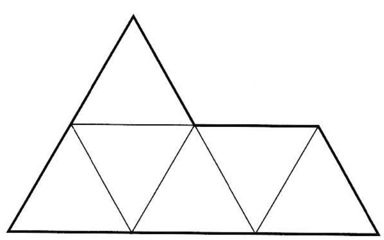


Figure 5

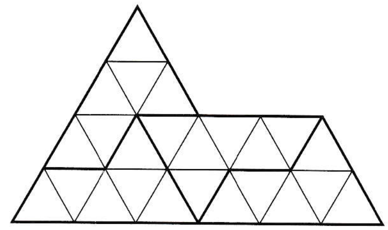


Figure 6

Sphinx. The *sphinx* figure (hexomino) consists of six equilateral triangles (figure 5) and can be decomposed into four similar copies (figure 6).

Problem 3. Find the fixed point for the similarity transformations that take the original sphinx onto each smaller figure.

Self-similar figures and mosaics

Consider a "good" self-similar figure F; that is, one that does not contain holes. Some examples are a triangle, parallelogram, or any other polygon that can be decomposed into equal polygons, each similar to the original one. Then an entire plane can be tiled by copies of F, without gaps or overlaps. A covering of the entire plane with non-overlapping tiles is called a *mosaic* or *tessellation*. If all the tiles of the mosaic are congruent, it is called *monodical*.

How can we obtain a monodical mosaic from a self-similar figure F? There are several ways. One way seems to be the simplest, but in fact conceals a tricky point. Consider, for example, a chair F of a certain size (figure 7a) and decompose it into four small chairs, as shown in figure

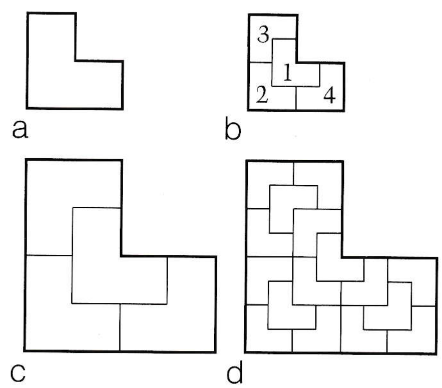
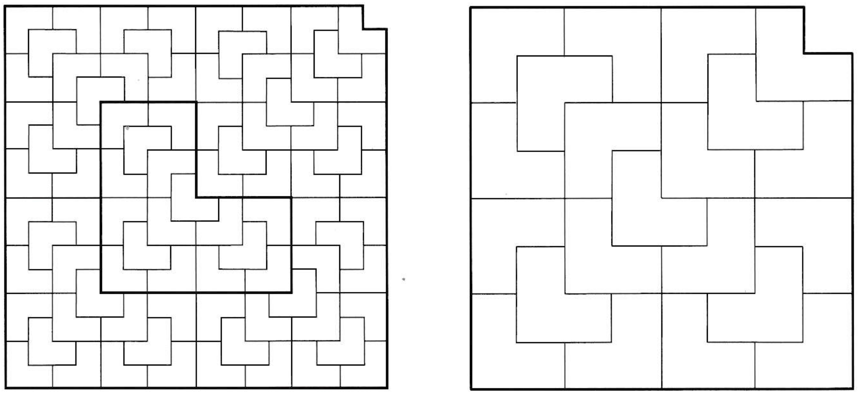


Figure 7

7b. Now double this picture to make each small chair equal to the initial one (figure 7c). Then cut each of the four chairs into four smaller chairs (figure 7d) and double the resulting picture once more. Repeating this procedure, we obtain an infinitely expanding chair-shaped domain consisting of congruent chairs.

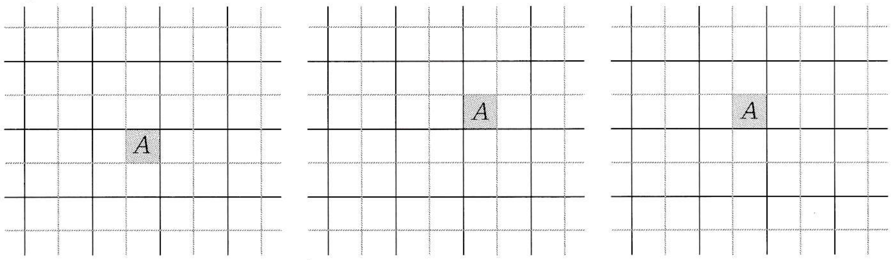
This procedure has a peculiar name: *deflation–inflation*. *Inflation* corresponds to the increase in the

[3]This observation depends on the fact that the centroid (intersection of the medians) divides each median in the ratio 2 : 1.



a b

Figure 8



a b c

Figure 9

size of the tiles; *deflation* corresponds to the decomposition of large tiles into smaller ones. Taken to the limit, this process (which is called the *d-process*) results in a monodical mosaic. However, the transition to the limit is the tricky point mentioned above. As a matter of fact, what is the limit here? Apparently, the d-process results in a sequence of domains that increase in size and are covered with congruent tiles. However, this sequence is not a sequence of fragments that grows as we add new tiles to the mosaic constructed in previous steps. Nevertheless, it can be proved that the plane can always be tiled by self-similar polygons.

A mosaic constructed from a self-similar figure F is called *self-similar* if

- the tiles of this mosaic (let's call them first-level tiles) can be joined into bigger tiles (second-level tiles) that are similar to the first-level tiles and such that the second-level tiles also constitute a mosaic (figure 8a);
- this "sequential integration" can be performed for any level (figure 8b).

For this reason, self-similar mosaics are also called hierarchical, referring to the hierarchy that exists between the tiles of preceding and succeeding levels. This hierarchy can be *strong* or *weak*. With a strong hierarchy, the mosaic of each succeeding level can be constructed from the tiles of the preceding level in a unique way. With a weak hierarchy, the mosaic tiles can be joined to obtain the tiles of the succeeding level in several different ways.

The strong or weak nature of the hierarchy is mainly determined by the figure F itself. For example, a square produces a weak hierarchy. Indeed, a mosaic composed of squares (figure 9a) can be integrated into its second-level mosaic in different ways. That is, a given square A can occur among the second-level tiles in different positions (figures 9b and 9c).

However, the chair, sphinx, and domino figures produce a strong hierarchy. Consider a chair-figured tile in its corresponding mosaic. Together with three other chairs, it constitutes a second-level chair, and each first-level tile uniquely determines three other complimentary tiles.

Thus, the second-level mosaic is uniquely determined. Since the second-level mosaic possesses the same properties as the first, the self-similar decomposition of the plane into chairs is strongly hierarchical.

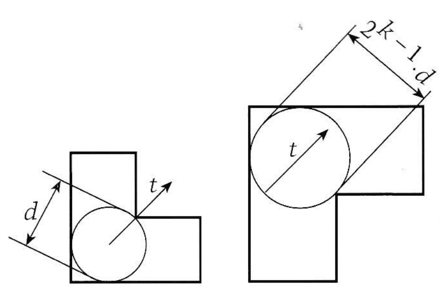
Properties of strongly hierarchical mosaics

Strongly hierarchical mosaics have a number of peculiar properties that differ from the properties of weakly hierarchical mosaics.

Aperiodicity. A mosaic for which at least one translation exists such that it maps the mosaic onto itself is called *periodic*. As we can see in the example of the square mosaic, weakly hierarchical mosaics can be periodic. Any tile of the square mosaic can be translated into any other tile together with the entire mosaic.

The most important property of strongly hierarchical mosaics is that they are *aperiodic*. Let us prove that such a mosaic cannot be periodic. Assume that a translation t exists which maps the entire mosaic onto itself. Then t moves a tile F_1 onto some other tile F_2. Because the next mosaic consisting of second-level tiles is uniquely determined, the translation t also maps the second-level mosaic onto itself. Again, by virtue of the fact that the second-level tiles are uniquely integrated into the third-level tiles, the translation t that maps the second-level mosaic onto itself also transfers the third-level mosaic onto itself, and in general it maps the mosaic of any kth level onto itself, for any k.

The tiles of the kth level are 2^{k-1} times bigger than the tiles of the first level. Thus, if the tiles of the first



a b

Figure 10

level contain a circle of diameter d (figure 10a), the tiles of the kth level contain a circle of diameter $2^{k-1} \cdot d$. For a large enough k, $2^{k-1} \cdot d$ will be greater than the length of the translation vector t (figure 10b). This means that the translation by t maps a circle of diameter $2^{k-1} \cdot d$ onto a circle that overlaps the original one. On the other hand, these circles must belong to different tiles of the kth level and thus cannot intersect. Why do they belong to different tiles? No bounded figure can be mapped into itself by translation; since different tiles of the kth-level mosaic do not overlap, the circles inside those tiles do not overlap either. Thus, we have arrived at a contradiction. So, we know that all strongly hierarchical mosaics are aperiodic.

Periodic mosaics provide a good model for crystals, while strongly hierarchical mosaics play an important role in the study of quasicrystals. In contrast to crystals, these structures are aperiodic. In particular, the famous Penrose patterns (figure 11), the best-known model of quasicrystals, are a direct generalization of strongly hierarchical mosaics.

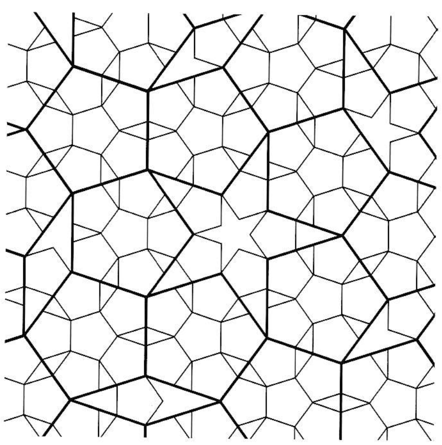


Figure 11

All Alike. Let's consider another peculiarity of strongly hierarchical mosaics: The mosaic of each succeeding level can be uniquely reconstructed from the previous level. Therefore, it may seem that a strongly hierarchical mosaic—for example, the chair—is determined uniquely. However, there are many self-similar mosaics composed of chair-shaped tiles. Moreover, there are even innumerably many of them. To be more precise, two (infinite) plane mosaics are considered identical if one of them can be matched with the other by a rigid motion of the plane. Otherwise, these mosaics are considered different.

Let us explain why an uncountable set of self-similar mosaics can be obtained from the chairs. Decompose a chair into four smaller chairs and assign a number 1, 2, 3, or 4 to each of them, as shown in figure 7b. Let a chair occur in a bigger chair under number a_1 at the first stage of the d-process. At the second stage, this chair occurs under number a_2 in the chair of the second level, and so on. Thus, the mosaic that grows from the given chair determines a sequence consisting of the numbers 1, 2, 3, and 4. The same mosaic can grow from any of its other tiles, but the sequence produced will be different. Since the mosaic is composed of a countable set of tiles, and there exist innumerable different sequences, there exist innumerably many different self-similar mosaics consisting of chairs.

Because there are infinitely many strongly hierarchical mosaics that can be constructed from a given self-similar tile, these mosaics cannot be enumerated by natural numbers as the elements of a sequence can be. However, they can be enumerated by real numbers.

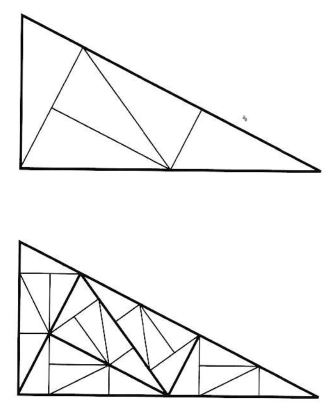
Assume that all mosaics from the uncountable family "Chair" have already obtained names in the form of real numbers. Suppose we want to make a family album of these mosaics. Every mosaic is infinite, and it cannot be placed on a finite photo. Thus, the portrait of each mosaic inevitably captures only a small part of it. Therefore, a given mosaic can have infinitely many portraits. Now assume that a photographer has chosen a portrait of each mosaic for his album, but he did not label the photos in time, and instead wrote them in afterward at random. However extraordinary it may seem, he did not get any wrong. The point is that *any finite fragment in a mosaic of the "Chair" family also occurs in any other mosaic of this family*, and it even occurs infinitely many times in each mosaic. Thus, although all strongly hierarchical mosaics of the same family differ globally, they look identical locally.

Conway's mosaics

Recall that a self-similar mosaic can be periodic. Strongly hierarchical mosaics are aperiodic. However, despite their aperiodicity, the tiles must have a finite number of different positions up to translation. For example, with dominos, every tile belongs to one of the two classes of parallel tiles. With chairs, there are four classes of tiles.

Problem 4. How many classes of parallel tiles are in the sphinx hierarchical mosaic?

It would be interesting to discover whether a mosaic consisting of identical tiles exists such that these tiles have infinitely many different orientations. In 1992 Conway suggested a self-similar, strongly hierarchical mosaic with tiles of equal triangles having infinitely many different orientations. The underlying idea is very simple. Consider a right triangle with legs equal to 1 and 2 and a hypotenuse of $\sqrt{5}$. This triangle decomposes into five equal and self-similar triangles (figures 12a, 12b). The acute angle of the triangle is $\alpha = \arctan(1/2)$. This decomposition induces self-similar mosaics called Conway mosaics. It can easily be seen that in a Conway mosaic, for any integer m and for any



a

b

Figure 12

tile, there exists another tile that is oriented at the angle $m\alpha$ with respect to the initial tile. Since the angle α is incommensurable with 2π (see problem 5), any two triangles oriented at the angle $m\alpha$ with respect to each other cannot be parallel. Therefore, triangle tiles in a Conway mosaic occur in infinitely many different orientations.

It turns out that all Conway mosaics are self-similar mosaics with a strong hierarchy. Therefore, there are infinitely many of them, and all of them include triangles with infinitely many different orientations.

Problem 5. Prove that the angle arctan (1/2) is incommensurable with π—that is, the equation n arctan $(1/2) = m\pi$ has no solution in natural numbers (m, n).

Let's rephrase the statement of the main property of Conway's mosaic: For any possible orientation of Conway's original triangle, Δ, and any small positive number ε, there exists a tile Δ_ε that is "almost parallel" to Δ, up to an accuracy of ε. That is, the angles between the corresponding sides of the triangles Δ and Δ_ε are less than ε. In other words, the tiles' orientations are distributed everywhere densely in the set of all possible orientations.

Later on, Conway (together with Charles Radine) constructed a mosaic in space consisting of equal prisms whose orientations are distributed everywhere densely in the set of all possible orientations. The orientation of a polyhedron can be determined using a triple of, say, mutually perpendicular vectors that are rigidly attached to this polyhedron. The property of being "everywhere dense" means that for any orientation of the polyhedron P and for any arbitrarily small $\varepsilon > 0$, there exists a three-dimensional tile P_ε such that the angles between the vectors of its triple and the vectors of the P-triple are less than ε.

The right triangular prism of height 2, whose base is the right triangle with legs 2 and $2\sqrt{3}$ and hypotenuse 4 at its base (figure 13a) serves as the initial object. The base triangle is decomposed into four similar triangles, as shown in figure 13b. Therefore, the initial prism can be decomposed into eight similar prisms, as shown in figure 13c. The two prisms, A and B, on the upper floor constitute a regular triangular prism. Therefore, this pair (taken as a whole) can be rotated by 120° and then returned to its place (figure 13d). The prisms C and D at the lower floor constitute a right-angled parallelepiped with a square face of 1×1. This pair can be turned by 90° and returned to its place (fig. 13d).

As a result, we obtain the construction of a Conway–Radine mosaic. Note that in this construction, the prisms that are turned at angles of 120° and 90° about mutually perpendicular axes with respect to each other are identical. Suppose we construct a self-similar mosaic consisting of the Conway–Radine prisms, using the inflation–deflation process. Then this mosaic will contain, together with each prism P, all prisms that are turned with respect to P by various angles obtained by all possible combinations of the form

$$g_1^{m_1} \cdot g_2^{m_2} \cdot g_1^{m_3} \cdot g_2^{m_4} \cdot \ldots \cdot g_1^{m_n} \cdot g_2^{m_{n+1}},$$

where g_1 and g_2 are turns by 120° and 90° about mutually perpendicular axes.

We can use the three mutually perpendicular edges that meet at the vertex of the right angle at the base of the prism as the orientation triple. It is relatively simple to prove (using the fact that the rotational axes of g_1 and g_2 are mutually perpendicular) that the set of different orientations is infinite. It is more difficult to establish that the set of orientations is everywhere dense. The proof requires the use of group theory.

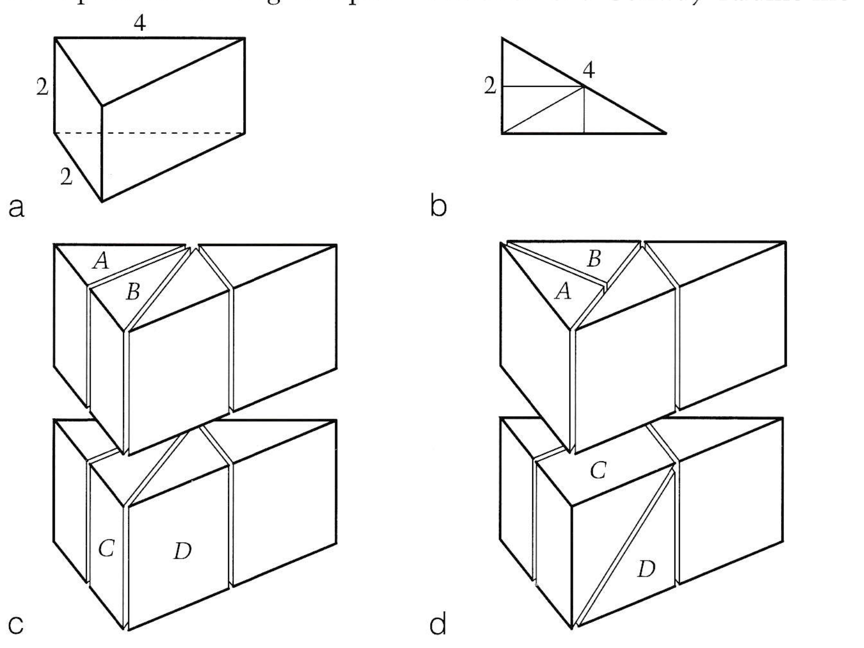


Figure 13

"Chaos" and self-similar mosaics

The game "Chaos" provides an unexpected and simple method of obtaining self-similar mosaics using the computer.

Let's look at the rules of this game. We choose an initial set of transformations. For the present case, we take those that map the Conway triangle onto its five constituent triangles. We denote these transformations by h_1, h_2, h_3, h_4, and h_5 (figure 12b). Let a random number generator produce numbers 1, 2, 3, 4, and 5. Mark an arbitrary initial point x_0 on the plane.

Step 1. Assume that the random number generator produced the number 2. Set $x_1 = h_2(x_0)$.

Step 2. Assume that the random number generator produced the number 1. Set $x_2 = h_1(x_1)$.

Step *n*. Assume that the random number generator produced the number α_n, where $\alpha_n = 1, 2, 3, 4$, or 5. Then set $x_n = h_{\alpha_n}(x_{n-1})$.

This procedure produces the sequence of points

$$X = \{x_0, x_1, \ldots, x_n, \ldots\}.$$

After two or three thousand steps, we will see Conway's triangle displayed. The reasons for the phenomenon are not simple, and are a subject for another article.

In the general case, let F be a self-similar figure, $F = F_1 \cup \ldots \cup F_m$, and $h_1, \ldots, h_m$ be the similarity transformations that transfer F into its constituent parts $F_1, \ldots, F_m$. The "Chaos" game makes it possible to obtain the figure F on the display by using the transformations h_i.

To obtain the mosaic, let us paint the point $x_n = h_{\alpha_n}(x_{n-1})$ depending on the value of α_n. Suppose we assign green to 1, red to 2, blue to 3, orange to 4, and gray to 5. We paint the point $x_n = h_{\alpha_n}(x_{n-1})$ the color that corresponds to the number α_n. Then we obtain on the display the color picture shown in figure 14a. This is the first fragment of the self-similar Conway mosaic. If we choose to paint x_n the color corresponding to α_{n-1}, we will obtain a more detailed colored portrait of the same mosaic (figure 14b). We will obtain a still more detailed portrait if we paint x_n the color corresponding to x_{n-2} (figure 14c), and so on.

a

b

c

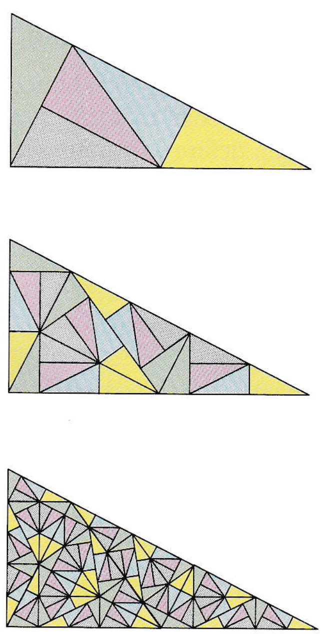


Figure 14

Conway's problem

Let's summarize our results. If a polygon is self-similar, then copies of this polygon can tile the entire plane. If the mosaic consisting of these polygons is strongly hierarchical, it is aperiodic—for example, the self-similar mosaics consisting of chairs. However, it would not be correct to think that one can compose only aperiodic mosaics of chairs. Figure 15 shows a simple periodic mosaic consisting of chairs.

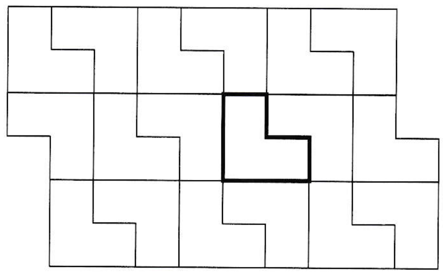


Figure 15

Thus, self-similar polygons can make up periodic mosaics along with aperiodic, strongly hierarchical mosaics.

John Conway raised a question: Does a polygonal or even curvilinear plane figure exist such that it can produce only aperiodic mosaics? Interestingly enough, an affirmative answer has recently been discovered in the form of the so-called Schmidt–Conway–Danzer biprism (figure 16a). This biprism can be glued from the pattern shown in figure 16b. Note that this development of the polyhedron is not of the usual type. It contains a rhombus, which is not a face of the polyhedron, but rather an auxiliary element of the construction.

The biprism is constructed as follows. First, take the triangular prism $ABCA_1B_1C_1$, whose lateral face ABB_1A_1 is a rhombus (with an acute angle α). Then, attach to this lateral face the same prism turned by 180° about the diagonal of the rhombus face. Note that the lateral edges of the second prism are at the angle α with the lateral edges of the first one. A pair of such prisms attached to one another constitutes the desired *biprism*.

a

b

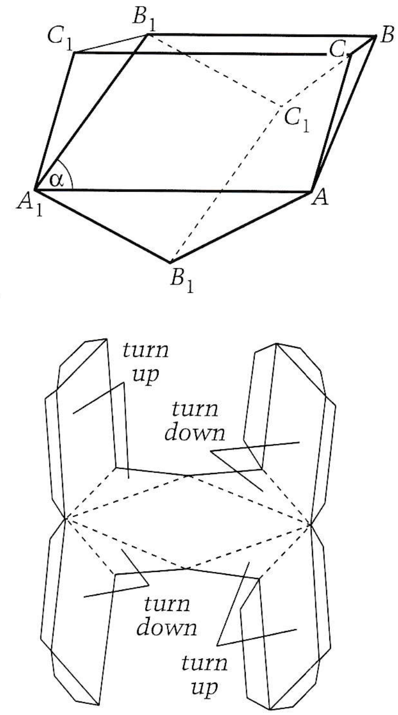


Figure 16

It is not difficult to verify that space can be tiled with such biprisms without gaps or overlaps. The construction of such mosaics is predefined in many respects. If we want to tile space with such biprisms, first we must construct a layer of them (figure 17a). In such a layer, all the biprisms are parallel to each other. Moreover, the layer is a

a

b

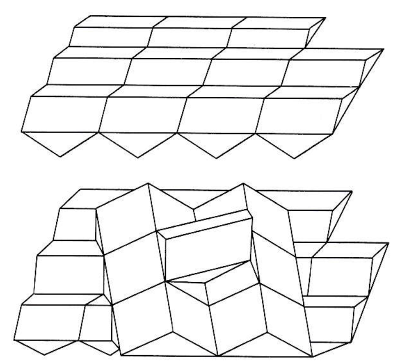


Figure 17

periodic family of biprisms. Then all of space can be filled with such layers (figure 17a). Each succeeding layer is obtained from the previous one by turning it about the axis perpendicular to the plane of the layer through an angle equal to the acute

CONTINUED ON PAGE 20

Catching up on rays and waves

A rhapsody on wavelengths and the Stefan–Boltzmann law

by Albert Stasenko

WHY DO MUSICAL INSTRUMENTS produce *musical* sounds and not the disorderly noise produced by, say, banging a spoon against a pan? Because musical instruments don't generate random sounds of every frequency. They emit sounds of only certain frequencies—so-called *monochromatic* ("single-colored") tones.

If the frequency of a sound is ν, the corresponding wavelength in the air is $\lambda = v/\nu$, where v is the speed of sound in the air. The length of a piano string, or a pipe in a pipe organ, determines the wavelengths of the sound generated. Figure 1 illustrates this idea for a string. It shows three variants (or *modes*) of standing waves on a string. Each of these standing waves contains an integral number of half-wavelengths:

$$l = i\frac{\lambda_i}{2},\ i = 1, 2, 3, \ldots \qquad (1)$$

The longest wavelength is $2l$ ($i = 1$), while all other standing waves have smaller wavelengths and larger i. The number i indicates how many half-wavelengths exist on the string.

Now what if, instead of a string, we have a square plate with an area of $l \times l$ (figure 2)? Then we can have the following numbers of half-wavelengths along each axis:

$$i = 2\frac{l}{\lambda_i}$$

along the x-axis ($i = 1, 2, \ldots$) and

$$j = 2\frac{l}{\lambda_j} \qquad (2)$$

along the y-axis ($j = 1, 2, \ldots$).

An interesting feature of these standing waves is that the wavelengths λ_i and λ_j can describe either independent waves or the same wave traveling at some angle α relative to the x-axis (the angled solid lines in figure 3). In the latter case,

$$\lambda_i = \frac{\lambda}{\cos\alpha} = \frac{2l}{i},\ \lambda_j = \frac{\lambda}{\sin\alpha} = \frac{2l}{j}.$$

Whenever a physicist encounters "sin" and "cos" in a formula, there is a keen desire to square them and add the squares together:

$$\cos^2\alpha + \sin^2\alpha = 1 = \frac{\lambda^2}{4l^2}\left(i^2 + j^2\right).$$

It's clear that this equation can be satisfied by more than one pair of numbers i, j. For example, the broken lines in figure 3 show another wave described by the same equation:

$$i^2 + j^2 = 4\left(\frac{l}{\lambda}\right)^2 = R^2. \qquad (3)$$

This is the equation for a circle of radius R in the i, j-plane (figure 4).

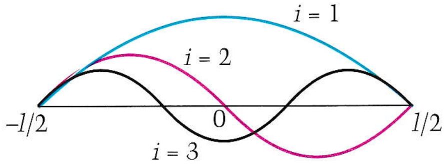


Figure 1

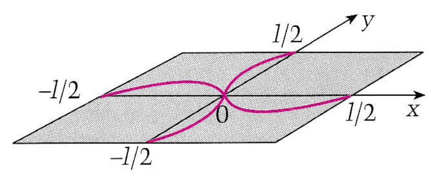


Figure 2

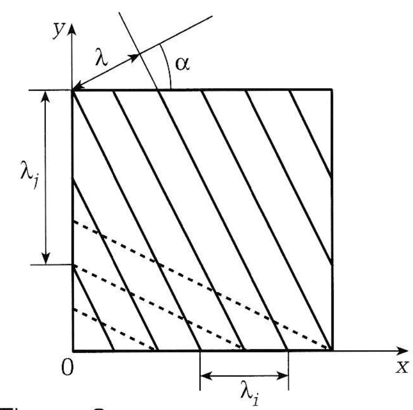


Figure 3

Art by Vadim Ivanyuk

R

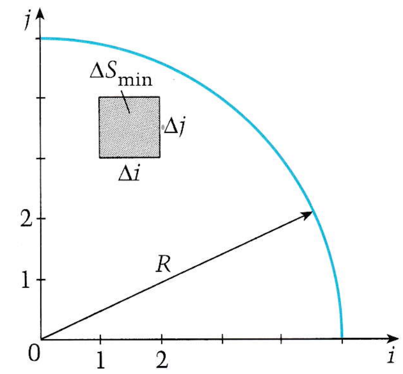


Figure 4

However, the abscissa and ordinate in this plot are integers; therefore, the area in this plot is "granulized," and its minimum value is $\Delta S_{min} = \Delta i \cdot \Delta j = 1$ (the shaded square). The radius of this circle isn't measured in meters—it belongs to the realm of dimensionless numbers. How many such square "granules" could be placed in one quarter of the circle? (Why only one quarter? Because the numbers i and j are positive.) Scientists call this quarter-circle the "first quadrant." To answer the question, we need to divide the area of the first quadrant by $\Delta S_{min} = 1$ (in other words, we can skip the division). Thus

$$N \approx \frac{\pi R^2/4}{1} = \pi\left(\frac{l}{\lambda}\right)^2.$$

The "≈" sign reminds us that it's not easy to cover a round expanse of floor with square tiles.

Let's move into three-dimensional space now and consider a spatial figure (say, a cube with edge length l). Sound waves can now travel along three axes (x, y, and z). We need to add a new equation to the system (2):

$$k = 2\frac{l}{\lambda_k} \quad (k = 1, 2, \ldots),$$

and equation (3) becomes

$$i^2 + j^2 + k^2 = 4\left(\frac{l}{\lambda}\right)^2 = R^2.$$

This is the equation for a sphere in the i, j, k-coordinate system (figure 5). This space is also "granulized" and has a minimum volume $\Delta V_{min} = \Delta i \cdot \Delta j \cdot \Delta k = 1 \cdot 1 \cdot 1 = 1$. Therefore, one-eighth of a sphere of radius R (guess why we consider only one-eighth of it?)—the "first octant"—contains the following number of such "granules":

$$N \approx \frac{(1/8)\left(4\pi R^3/3\right)}{1} = \frac{4\pi}{3}\left(\frac{l}{\lambda}\right)^3. \quad (4)$$

The smaller λ is, the larger N is. Recall that each "granule" (the set of three numbers i, j, and k) describes an individual standing wave. Therefore, we have found the total number of modes—that is, the number of standing waves with wavelengths less than l—generated inside a cube with edge length l.

However, sounds are not the only things that can be musical. In a sense, electromagnetic waves and visible light can also be "musical." This "musicality" is called *color*, and any color is characterized by its own frequency ν and wavelength $\lambda = c/\nu$, where c is speed of light. In this case, a laser producing a monochromatic wave could be considered analogous to an organ pipe or piano string. If the distance between the two parallel mirrors of a laser is l, the laser generates a wave whose wavelength is described by equation (1).

Is it difficult to construct a cube filled with electromagnetic waves? Not at all. We only need to pump everything out of a cube of volume l^3—air, water vapor, carbon dioxide, and so on. Will the cube be empty? Paradoxically, no. It will be filled with so-called "equilibrium radiation" corresponding to the temperature T of the cube's walls. At this temperature the walls emit and absorb the same amount of energy per unit time. Every cubic centimeter of the space inside the cube is permeated by electromagnetic waves traveling in every direction. These are waves of every sort—ultraviolet waves, visible light, infrared radiation... Of course, all have wavelengths less than $2l$.

If such a "stove" is heated only to room temperature, it will be a very weak "radio station" that mainly gives off "warm" (infrared) radiation. An open-hearth furnace, on the other hand, heated to about 1,000 K produces not only infrared radiation but visible light as well. The wavelengths of the electromagnetic waves in this range vary from a fraction of a micron to a few microns, so the distance between adjacent spectral lines (with wavelengths λ_i and λ_{i+1}) given by equation (1) is very small. Therefore, the set of wavelengths (or frequencies) can be considered continuous rather than discrete. Equation (4) says that the number of equilibrium electromagnetic waves filling the volume l^3 of a "stove" is

$$N(\lambda) \approx \frac{4\pi l^3}{3\lambda^3} = \frac{4\pi l^3}{3c^3}\nu^3 = N(\nu). \quad (5)$$

Every photon of frequency ν carries an energy $h\nu$ (h is Planck's constant). The equilibrium electromagnetic radiation is sometimes called the "photon gas." It's similar to conventional gas in that the photons travel in all directions like molecules. However, unlike molecules, the photons do not collide with one another—they only "strike" the walls of the vessel (our "stove"). In addition, the speed of all the photons is the same (it equals the speed of light), so physicists say they are distributed in frequency (while the molecules in a gas are distributed in speed). So what is the mean energy of the photons?

First, let's consider a molecular gas. The number density of its molecules is n and the mass of each molecule is m. It's known that the

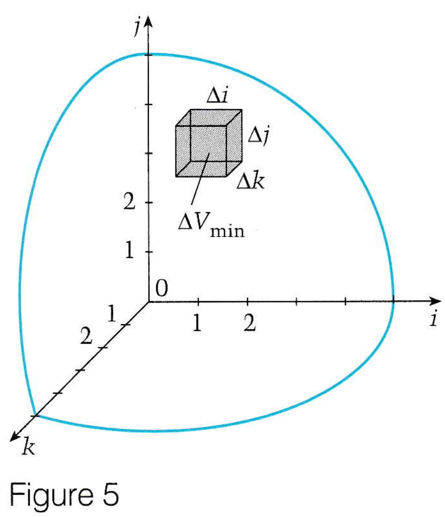


Figure 5

mean kinetic energy of a molecule of gas at a temperature T is proportional to $k_B T$, where k_B is Boltzmann's constant:

$$\frac{m\overline{v^2}}{2} \sim k_B T \qquad (6)$$

Thus the energy density of this gas is

$$n\frac{m\overline{v^2}}{2} \sim nk_B T = P,$$

where P is pressure.

In the Earth's atmosphere the density of the gas varies with altitude according to Boltzmann's formula:

$$n = n_0 e^{-\frac{mgy}{k_B T}}.$$

This formula gives the characteristic altitude at which the density of air decreases e-fold compared to that on the Earth's surface:

$$H_e = \frac{k_B T}{mg} = \frac{RT}{Mg} = \frac{8.31\cdot 300}{29\cdot 10^{-3}\cdot 9.8}\,\text{m} \approx 8.8\ \text{km}.$$

At this altitude the potential energy of a molecule is $mgH_e = k_B T$. It's curious that this value is also equal to the mean potential energy of molecules in an isothermal atmosphere:

$$\overline{mgy} = k_B T \quad (\text{or } H_e = \overline{y}). \qquad (7)$$

According to the mathematical definition of a mean value,

$$\overline{mgy}N = mg\int_0^\infty y\,dn(y),$$

where N is the total number of molecules in the column of air above a unit area:

$$N = \int_0^\infty dn(y).$$

This relationship can be rewritten as

$$\overline{mgy} = mg\frac{\int_0^\infty y\,dn(y)}{\int_0^\infty dn(y)}.$$

The denominator of this expression is $n(\infty) - n(0) = 0 - n_0 = -n_0$ (here we took into account that $n(\infty) = 0$—that is, at an infinite height the density is zero). The numerator can be obtained by means of the identity

$$d(yn) = ydn + ndy,$$

from which we get

$$\int_0^\infty y\,dn(y) = \int_0^\infty d(yn) - \int_0^\infty n\,dy$$

$$= yn\Big|_{y=0}^{\infty} - n_0\int_0^\infty e^{-\frac{mgy}{k_B T}} d\left(y\frac{mg}{k_B T}\right)\left(\frac{k_B T}{mg}\right)$$

$$= 0 - n_0\frac{k_B T}{mg}.$$

Finally,

$$\overline{mgy} = k_B T.$$

Equations (6) and (7) show that the mean values of both the kinetic and potential energies of the gas molecules are of the order of $k_B T$. This result is also true for any other case where thermodynamic equilibrium is present for a large number of chaotically moving molecules.

What can we say about the equilibrium radiation in our "stove"? According to equation (1), the longest wavelength is $\lambda = 2l$, so for a stove with a side length $l \sim 1$ m we have $\lambda_{max} \sim 2$ m. This lies in the radio range. The shortest wavelength can be anything down to that of X-rays. Of course, nobody uses a stove as an X-ray generator or radio station. However, for a given temperature a certain value ν_* will be most characteristic ("will be found most frequently among the photons," "will be most likely," "will be the mean"—choose your own terminology). We may expect that the corresponding energy of equilibrium photon radiation is also of the order of $k_B T$—that is,

$$h\nu_* \sim k_B T, \qquad (8)$$

while the frequency distribution of the radiation energy is described by a bell curve (figure 6). This curve shows that there is only a small amount of energy at very long and very short frequencies. In contrast,

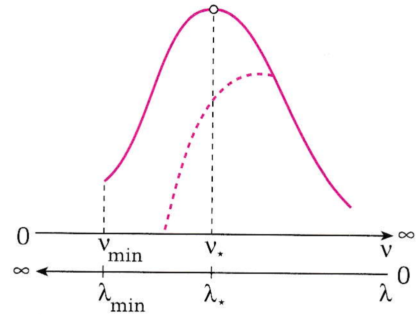


Figure 6

the largest portion of the energy is carried by photons with frequencies near the characteristic frequency $\nu_* \sim k_B T/h$. Thus the total energy of the equilibrium radiation in the volume of the stove can be found from the formula

$$U \sim N(\nu_*)\cdot h\nu_* \sim \frac{l^3}{c^3}h\nu_*^4.$$

Correspondingly, the energy density (energy per unit volume) is described by the formula

$$u = \frac{U}{l^3} \sim \frac{1}{c^3 h^3}(k_B T)^4. \qquad (9)$$

For the sake of this formula—

$$u = \alpha T^4$$

—we've covered a lot of ground. That's because we derived it rather than simply cite the Stefan–Boltzmann law. Note that these formulas describe only the dimensionality and qualitative dependence of the values sought. The correct frequency distribution of photons is given by Planck's formula, which contains the dimensionless factor $1/(e^{h\nu/k_B T} - 1)$. However, here we're only trying to estimate the physical values (which is a very important step in exploring any problem). The above coefficient is dimensionless, so it will not change the dimension of our formula for u. We're on the right track.

We can take another step and calculate the energy radiated per unit time from a unit surface area of a body in thermal equilibrium (that is, at a constant temperature). As noted earlier, such a body radiates and absorbs the same amount of energy. When we multiply the energy density u by its rate of propagation (the speed of light c), we get the density of the energy flow uc (its dimensionality is $\text{J/m}^3(\text{m/sec}) = \text{J/(m}^2 \cdot \text{sec)}$). The photons propagate in every direction, and 1/6 of the photons travel to the *surface* of the body (because it's one of six possible directions: forward, backward, up, down, right, and left). Thus the density of the energy flow is

$$q = \frac{1}{6}uc \sim \frac{k_B^4}{c^2h^3}T^4. \qquad (10)$$

This is another form of the Stefan–Boltzmann law.

If we look a little deeper into the problem, we might guess that the correct coefficient is 1/4, not 1/6. But this is a fine point that doesn't concern us here. Our aim was to obtain not only the Stefan–Boltzmann law in the form $u = \alpha T^4 \sim q = \sigma T^4$ (where σ is the Stefan–Boltzmann constant), but also the rather important nontrivial relationship between the proportionality factors and fundamental physical constants:

$$\alpha \sim \frac{k_B^4}{c^3h^3}, \quad \sigma \sim \frac{k_B^4}{c^2h^3}.$$

The precise value of the Stefan–Boltzmann constant is $\sigma = 5.67 \cdot 10^{-8}$ $\text{J/(m}^2 \cdot \text{s} \cdot \text{K}^4)$. Note that these combinations of fundamental constants could be obtained (as has happened many a time in the pages of *Quantum*) by dimensional analysis, provided the set of related values is known (here they are h, k_B, and c). In this article we took a step further and showed how to obtain the formulas by playing with basic laws rather than basic constants.

Clearly the dependence of q on T is very steep: if we double the temperature, the density of radiation energy increases by a factor of 16!

Now that we've obtained such a powerful law, it's tempting to use it right away. For example, we can calculate the temperature at the Sun's surface knowing only its angular diameter $\theta_S = D_S/L$ (D_S is the Sun's diameter, L the distance between Earth and the Sun) and its mean temperature. Indeed, the energy radiated from the entire surface of the Sun per unit time is

$$Q_S = q_S \cdot 4\pi R_S^2 = \sigma T_S^4 \pi D_S^2.$$

The disk of the Earth, whose area is πR_E^2, receives only a small fraction of this energy, which is equal to $(\pi R_E^2)/(4\pi L^2)$. All this "intercepted" energy is radiated into space from the entire surface of the Earth $4\pi R_E^2$. Equating the solar energy striking the Earth to the energy radiated by the Earth, we get

$$\sigma T_S^4 \pi D_S^2 \frac{\pi R_E^2}{4\pi L^2} = 4\pi R_E^2 \sigma T_E^4.$$

Note that we don't need the precise value of the Stefan–Boltzmann constant, because it cancels out. Thus we have

$$T_S = T_E \sqrt[4]{\frac{16}{(D_S/L)^2}} = T_E \frac{2}{\sqrt{\theta_S}}.$$

Plugging $T_E \approx 300$ K and $\theta_S \approx 0.5 \approx 10^{-2}$ rad into this formula, we get

$$T_S \approx 20T_E = 6000 \text{ K}.$$

Now that we know the temperature at the Sun's surface, we can calculate the area of a solar "sail" that can generate a force of one newton to propel a spacecraft. We'll assume that the spacecraft and the Earth are traveling at the same distance L from the Sun. The surface of the sail is covered by an ideally reflecting layer. Every photon that hits the sail perpendicular to its surface is reflected back and thereby changes its own momentum by $h\nu/c - (-h\nu/c) = 2h\nu/c$. Since the energy falling on a sail of area A per unit time is

$$Q_\downarrow = \sigma T_S^4 \pi D_S^2 \frac{A}{4\pi L^2},$$

(we showed earlier how to derive such a formula), we multiply this value by 2 and divide by c to obtain the change in momentum of all photons striking the sail (which is the propulsive force F):

$$F = \frac{2Q_\downarrow}{c} = \frac{2}{c}\sigma T_S^4 \theta_S^2 \frac{A}{4},$$

from which we get

$$A = \frac{4Fc}{2\theta_S^2 \sigma T_S^4} \approx \frac{4 \cdot 1 \cdot 3 \cdot 10^8}{2 \cdot 10^{-4} \cdot 5.67 \cdot 10^{-8} \cdot 6000^4} \text{m}^2$$

$$\sim 10^5 \text{ m}^2.$$

This is almost ten hectares, to speak in agricultural terms.

Knowing the temperature of the Sun, we can now make the relationship (8) more precise. Since $\nu_* = c/\lambda_*$, we can rewrite it as

$$\lambda_* T \sim \frac{hc}{k_B}.$$

We see that the product of the temperature and the characteristic radiation wavelength is some constant composed of fundamental physical constants. Since the temperature of the Sun is about 6,000 K and the characteristic wavelength of visible light is about 0.5 μm, this constant is about $0.5 \cdot 10^{-6}\ \text{m} \cdot 6000\ \text{K} = 3 \cdot 10^{-3}\ \text{m} \cdot \text{K}$. The relationship between the temperature and the characteristic radiation wavelength is known as the Wien displacement law, which is one of the universal physical laws.

From this law it follows that at room temperature all bodies emit electromagnetic radiation predominantly at the wavelength $\lambda_* = 3 \cdot 10^{-3}/300$ m = 10 μm—that is, in the infrared range. Therefore, this radiation cannot be seen in the dark (by the human eye). However, if there were such a thing as an "invisible man," an infrared camera would detect this "warm" object quite easily against the background room-temperature radiation.

We can draw one more conclusion from our reasoning. A very small stove or a speck of dust heated to a certain temperature under conditions of thermal equilibrium should radiate a number of wavelengths limited by condition (1): the longest wavelength will be about the size of the speck. Therefore, the spectrum of radiated frequencies will

CONTINUED ON PAGE 20

HOW DO YOU FIGURE?

Challenges

Math

M296

Exacting equation. Let $\alpha^3 - \alpha - 1 = 0$. Find the exact value of the expression

$$\sqrt[3]{3\alpha^2 - 4\alpha} + \alpha\sqrt[4]{2\alpha^2 + 3\alpha + 2}.$$

M297

Copper conundrum. Two pieces of metal have masses of 1 kg and 2 kg, respectively. They are alloys of copper with some other metals. The two pieces were melted down and reformed into two new pieces. One of the new pieces had a mass of 0.5 kg and was 40% copper; the other had a mass of 2.5 kg and was 88% copper. What was the percentage of copper in each original piece?

M298

Linear thinking. Let the perpendicular to side AD of the parallelogram $ABCD$ passing through vertex B intersect line CD at point M, and the perpendicular to side CD passing through vertex B intersect line AD at point N. Prove that the perpendicular dropped from B onto diagonal AC passes through the midpoint of segment MN.

M299

Circular logic. A circle lies entirely inside a given angle. Construct another circle, tangent to the first and to the sides of the given angle. How many such circles are there?

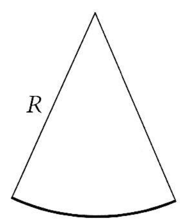
M300

Think small. Let a line m be perpendicular to a plane L. Three spheres that are tangent in pairs are also tangent to line m and plane L. The radius of the largest sphere is 1. Find the minimum possible radius of the smallest sphere.

Physics

P296

Pendulum puzzle. A wire arc of length L and radius R is suspended on two light inextensible strings of the same length R. Find the period of small oscillations for such a pendulum if the strings and the arc always remain in the same plane. (M. Yermilov)



P297

Extraterrestrial ozone. According to some estimates, the mass of ozone (O_3) in the Venusian atmosphere is $\alpha = 10^{-5}$ percent of the entire atmospheric mass. What would the thickness of the ozone layer be if it were collected at the planet's surface and had a temperature and pressure equal to that at the surface of Venus? The acceleration due to gravity on Venus is $g = 8.2$ m/s^2, and the temperature at its surface is $T = 800$ K. (A. Sheronov)

P298

Two-plate special. A parallel plate capacitor of capacitance C is composed of two large conducting plates, each of which is a double layer made of electrically connected thin sheets of foil. The plates are charged with charges Q and $2Q$ of the same polarity. The outer foil layer of the plate with the largest charge is carefully disconnected, moved away parallel to the other plates, and positioned as the third layer on the outer side of the plate with charge Q. A very narrow gap is left between this third layer and the plate, which prevents any electric contact between them. What work must be expended in this transformation? (All the actions are done at a distance so as not to influence the distribution of charges on the plates.) (A. Zilberman)

P299

Spot-light. A plano-convex lens made of glass has a refractive index $n = 1.5$ and a diameter $D = 5$ cm. The radius of the convex spherical surface $R = 5$ cm. A broad parallel beam of light hits the flat side of the lens along its optic axis. Calculate the size of the light spot formed on a screen set behind the lens perpendicular to the incident beam. The screen is positioned so as to obtain the smallest light spot for a narrow beam (restricted by a diaphragm) directed along the optic axis.

(A. Zilberman)

P300

Con-fusion? The nuclei of deuterium D($^2_1$H) and tritium T($^3_1$H) can fuse according to the reaction D + T $\rightarrow$ $^4_2$He + ^1_0n to produce a neutron and alpha-particle ($^4_2$He). In addition, each pair of interacting nuclei releases energy $E = 17.6$ MeV. What energies are carried away by the neutron and the alpha particle? The kinetic energy of the nuclei before nuclear fusion is negligible.

(Y. Samarsky)

ANSWERS, HINTS & SOLUTIONS ON PAGE 51

The Borsuk–Ulam theorem

Horsing around with continuous functions on a circle

by M. Krein and A. Nudelman

THE WEATHER IS CAPRICIOUS. The parameters that describe it (for example, pressure, temperature, and humidity) vary continuously over time and from place to place. The isothermal and isobaric curves on weather maps take whimsical (and, alas, often unpredictable) shapes. Yet no matter how convoluted the weather map looks, the following theorem is true.

Weather theorem. *At any moment there exists a pair of diametrically opposed points on the Earth ("antipodes") where both the temperature and the pressure are identical.*

Although we've couched this proposition in meteorological terms, it is actually a property of continuous functions defined on a sphere rather than properties of the atmosphere. The theorem lies within the realm of topology, a division of mathematics that deals, among other things, with functions or collections of functions that are continuous for certain sets.

Some properties of such functions are determined by the structure of the set on which they are defined. For example, for numerical functions that are studied in high school the following theorem holds:

Theorem of the zero of a function. *If a function f is continuous on the interval $[a, b]$ and takes values with opposite signs at its endpoints, then there exists a point x_0 between a and b such that $f(x_0) = 0$.*

We won't prove this theorem—it may seem obvious geometrically, but strange to say, the proof is far from elementary.

For the "zero theorem" to be true, both the continuity of the function and the *connectivity* of the segment are essential. The reader is invited to explore why we need continuity. By "connectivity" we mean that there are no "gaps" in the segment. For example, the function

$$f(x) = \sqrt{\left(x^2 - 1\right)\left(4 - x^2\right)} + 2x$$

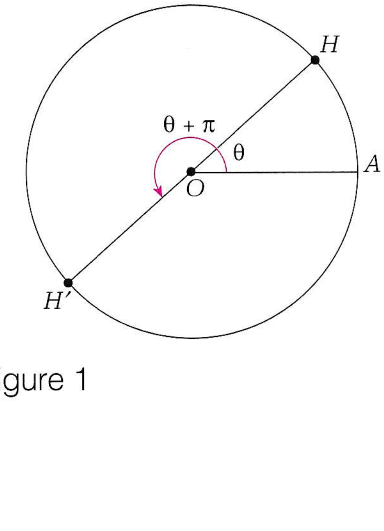


Figure 1

is continuous on its domain $[-2, -1] \cup [1, 2]$, is negative on $[-2, -1]$, and positive on $[1, 2]$. However, it doesn't have a value of zero at any point.

In this article, we'll examine some properties of pairs of continuous functions defined on a sphere. But first we'll look at a simpler case: an unexpected property of continuous functions defined on a circle.

The case of a circle: a circus horse performs

Suppose that a circus horse begins running smoothly around a ring from a point A, and stops smoothly at the same point, after making a full circle. It turns out that no matter how the speed of the horse varies, *there exists a pair of diametrically opposite points where the horse's speed is the same.*

It goes without saying that the horse isn't the cause of this—in fact, we'll turn the horse into a point later on. Rather, it's a property of continuous functions.

We can determine the position of the horse H on the circle by the magnitude of the angle θ $(0 \le \theta \le 2\pi)$ that the radius OH forms with the radius OA drawn from the starting point (figure 1). The corresponding

Art by Leonid Tishkov


Fantastic Power of Mathematician
Tishkov 2000

speed will be denoted by $v(\theta)$. We assume that the function v is continuous on $[0, 2\pi]$ (the horse stops and starts running smoothly), and according to the statement of the problem $v(0) = 0$, $v(2\pi) = 0$. The point H' diametrically opposed to H is determined by the angle $\theta + \pi$ (where $0 \le \theta \le \pi$). We need to prove that there exists $\theta_0 \in [0, \pi]$ such that $v(\theta_0 + \pi) = v(\theta_0)$.

Consider the function $u(\theta) = v(\theta + \pi) - v(\theta)$. We want to find $\theta_0 \in [0, \pi]$ such that $u(\theta_0) = 0$. The function u is continuous on the interval $[0, \pi]$ (as the difference of two continuous functions), and its values at the endpoints of this interval have opposite signs (if $v(\pi) \neq 0$) or are both equal to zero (if $v(\pi) = 0$). Indeed, $u(0) = v(\pi) - v(0) = v(\pi)$ and $u(\pi) = v(2\pi) - v(\pi) = -v(\pi)$. In the case when $v(\pi) = 0$, we can set $\theta_0 = 0$; otherwise, the existence of θ_0 follows from the theorem of the zero of a function.

We can see that the nonnegativity of the function $v(\theta)$ is not needed to prove the theorem. (That is, the horse may sometimes run in the opposite direction around the ring.) From the conditions $v(0) = 0$ and $v(2\pi) = 0$, only the equality $v(0) = v(2\pi)$ is essential (the horse may pass the initial point A with a nonzero speed. What matters is that it return to this point with the same speed).

It's clear that v may be considered a function of the point H rather than a function of the number θ. Therefore, our theorem can be formulated as follows.

The "horse" theorem. *If a function is defined on a circle and is continuous, there exist two diametrically opposite points on the circle where this function takes on equal values.*

Mathematical formulations of the weather theorem

Let's rephrase the weather theorem in mathematical terms. At a given moment in time, each point P of the Earth's surface S can be characterized by two numbers—the pressure $f(P)$ and the temperature $g(P)$. Thus two functions are defined on the sphere S. We assume they are continuous. Indeed, the values of these functions cannot vary too much when the location of point P changes slightly.

A precise definition of the continuity of a function defined on an arbitrary point set X (located on a line, on a plane, or in space) can be formulated as follows. The function φ is called *continuous* at the point $P_0 \in X$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that, for any point $P \in X$ whose distance from P_0 is less than δ, the inequality $|\varphi(P) - \varphi(P_0)| < \varepsilon$ holds. The function is *continuous on the set* X if it is continuous at every point of this set.

We will denote by P' the endpoint of the diameter of the sphere whose other point is P.

Leaving aside the "meteorological" meaning of the functions f and g, we'll formulate the theorem in its general form.

Borsuk–Ulam[1] theorem. *If the functions f and g are defined on a sphere S and are continuous, then there exist diametrically opposed points P_0 and P_0' on the sphere such that $f(P_0') = f(P_0)$ and $g(P_0') = g(P_0)$.*

Let's introduce two functions $F(P) = f(P') - f(P)$ and $G(P) = g(P') - g(P)$. Both these functions are continuous on S and are *antisymmetric*: $F(P') = -F(P)$ and $G(P') = -G(P)$. For instance, $F(P') = f((P')') - f(P') = f(P) - f(P') = -F(P)$. For the points $P_0 \in S$ such that $f(P_0') = f(P_0)$ and $g(P_0') = g(P_0)$ and only for such points, $F(P_0) = 0$ and $G(P_0) = 0$.

Therefore, the Borsuk–Ulam theorem would follow from the following theorem.

Common zero theorem. *If the functions F and G are continuous and antisymmetric on a sphere S, then there exists a point P_0 at which both these functions are equal to zero: $F(P_0) = G(P_0) = 0$.*

Theorem of the zeros of vector fields

Zero theorem gives the conditions that are sufficient for a function that is continuous on an interval to have a zero on this interval. We now formulate a theorem that gives a criterion for the existence of a zero that is common to two functions that are continuous on a circle.

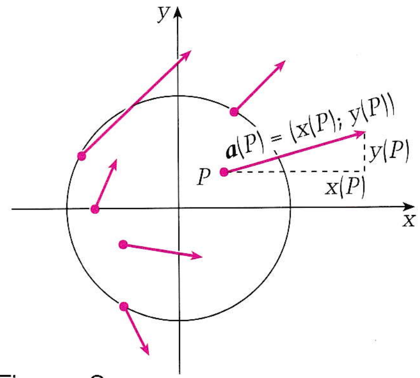


Figure 2

Let every point Q of a set K (for our purposes, this will usually be a circle) in a plane be assigned a vector $\boldsymbol{a}(Q)$ in the same plane. In this case, we say that a *two-dimensional vector field* is defined on K (figure 2). A vector in a plane is defined by two coordinates. Therefore, the definition of a two-dimensional vector field is equivalent to the definition of two functions: $\boldsymbol{a}(Q) = (x(Q), y(Q))$.

The vector field is said to be *continuous at the point* Q_0 if both functions $x(Q)$ and $y(Q)$ are continuous at this point. The vector field is *continuous on a given set* if it is continuous at every point of this set. The vector field is said to be *degenerate* on the set K if $\boldsymbol{a}(Q_0) = 0$ for a certain point $Q_0 \in K$. We want to obtain conditions that are sufficient for a continuous vector field in a circle to be degenerate.

First, we give a vector interpretation of the zero theorem. A vector defined on a line is given by a single coordinate. Therefore, a function defined on an interval may be interpreted as a *one-dimensional vector field* (consisting of vectors that are oriented along the given line). Thus, the theorem of the zero of a function may be formulated as follows.

Theorem of the zero of a one-dimensional vector field. *If a continuous one-dimensional vector field is defined on an interval and the corresponding vectors at the endpoints of this segment have opposite di-*

[1] Karol Borsuk (born in 1915) is a Polish mathematician; Stanislaw Ulam (1909–1984) was an American mathematician.

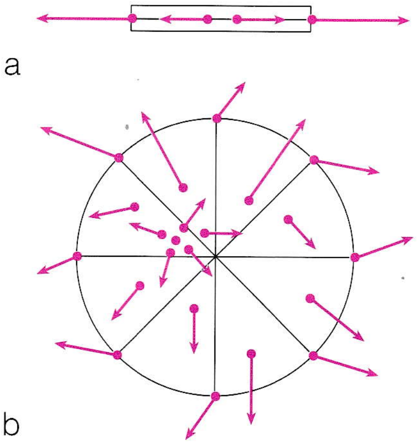


Figure 3

rections, then this vector field is degenerate (figure 3a).

This theorem can be extended for the case of a two-dimensional vector field as follows.

Theorem of the zero of a two-dimensional vector field. *If a continuous two-dimensional vector field is defined on a circle and the corresponding vectors at any diametrically opposed points on its circumference have opposite directions, then this vector field is degenerate* (figure 3b).

Proofs of the "one-dimensional zero theorem" and the "two-dimensional zero theorem" would take us well beyond the high school curriculum. But while the one-dimensional theorem is easy to grasp visually, such is not the case for two-dimensional vector fields. So we'll provide a "plausible reason" for you to accept the validity of this theorem.

Let O be the center of a circle K and r its radius. We'll denote the circumference of K by C, and for any number a, we'll let C_a denote the circumference of the circle centered at O with radius a. Let $\boldsymbol{a}$ be a continuous vector field in K. Suppose that $\boldsymbol{a}(Q) \neq 0$ for all $Q \in C_a$. Then the vector $\boldsymbol{a}(Q)$ will change its direction somehow as the point Q moves along C_a. Let's denote by $\nu(C_a)$ the number of revolutions that the vector $\boldsymbol{a}(Q)$ applied to a point I performs in the counterclockwise direction as point Q passes along C_a counterclockwise. The integer $\nu(C_a)$ can be positive, negative, or zero. To be precise, we note that if $\boldsymbol{a}(Q)$ makes several revolutions counterclockwise and several revolutions clockwise as Q moves along an arc of C_a, these revolutions "cancel out." We suggest that you try to find $\nu(C)$ for the fields depicted in figure 4 (the field of velocities of a rotating circle and the field corresponding to parallel translation).

Let $\boldsymbol{a}(Q)$ (where $Q \in K$) be a vector field satisfying the conditions of the theorem. Now let's assume the opposite—that is, that $\boldsymbol{a}(Q) \neq \mathbf{0}$ for all $Q \in K$. Then, in particular, $\boldsymbol{a}(0) \neq \mathbf{0}$. Due to the continuity of the field, the direction of all vectors $\boldsymbol{a}(Q)$ is close to the direction of $\boldsymbol{a}(0)$ in the vicinity of 0. Therefore, for a small $\varepsilon_0 > 0$ we have $\nu(C_\varepsilon) = 0$, since the direction of the vectors on C_{ε_0} is "almost the same" as the direction of $\boldsymbol{a}(0)$ (see figure 5), and the vector $\boldsymbol{a}(Q)$ makes no full revolutions as Q moves along C_{ε_0}. We now gradually increase ε to its extreme value $\varepsilon = 1$. Since $\boldsymbol{a}(Q) \neq \mathbf{0}$, the number $\nu(C_\varepsilon)$ is defined for all values of ε, $0 < \varepsilon \leq 1$. That is, the function $\nu(C_\varepsilon)$ (considered a function of ε) is defined for all ε on $(0, 1]$. This function is continuous (since our field is continuous) and takes integer values only. However, the integer-valued function can change only by steps (each step is greater than or equal to 1). Therefore, it is either discontinuous or constant. Thus, in our case, $\nu(C_\varepsilon)$ = const, from which we get $\nu(C) = \nu(C_{\varepsilon_0}) = 0$.

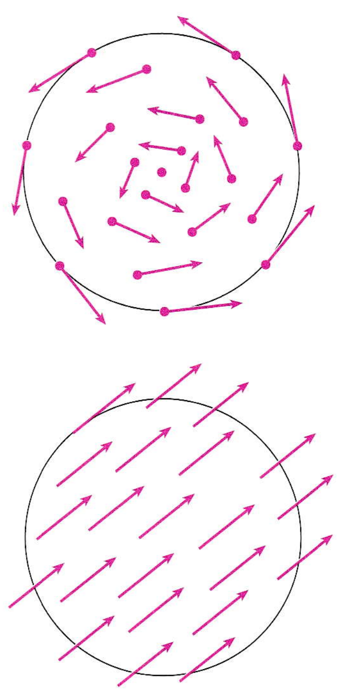


Figure 4

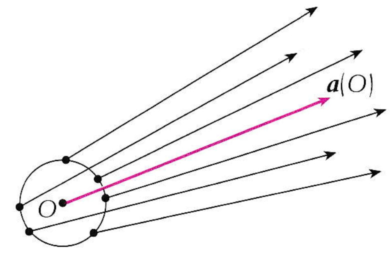


Figure 5

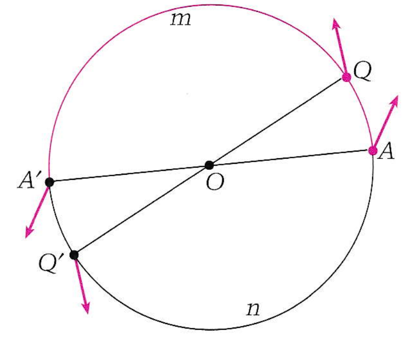


Figure 6

Now consider a point A on the circumference C and its antipodal point A' (figure 6). Since the directions of the vectors $\boldsymbol{a}(A)$ and $\boldsymbol{a}(A')$ are opposite to one another, the vector $\boldsymbol{a}(Q)$ makes an odd number of half-revolutions as Q moves from A to A' along the arc AmA'. As Q moves further from A' to A along the arc $A'nA$, the vector $\boldsymbol{a}(Q)$ makes the same number of half-revolutions in the same direction. Therefore, $\nu(C)$ is an odd number, which contradicts the equality $\nu(C) = 0$ proved earlier. This completes our reasoning.

Proof of the theorem of the common zero

Let the functions F and G be continuous and antisymmetric on the sphere S:

$$F(A') = -F(A), \; G(A') = -G(A).$$

Construct a plane passing through the center O of S. The cross-section is a circle K, and we denote its circumference by C. Let's introduce a rectangular coordinate system in the secant plane whose origin is the center of the sphere. Assign the vector $\boldsymbol{a}(Q)$ with coordinates $F(P)$ and $G(P)$

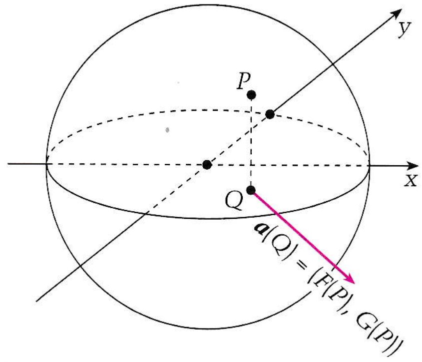


Figure 7

to every point $Q \in K$, where P is the point on the upper hemisphere that projects onto Q (figure 7). We suggest that you verify that this vector field is continuous on K. By virtue of the antisymmetric property of the functions F and G, we find, for all $Q \in C$, that

$$\boldsymbol{a}(Q) = -\boldsymbol{a}(Q').$$

By the zero theorem for two-dimensional fields, there exists a point $Q_0 \in K$ such that $\boldsymbol{a}(Q_0) = \mathbf{0}$. Therefore, $F(P_0) = 0$ and $G(P_0) = 0$, where P_0 is the point on the sphere that projects onto the point Q_0. Thus, the theorem is proved.

A geographical consequence

There is a consequence of the Borsuk–Ulam theorem that is rather distressing for geographers. The location of a point on Earth is given by geographic coordinates: latitude θ and longitude φ. These may be considered functions of a point on the terrestrial sphere. In this coordinate system, the poles have a peculiar property: the latitude of the poles is 90° (N or S), and they can be assigned an arbitrary longitude. So if we go to the North Pole along a meridian and then continue moving along another meridian upon reaching the pole, our motion will be continuous and the latitude will vary continuously, but the longitude undergoes a discontinuity. If we assign the plus sign to East longitudes and the minus sign to West longitudes, then the longitude undergoes a discontinuity when it crosses the meridian that is antipodal to the Greenwich meridian. The question arises: *is it possible to introduce a coordinate system on the whole sphere such that the coordinates are continuous functions of the corresponding point on the sphere?* Naturally, different points must have different coordinates.

It follows from the Borsuk–Ulam theorem that this is *impossible.* Indeed, if continuous coordinates $x(P)$ and $y(P)$ were given on the sphere, a pair of antipodal points $P_0, P_0' \in S$ would exist for which $x(P_0) = x(P_0')$ and $y(P_0) = y(P_0')$.

Some generalizations

If we thoroughly analyze the proof of the "horse" theorem, we'll see that it isn't so important that the points H and H' be diametrically opposed. The reasoning remains valid if we take an arbitrary point O_1 inside the circle instead of O and interpret H and H' as the opposite endpoints of a chord passing through O_1. This also applies to the zero theorem for the two-dimensional field and to the Borsuk–Ulam theorem (figure 8). It's only important that the property $(P')' = P$ be valid for the new definition of the point P'.

Any continuous mapping $P \to P'$ that possesses this property is called an *involution.* The Soviet mathematician A. Fet used ingenious and powerful topological methods to prove that the Borsuk–Ulam theorem (even its n-dimensional version) remains true for an arbitrary involution $P \to P'$ on the sphere. Here is the precise formulation of this theorem for three-dimensional space.

Let an arbitrary involution $P \to P'$ be given on a sphere S. For any pair of continuous functions $f(P)$ and $g(P)$ defined on S, there exists a point $P_0 \in S$ such that $f(P_0') = f(P_0)$ and $g(P_0') = g(P_0)$.

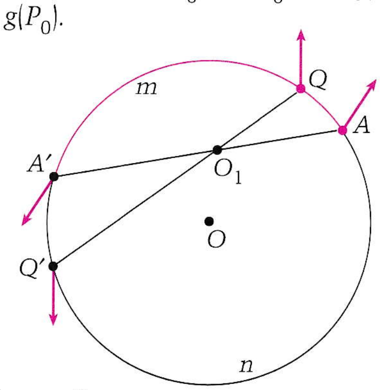


Figure 8

* * *

We recommend that readers who wish to gain a deeper understanding of the basic concepts of topology read the excellent book by W. G. Chinn and N. E. Steenrod, *First Concepts of Topology: The Geometry of Mappings of Segments, Curves, Circles, and Disks* (New York: Random House, 1966).

CONTINUED FROM PAGE 9

angle of the rhombus. Then the layer is translated. Therefore, if the angle of the rhombus is incommensurable with π (that is, if a cannot be written as $(m/n)\pi$ for integers m and n), then no two biprisms from different layers can be parallel. On the other hand, any translation that maps a layer onto itself cannot map any other layer onto itself. Thus, if the angle a is incommensurable with π, there are no translations that map the decomposition described onto itself.

However, if we ask the same question about a plane figure, the answer is not known. It is possible that there are no such *aperiodic* tiles on the Euclidean plane. An analogue of the aperiodic tile on the Lobachevskian plane has already been found. It would be wonderful if a reader of *Quantum* discovered an aperiodic tile on the Euclidean plane.

CONTINUED FROM PAGE 14

be "cut off" on the low-frequency side, as shown qualitatively by the dashed line in figure 6—that is, they are shifted toward the "violet" portion of the spectrum. If the specks of dust could be heated to the temperature of the Sun's surface, the smaller they are, the bluer they'd look (in the visible range of the spectrum).

All these considerations come into play in many fields of science and technology—for example, in studies of the energy balance in planetary atmospheres, metallurgical furnaces, rocket jets, and so on. So we see that the waves generated by a piano string have reverberated far into diverse areas of research.

Algebraic and transcendental numbers

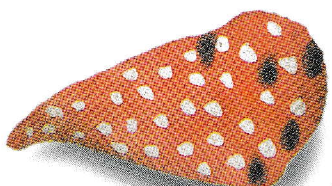
Thought-provoking for thousands of years

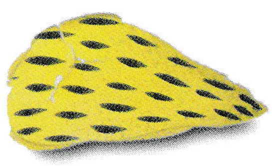
by N. Feldman

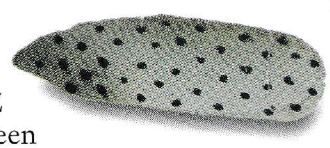
NATURAL NUMBERS, INTEGERS, RATIONAL numbers, real numbers, and complex numbers—this expanding chain,

$$\mathbf{N} \subset \mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R} \subset \mathbf{C},$$

has been familiar to mathematicians for quite some time.







Perhaps you've read the articles in *Quantum* describing how hard it was for negative ($\mathbf{N} \subset \mathbf{Z}$) and complex ($\mathbf{R} \subset \mathbf{C}$) numbers to be accepted as links in the chain. This article will have something to say about another part of the chain—the inclusion $\mathbf{Q} \subset \mathbf{R}$.

You've certainly heard about *irrational numbers* (that is, numbers that cannot be represented as a fraction m/n, where $m \in \mathbf{Z}$ and $n \in \mathbf{N}$). These numbers had already been discovered in antiquity. The fact that the diagonal of a square is incommensurable with its side (translated into the language of algebra, this means that $\sqrt{2}$ is an irrational number) was one of the most exciting (and disturbing) scientific discoveries of the time. Nowadays, a proof of this fact is given in elementary textbooks. If you have any feel for the mathematical way of thinking, the elegance and deceptive simplicity of this proof cannot leave you cold.

Besides being classified as rational and irrational numbers, real numbers can also be classified as algebraic (such as 2/3, $\sqrt{2}$, $5 - \sqrt{7}$, or $\sqrt[3]{4}$) and transcendental numbers (such as π, e, or lg 2). This classification isn't as well known, but it's very important nonetheless. In this article we'll examine these two classes of numbers, their properties, and their (ongoing) history.

Art by Vasily Vlasov

Algebraic numbers

Every rational number a = a/b ($a \in \mathbf{Z}$, $b \in \mathbf{Z}$) is the root of a polynomial with integer coefficients—for example, of the polynomial $bx - a$. Any irrational number of the form $\sqrt[n]{a}$ ($a \in \mathbf{Z}$) is the root of a polynomial of this kind as well—for example, of the polynomial $x^n - a$. We now want to consider only the numbers of this sort—that is, roots of polynomials with integer coefficients.

By definition, a real number is called *algebraic* if it is the root of a polynomial with integer coefficients that is not equal to an identically zero polynomial.[1] We denote the set of all algebraic numbers by **A**. As we've already seen, $\mathbf{Q} \subset \mathbf{A} \subset \mathbf{R}$. To get a better feel for the concept of algebraic numbers, prove the following propositions.

1. If $\alpha \in \mathbf{A}$ ($\alpha \neq 0$), then $1/\alpha \in \mathbf{A}$.
2. If α is the root of a polynomial with rational coefficients, then $\alpha \in \mathbf{A}$.
3. If $\alpha \in \mathbf{A}$ and $a \in \mathbf{Q}$, then $a\alpha \in \mathbf{A}$ and $a + \alpha \in \mathbf{A}$.

It can be proved that if $\alpha \in \mathbf{A}$ and $\beta \in \mathbf{A}$, then $\alpha + \beta \in \mathbf{A}$, $\alpha - \beta \in \mathbf{A}$, $\alpha \cdot \beta \in \mathbf{A}$, and $\alpha/\beta \in \mathbf{A}$ (in the latter case, β cannot equal zero). In other words, *arithmetic operations do not take us out of the set of algebraic numbers.*

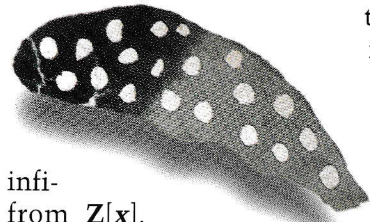
---

[1] The set of polynomials with integer coefficients is denoted by $\mathbf{Z}[\mathbf{x}]$. In this article, we consider only polynomials that are distinct from the zero polynomial without specifying this fact every time. For readers who are familiar with complex numbers, we note that it is possible to introduce and study *complex algebraic numbers* as well.

(The proof of this fact is more complex than the solution of the first three problems, so I won't give it here.) Thus, the set **A** with the operations + and ×, as well as the set **Q** with the same operations, form a *number field*—that is, a number set with all four arithmetic operations defined on it (except for division by 0, of course), and all the operations act as they usually do in arithmetic.

The question naturally arises: are there any real numbers that are not algebraic? To give an answer, we need a notion that is explained in the next section.

The degree of an algebraic number



If α is a root of the polynomial $P(x)$, then it is also a root of the polynomial $P(x)Q(x)$, where $Q(x)$ is an arbitrary polynomial. Therefore, every algebraic number α is the root of an infinite set of polynomials from $\mathbf{Z}[x]$. Clearly we can find polynomials of minimum degree among them. If this minimum degree is n, we say that α is an algebraic number of *degree n* and write $\deg \alpha = n$. We can see that $\deg \alpha = 1$ if and only if $\alpha \in \mathbf{Q}$. It's also clear that the degree of an irrational number of the form $\sqrt{a}$ ($a \in \mathbf{Z}$) is 2—that is, $\deg \sqrt{a} = 2$.

To go any further, we'll need the following simple, yet important, theorem.

The remainder theorem (1779). *The remainder when the polynomial $P(x)$ is divided by $x - \gamma$ is $P(\gamma)$.*

Proof. Let us divide $P(x)$ by $x - \gamma$. The remainder is a constant, which we denote by c:

$$P(x) = (x - \gamma)\, P_0(x) + c,$$

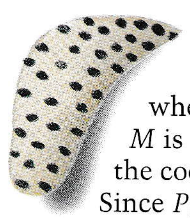
where $P_0(x)$ is a polynomial. Plugging $x = \gamma$ into this formula, we find that $c = P(\gamma)$.

On the basis of this theorem, we can easily prove the following lemma.

Lemma. *If an algebraic number α of degree $n \ge 2$ is a root of the polynomial $P(x) \in \mathbf{Z}[x]$ of degree n, then $P(x)$ has no rational roots.*

Proof. Assume, on the contrary, that $P(a/b) = 0$, where $a \in \mathbf{Z}$ and $b \in \mathbf{N}$. By the remainder theorem, the remainder when $P(x)$ is divided by $x - a/b$ is 0. Therefore, $P(x)$ is divisible by $x - a/b$:

$$P(x) = \left(x - \frac{a}{b}\right) P_0(x),$$



where $P_0(x)$ obviously has rational coefficients. If M is a common multiple of the denominators of the coefficients of $P_0(x)$, then $P_1(x) = M\,P_0(x) \in \mathbf{Z}[x]$. Since $P(\alpha) = 0$ and $\alpha \ne a/b$ (the degree of α is greater than 1), we have $P_0(\alpha) = 0$. Therefore, $P_1(\alpha) = 0$. However, the degree of the polynomial $P_1(\alpha)$ is $n - 1 < n = \deg \alpha$. So we have arrived at a contradiction.

The decisive step in the search for numbers that are not algebraic was the following theorem.

Liouville's theorem

At first glance, the formulation of this theorem is unrelated to the existence of "nonalgebraic" numbers.

Liouville's theorem (1844). *If α is an algebraic number of degree $n \ge 2$, then there exists a number $c > 0$ such that, for any $p \in \mathbf{Z}$ and $q \in \mathbf{N}$,*

$$\left|\alpha - \frac{p}{q}\right| \ge \frac{c}{q^n}.$$

This theorem in effect says that an irrational algebraic number α cannot be approximated by rational fractions "very well." Therefore, if we find an irrational number that *can* be approximated "very well" by rational numbers, it is not algebraic.

Proof. Let α be an algebraic number of degree $n \ge 2$. Then there exists a polynomial

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 x^n, \ (a_n \ne 0)$$

with integer coefficients such that $P(\alpha) = 0$. Denote by H the greatest of the absolute values among $|a_1|$, $|a_2|$, ..., $|a_n|$. We'll show that the number

$$c = \frac{1}{n^2 H\left(1 + |\alpha|\right)^{n-1}}$$

possesses the desired property. Notice that $c < 1$. Let's take an arbitrary $p \in \mathbf{Z}$ and $q \in \mathbf{N}$. Then

$$P\left(\frac{p}{q}\right) = \frac{a_n p^n + a_{n-1} p^{n-1} q + \ldots + a_1 p q^{n-1} + a_0 q^n}{q^n} = \frac{a}{q^n},$$

where we have denoted the numerator of the fraction, which is an integer by a.

By the lemma,

$$P\left(\frac{p}{q}\right) \ne 0.$$

Therefore, $a \ne 0$. Since $a \in \mathbf{Z}$, we have $|a| \ge 1$. Therefore,

$$\left|P\left(\frac{p}{q}\right)\right| \ge \frac{1}{q^n}.$$

Since $P(\alpha) = 0$, we obtain

$$\frac{1}{q^n} \le \left|P\left(\frac{p}{q}\right)\right| = \left|P(\alpha) - P\left(\frac{p}{q}\right)\right|$$

$$= \left|a_n\left(\alpha^n - \left(\frac{p}{q}\right)^n\right) + a_{n-1}\left(\alpha^{n-1} - \left(\frac{p}{q}\right)^{n-1}\right) + \ldots + a_1\left(\alpha - \frac{p}{q}\right)\right|.$$

If

$$\left|\alpha - \frac{p}{q}\right| \ge 1,$$

then we have

$$\left|\alpha - \frac{p}{q}\right| \ge 1 \ge \frac{1}{q^n} > \frac{c}{q^n}$$

and the assertion of the theorem holds. On the other hand, if

$$\left|\alpha - \frac{p}{q}\right| < 1,$$

then

$$\left|\frac{p}{q}\right| < |\alpha| + 1.$$

Obviously, $|\alpha| < |\alpha| + 1$. Then, for any $1 \le k \le n$ ($k \in \mathbf{N}$), we obtain

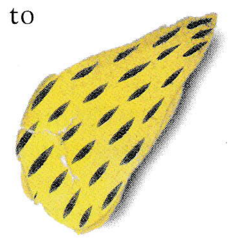
$$\left|\alpha^k - \left(\frac{p}{q}\right)^k\right| = \left|\alpha - \frac{p}{q}\right| \cdot \left|\alpha^{k-1} + \alpha^{k-2} \cdot \frac{p}{q} + \ldots + \left(\frac{p}{q}\right)^{k-1}\right|$$
$$\le \left|\alpha - \frac{p}{q}\right| k(|\alpha| + 1)^{k-1} \le \left|\alpha - \frac{p}{q}\right| n(|\alpha| + 1)^{n-1},$$

from which we get

$$\frac{1}{q^n} \le \left|\alpha - \frac{p}{q}\right| n^2 (|\alpha| + 1)^{n-1} H = \left|\alpha - \frac{p}{q}\right| \cdot \frac{1}{c}, \text{ and } \left|\alpha - \frac{p}{q}\right| \ge \frac{c}{q^n}.$$

Thus the theorem is proved.

It's possible to reformulate Liouville's theorem so as to eliminate the condition deg $\alpha \ge 2$.



Theorem. *If α is an algebraic number of degree n, then there exists a number $c_0 > 0$ such that, for any $p \in \mathbf{Z}$ and $q \in \mathbf{N}$ for which $\alpha \ne p/q$,*

$$\left|\alpha - \frac{p}{q}\right| \ge \frac{c}{q^n}. \qquad (1)$$

Proof. The case deg $\alpha \ge 2$ has already been considered above. Let deg $\alpha = 1$—that is, $\alpha = a/b$ ($a \in \mathbf{Z}$, $b \in \mathbf{N}$). Then, the number $c' = 1/b$ possesses the desired property. Indeed, if $p/q \ne a/b$, then $|pb - qa| \ne 0$, so $|pb - qa| \ge 1$. Therefore,

$$\left|\alpha - \frac{p}{q}\right| = \left|\frac{a}{b} - \frac{p}{q}\right| = \frac{|pb - qa|}{bq} \ge \frac{1}{bq} = \frac{c'}{q}.$$

Setting c_0 equal to the minimum of the numbers c and c', we obtain the desired inequality.

Liouville's theorem can also be proved by examining the difference $P(\alpha) - P(p/q)$ and using Lagrange's mean value theorem. Try to find this proof!

Approximating algebraic numbers with rational numbers

We say that the number α allows *approximations of order m* if, for a certain constant γ, there exist infinitely many rational fractions p/q that satisfy the inequality

$$0 < \left|\alpha - \frac{p}{q}\right| < \frac{\gamma}{q^m}. \qquad (2)$$

Liouville's theorem shows that *algebraic numbers of degree n do not allow approximations of an order greater than n*. Indeed, if α allows approximations of order m, then it follows from (1) and (2) that, for an infinite sequence of natural numbers q, the inequality

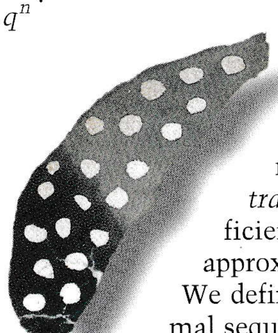
$$\frac{c}{q^n} < \frac{\gamma}{q^m},$$

which can be transformed into

$$\frac{1}{q^{m-n}} > \frac{c}{\gamma}$$

holds. However, this is impossible for $m > n$ and sufficiently large q.

Example of a transcendental number[2]



Now we have an instrument for constructing real numbers that are not algebraic (such numbers are called *transcendental*). For this purpose, it's sufficient to construct a number that admits approximations of an arbitrarily high order. We define such a number as an infinite decimal sequence $\alpha = 0.a_1a_2a_3 \ldots$, where

$$a_i = \begin{cases} 1, \text{ if } t = m! & (m = 1, 2, \ldots), \\ 0, \text{ if } t \ne m!. \end{cases} \qquad (3)$$

(Here $m!$ denotes the product $1 \cdot 2 \cdot 3 \cdot \ldots \cdot (m-1) \cdot m$, which is called the "$m$ factorial.")

In particular, $a_1 = a_2 = a_6 = a_{24} = a_{120} = a_{720} = \ldots = 1$ and $a_3 = a_4 = a_5 = a_7 = \ldots = a_{23} = a_{25} = \ldots = a_{119} = a_{121} = \ldots = 0$. Then, for any $m > 1$,

$$\alpha = 0.\,a_1a_2a_3 \ldots a_{(m-1)!} + 0.0 \ldots 0a_{m!}a_{m!+1} = \frac{p_m}{q_m} + \beta_m,$$

where

[2] If you are familiar with the concept of countability, you can easily prove that the set of algebraic numbers is countable. If you also know that the set of all real numbers is uncountable, you can immediately conclude that transcendental numbers do exist. However, this reasoning does not provide a single concrete example of a transcendental number.

$$p_m = a_1a_2a_3 \dots a_{(m-1)!}, \quad q_m = 10^{(m-1)!}, \quad \beta_m = 0.0\dots0a_{m!}a_{m!+1}\dots,$$

$$0 < \beta_m = 10^{-m!} \cdot a_{m!}a_{m!+1}\dots = 10^{-m!} \cdot 1,\dots < 2 \cdot 10^{-m!} = \frac{2}{(q_m)^m}.$$

Thus

$$0 < \left|\alpha - \frac{p_m}{q_m}\right| < \frac{2}{(q_m)^m}, \quad m = 1, 2, \dots,$$

which means that α allows approximations of any order whatever. Therefore, it cannot be algebraic.

Exercises

4. Prove that α is transcendental if in equation (3)

$$a_t = \begin{cases} 1 \text{ for } t = m^m & (m = 1, 2, \dots), \\ 0 \text{ for } t \neq m^m. \end{cases}$$

5. Find several more transcendental numbers using Liouville's theorem.

Dirichlet's theorem

In 1955 the English mathematician Klaus Roth proved that *no irrational algebraic number can be approximated to any order greater than* 2.

At the same time, every irrational number can be approximated to an order of 2. This fact was proved by the German mathematician Peter Dirichlet using a principle that now bears his name. This principle is simple yet fruitful: *if n items are distributed among n – 1 boxes, then at least one box contains* 2 *or more items.*

Exercise 6. Construct several transcendental numbers using the Roth theorem.

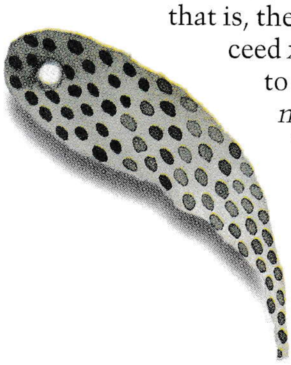
Dirichlet's theorem (1824). *For any real number* α *and any natural number m, there exist* $p \in \mathbf{Z}$ *and* $q \in \mathbf{N}$ *such that* $q \leq m$ *and*

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q \cdot m}. \tag{4}$$

Proof. The interval [0, 1) is a union of m intervals

$$\left[0, \frac{1}{m}\right), \left[\frac{1}{m}, \frac{2}{m}\right), \dots, \left[\frac{m-2}{m}, \frac{m-1}{m}\right), \left[\frac{m-1}{m}, 1\right). \tag{5}$$

Consider the numbers $\{k\alpha\}$ ($k = 0, 1, \dots, m$) (where $\{x\}$ denotes the fractional part of x. We recall that, by definition, $\{x\} = x - [x]$, where $[x]$ is the integer part of x—that is, the greatest integer that does not exceed x). Each of these numbers belongs to one of the intervals (5). We have $m + 1$ numbers and m intervals. Therefore, by Dirichlet's principle, at least one of the intervals (5) contains two or more numbers. Let these numbers be $\{k_1\alpha\}$ and $\{k_2\alpha\}$ ($k_1 > k_2$). Then



$$\frac{1}{m} > \left|\{k_1\alpha\} - \{k_2\alpha\}\right| =$$

$$\left|k_1\alpha - [k_1\alpha] - k_2\alpha + [k_2\alpha]\right| = \left|(k_1 - k_2)\alpha - ([k_1\alpha] - [k_2\alpha])\right|.$$

If we now set $q = k_1 - k_2$ and $p = [k_1\alpha] - [k_2\alpha]$, we obtain the desired inequality by dividing the above inequality by q and taking into account the fact that $0 \leq k_2 < k_1 \leq m$.

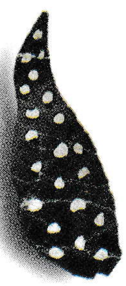
Corollary. *Any irrational number* α *can be approximated to an order of* 2.

Proof. For any $m \in \mathbf{N}$, there exist $p \in \mathbf{Z}$ and $q \in \mathbf{N}$ such that $q \leq m$, and inequality (4) holds. Since $q \leq m$ and α is irrational, from (4) we obtain

$$0 < \left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}. \tag{6}$$

It follows from (4) that the quantity $|\alpha - p/q|$ becomes arbitrarily small as m increases. Since this quantity cannot be zero, the fraction p/q comes ever closer to α as m increases. Thus (6) holds for an infinite number of rational numbers p/q.

Famous transcendental numbers



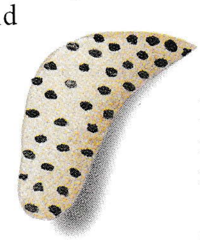
Although Liouville's and Roth's theorems allow us to construct infinitely many transcendental numbers, they have been of no use so far in directly proving the transcendence of such well-known numbers as π, e, ln 2, lg 2, and so on. These numbers have been attracting attention for centuries.

The number π is especially famous. The mathematicians of ancient Greece posed the problem of squaring the circle: *given any circle, construct a square of equal area with a straightedge and compass.* This problem is reduced to constructing a segment of length π, given a segment of unit length. For 2,000 years all attempts to solve this notorious problem have failed. Eventually such a solution was shown to be impossible, and to establish that fact it's sufficient to prove the transcendence of π (in fact, it's sufficient to prove that π is not an algebraic number of a certain type).

The irrationality of the numbers e and π was proved by J. Lambert in 1766. In 1873 C. Hermit proved the transcendence of e. The method he developed for this purpose continues to play an important role in number theory. In 1882 F. Lindemann improved on Hermit's method and proved that π is transcendental. He also proved that the number e^α is transcendental for $\alpha \in \mathbf{A}$ ($\alpha \neq 0$). This fact implies that *natural logarithms of all algebraic numbers distinct from* 1 *are transcendental* (try to prove this).

In 1748 Euler suggested that *if a, $b \in \mathbf{Q}$, and $\log_a b$ is irrational, then it is also transcendental.* Certainly, it is clear that $\log_a b$ can be rational—for example, $\log_4 8 = 3/2$. This conjecture was not proved in the 18th or 19th centuries.

In 1900, at the International Congress of Mathematicians in Paris, David Hilbert formulated twenty-three problems that he thought would stimulate the development of mathematics. The seventh problem was as follows: if α and β are algebraic numbers, α is not 0 or 1, and β is irrational, then α^β is transcendental. In particular, Hilbert suggested that someone prove that $2^{\sqrt{2}}$ and e^π are transcendental (the second number can be reduced to the form α^β, where $\alpha, \beta \in \mathbf{A}$; however, this requires some knowledge of functions of a complex argument).



Exercise 7. Prove that Hilbert's proposition implies Euler's hypothesis.

The first partial solution of Hilbert's seventh problem was obtained in 1929 by a postgraduate student at Moscow University, A. O. Gelfond. Among other things, he proved the transcendence of e^π. A year later the Soviet mathematician R. O. Kuzmin showed that Gelfond's method with certain improvements could be used to prove the transcendence of the numbers α^β when α is an algebraic number different from 0 or 1 and $\beta = \sqrt{d}$, where d is a natural that is not a perfect square. In particular, he proved the transcendence of $2^{\sqrt{2}}$.

A complete solution of the Hilbert's seventh problem was given by A. O. Gelfond in 1934 by means of a new method, which was called Gelfond's second method.

Gelfond's theorem. *Let $\alpha, \beta \in \mathbf{A}$; α is not 0 or 1, and β is irrational. Then α^β is transcendental.*

Exercise 8. Prove that if the numbers α, β, and ρ are such that the expression $\log_\rho\alpha/\log_\rho\beta$ is defined and $\alpha, \beta \in \mathbf{A}$, then the number $\log_\rho\alpha/\log_\rho\beta$ is transcendental or rational.

Gelfond's second method makes it possible to prove many other theorems. An improvement of this method by A. Baker in 1966 led to significant advances in number theory. Work in this area is far from finished.

Bulletin board

Cyberteaser winners

The following are the names of first ten people to submit a correct answer to this month's Cyber-Teaser—Wrong-Way Mercury. We're sure the problem generated heated debate among our contestants, but cooler heads prevailed and the solution was found.

Dimitrios Vardis (Ithaca, Greece)
Eu Jin Teoh (Johor Bahru, Malaysia)
Maxim Bachmutsky (Kfar-Saba, Israel)
Oleg Ivrii (Toronto, Ontario)
Mikhail Agladze (Ithaca, New York)
Theo Koupelis (Wausau, Wisconsin)
Chris Ridgers (Cambridge, England)
Howard Brown (Idaho Falls, Idaho)
Clarissa Lee (Selangor, Malaysia)
Igor Astapov (Kingston, Ontario)

Our congratulations to the winners, who will receive a copy of this issue of *Quantum* and the coveted *Quantum* button. Everyone who submitted a correct answer (up to the time the answer is posted on the web) is entered into a drawing for a copy of *Quantum Quandaries*, a collection of 100 *Quantum* brainteasers. Our thanks to everyone who submitted an answer—right or wrong. You will find our next CyberTeaser at: http://www.nsta.org/quantum.

Geometric su

WHEN YOU STUDY GEOMetry in school, you often have to prove theorems that are pretty obvious—for example, the fact that when two parallel lines are intersected by a third line, the alternate interior angles are equal. However, the fact that three altitudes in a triangle meet at a point is not so obvious. In fact, mathematicians in ancient Greece didn't know this fact, even though they were excellent geometers and knew how to prove that three bisectors, as well as three medians, of a triangle meet at a point.

When a mathematical fact is unexpected, it gives that fact a certain charm. This adds to the beauty of mathematics, along with unexpected and elegantly brief proofs.

In this article, we'll acquaint you with some surprising geometrical facts.

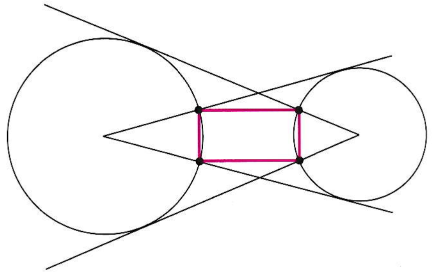


Figure 1

Let's construct two circles and draw tangents from the center of each circle to the other circle (figure 1). Now connect the points where these tangents intersect the circles. The quadrilateral obtained turns out to be a rectangle! We don't know who first discovered this unexpected fact. Try to prove it—it's not very difficult.

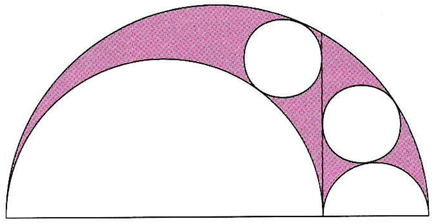


Figure 2

We owe the next geometric surprise to Archimedes. While studying lunes formed by circles, he discovered that two circles inscribed in "curvilinear triangles" (figure 2) are equal. The figure obtained from the largest semicircle by removing the other two semicircles reminds me of a medieval battle-ax. Archimedes, who lived much earlier, thought this figure looked like the cobbler's knife called *arbelos*; so this theorem is known in mathematics as the arbelos theorem.

It's interesting that many surprising geometrical facts can be seen on the walls of Japanese temples. Japanese mathematicians discovered them several centuries ago. In 1800 an inscription was made on the wall of a Japanese temple which made the following observation.

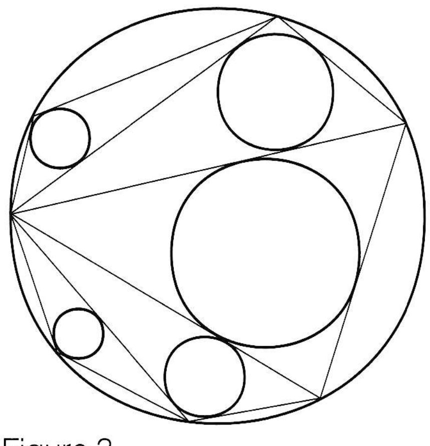


Figure 3

Let's divide an inscribed polygon into triangles by drawing all the diagonals through one of its vertices (figure 3). Then inscribe a circle in each of the triangles obtained. It turns out that the sum of the radii of these circles is a constant, and is independent of the choice of vertex of the polygon.

Later it was proved that the same sum of the radii is obtained for any decomposition of the inscribed polygon into triangles (figure 4).

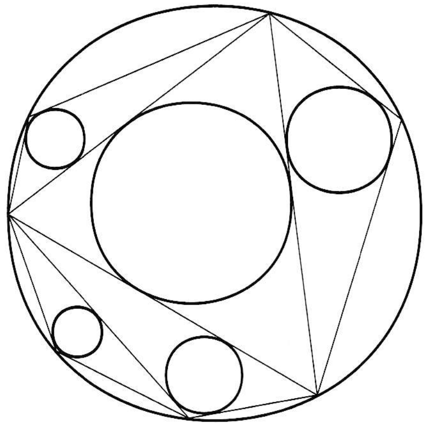


Figure 4

You have certainly dealt with quadrilaterals that are inscribed in a circle or circumscribed about a circle. However, many interesting properties of these quadrilaterals are less well known. One such property was discovered by Claudius Ptolemy, who lived in the second century. He is known as an outstanding astronomer, but he also contributed to the development of mathematics. Ptolemy discovered that the sum of the products of the lengths of the opposite sides of an inscribed quadrilateral (figure 5) is equal to the product of the lengths of its diagonals. Ptolemy used particular cases

c surprises

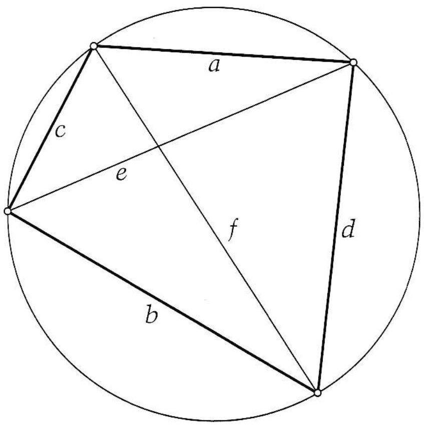


Figure 5

of this theorem, now known as Ptolemy's theorem, in his astronomical calculations.

Another interesting theorem involving a circumscribed quadrilateral belongs to Isaac Newton. He noticed that the center of the circle inscribed in a quadrilateral lies on the line connecting the midpoints of its diagonals (figure 6).

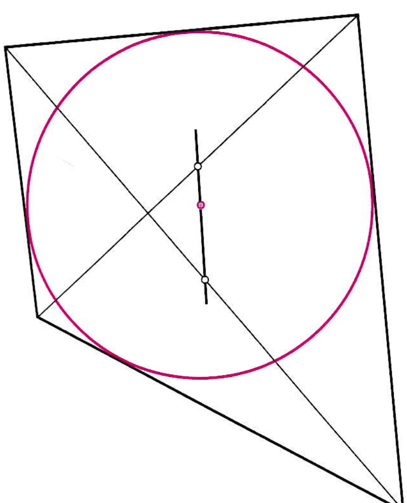


Figure 6

Continuing the list of prominent people who discovered unexpected properties of geometrical figures, we'll mention Napoleon Bonaparte, who was a serious student of geometry and even read a paper at the Paris Academy of Science. The theorem ascribed to Napoleon is as follows. Let's construct equilateral triangles on the sides of an arbitrary triangle ABC (figure 7) and mark their centers O_1, O_2, and O_3. It turns out that triangle $O_1O_2O_3$ is equilateral.

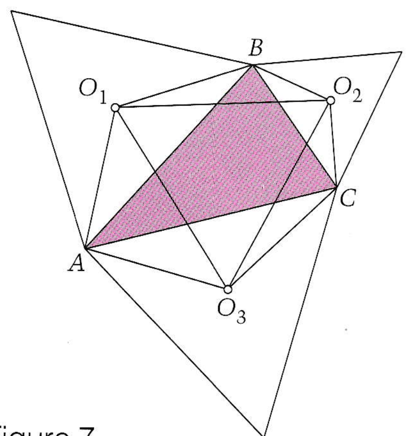


Figure 7

The proof of this fact is simple and elegant. Let's connect the points O_1, O_2, and O_3 with the nearest vertices of triangle ABC. Then rotate two of the three triangles obtained about points O_1 and O_2 as shown in figure 8. The triangle composed of these three triangles has the same

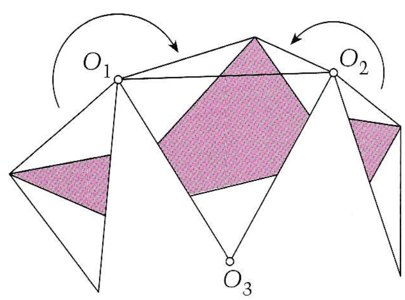


Figure 8

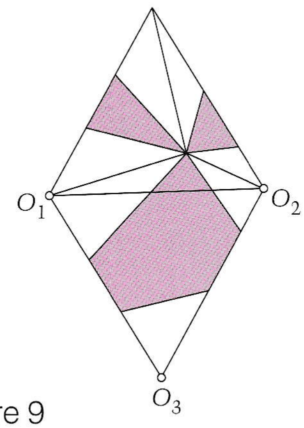


Figure 9

sides as triangle $O_1O_2O_3$, and it's not hard to determine that its angles are equal to 60°. Thus the upper triangle in figure 9 is equilateral; therefore, triangle $O_1O_2O_3$ is equilateral as well.

I'd like to round out this collection of surprising mathematical facts with an elegant miniature by the Moscow mathematician V. V. Proizvolov. Consider a strip formed by two parallel lines. Let's superimpose on this strip a square whose sides are equal to the width of the strip. Then connect "crosswise" the points where the sides of the square intersect the boundaries of the strip (figure 10). The angle formed by these lines is 45°. Surprising, isn't it?

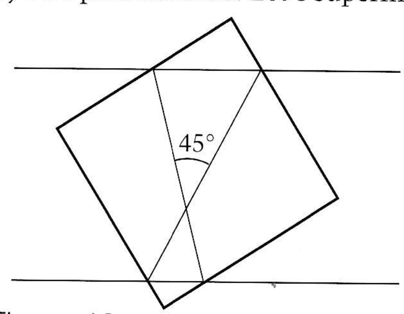


Figure 10

—A. Savin

PHYSICS CONTEST

Batteries and bulbs

by Larry D. Kirkpatrick and Arthur Eisenkraft

> *The habit of analysis has a tendency to wear away the feelings.*
> *—John Stuart Mill*

THE RULES OF BASEBALL ARE the same for everyone—from the smallest Little Leaguer to the biggest Major Leaguer. However, we expect the expertise of the player to increase with age. The laws of physics are the same for everyone. We expect that the problems adults can tackle are more difficult than the ones we give children. That's usually true—but not always.

Given a flashlight battery, a flashlight bulb, and a single piece of wire, hold them together to make the bulb light. We have seen adults take more than an hour to light the bulb! And yet, this is the first activity in a lesson on circuit electricity for fifth graders. Experience has shown us that fifth graders are much more successful at this task than adults. Experience has also shown us that studying Ohm's law does not guarantee that one can successfully analyze circuits containing batteries and bulbs. Elementary education majors who have studied *Batteries and Bulbs* in physical science courses at college have often reported that their friends and spouses in electrical engineering did not have the conceptual understanding to help them with their homework.

Batteries and Bulbs was developed and written by the Elementary Science Study project in the mid 60s. Gerry Wheeler, currently the Executive Director of the National Science Teachers Association, wrote the final version of this popular unit in 1968. It stresses the development of a logical framework for understanding electric circuits and was an early example of the kind of inquiry supported by the National Science Foundation.

After unsuccessfully trying to light the bulb using arrangements such as that shown in figure 1, most students discover that they must use two parts of each of the objects: the two ends of the wire, the two ends of the battery, and the two metal parts of the bulb. The two parts of the bulb are the metal tip and the metal around the base. Whenever all of these six parts are connected in pairs—no matter how you do it—the bulb lights. One such way is shown in figure 2. Can you find the other three ways of doing this?

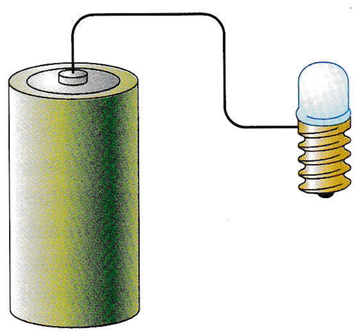


Figure 1

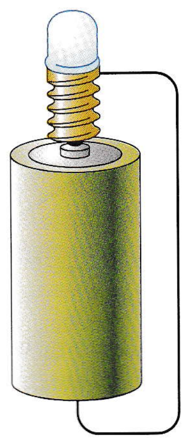


Figure 2

All four ways have one thing in common—a continuous conducting path allows charges to flow from one end of the battery through the light bulb to the other end of the battery. This path is known as a *complete circuit*. By examining a broken flashlight bulb you can see that there is a continuous conducting path from one metal part through the light bulb to the second metal part as shown in figure 3. (The com-

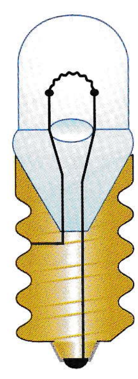
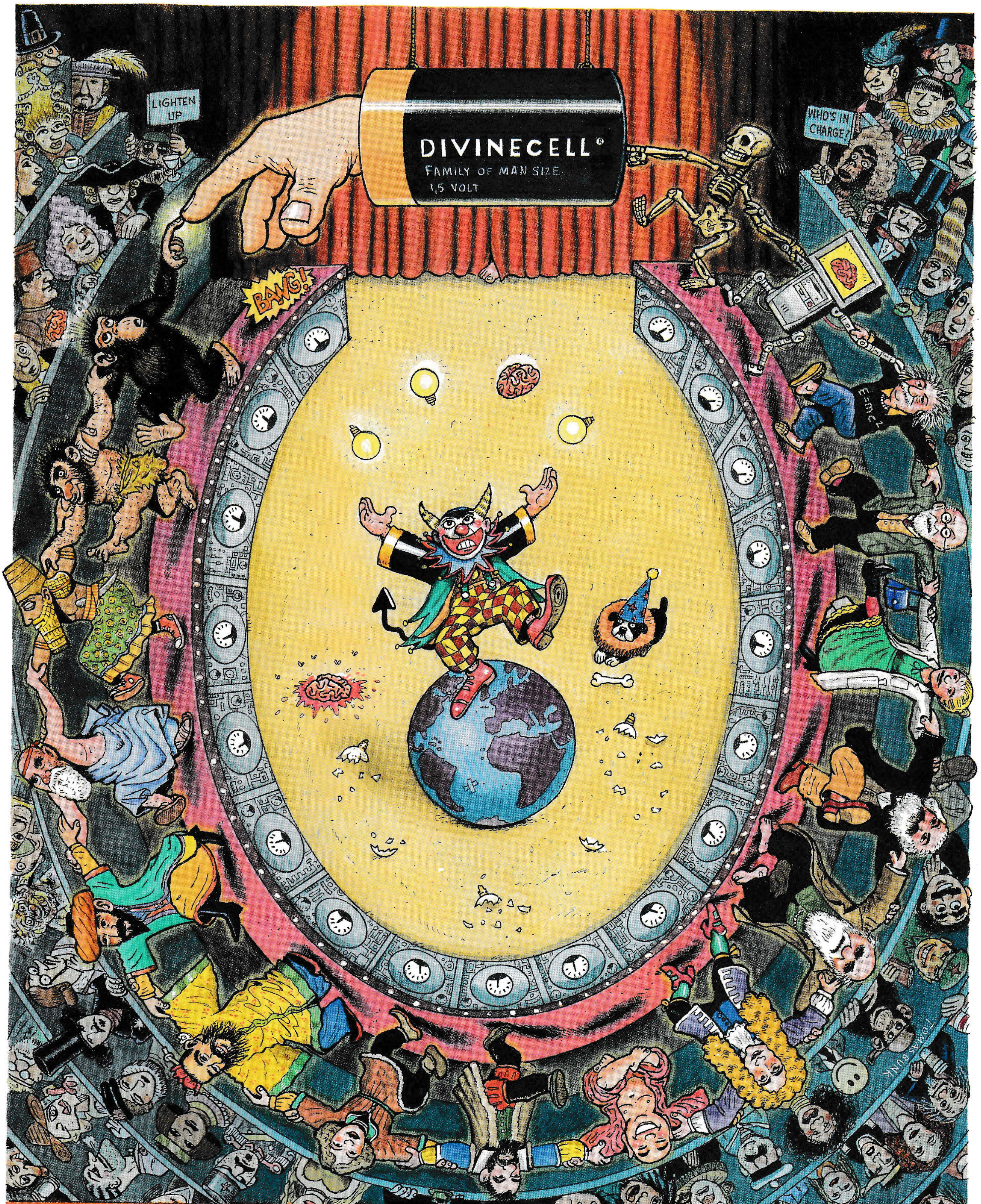


Figure 3

Art by Tomas Bunk



CIRCUIT MAXIMUS

plete circuit is preserved when bulbs are screwed into sockets as shown in the rest of the figures.)

Combining the concept of a complete circuit with the law of conservation of charge leads to the conclusion that electricity flows from one end of the battery and back into the other. All of the charge that leaves one end returns to the other end. Charge does not get lost along the way.

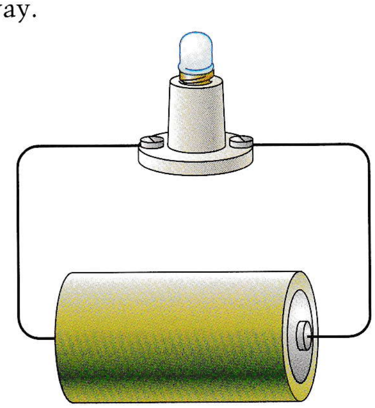


Figure 4

If we light one bulb with one battery as shown in figure 4, we find that the battery runs down in something less than a day. (The actual time depends on the type of battery and the type of bulb.) However, if we connect the wire directly between the two ends of the battery, the battery runs down in less than an hour and the wire usually gets too hot to touch. We infer that the current is larger through the wire than through the bulb. We say that the bulb provides more *resistance* to the flow of charge than the wire. Let's denote the brightness of the single bulb connected to a single battery as the standard brightness.

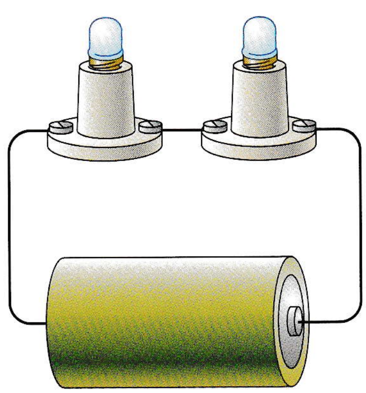


Figure 5

Let's now look at what happens if we use a single battery to light two identical bulbs. We start by connecting the bulbs as shown in figure 5, an arrangement known as *series*. Because there is only a single path through the two bulbs, whatever charge flows through one of the bulbs must flow through the other bulb. If we use identical bulbs, we notice that the two bulbs have the same brightness. We also notice that these bulbs are dimmer than the standard brightness. If we leave the bulbs lit, we discover that the battery lasts longer than the battery in the standard circuit. From this we infer that there is less current in the circuit and, therefore, the resistance of the two bulbs in series is greater than the resistance of a single bulb. (We assume that you are already familiar with the concept of resistance. If not, we would spend time developing this concept more carefully.) From these observations we infer that the brightness of a bulb is a rough measure of the current passing through the bulb. We will assume this from now on.

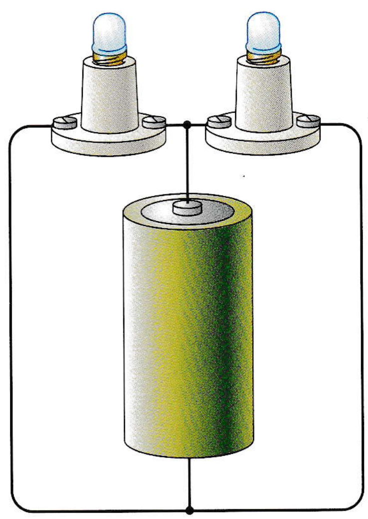


Figure 6

We can also connect two identical bulbs to the battery so that each bulb is on its own path from one end of the battery to the other, an arrangement known as *parallel*. (Note that the paths may share some of the same wires, as seen in figure 6.) In this case, each bulb has the standard brightness. The current through one bulb does not pass through the other bulb. You can check this by disconnecting either bulb and noticing that the other bulb is not affected. This means that the current through the battery must be twice that in the standard circuit and the total resistance of the combination must be one-half the resistance of a single bulb. You can verify this experimentally by letting the battery run down. It does so in approximately one-half the time. In general, adding a path in parallel always reduces the resistance of the combined paths.

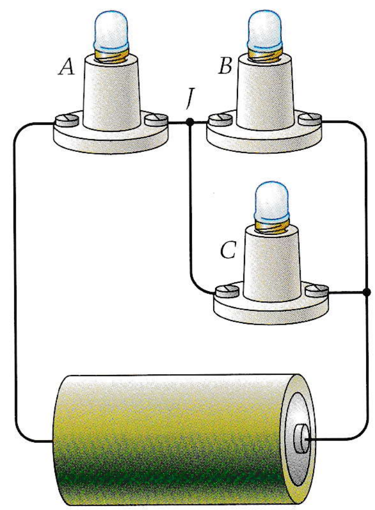


Figure 7

Let's use these ideas to analyze the circuit in figure 7 containing three identical bulbs. Which of the three bulbs is brighter and why? How do the brightnesses of the other two bulbs compare to each other? Notice that the entire current from the battery must pass through bulb *A*. Therefore, it must be the brightest. At junction *J*, the current must split. Because each path following the junction contains a single bulb, the two paths are equivalent and the current must split equally. Conservation of charge tells us that the currents through bulbs *B* and *C* are each one half of the current through bulb *A*. Therefore, bulbs *B* and *C* are equally bright but dimmer than bulb *A*.

We can check our understanding of the model by answering the following questions about this circuit. (1) What happens to the brightness of the bulbs when bulb A is removed from its socket? (2) What happens to the brightness of the bulbs when bulb C is removed from its socket? (3) What happens to the brightness of the bulbs when a wire is connected across the two terminals of socket *A*? (4) What happens to the brightness of the bulbs when a wire is connected across the two terminals of socket *C*? Be sure to write down your answers to these questions before you read on.

Now that you've committed your answers to writing, we are ready to look at the answers to these questions.

(1) Bulbs *B* and *C* will go out as the single path to the battery has been broken and there is no current.

(2) After the removal of bulb *C*, bulbs *A* and *B* are wired in series and are equally bright. Removing bulb *C* removes a parallel path to the right of the junction and therefore increases the resistance of this part of the circuit. This, in turn, reduces the current from the battery. Therefore, bulb *A* becomes dimmer. Two competing effects determine the brightness of bulb *B*. There is less current from the battery but it all passes through bulb *B*. Qualitative arguments do not tell us the answer, but observation tells us that bulb *B* gets brighter.

(3) Connecting a wire across the terminals of socket *A* provides a very low-resistance path around bulb *A*, so bulb *A* goes out. This also reduces the resistance in the circuit, so there is more current from the battery. Therefore, bulbs *B* and *C* brighten.

(4) Connecting a wire across the terminals of socket *C* provides a very low-resistance path around both bulb *B* and around *C*. Therefore, they both go out. Because this also reduces the resistance of the circuit, bulb *A* brightens.

A. For the first part of our contest problem, examine the circuits shown in figures 8 and 9. In each case, which bulbs are the brightest and which bulbs are the dimmest? Repeat the questions asked above for each of these circuits.

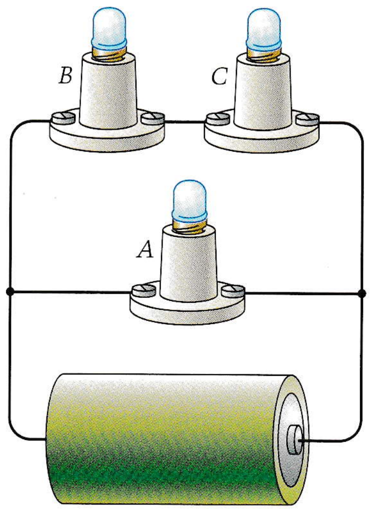


Figure 8

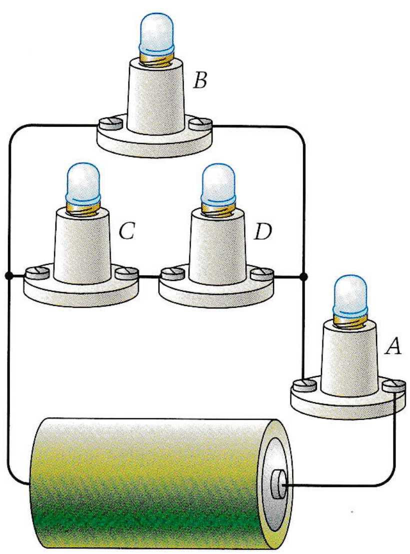


Figure 9

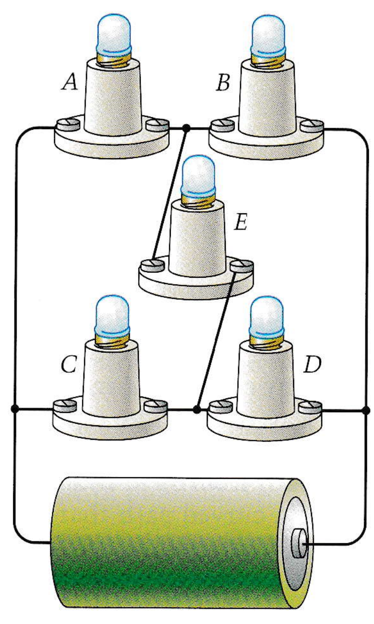


Figure 10

B. The second part of our contest problem is a modification of one of the questions on the exam given to select the members of the 2000 US Physics Team. Which of the identical bulbs in the circuit in figure 10 are the brightest? Which are the dimmest? What happens to the brightness of the bulbs for each of the following? (1) Bulb *A* is removed from its socket. (2) Bulb *E* is removed from its socket. (3) Bulbs *A* and *E* are both removed from their sockets. (4) Bulbs *A* and *D* are both removed from their sockets. (5) A wire is connected across the terminals of socket *A*. (6) A wire is connected across the terminals of bulb *E*. (7) Wires are connected across the terminals of sockets *C* and *E*. (8) Wires are connected across the terminals of sockets *A* and *D*.

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

Tunnel trouble

The January/February 2000 contest problem concerned gravity and the construction of gravity tunnels. Art Hovey of Amity Regional HS in Connecticut provided a solution to all parts, and a number of his students (Brian Chin, Alex Rikun, Josh Leven, and Victoria Buffa) were able to present solutions to parts A and B.

Part A asked for the force a hollowed-out lead sphere exerts on a small sphere of mass *m* that lies at a distance *d* from the center of the lead sphere on the straight line connecting the centers of the spheres and the hollow.

There are three equivalent ways of looking at the solution to this problem. The first is to fill in the missing mass of the hollow in the sphere and add an equivalent mass on the opposite side of the small sphere. The difference of the two forces is the desired force. A second approach is to calculate the force of the sphere as if it were solid and subtract the force due to the mass imagined to fill the spherical hollow. The third approach is to calculate the force of the sphere as if it were solid and add a second force due to a "negative mass" filling the hollow. (The positive mass and the negative mass add together to produce the hollow.) Let's use the second approach.

$$F_1 = \frac{GMm}{d^2},\ F_2 = \frac{GM'm}{(d - R/2)^2},$$

where $M' = 1/8M$ because the spherical hollow has 1/2 the radius of the sphere. Therefore the force on the small sphere is:

$$F_1 - F_2 = \frac{GMm}{d^2}\left(1 - \frac{1}{8(1 - R/2d)^2}\right).$$

Part B asked for an analysis of a tunnel drilled along a chord of the Earth connecting points A and B as shown in figure 11. At the position shown, there is a component of the gravitational force along the tunnel.

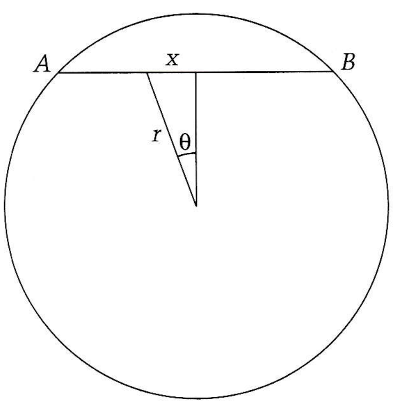


Figure 11

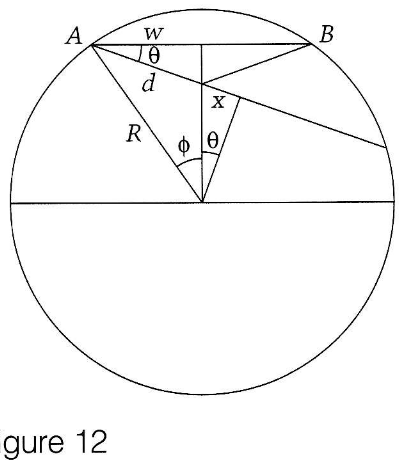


Figure 12

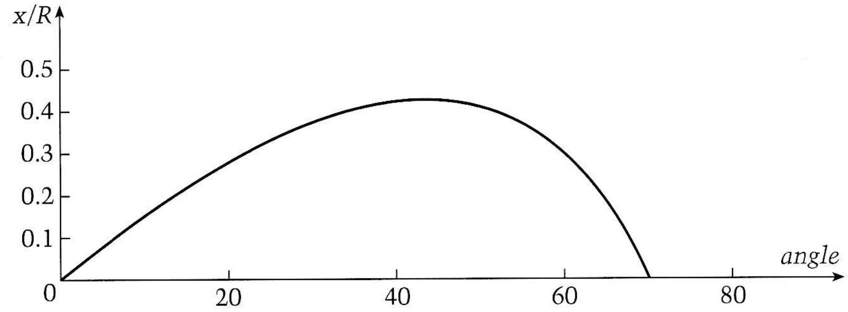


Figure 13

This force is proportional to the mass that lies inside the sphere of radius r.

$$F = \frac{GM'm}{r^2}\sin\theta = \frac{G\rho\frac{4}{3}\pi r^3 m}{r^2}\sin\theta$$

$$= \frac{G\rho\frac{4}{3}\pi r^3 m}{r^2}\frac{x}{r} = \left(\frac{4\pi G\rho m}{3}\right)x = kx.$$

When the displacement is to the right, the force is to the left, so the correct form of the equation is

$$F = -kx.$$

Once again, we see that the path through the tunnel is simple harmonic motion with the same period (84 minutes) for a tunnel along an Earth diameter and also equivalent to the period of an orbiting satellite.

The tunnel would not be particularly feasible due to the difficulties of drilling through the Earth and the presence of friction, heat, and air resistance. If the Earth's molten core doesn't present enough difficulties, we will also have to worry about the walls melting and collapsing.

Part C asked if the straight tunnel provides for the fastest journey from A to B? We found that the period for any chord is 84 minutes, or a one- way travel time of 42 minutes. However, a chord is not the fastest path from A to B. It is best to travel a curved path that passes nearer the center of the Earth. Finding that curved path requires the use of the calculus of variations. Let's find a path with two straight segments that takes less time.

Consider the path from A along d and then to B as shown in figure 12. Since every cord requires 42 minutes for the trip, path w will require 1/2 that time, or 21 minutes. Path $d + x$ will also require 21 minutes, showing that path d requires less than 21 minutes.

We can find the minimum path by finding the path $d + x$ that maximizes x. First, we do some trigonometry.

$$\sin\phi = \frac{w}{R},\ \sin(\phi + \theta) = \frac{d + x}{R},$$

$$\cos\theta = \frac{w}{d}.$$

We now write down an expression for x.

$$x = R\sin(\phi + \theta) - d,$$

$$x = R\sin(\phi + \theta) - \frac{w}{\cos\theta}$$

$$= R\sin(\phi + \theta) - \frac{R\sin\phi}{\cos\theta}.$$

At this point we could take the derivative of this equation and set it equal to zero, but a simple solution does not emerge. Alternatively, we can solve it numerically using a spreadsheet and finding θ for any given ϕ. As a specific example, let's choose $\phi = 20°$.

From the graph in figure 13, we obtain a maximum value for θ of 45° and a corresponding distance $x = 0.43R$. The resulting time savings can be determined by analyzing the equations for the simple harmonic oscillation... but that's another problem.

AT THE BLACKBOARD I

Sharing a point

by I. Sharygin

Art by Sergey Ivanov

FROM TIME TO TIME YOU may encounter problems where you need to prove that three or more lines meet in a point. For example:

Problem. Three isosceles triangles are constructed on the sides of triangle ABC as shown in figure 1. Prove that the perpendiculars

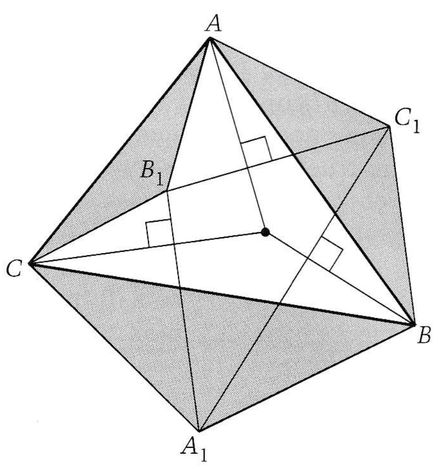


Figure 1

dropped from the points A, B, and C onto the lines B_1C_1, C_1A_1, and A_1B_1 meet in a point.

Here's a method that is useful in solving such problems: prove that two of the given lines intersect in a point that satisfies a certain condition, and then prove that all points of the third line and only these points satisfy this condition. The following well-known theorems can be proved by this method: three bisectors of the internal angles of any triangle meet in a point, and three perpendicular bisectors of any triangle meet in a point.

Similarly, if we have to establish that three or more points belong to a straight line, we can try to prove that all the given points satisfy a condition and then prove that all the points of a line and only such points satisfy this condition (this line of reasoning can be used for circles as well).

Now let's see how we can find the locus of points that helps us solve this type of problem.

Formulating the propositions

Proposition 1. *Let A_1 and A_2 be two fixed (different) points in a plane, and let k, k_1, and k_2 be real numbers. Then the locus of points M such that*

$$k_1(A_1M)^2 + k_2(A_2M)^2 = k \qquad (*)$$

is as follows:

(a) a circle, a single point, or the empty set if $k_1 + k_2 \neq 0$;

(b) a perpendicular to the segment A_1A_2 if $k_1 + k_2 = 0$ (and $k_1 \neq 0$).

A generalization of proposition 1 for several points holds.

Proposition 2. *Let $A_1, A_2, \ldots, A_n$ be fixed points in a plane, and let $k_1, k_2, \ldots, k_n$ (all $k_i \neq 0$) be real numbers. Then the locus of points M such that the sum*

$$k_1(A_1M)^2 + k_2(A_2M)^2 + \dots + k_n(A_nM)^2$$

is a constant is as follows:

(a) a circle, a single point, or the empty set if $k_1 + k_2 + \dots + k_n \neq 0$;

(b) a line or the entire plane if $k_1 + k_2 + \dots + k_n = 0$.

Using assertion 1b, we can prove the following useful condition.

Proposition 3. *Let perpendiculars be dropped from points* A_1, B_1, *and* C_1 *onto the sides BC, AC, and AB, respectively, of triangle ABC. In order for these perpendiculars to meet in a point, it is necessary and sufficient that the following equation holds:*

$$(A_1B)^2 - (BC_1)^2 + (C_1A)^2 - (AB_1)^2 + (B_1C)^2 - (CA_1)^2 = 0. \quad (1)$$

This proposition implies another.

Proposition 4. *Let perpendiculars dropped from the vertices* A_1, B_1, *and* C_1 *of triangle* $A_1B_1C_1$ *onto sides BC, AC, and AB of triangle ABC meet in a point. Then the perpendiculars dropped from points A, B, and C onto lines* B_1C_1, A_1C_1, *and* A_1B_1 *also meet in a point.*

Try to prove all these propositions. In the next section we'll see how we can use proposition 3 or proposition 4 to solve the problem formulated at the beginning of the article; then we'll prove the propositions themselves.

Solution of the problem

According to proposition 3, it's sufficient to verify that

$$(AB_1)^2 - (B_1C)^2 + (CA_1)^2 - (A_1B)^2 + (BC_1)^2 - (C_1A)^2 = 0.$$

(See figure 1. Note that the equation in the statement of proposition 3 has here been multiplied by –1.) Indeed, this equality holds since

$$AB_1 = B_1C,\ CA_1 = A_1B,\ BC_1 = C_1A.$$

We can also use proposition 4. In this case it's sufficient to note that the perpendiculars dropped from points A_1, B_1, and C_1 onto the sides of triangle ABC pass through the midpoints of the sides of ABC and, therefore, meet in a point that is the center of the circle circumscribed about triangle ABC.

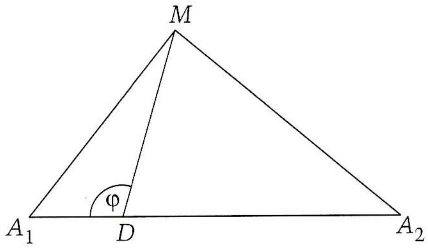


Figure 2

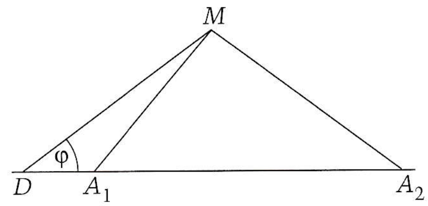


Figure 3

Now let's prove the propositions we've formulated.

Proof of the propositions

Proposition 1a. Let $k_1 > 0$ and $k_2 > 0$. Consider the point D that divides the segment A_1A_2 in the ratio $k_2 : k_1$. Then $k_1(A_1D) = k_2(A_2D)$. Let $\angle MDA_1 = \varphi$ (see figure 2). Choose any point M in the plane, and draw MA_1, MD, MA_2. We can then apply the law of cosines to triangles MDA_1 and MDA_2:

$$(A_1M)^2 = (A_1D)^2 + (MD)^2 - 2MD \cdot DA_1 \cos\varphi,$$

$$(A_2M)^2 = (A_2D)^2 + (MD)^2 + 2MD \cdot DA_2 \cos\varphi.$$

Multiply the first equation by k_1, the second by k_2, and add them to obtain

$$k_1(A_1M)^2 + k_2(A_2M)^2 = k_1(A_1D)^2 + k_2(A_2D)^2 + (k_1 + k_2)(MD)^2. \quad (2)$$

Therefore, if M equation satisfies (*), then

$$(MD)^2 = \frac{k - k_1(A_1D)^2 - k_2(A_2D)^2}{k_1 + k_2} = C.$$

The right-hand side is independent of M, so it is constant. Thus, $(MD)^2$ is constant. Therefore, if $C > 0$, then point M lies on the circle of the radius $\sqrt{C}$ centered at D. If $C = 0$, then M coincides with D; and if $C < 0$, there are no points M satisfying the conditions of the problem.

The converse assertion—that is, *every point M of the set obtained satisfies the equation* $k_1(A_1M)^2 + k_2(A_2M)^2 = k$—can be easily verified. It's sufficient to substitute the expression for $(MD)^2$ in equation (2).

We've considered the case $k_1 > 0$, $k_2 > 0$. The case $k_1 < 0$, $k_2 < 0$ can be reduced to the previous one by reversing the signs of k_1, k_2, and k. In the case $k_1 > 0$, $k_2 < 0$ (or $k_1 < 0$, $k_2 > 0$), our reasoning can follow that of the first case. However, in this case, we must take D outside segment A_1A_2 (see figure 3—try to perform all the computations). Equation (2) remains true for all cases—we'll make use of this fact later.

Proposition 1b. The relation

$$k_1(A_1M)^2 - k_1(A_2M)^2 = k$$

is equivalent to the relation

$$(A_1M)^2 - (A_2M)^2 = \frac{k}{k_1}.$$

Choose any point M on the plane, and let D be the projection of M onto the line A_1A_2. Then, we have by Pythagorean theorem (see figures 4 and 5)

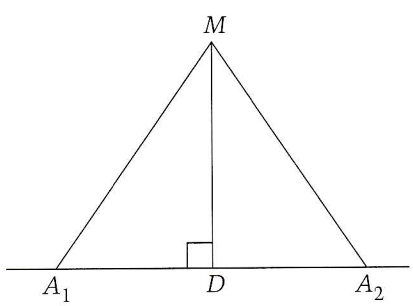


Figure 4

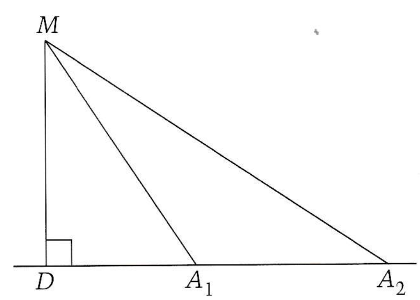


Figure 5

$$(A_1M)^2 = (A_1D)^2 + (MD)^2,$$
$$(A_2M)^2 = (A_2D)^2 + (MD)^2.$$

Therefore,

$$(A_1M)^2 - (A_2M)^2 = (A_1D)^2 - (A_2D)^2 = \frac{k}{k_1}.$$

Thus, the problem is reduced to finding the points D on the line A_1A_2 that satisfy this equation. It is clear that such a point is unique, and it can be easily found. Thus, the point M must lie on the perpendicular to A_1A_2 erected at point D. The details of this discussion are left for the reader.

The converse proposition is also true: *for any point on the perpendicular to A_1A_2, the difference of the squares of the distances to A_1 and A_2 is constant.* The proof is left to the reader.

Thus, proposition 1 is proved.

Proposition 2. We conduct the proof by induction. For $n = 2$, this proposition has already been proved. (For $n = 2$, the locus of points coincides with the entire plane if $k_1 + k_2 = 0$ and A_1 coincides with A_2. Then, for all points of the plane, $(A_1M)^2 - (A_2M)^2 = 0$.)

Now we assume that proposition 2 holds for a certain n and prove that it is true for $(n + 1)$. Notice that if $n \geq 2$ and all $k_1, k_2, \ldots, k_n, k_{n+1}$ are distinct from zero, then there exist two of them such that their sum is not zero. Let them be k_1 and k_2. Consider the point D constructed in the proof of proposition 1a and apply formula (2). Then the equation

$$k_1(A_1M)^2 + k_2(A_2M)^2 + k_3(A_3M)^2 + \ldots + k_{n+1}(A_{n+1}M)^2 = k$$

can be written as

$$(k_1 + k_2)(DM)^2 + k_3(A_3M)^2 + \ldots + k_{n+1}(A_{n+1}M)^2 = k - k_1(A_1D)^2 - k_2(A_2D)^2.$$

On the right-hand side of this equation, we have a constant, the number of points on the left-hand side is reduced by one, and the sum of the coefficients remains the same. By the induction hypothesis, proposition 2 holds for the last equation. Therefore, it holds for $(n + 1)$ points. Thus, proposition 2 is proved.

Proposition 3. Necessity. Let P be the point of intersection of the perpendiculars dropped from points A_1, B_1, and C_1 onto BC, AC, and AB, respectively. The following relations follow from proposition 1b:

$$(A_1B)^2 - (CA_1)^2 = (PB)^2 - (CP)^2,$$
$$(B_1C)^2 - (AB_1)^2 = (PC)^2 - (AP)^2,$$
$$(C_1A)^2 - (BC_1)^2 = (PA)^2 - (BP)^2.$$

Adding up these equations, we see that condition (1) is satisfied.

Sufficiency. Let condition (1) be satisfied and let P be the point of intersection of the perpendiculars dropped from points A_1 and B_1 onto BC and AC, respectively. It follows from proposition 1b that

$$(A_1B)^2 - (CA_1)^2 + (B_1C)^2 - (AB_1)^2 = (PB)^2 - (AP)^2.$$

Condition (1) implies that the left-hand side of this equation equals $(B_1C)^2 - (C_1A)^2$. That is, $(BC_1)^2 - (C_1A)^2 = (PB)^2 - (AP)^2$, which means that the point P lies on the perpendicular dropped from C_1 onto AB, which was to be proved.

Proposition 4. The validity of this proposition follows from the fact that condition (1) is symmetric with respect to A and A_1, B and B_1, and C and C_1.

Exercises.

1. Use Proposition 3 to prove that all three altitudes in the triangle meet in a point.

2. Three pairwise intersecting circles are given. Prove that all common chords of any two of these circles meet in a point.

3. Prove that if the perpendiculars dropped from the points $A_1, A_2, \ldots, A_n$ onto the lines $B_1B_2, B_2B_3, \ldots, B_nB_1$ meet in a point, then

$$(B_1A_1)^2 - (A_1B_2)^2 + (B_2A_2)^2 - (A_2B_3)^2 + \ldots + (B_nA_n)^2 - (A_nB_1)^2 = 0.$$

4. An *escribed* circle of a triangle is a circle that is tangent to one side of the triangle and to the extensions of the other two sides (so that its center lies outside the triangle). Prove that the three perpendiculars to the sides of a triangle at the points of tangency of one of its escribed circles all meet in a point.

5. Let the distances from a point M to the vertices A, B, and C of a triangle ABC be a, b, and c, respectively. Prove that for any $d \neq 0$, the distances to the vertices A, B, and C from any point of the plane (taken in the same order) can never be

$$\sqrt{a^2 + d},\ \sqrt{b^2 + d},\ \sqrt{c^2 + d}.$$

6. Let an equilateral triangle ABC and an arbitrary point D be given. Let A_1, B_1, and C_1 be the centers of the circles inscribed in triangles BCD, ACD, and ABD, respectively. Prove that the perpendiculars dropped from A, B, and C onto B_1C_1, A_1C_1, and A_1B_1, respectively, meet in a point.

7. Let A_1, A_2, A_3, and A_4 be arbitrary points in a plane. Prove that there exist four numbers x_1, x_2, x_3, and x_4 (not all of them equal to zero) such that $x_1(A_1M)^2 + x_2(A_2M)^2 + x_3(A_3M)^2 + x_4(A_4M)^2$ is constant for any point M of this plane.

8. Let a triangle ABC be given. Consider all pairs of points M_1 and M_2 such that $AM_1 : BM_1 : CM_1 = AM_2 : BM_2 : CM_2$. Prove that all lines M_1M_2 meet in a point.

9. A circle is tangent to side AB of triangle ABC and to the extensions of sides AC and CB at points M and N, respectively. Another circle is tangent to side AC and to the extensions of sides AB and BC at points P and K, respectively. Prove that the intersection point of lines MN and PK lies on the altitude of triangle ABC drawn from vertex A.

10. Two segments, AB and CD, are given. Find the locus of points M such that the sum $S_{\Delta ABM} + S_{\Delta CDM}$ is constant.

11. Use the previous problem to prove that the midpoints of the diagonals of any circumscribed quadrilateral and the center of the circle inscribed in it lie on a straight line (Newton's problem) (see "Kaleidoscope").

12. Prove that the locus of points—such that the ratio of the distances from these points to two fixed points of the plane is a constant different from 1—is a circle (called a *circle of Apollonius*).

The little house on the tundra

by A. Tokarev

THE GREAT 16TH-CENTURY Italian architect Andrea Palladio (1518–1580) thought that any building worthy of public approval must satisfy three requirements. These are usefulness (and comfort), beauty, and durability. Leaving aside the principles of usefulness and beauty, let's talk about durability—in other words, the reliability and safety of buildings.

Just about everybody knows that the construction of any building begins with laying the foundation. A good, solid foundation is a token of further success. But just what is a "strong foundation"?

Builders face many challenging tasks, and one of them is laying the foundation. This is especially difficult in permafrost areas. Many buildings in these areas are subject to cracking due to settling of the foundation in soil that has melted.

Can we prevent thawing of the ground under buildings erected in permafrost, or at least minimize it?

Let's try to solve this problem using the simplest physical laws and rules. We begin by analyzing the conditions of the problem. Why does the ground under a building start to thaw? Clearly because the foundation transfers heat to it. So we need to focus on the foundation.

The first thing to do is decrease the area of contact between the ground and the foundation. This is why some buildings in permafrost areas are built on piles instead of a conventional solid foundation. But this isn't enough.

Are there any other ways to reduce the flow of heat to the ground? In this particular case, heat is transferred by thermal conduction only; therefore, we should make the piles out of material with the lowest possible thermal conductivity. What material could that be?

The thermal conductivity of various substances is described by a special physical parameter, the coefficient of thermal conductivity. Naturally this coefficient is different for different substances. Thermal conductivity is highest for metals (which correspondingly have the highest coefficients of thermal conductivity), while it's lower in liquids and much lower in gases. For the purposes of our analysis it isn't important how the coefficient of thermal conductivity is determined and in what physical units it's measured. We need only compare the capacity of various substances to transfer heat. So we take a reference book and construct a grid (table 1) that shows the coefficients of thermal conductivity relative to that of water.

substance	*relative thermal conductivity*
gasoline	0.2
cotton	0.07
water	1
air	0.04
felt	0.1
wood	0.2–0.6
iron	122
kerosene	0.2
brick	1.1
ice	3.7
cork	0.07
alcohol	0.3

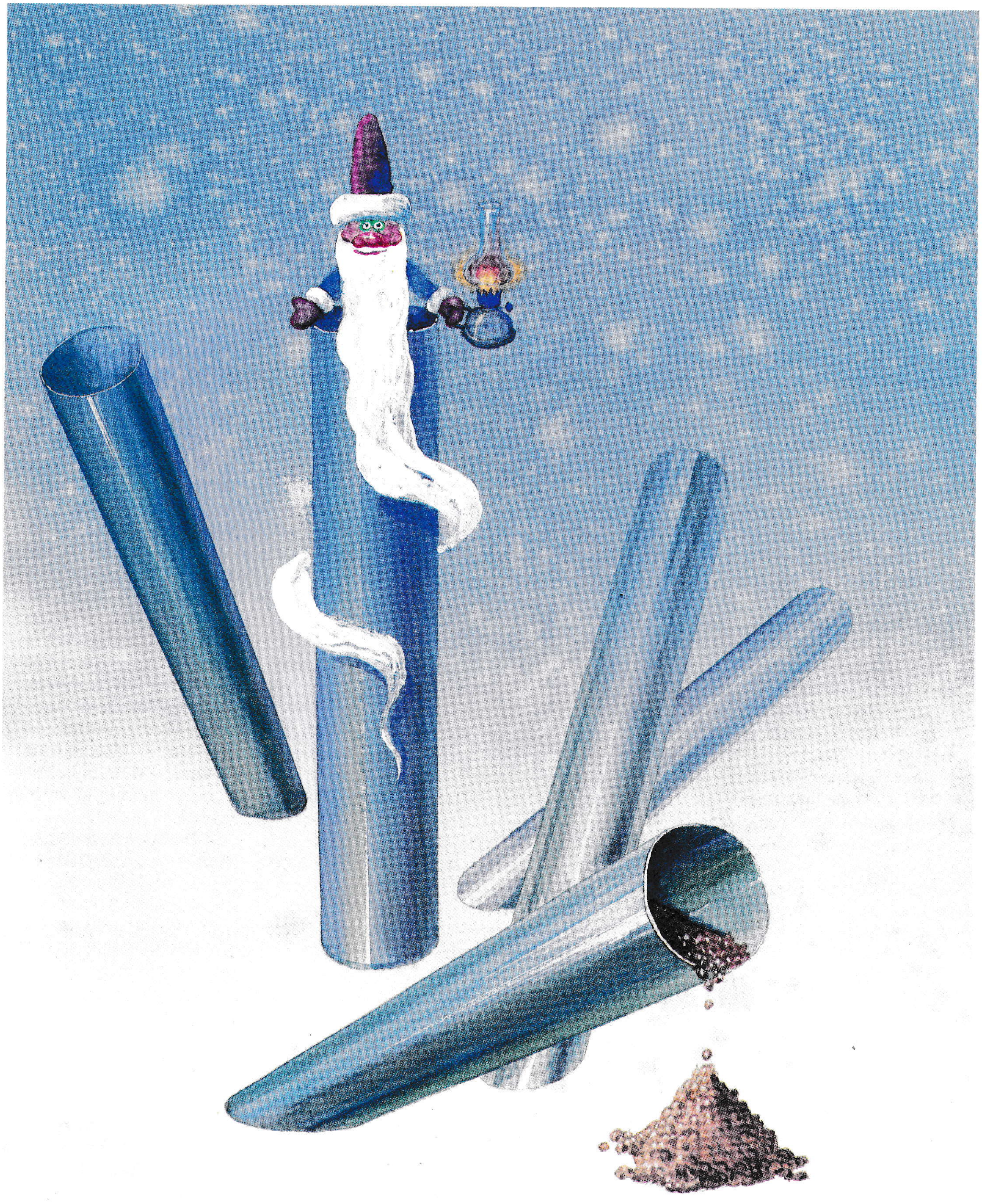
Table 1

Take a good look at table 1. Among solid substances, cotton and cork have the lowest values of thermal conductivity, while wood and brick are next. Now, a pile must be durable and strong, so the substances with the lowest values won't make good piles. But what if we take some steel pipe and stuff it with cotton, felt, or some other porous substance that contains a lot of air? I think we've found a solution!

Piles should be made of a durable solid material and filled with a porous substance.

Analyzing our result, we arrive at this conclusion: due to its low thermal conductivity, a pile of this design will actually decrease the flow of heat from the surrounding air to

Art by Pavel Chernusky



the ground (that is, downward) during the warm season. However, a properly constructed pile can do even more. It would be nice if in winter, when the temperatures are well below freezing, the piles could cool the ground (that is, they could transmit heat upward). This would add strength to the soil and decrease its thawing during the following summer.

Let's try to formulate more precisely the details of the physical processes in a pile and the surrounding soil in summer and winter. During the warm season, the upper part of a pile is heated due to contact with the warm air. Gradually the lower part of the pile, buried in the ground, is also warmed. The less the lower part of the pile (and the soil surrounding it) is heated, the better. In winter the air cools the upper part of the pile. Gradually the lower part of the pile and the adjacent soil are also cooled. The colder the ground gets, the better.

Therefore, a pile should have the following properties:

(a) if the temperature of the upper part of the pile is higher than the temperature of the lower part, the pile should conduct heat very weakly;

(b) if the upper part of the pile is colder than the lower part, the pile should conduct heat efficiently.

In other words, downward heat transfer should be small, but the upward heat transfer should be large. The pile should be a "heat semiconductor."

It's known that heat exchange in solids is due entirely to thermal conductivity, which doesn't depend on direction. So an ideal pile cannot be completely solid. Our previous model of a metal pipe filled with a porous material is of no use either, because its porous interior will conduct heat weakly not only in summer but also in winter, when it's necessary to cool the ground.

What if we fill a strong and durable pipe (or some other empty metal shell) with a fluid—that is, a liquid or a gas? In this case, the heat is transferred not only by molecular thermal conductivity but also by convection. How would such a pile work?

In winter, the upper layer of fluid will be cooled. The cold fluid has higher density than the warm fluid, so it will sink. Warmer, less dense fluid layers will rise and release heat to the surrounding air. Then this portion of the fluid will be replaced by cold fluid from the bottom, and so on. As a result, the lower part of the pile and the adjacent soil will be cooled to the temperature of the surrounding air. Notice that we don't need to construct any special refrigerators and waste energy to cool the foundation—everything is done "at the expense" of the naturally occurring cold winter air.

In summer, the upper layers of the fluid in the pile will be heated by the air. But being less dense, they'll stay in the upper part of the pile. As a result, no convection occurs in summer; the heat will be transferred downward only because of thermal (molecular) conductivity, which is very small in fluids.

Since this type of pile conducts heat poorly in summer, the temperature of the surrounding soil will increase only insignificantly.

Thus we've arrived at another important conclusion: the pile, constructed of a durable material, should be filled with a fluid (gas or liquid).

One question remains: With what exactly should we fill the pile? In winter, the properties of the fluid aren't crucial, since there's plenty of time to cool the ground to the ambient temperature. In summer, however, it's very important that the ground be heated as little as possible. Therefore, we should use a fluid whose temperature rises the least when heated. From this well-known formula—

$$Q = cm(t_2 - t_1)$$

—we find that the change in temperature $(t_2 - t_1)$ depends not only on the amount of heat transferred but also on the heat capacity c of the fluid and its mass m.

Since the mass of a liquid is always greater than the mass of a gas in the same volume, we prefer to use a liquid as our filler for the pile. Now let's compare the heat capacities of various liquids (we'll use a reference book again). We can see from table 2 that not only density but also heat capacity is higher in liquids compared to gases. Thus the piles should be filled with a liquid.

substance	*heat capacity, kJ/(kg · K)*
gasoline	1.4
water	4.2
air	1
glycerin	2.4
kerosene	2.1
motor oil	1.9

Table 2

What liquid is the best for this purpose? Although water is cheap and readily available, it won't do: in winter it freezes. Both glycerin and motor oil thicken at low temperatures, which makes convection inefficient. This leaves gasoline and kerosene from our list. The freezing point of both these liquids is less than –50°C, so either could withstand the low Alaskan and Siberian temperatures. Of the two we should probably choose kerosene, because it's cheaper and has a higher heat capacity than gasoline.

At long last we can formulate the "final answer" to the problem:

To decrease thawing of the ground beneath buildings in permafrost areas, they must be erected on piles; the piles should be made of a hollow durable material and filled with kerosene.

We should note that this method of strengthening frozen foundations (decreasing thawing of the ground under buildings) isn't just idle speculation. It was calculated theoretically and tested by the construction industry in permafrost areas. As expected, construction costs were reduced significantly.

The enigmatic magnetic force

by E. Romishevsky

IT IS KNOWN FROM experience that, in general, the force acting on a point electric charge q placed in electric and magnetic fields depends on the position of the charge and on its velocity. Usually, this force is resolved into two components: the electric force $\mathbf{F}_e = q\mathbf{E}$, which is independent of the motion of the charge, and the magnetic force $\mathbf{F}_m$, which depends on the velocity of the charge. In this article we discuss the nature of this magnetic force and its interplay with the electric force.

At every point in space the magnetic force is perpendicular to the velocity of the electric charge. The magnetic force is also perpendicular to a special direction, which is defined at every point as well. The magnitude of the magnetic force is proportional to that component of the velocity of the charge which is normal to the "special direction" mentioned above.

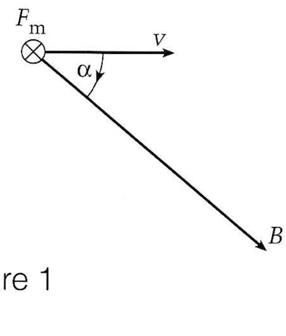


Figure 1

This property of the magnetic force can be described in another way using the concept of the magnetic field. The direction of the magnetic field coincides with the special direction in space.

The magnitude and direction of the magnetic force are determined by the formula

$$\mathbf{F}_m = qvB \sin \alpha \, \boldsymbol{\zeta},$$

where v and B are the magnitudes of the velocity and the magnetic field, while the unit vector $\boldsymbol{\zeta}$ (according to the right-hand rule) serves to indicate the direction of the magnetic force. This direction coincides with the advance of a right-hand screw whose head lies in the plane of the vectors $\mathbf{v}$ and $\mathbf{B}$, and which is turned through the smaller angle from the vector $\mathbf{v}$ to the vector $\mathbf{B}$ (figure 1). The magnetic force $\mathbf{F}_m$ is normal to both of the vectors $\mathbf{v}$ and $\mathbf{B}$.

The total electromagnetic force $\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m$ acting on a particle with a charge q is called the "Lorentz force." By measuring the Lorentz force acting on a test charge of known sign (positive or negative), one can determine the magnitudes and directions of the vectors $\mathbf{E}$ and $\mathbf{B}$.

Note that the magnetic force does not affect an electric charge at rest. Another important feature of the magnetic force is its direction: it is always normal to the velocity, so that it performs no work while acting on a charge. Therefore, in a constant magnetic field the kinetic energy of a charged particle does not change, whatever motion this particle undergoes.

As an example, consider the motion of two particles with opposite charges $+q$ and $-q$ that have different masses $M_1 = 2m$ and $M_2 = m$. Initially, the velocities of these particles have the same value $\mathbf{v}_0$, whose direction is perpendicular to the boundary of a homogeneous mag-

Art by Jose Garcia

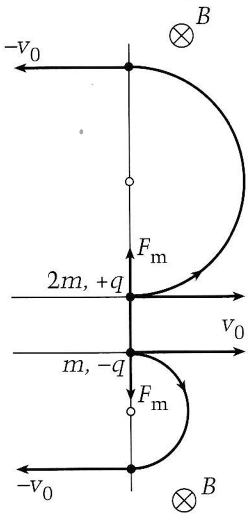


Figure 2

netic field **B** (figure 2; the vector **B** is normal to the plane of the page and directed away from the reader). When the positively charged particle enters the magnetic field, the magnetic force $F_m = qv_0B$, which is initially directed upward. The negative particle experiences the same magnetic force, but this force is initially directed downward. Each particle describes a semicircle, after which it leaves the region of the magnetic field. The radius of each circle can be found from Newton's second law:

$$qv_0B = \frac{Mv_0^2}{R},$$

whence

$$R = \frac{Mv_0}{qB}.$$

The angular velocity of the particle and its period are

$$\omega = \frac{v_0}{R} = \frac{qB}{M}$$

and

$$T = \frac{2\pi}{\omega} = \frac{2\pi M}{qB}.$$

Clearly, the positive particle ($M_1 = 2m$) describes a semicircle whose radius is twice that of the negative particle ($M_2 = m$), which moves in the opposite direction. The heavier positive particle will return to the no-field region in a half-period which is twice as large as the corresponding interval of time for the lighter negative particle. Thus, a homogeneous magnetic field is capable of separating, in time and space, particles that move in the same beam but that have different masses and charges. This property is used in mass spectrometers, which can separate isotopes (atoms of the same charge but different masses).

Moving charges (that is, electric currents) generate magnetic fields. Numerous experiments with magnetic forces yielded a simple law that gives the magnetic field **B** generated by a point charge q moving with constant velocity **v** that is much less than the speed of light c. This law can be written as

$$\mathbf{B} = \frac{1}{4\pi\varepsilon_0 c^2}\frac{qv\sin\alpha}{r^2}\boldsymbol{\zeta},$$

where α is the angle between the velocity **v** of the charge and the radius vector **r** drawn from the charge to the observation point; $\boldsymbol{\zeta}$ is a unit vector obtained by applying the right-hand rule to the vectors **v** and **r** (figure 3). The constant $1/(\varepsilon_0 c^2)$ is usually denoted by μ_0 and is called the magnetic permeability of free space.

By multiplying both sides of this formula by the number of electrons $\Delta N = n\Delta lS$ in a segment of wire of length Δl, electron density n, and cross-sectional area S carrying an

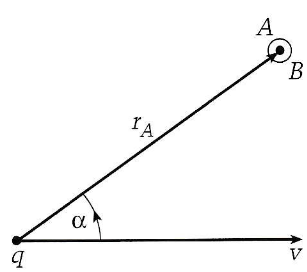


Figure 3

electric current $I = qnvS$, we obtain the famous Biot-Savart law for the contribution $\Delta\mathbf{B}$ to the magnetic field generated by an electric current element $I\Delta l$:

$$\Delta\mathbf{B} = \frac{\mu_0}{4\pi}\frac{I\Delta l\sin\alpha}{r^2}\boldsymbol{\zeta}.$$

In this case the lines of magnetic field are concentric circles drawn around the trajectory of the moving charges (figure 4). The magnitude of

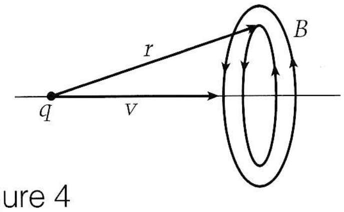


Figure 4

the magnetic field decreases with distance as $1/r^2$, just like the magnitude of the electrostatic field generated by a point charge. The analogy between the electric and magnetic fields is not universal: the magnetic field has no "sources" and "sinks," so that the magnetic lines are always closed. Such a physical vector field has specific features and is referred to as a vortical or solenoidal field.

Now let us consider another example. Suppose that two fairly massive point particles 1 and 2 with equal charge q move parallel to each other with the same nonrelativistic velocity **v** (figure 5). Each particle is affected by a repulsive electric (Coulomb) force $F_e = qE$ and an attractive magnetic force $F_m = qvB$ (the velocity of one particle is normal to the magnetic field generated by the other particle). Let us compare these two components of the total electromagnetic (Lorentz) force that act, say, on particle 2:

$$\frac{F_{m2}}{F_{e2}} = \frac{qvB_{21}}{qE_{21}},$$

where B_{21} and E_{21} are the magnetic and electric fields generated by charge 1 at the position of charge 2. Inserting the corresponding expres-

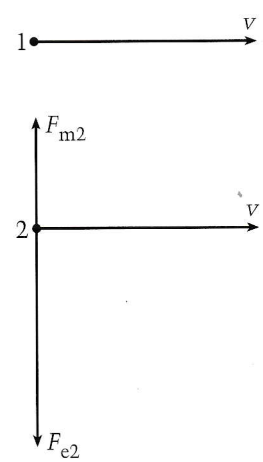


Figure 5

sions in this ratio, we get

$$F_{m2}:F_{e2} = qv\frac{qv}{4\pi\varepsilon_0 c^2 r^2}:q\frac{q}{4\pi\varepsilon_0 r^2}$$
$$= v^2 : c^2.$$

This ratio shows that at nonrelativistic speeds the magnetic force produced by moving charges is much weaker than the electric force acting on the charges. In other words, under these conditions the magnetic force is a minor contribution to the total electromagnetic force.

What will happen if we choose an inertial reference frame that moves with the same velocity **v** as our particles? In this moving reference frame the particles are at rest, so that their magnetic fields and magnetic forces disappear!

Well, this paradox could have been expected: the magnetic component of the Lorentz force depends on the velocity of a charged particle, and this velocity changes when one reference frame is replaced by another. At the same time, the total Lorentz force, just like any other force, does not depend on the choice of nonrelativistic inertial frame. Therefore, in the reference frame in which the magnetic component of the Lorentz force disappears, the electric component of this force must change to compensate for such a loss. In other words, dividing the total Lorentz force into electric and magnetic components is meaningless without specifying a reference frame.

The last example raises the question of whether it is reasonable to study and take into account such relatively small magnetic forces. Of course, it is worthwhile, and here are the reasons why.

First, the ratio obtained is valid also at relativistic speeds $v \sim c$. In this case the magnetic forces are comparable with the electric ones. For example, they play a major role in a rapidly moving beam of charged particles.

Second, there are situations where a "negligible" magnetic force is really a single unbalanced force in a physical system. This is the case for electrons moving in a conducting wire. Here, there are no net electric forces as a result of the almost ideal balance of the negative and positive charges in a conductor. Recall what a huge number of charged particles participate in generating an electric current in metals—about 10^{23} elementary charges in one cubic centimeter! This enormous number produces a very large magnetic force—for example, in electric motors.

Third, sometimes the electric charges move under the action of various combinations of electric and magnetic fields generated by different sources. In general, the relationships between electric and magnetic forces can be quite versatile, including the case when the magnetic force dominates over the electric one. Therefore, magnetic knowledge is power in itself, isn't it?

Can you carry water in a sieve?

by A. Dozorov

DO YOU KNOW THE STORY of the little ant who was in a big hurry to get home? Many creatures helped him. For example, a water strider carried him across a river. Maybe you've seen this insect. The water strider stands calmly on the surface of the water, which sags slightly under its weight. Why doesn't this insect sink? And can water really "sag"?

It turns out that the surface layer has a number of unusual properties. We can investigate them in some simple experiments.

1. *The water surface can support various objects.*

Pour some water into a saucer. Take a needle and place it carefully on the surface—it doesn't sink. If this experiment failed, don't give up. Rub the needle with your fingers (or oil it slightly, or rub it with a candle). Repeat the experiment and look closely at the surface. Now do you see that the surface is bent? It looks as if the needle is lying on a film.

A rather good comparison is that surface layer of a liquid is similar to a stretched piece of cellophane (although the specific properties of the surface layer are quite different from those of stretched cellophane). Let's try to guess why.

A molecule within a liquid is surrounded by other molecules, which pull it equally in every direction. By contrast, molecules of the surface layer have no molecules above them, so they are attracted only by the molecules below them. It looks as if the liquid "tries" to have the minimum number of molecules at its surface. As a result, the surface layer of a liquid is "stretched" much like cellophane. Let's call this surface a "film" (in quotation marks).

We see that not only water striders but even denser bodies (such as a metal needle) can stay on the surface and not sink. They don't swim or float in the usual sense of the words—they're held up by the surface tension of the liquid. However, if we test thicker and thicker needles, we'll eventually find one whose weight is greater than the supporting force of the surface tension. Of course, this needle (and heavier ones) will sink to the bottom. It's interesting that the length of the needle has virtually no effect on its ability to "float."

2. *The tension of a surface "film" depends on the liquid used.*

Place a needle on the water surface. Take a wooden matchstick and cut off the head. Rub the end of the match with soap and touch the water about 1 cm from the side of the needle. The needle will immediately "jump away" from the match. Why? Well, you created a soap solution near the needle when you dipped the matchstick in the water. The molecules of the soap solution don't attract the needle as strongly as the pure water molecules on the opposite side. Since unbalanced forces act on the needle, it moves in the direction of the greater force. In other words, the surface tension of pure water is greater than that of a soap solution.

You can use this trick to pilot the needle all around the saucer. Maneuver it to the very edge of the saucer and watch what the needle does. Keep in mind that the soap spreads across the water surface very quickly, so don't forget to change the water in the saucer from time to time. You can replace the needle with a match and repeat the experiment. (A needle that keeps sinking is an unnecessary complication in your experiments.)

Take two matches and place them carefully onto the water surface parallel to each other. What happens? The matches are drawn toward each other (figure 1). Pull them apart and touch the water on both sides of this pair with the tip of a third match, which you've rubbed with soap, as before. What do you see now?

Using the same principle, you can make a number of simple toys and entertain kids with them. Slit the end of a match and skewer a piece of paper with it (figure 2). Soak the paper in the soapy soup that often collects in soap dishes (or make some "soup" yourself, if there isn't any). Now lay this match "ship" on the water—it starts "sailing." Did you

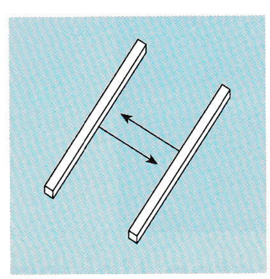


Figure 1

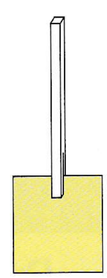


Figure 2

notice the direction in which it sailed? Now slit a matchstick at both ends and slip pieces of paper into the slots as shown in figure 3. Soak these pieces of paper in the soap soup. This match will rotate on the surface like a propeller. Figure 4 shows a "gun" cut out of thick paper. To "fire" the gun, touch the water surface at point A with the end of a match that you've rubbed with soap.

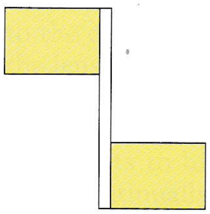


Figure 3

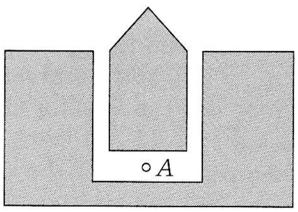


Figure 4

Try testing other substances instead of soap. Here's a nostalgic scene from summer camp: a bunch of kids are gathered around a little puddle. They've taken wood chips and rubbed their ends with resin from fir trees. Now they're holding races—the chips slide swiftly along intricate paths on the water surface.

3. *Surface tension can raise the liquid rather high.*

Take a glass tube with a very narrow internal diameter (much less than 1 mm)—a so-called "capillary tube" (or just "capillary"). Lower one end into a jar of water and watch the water rise in the capillary to a height greater than the water level in the jar. The thinner the capillary, the higher the water is lifted in it. If you've ever had blood drawn from your finger at the doctor's office, you've seen how the nurse collects it in a capillary tube. This "capillary action" can be observed everywhere. You can see it in tea rising in the tiny holes in a sugar cube, in kerosene rising in the wick of an oil lamp, in water absorbed from soil by the roots of various plants, and so on.

A more modest experiment can be made with a thicker tube. Put some water in it and plug the bottom with your finger (figure 5). You'll see that the water surface in the tube is curved. This curvature—called a "miniscus"—is explained by the fact that water molecules are attracted more strongly to the walls of the tube than to each other. In this case we say that the liquid wets the surface of a container.

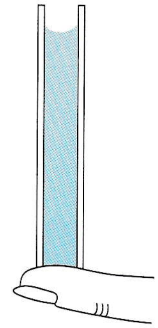


Figure 5

Now let's do one more experiment. Pour tea from a cup, but leave some tea and a few tea leaves at the bottom. Carefully touch the surface of the liquid with a teaspoon or a match and watch how the surface quickly "crawls" upward, drawing the tea leaves along with it.

4. *Not all liquids "cling" to the walls, and it doesn't happen in every pipe.*

There are cases when a liquid in a capillary doesn't rise—not only that, but the miniscus is curved in the other direction (it's convex). Why is that? It's because this particular liquid doesn't wet the surface of the walls—the mutual attraction between molecules in the liquid is stronger than the attraction between these molecules and the walls of the tube. This is how mercury behaves in a capillary (figure 6).

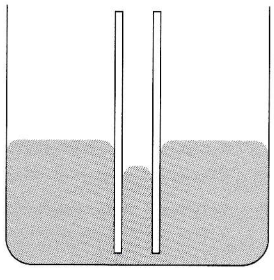


Figure 6

Collect some water in a pipette. Carefully release one drop onto clean glass and another onto a piece of buttered bread. The first drop spreads out on the glass, while the other one maintains its round shape. So we conclude that water wets glass but doesn't wet butter.

And now the moment of truth: how would you answer the question posed in the title of this article? Can you carry water in a sieve? Well, let's take a sieve and spread butter on it, or even better, rub it with a candle. Pour some water into it—the water doesn't run out! It's supported by a surface "film" that forms because the water doesn't wet the edges of the sieve's tiny openings. If you don't have a sieve, you can do this experiment with a can with a small hole punched in the bottom.

As we saw, a liquid that doesn't wet the surface doesn't spread out but collects itself into a drop. In this case, the smaller the drop, the nearer its shape approximates a sphere. Why? Due to the strong mutual attraction between molecules in such a liquid, the drop assumes the shape with the least surface area. As a rule, this is a sphere, which is easily (and often) demonstrated in the weightlessness aboard an orbiting spacecraft. If an astronaut releases water from a container (one can't "pour it" as one does on Earth), it immediately assumes a spherical shape.

The brief weightlessness of molten drops of metal as they fall from a high tower has been used for generations to produce pellets. The drops become spherical as they fall and stay that way long enough to solidify as spheres.

You can do a similar experiment at home. Tilt a burning candle and pore the melted wax into a basin filled with cold water—you'll get small wax pellets. Hold the candle as close to the water as you can, so that the wax solidifies right at its surface.

5. *Sometimes surface tension is so strong you can literally "feel" it.*

Take two identical plates of glass. Clean them carefully and put one on top of the other. You can easily separate them again. Now wet one plate with water and put one atop the other again. Try to pull the plates apart (without sliding one over the other). It's not so easy, is it? That's surface tension at work.

Ernst Abbe and "Carl Zeiss"

by A. Vasilyev

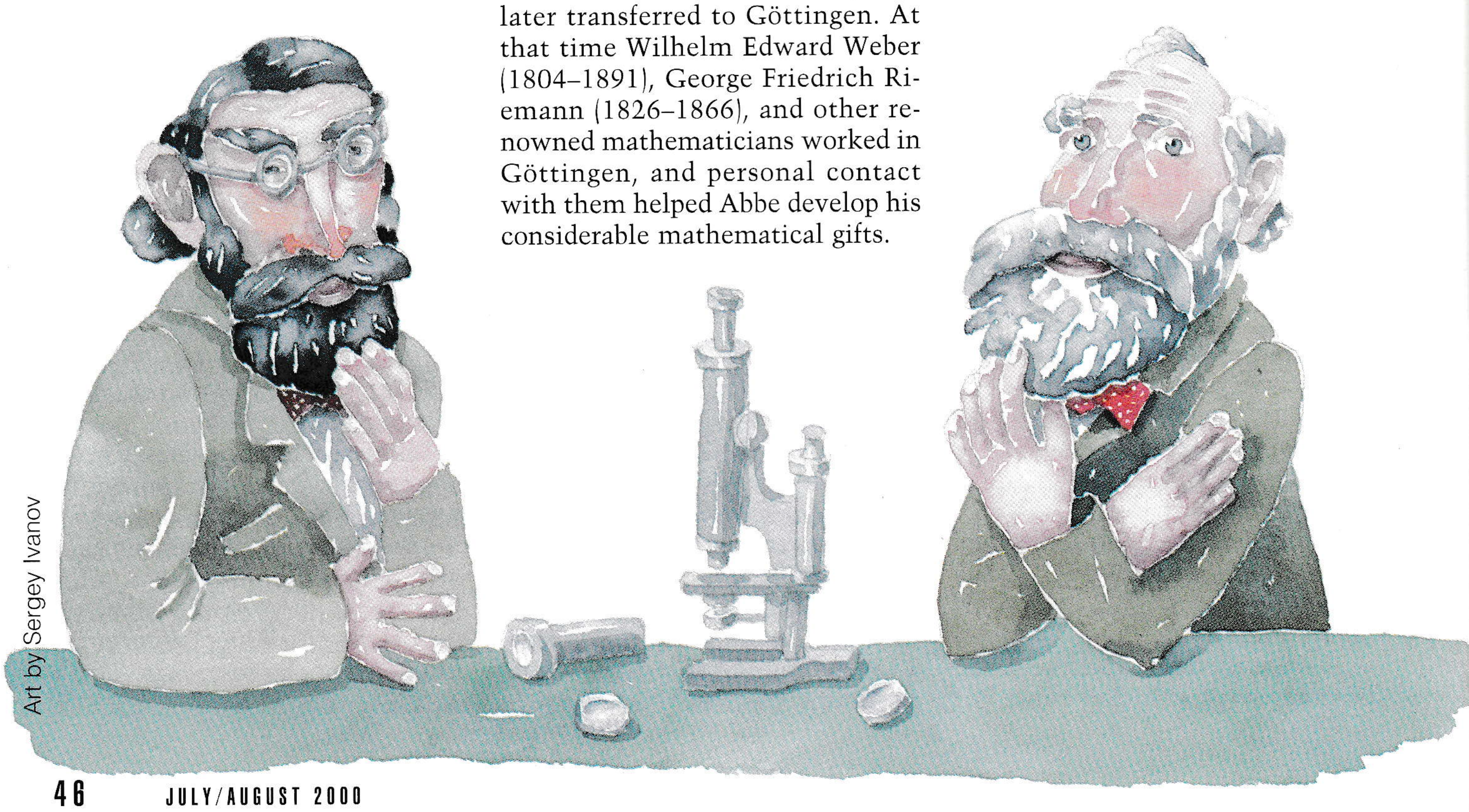
ANYONE IN THE FIELD OF optics knows the name of the German scientist Ernst Abbe (1840–1905). Thanks to the work of Abbe and Carl Zeiss (1816–1888), a brilliant engineer and innovator in the commercial production of optical instruments, standards in optics were raised to a level that has remained essentially unchanged to the present day.

The period from the middle of the 19th century to the beginning of the 20th was a time of revolutionary discoveries in various fields of natural science, which enriched humankind with sophisticated new tools and methods of investigation. The demands of science and technology led to the invention of devices for observing various objects, which resulted in the rapid development of applied optics and optical engineering. The production process for scientific instrumentation was radically improved. Small workshops were replaced by scientific and manufacturing conglomerates such as "Carl Zeiss" (its modern name). Abbe's work in this firm contributed to its prosperity and made it possible to produce outstanding optical instruments.

Ernst Abbe was born in Eisenach in 1840, but he finished primary and secondary school in Jena, where he entered the local university. He later transferred to Göttingen. At that time Wilhelm Edward Weber (1804–1891), George Friedrich Riemann (1826–1866), and other renowned mathematicians worked in Göttingen, and personal contact with them helped Abbe develop his considerable mathematical gifts.

He defended his doctoral thesis in 1861, and in 1863 he became an assistant professor at the University of Jena. Abbe lived in Jena for 35 years and brought worldwide fame to the town. As a professor at the university, Abbe focused all his attention on optics: the theory of optical instruments, analytical and mathematical optics, and the technology of optical experimentation.

The period of Abbe's life from 1866 to 1888 was closely tied to the work of the legendary German optical engineer Carl Zeiss. In 1846 Zeiss founded a workshop in Jena that initially produced magnifying glasses and primitive microscopes. Very



Art by Sergey Ivanov

soon, however, Zeiss microscopes received high praise and were widely used by scientists and engineers due to the exceptional workmanship of the lenses. Beginning in 1858, Zeiss produced sophisticated microscopes, and later he added other optical instruments to his list of products. Zeiss continually strove to "base the practical design of microscopes entirely on scientific theory," so he invited a number of outstanding specialists in applied optics to work with him, Ernst Abbe being the brightest star among them.

By that time, optical tools had been manufactured in Europe for three centuries. However, production was based mainly on intuition and traditional workmanship. One of the inventors of a two-lens microscope with a biconvex objective and a biconcave eyepiece was the great Italian scientist Galileo Galilei (1564–1642). The father of the modern microscope is Cornelius Drebbel (1572–1634), whose microscope consisted of a biconvex objective and a plano-convex eyepiece. A fundamental improvement was made by Robert Hooke (1635–1703), who in 1663 inserted the third "collecting" lens between the objective and the eyepiece. The next step was taken in 1716 by Hertel, who added a rotating stage with a mirror beneath it to reflect light. This led to better illumination of the object and a clearer image. This is essentially the microscope design that is used today.

Any further improvement in image quality would have to be made by eliminating defects in the optical system—above all, spherical and chromatic aberrations. Because of spherical aberration, paraxial rays (rays traveling near the optic axis) pass through different parts of a lens and cross the optic axis at different points, causing the image of a point source to look like a nonhomogeneously illuminated disk. Chromatic aberration causes a ray of white light to split into a number of rays of different colors, which cross the optic axis at different points because the focal length of a lens depends on the wavelength

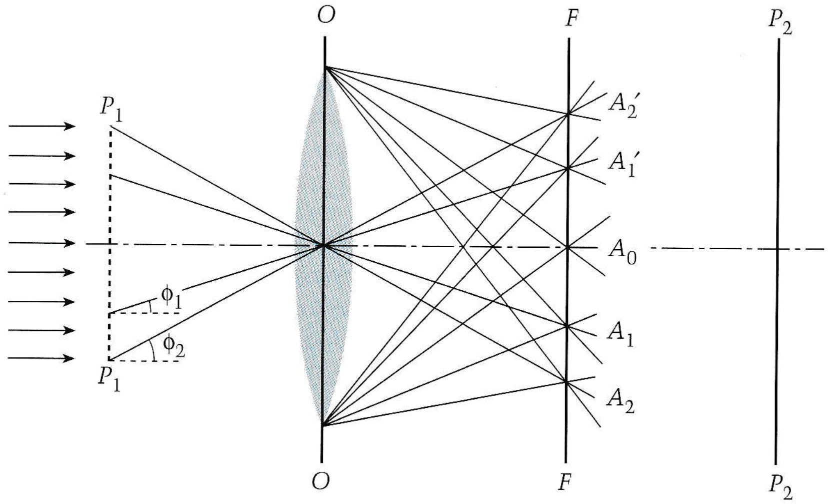


Figure 1. *Abbe's microscope diffraction theory: P_1P_1 is the plane of the object, FF is the focal plane of objective OO, P_2P_2 is the image plane, ϕ_k is the diffraction angle, and A_k is the diffraction maxima in the focal plane.*

of the incident light. This phenomenon is known as dispersion.

From the 17th through the 19th centuries, investigators tried not only to improve the quality of the images formed by microscopes, but also to construct a microscope with the greatest possible magnification. It's known that the magnification of a microscope increases as the focal length of its objective decreases, so opticians started to work with short-focal-length objectives. In addition, the resolving power of a microscope depends on its aperture—that is, the angle between the outermost beams from the object to the edge of the objective. An aperture of almost 180° was achieved by the middle of the 19th century. However, the short-focal-length and wide-aperture objectives suffered from even greater aberrations.

Attempts were made to improve the performance of microscopes and calculate their magnification on the basis of geometrical optics. It turned out that geometry could not completely explain the process of image formation in microscopes. This failure directed Abbe's attention to physical optics.

Abbe published his studies on microscope design in 1873. In these papers he considered the role played by the objective and the eyepiece in image formation. For the first time in optics, he classified the aberrations. But Abbe's biggest achievement was discovering the limits imposed on designers of optical systems by the wave nature of light.

Abbe explained how a lens forms the image. First an interference pattern is formed in the plane perpendicular to the axis of the lens. This is a system of alternating maxima and minima of illumination, which plays the role of a diffraction grating. The light flux passes from the lens through this grating and interacts with it. Only then does an image appear a short distance from the plane of the grating, which can be seen on a piece of frosted glass or photograph. This is how an image is formed with one lens. In a microscope, however, according to Abbe's theory, the image is obtained in two stages, shown schematically in figure 1.

In the first stage, the light illuminating an object P_1P_1 falls on the microscope's lens after being scattered and diffracted by the details of the object, so that the structure of the light beam is determined by the object. After passing through the objective of the microscope, the light beam produces a diffraction pattern in the focal plane *FF*, which is a system of illumination maxima whose angular sizes depend on the structural details of the object. The directions to these maxima are determined by the condition $nd \sin \phi$

$= k\lambda$, where n is the refractive index of the medium, d the characteristic size of the object's details, ϕ the diffraction angle, $k = 0, 1, 2, \ldots$ the number of a particular maximum, and λ the wavelength.

In the second stage, the illumination maxima are considered point sources emitting coherent beams of light. These beams mutually interfere behind the focal plane of the objective and produce an image of the object in the P_2P_2 plane. Abbe called the pattern in the focal plane of the objective the *primary image* and the pattern in the linked plane the *secondary image*.

To obtain the correct image of an object, the secondary image must be formed as a result of the interaction of the beams emitted by all the maxima of the primary image. Of particular importance are the first-order maxima situated at small angles to the principal axis and produced by the largest and usually most important details of the object being examined. The maxima corresponding to large angles are produced by smaller details of the object. Minute details of the object (smaller than the wavelength of light) cannot be seen at all, because the waves diffracting off such small details do not reach the screen, even through an objective with the largest possible aperture. This sets a limit on the resolving power of a microscope: $d \geq \lambda = \lambda_0/n$, where λ_0 is the wavelength of light in vacuum.

Usually there are no obstacles for light inside a microscope, so the number of diffraction maxima passing through the objective is limited only by its mount. The smaller the object or its detail, the larger the diffraction angles it produces (half this angle is called the aperture u) and the larger the opening of the objective must be.

If the aperture is less than the diffraction angle ϕ_1 corresponding to the first-order spectra (that is, if $\sin u < \sin \phi_1 = \lambda_0/d$), only rays from the central maximum will pass through the objective into the microscope, and we won't see an image corresponding to details whose size is of the order of d. The larger $\sin u$ is relative to λ_0/d, the more high-order spectra will contribute to the image formation, and greater detail will appear in the image.

Usually an object is illuminated not only by light beams passing along the optic axis, but also by beams at larger angles, and this improves the resolving power. If the illuminating beam makes an angle α with the microscope's axis and diffracts at an angle α_0, the condition for the maxima takes the form $\sin \alpha_0 - \sin \alpha = k\lambda/d$.

In order for the first spectrum to enter the objective completely, the following requirements must be met: $\alpha = -u$, $\alpha_0 = u$, $k = 1$. Also, $2 \sin u \geq \lambda_0/(nd)$ or $d \geq \lambda_0/(2n \sin u)$. Abbe called the value $A = n \sin u$ the "numerical aperture." According to Abbe's theory, the numerical aperture determines several important properties of a microscope—for example, the brightness of the image and the degree of similarity between the object and its image. The larger the numerical aperture of a microscope, the smaller the details in the object that it can resolve. Abbe's theory says that it is impossible to see objects in a microscope that are smaller than half the wavelength of the light illuminating them. Abbe confirmed the validity of his theory by experiment (in which the objects examined were the absorbing gratings), and in 1887 he formulated a strict mathematical theory of the microscope.

In order to improve the resolving power of microscopes, Abbe tried to increase the numerical aperture. There were three ways to do this: increase the angular aperture, increase the refractive index of the medium, or decrease the wavelength of the light beam. Even at the beginning of his optical research Abbe realized that microscopes had reached their limit in angular aperture and that this was a dead end.

The second approach looked more promising: Abbe proposed increasing the medium's refractive index n by filling the open space between the object and objective with a substance whose refractive index is greater than that of air. In 1878 Abbe and Stephenson made a microscope in which cedar oil was placed between the object and objective. Their efforts met with success: this instrument improved the resolving power by one-third.

Of particular interest are Abbe's ideas about improving the resolving power of microscopes by decreasing the wavelength of the light used to form the image—specifically, the possibility of using ultraviolet light. This idea was realized in one of the microscopes made by Abbe's colleagues in the Carl Zeiss firm not long before the death of the great inventor. Later such microscopes helped in studying the structure of DNA and RNA, the large information-bearing molecules in living organisms.

Abbe also devoted a great deal of attention to correcting aberrations in optical systems. Since the various zones of a simple lens produce an image of a plane element with different magnifications, the images of a point source formed by various zones coincide only at the optic axis of the system, while the sharpness of the image degrades sharply outside this axis. Abbe showed that all the zones of an optical system magnify an object to the same degree as long as the "sine condition" is met. This requires that for all rays emerging from a point on the axis of the optical system and then collecting after refraction at the point of an image, the ratio between the sines of the angles of the respective rays with the optic axis must be constant:

$$\sin u_1/\sin u_2 = Kn_2/n_1,$$

where n_1 and n_2 are the refractive indices of the media on the object and image sides, and K is the magnification of the optical system.

Two points that have no spherical aberration, and for which Abbe's sine condition is valid, are called aplanatic. Abbe showed that only one pair of aplanatic points exists on the axis of an optical system. He also found a simple method for determining the degree to which the sine condition is satisfied. He drew the pat-

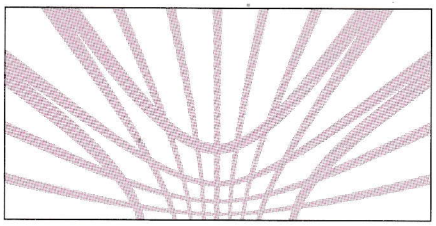


Figure 2. *Abbe's pattern for testing the Abbe sine condition.*

tern shown in figure 2, which is viewed with the optical system being tested. If the sine requirement is met, it is possible to find a location for the pattern such that the observer sees it as a rectangular grid. Abbe tested many microscope objectives made by trial-and-error by the old masters and found that the sine condition was valid for all the good objectives. Today, Abbe's sine condition is always taken into account in the design of any optical system.

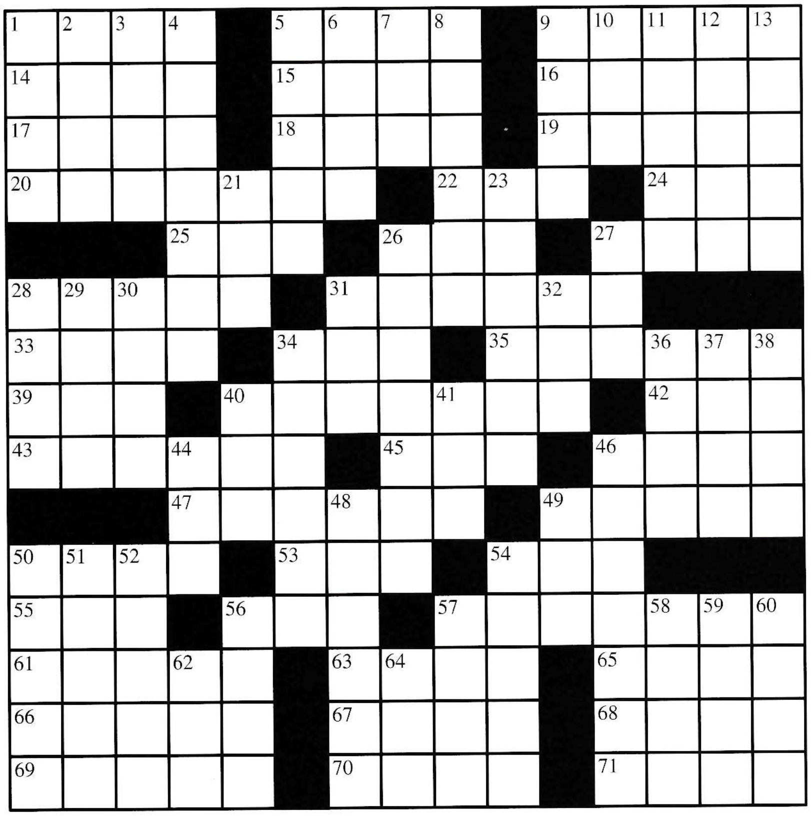
In his struggle with chromatic aberration, Abbe spared no effort in persuading the glass workshops to produce new kinds of optical glass with certain properties. To compare the properties of various types of optical glass, Abbe proposed the following method: select a number of reference wavelengths in the visible range of the spectrum and use the concept of relative dispersion (defined as the ratio of the refractive indices corresponding to the chosen wavelengths). In applied optics the value γ is known as the Abbe number. In 1873 Abbe managed to make the first objective in history that was achromatic for three colors. The coincidence of the foci for rays of three different wavelengths was achieved by using various types of glass with different Abbe numbers. Abbe called such objectives "apochromatic." In 1886 Abbe managed to design and produce an apochromatic objective in which both spherical and chromatic aberrations were virtually eliminated. It was a triplet whose outer lenses were single lenses while the inner lens consisted of three lenses glued together, each of which had a different Abbe number.

In designing optical systems Abbe always started from a theoretical analysis. It was theory that led him to the idea that an optical system must include special diaphragms that limit passage of light rays. He showed that to form an image, an optical system needs only those rays that pass through the device to the image without a delay, whereas rays that pass through only a part of the optical system (held back, perhaps, by the lens mount) are not only useless, they're harmful.

Abbe performed a great service by developing and constructing a number of new optical instruments, as well as organizing scientific research aimed at producing new types of optical glass. The Carl Zeiss firm designed and produced prismatic binoculars, new types of photographic lenses, refractometers (devices that measure the refractive index of a substance), and various optical devices to measure angular and linear values. All these instruments raised the standards of the optics industry to a higher level.

Criss × cross science

by David R. Martin



ACROSS

1 Pronto: abbr.
5 Quartz deposit in a furnace
9 "March King"
14 Point of minimum disturbance
15 Adams or McClurg
16 Type of circle or tube
17 Shebat follower
18 Hammer or wrench
19 Like Neon
20 Type of flow
22 Eisenhower's nickname
24 Author ___ Rand
25 State of matter
26 Status ___
27 Width times length
28 Cooks
31 1946 chemistry Nobelist James ___ (1887–1955)
33 ___ Descartes
34 Southern constellation
35 Type of engine
39 Abscisic acid: abbr.
40 Nickel-chromium alloy
42 Mine output
43 Electric doublet
45 Small amount
46 Pelvic bones
47 ___ paper (pH indicator)
49 Mitosis star
50 Chemical connection
53 Wave guide mode: abbr.
54 Narrow coastal inlet
55 Actress Gardner
56 Ferrous sulfide
57 Half-spin particle
61 Plane detector
63 Unlock
65 Famous lion
66 Sooty dirt
67 Musical sound
68 Baseballer Nolan ___
69 Eagle nest
70 Volt or coulomb preceder
71 Red stone

DOWN

1 ___ sphincter (certain muscle)
2 Sodium carbonate
3 Swedish botanist ___ Afzelius (1750–1837)
4 Orbital extremum
5 Alpha followers
6 Olfactory sensation
7 Silicon monoxide
8 Element 2
9 Trigonometric function
10 Yoko ___
11 Release
12 Physician Hans Hugo Bruno ___ (1907–1982)
13 Competition site
21 Sci. org.
23 Biochemist ___ Bloch (1886–1944)
26 This magazine
27 Appendage
28 Nail
29 Phosphatase unit
30 Sack or weed preceder
31 ___ Lanka
32 Sense organ
34 Minor building wing
36 Bump
37 Buffalo's lake
38 Fissure
40 Samuel's teacher
41 Biological duct
44 Aged
46 $CH_3(CH_2)_2CH_3$ and $CH_3CH(CH_3)_2$, e.g.
48 Strong nuclear particles?
49 It's mostly nitrogen
50 Large raft
51 Reproductive organ
52 Zenith's opposite
54 Civet relative
56 Gibbs ___ energy
57 Greek cheese
58 1977 Chem. Nobelist ___ Prigogine
59 Esker
60 A logic circuit
62 Friend: Fr.
64 Variable resistor (slang)

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE MAY/JUNE PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M296

It follows from the given equation that $\alpha = \alpha^3 - 1$. Therefore, $3\alpha^2 - 4\alpha = 3\alpha^2 - 3\alpha - \alpha^3 + 1 = (1 - \alpha)^3$. Thus, the cube root is equal to $(1 - \alpha)$. Let's transform the expression under the second root. We have $2\alpha^2 + 3\alpha + 2 = \alpha^2 + (\alpha^2 + \alpha) + 2(\alpha + 1) = \alpha^2 + \alpha(\alpha + 1) + 2(\alpha + 1) = \alpha^2 + \alpha^4 + 2\alpha^3 = (\alpha^2 + \alpha)^2 = \alpha^2(\alpha + 1)^2 = \alpha^8$. The second term in the sum is $\alpha^3 = 1 + \alpha$. Thus, the given expression equals 2.

M297

Let us ask how much of each old piece could have gone into the making of the larger new piece. If all of the old 2 kg piece were used, then the larger new piece must contain 0.5 kg of the 1 kg piece. If all of the old 1 kg piece were used, then 1.5 kg of the old 2-kg piece was used. In either case, the new 2.5 kg piece contains at least 0.5 kg of each old piece.

The reader is invited now to show that at least one of the original pieces contains no more than 40% copper.

Armed with these two propositions, we can now ask: what is the maximum possible percentage of copper in the 2.5-kg piece? We obtain it if we melt 0.5 kg containing 40% copper with 2 kg of 100% copper. That is, the maximum possible value is

$$\frac{0.5 \cdot 0.4 + 2}{2.5} 100\% = 88\%.$$

This is exactly the percentage of copper given in the condition of the problem for the 2.5-kg piece. Thus the original pieces were 40% and 100% copper, respectively.

M298

For definiteness, consider the case when angle ABC is obtuse (see figure 1). (The case when this angle is acute can be treated similarly.) Let P be the point symmetric to M about point B. We first prove that triangles ABC and BPN are similar. Note that $\angle ABC = \angle PBN$, since they are obtained by adding a right angle to the angle CBN. Triangles CBM and ABN are similar right triangles. Therefore,

$$\frac{BP}{BN} = \frac{BM}{BN} = \frac{BC}{AB}.$$

Thus triangles BPN and ABC are similar.

Triangle BPN can be obtained from triangle ABC by rotating ABC by an angle of 90° followed by a dilation centered at B. Under this transformation, the line AC goes to the perpendicular line PN. Thus BN is perpendicular to AC. But $BT \perp AC$ by construction, so $BT \parallel PN$, and BT passes through the midpoint of side PM in triangle PMN. A theorem of elementary geometry says that a line parallel to one side of a triangle and passing through the midpoint of a second side must pass through the midpoint of the third side as well. It follows that BT passes through the midpoint of MN.

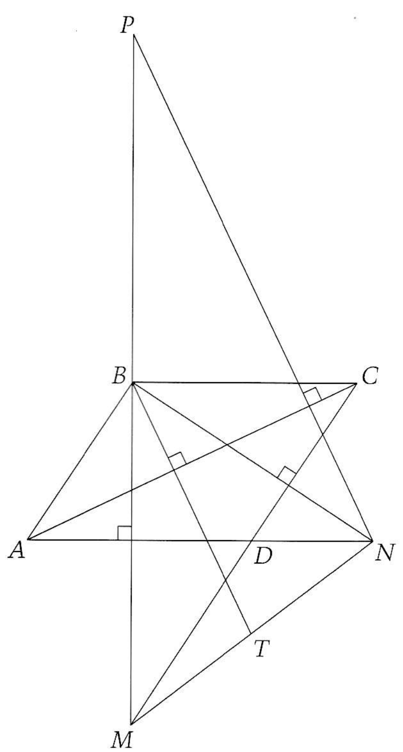


Figure 1

M299

Suppose that we have already constructed the desired circle C (see figure 2a). We draw a third circle, centered at C, with a radius equal to the sum of the radii of the other two circles. This new circle will then pass through the center of the given circle, and is inscribed in an angle (A_1 in the figure) with its parallel to

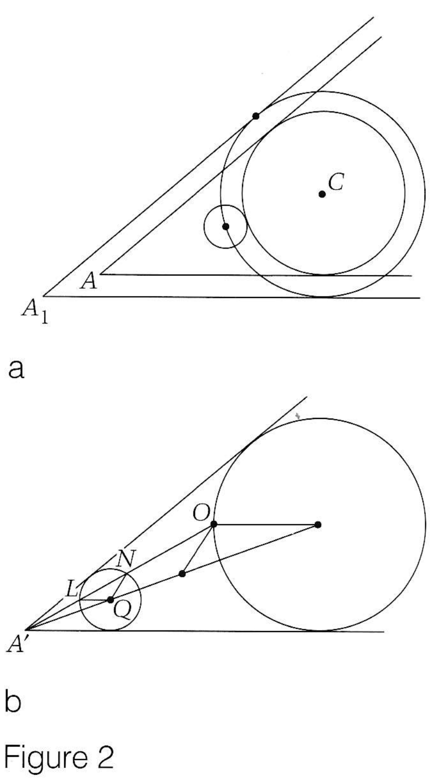


Figure 2

the given angle at a distance equal to the radius of the original circle.

If we can construct this third circle, we can then shrink its radius by a known amount to find a circle which solves our original problem. So we have reduced our problem to that of constructing a circle (the new one) tangent to a given angle (angle A_1), and passing through a given point inside the angle (the center of the original circle).

This problem can be solved by similarity, as shown in figure 2b.

(1) First, we inscribe an arbitrary circle ω centered at Q in the given angle.

(2) We find the points where the line $A'O$ intersects the circle ω; call them L and N.

(3) Draw the lines parallel to LQ and NQ through point O and find the intersections of this line with the bisector $A'Q$ of angle A'. These points are the centers of the desired circles.

The original problem has, in general, four solutions. We have shown how to get the two new circles that are tangent externally to the given circle. The construction of the circles that are tangent internally is left to the reader, as is the analysis of the special cases which result when the given circle is itself tangent to one or both sides of the given angle.

M300

We first obtain a simple and useful formula.

Lemma: Suppose two spheres, with radii x and y, are tangent externally, and each is also tangent to some plane, at points A and B, respectively. Then,

$$AB = 2\sqrt{xy}.$$

Proof: Let the centers of the spheres be O_1 and O_2. Figure 3a shows a cross section of the situation, taken through the plane determined by the O_1, O_2, A, and B. (The reader can prove that these four points are in fact coplanar, and is invited to recall that the line connecting the centers of two tangent circles passes through their point of tangency.) Note that triangle O_1O_2C is a right triangle, $CO_2 = AB$, $O_1O_2 = x + y$, and that $CO_1 = |x - y|$. Then, by the Pythagorean theorem,

$$CO_2 = \sqrt{O_1O_2^2 - CO_1^2}$$
$$= \sqrt{(x+y)^2 - (x-y)^2} = 2\sqrt{xy}.$$

(Note that the result for this lemma is really a property of tangent circles, not tangent spheres.)

Before proceeding to the stated problem, we solve a simplified version. Suppose we remove the mid-sized sphere (in the original problem statement). What then is the minimum possible radius for a sphere tangent to plane L, line m, and the unit sphere touching plane L and line m?

In figure 3b, we again see a cross perpendicular to plane L through the centers of the two given spheres (and through line m). It is clear from this figure that the center of the minimum sphere lies on the plane of this cross-section. If the radius of this minimal sphere is r, then our lemma shows that $MK = 2\sqrt{r}$, and we know that $OM = 1$, $KO = r$. But $OM = MK + KO$, so $1 = r + 2\sqrt{r}$. From this equation, we find that $r = 3 - 2\sqrt{2}$.

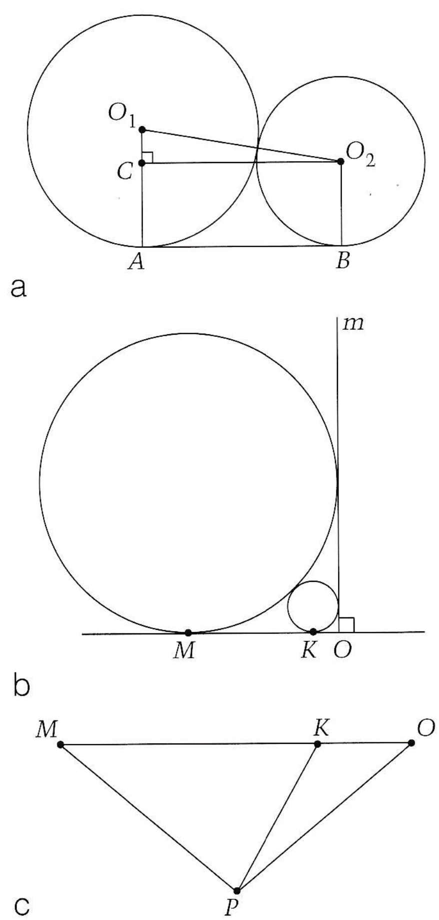


Figure 3

Now let us turn to the given problem by re-introducing the middle sphere. It's clear that the radius of the smallest sphere cannot be any less than the value for r found above. We will show that we can in fact introduce a middle sphere such that the radius of the smaller sphere is exactly r.

If such a middle sphere exists, then it must touch plane L at some point P. We will find a point P and a radius R for this middle sphere, which makes it touch the other two spheres in our auxiliary problem.

Figure 3c shows the situation on plane L. Points M, K, and O are as in figure 3b, and point P is the point where our new sphere is tangent to plane L. If there is such a point P, then it is not hard to show that $OP = R$, and our lemma tells us that $MP = 2\sqrt{R}$, and $KP = 2\sqrt{Rr}$. As before, $OM = 1$ and $OK = r$.

We apply the law of cosines, letting $\cos \angle POK = \lambda$. In triangle POK, we have $4Rr = R^2 + r^2 - 2\lambda Rr$. In triangle POM, we have $4R = R^2 + 1 - 2\lambda R$. Multiplying the second equation by r, and subtracting it from the first equation, we find that $R = \sqrt{r}$.

Now we can construct our middle sphere. It is not hard to see that if we assign the line segments lengths according to the computations above that $MO < MP + OP$. Thus we can construct triangle MOP to find point P, and the radius R will be the correct value so that a sphere tangent to plane L at P with radius R will be tangent to the other two spheres. Since $R < 1$ and greater than $r = 3 - 2\sqrt{2}$, this last value for r is the smallest possible.

Physics

P296

Let's displace the arc through a very small angle ϕ. The restoring torque of the force of gravity relative to the pivot point is determined by the "surplus" of mass m on one side and the "deficit" on the other:

$$2mgR\sin\alpha = 2\left(\frac{M}{L}R\phi\right)gR\sin\alpha,$$

where 2α is the angle between the strings and M is the mass of wire arc.

The moment of inertia of the system relative to the pivot point can readily be calculated since all the parts having mass are located the same distance R from the reference point:

$$I = MR^2.$$

The next step is to write down Newton's second law for rotational motion:

$$MR^2\phi'' = -\frac{2MR^2 g\sin\alpha}{L}\phi,$$

from which we obtain the period of oscillation:

$$T = 2\pi\sqrt{\frac{L}{2g\sin\alpha}},$$

where the angle α is

$$\alpha = \frac{L}{2R}.$$

When the arc is small, we can set $\sin\alpha \cong \alpha$ and obtain the usual expression for the period of a mathematical pendulum as expected.

This problem has an elegant solution that is based only on energy conservation and doesn't need the value of the moment of inertia. *Hint*: compare the maximum values of the potential and kinetic energies and recall a similar relationship for harmonic oscillations.

P297

Let m be the mass of the Venusian atmosphere, $M = 48$ g/mole the molar mass of ozone, and $R = 8.31$ J/(mol · K) the gas constant. At the planet's surface the ozone layer occupies a volume $V = 4\pi r^2 h$ at a pressure P and temperature T. According to the statement of the problem, $P = mg/4\pi r^2$, where r is the radius of the planet. On the other hand, the equation of state for ozone is

$$PV = \frac{\alpha m}{M}RT.$$

Plugging the formulas for V and P into this equation, we obtain the thickness of the ozone layer:

$$h = \frac{\alpha RT}{gM} \approx 1.7\cdot 10^{-3}\ \text{m}.$$

P298

For the given charges, the electric field outside the capacitor is not zero (in contrast to the case when the total charge on the plates is zero). Any rearrangement of the plates modifies only the internal field in the capacitor; it doesn't disturb the external field. The outer foils of both plates (figure 4) collect equal charges of the same polarity, each of which is half the net charge of the capacitor (in a "correctly" charged capacitor this half-sum is zero). In our case the half-sum is $1.5Q$.

Therefore, the inner foils of the plates carry charges $-0.5Q$ and $+0.5Q$. The internal field is determined only by these residual charges, because in this region the fields generated by the external charges cancel. The energy of the field located between the plates can be calculated as usual:

$$W_1 = \frac{(Q/2)^2}{2C} = \frac{Q^2}{8C}.$$

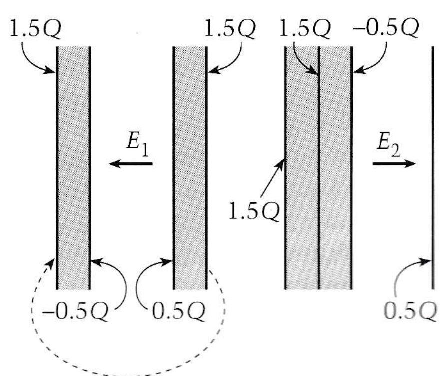


Figure 4

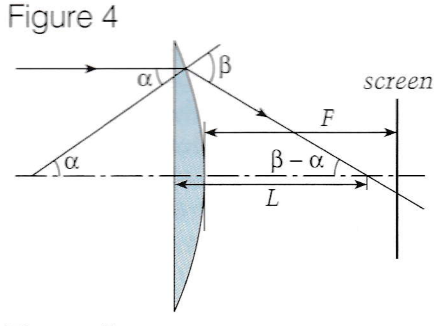


Figure 5

After the outer foil of the plate with charge $2Q$ is disconnected, the charge of the outer plate remains on it, and we carry this charge onto the other side of the capacitor. Now the charges of the plates of the modified capacitor become $2.5Q$ and $0.5Q$. The field between the plates changes direction (which is not essential for the energy calculation) and increases two-fold. Therefore, the energy of the field located between the plates increases by a factor of four and becomes

$$W_2 = \frac{Q^2}{2C}.$$

The outer field doesn't change, so our work was expended on increasing the internal field between the plates. Thus the work necessary for the charge transfer is

$$W = W_2 - W_1 = \frac{Q^2}{2C} - \frac{Q^2}{8C} = \frac{3Q^2}{8C}.$$

P299

According to the statement of the problem, the lens is placed in a way that simplifies our calculations: the parallel beam hits the flat side of the lens perpendicular to its surface and doesn't refract. Therefore, we should consider refraction only at the spherical boundary between the glass and air. To begin, we find the thickness d of the lens along the optic axis (where it's thickest):

$$R^2 = \left(\frac{D}{2}\right)^2 + (R-d)^2,$$

from which we get

$$d = 0.67\ \text{cm}.$$

The thickness of the lens is important because we'll measure the distances from various points on the surface of the lens. Thus a narrow (diaphragm-restricted) pencil of light parallel to the optic axis is focused at a distance

$$F = \frac{R}{n-1} = 10\ \text{cm}.$$

Now let's consider the ray farthest from the optic axis (figure 5). The angle of incidence for this ray measured relative to the radius drawn to the point of refraction at the spherical surface is $\alpha = 30°$, since $\sin\alpha = (D/2)/R = 0.5$. The angle of refraction can be obtained from Snell's law: $\sin\beta = n\sin\alpha = 0.75$, from which we get $\beta = 48.6°$. Simple calculations yield the point on the principal axis crossed by the ray after refraction. It's located at a distance $L = (D/2)\cot(\beta - \alpha)$ from the flat side of the lens. Taking into account the thickness of the lens, we find that the outermost rays of the beam intersect 3.2 cm from the screen, and so the diameter of the light spot is about 2.2 cm.

It would be interesting to investigate the question: Are there rays that produce a spot with a larger diameter than that produced by the outermost rays considered here?

P300

Since momentum and energy of the system before nuclear fusion are zero, the newly formed particles fly off in opposite directions with numerically identical momenta (this follows from the law of conservation of momentum):

$$p_n = \sqrt{2m_n E_n} = p_\alpha = \sqrt{2m_\alpha E_\alpha}.$$

Energy conservation requires

$$E = E_n + E_\alpha.$$

Solving this system of equations simultaneously, we get

$$E_\alpha = \frac{m_n}{m_\alpha + m_n} E \approx 3.5 \text{ MeV},$$

$$E_n = \frac{m_\alpha}{m_\alpha + m_n} E \approx 14.1 \text{ MeV}.$$

Brainteasers

B296

The drawing is incorrect. The quadrilaterals shown would not lie in a plane.

To see this, note that we can say the following about any three planes: either (1) they intersect in a point, in which case the lines of intersection of any two of them go through this point, or (2) they do not intersect in a point, in which case the lines of intersection of some pair of them are parallel. But the 'lines of intersection' of the 'planes' of the three quadrilaterals shown do not intersect in a point, and no two are parallel (figure 6).

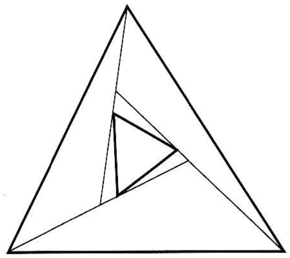


Figure 6

Challenge: What could the teacher's pentahedron have looked line?

B297

This situation is possible (see figure 7).

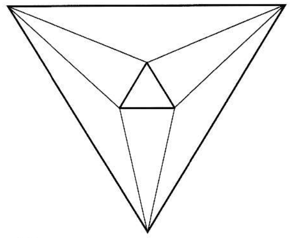


Figure 7

B298

Suppose there were more chemists than alchemists. Since the total number of participants was 100 (an even number), the number of chemists was not less than 51; then there were not more than 49 alchemists. Thus there was at least one chemist among those who answered the question, and she must have said that chemists were more numerous. Or suppose there were more alchemists than chemists. Then there was at least one alchemist among those who answered the question, and he must have said that chemists were more numerous. Therefore, 50 chemists and 50 alchemists attended the conference.

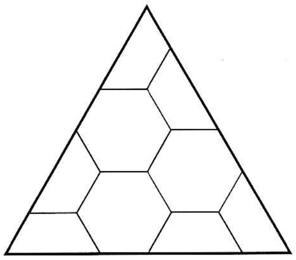


Figure 8

B299

See the figure 8.

B300

When the thermometer was taken out of the molten tin, the temperature of the glass bulb dropped sharply. The bulb shrank and squeezed the mercury, causing it to rise in the glass column. To quantify this effect, compare the coefficients of thermal expansion for glass and mercury.

Corrections

Vol. 10, No. 4

p. 7, col. 2: Third line after the second display equation:

for "x_i" read "ξ_i".

Vol. 10, No. 5

p. 36, col. 2: The formulas should be:

$$y = a\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a}.$$

$$\left(-\frac{b}{2a}, \frac{4ac - b^2}{4a}\right).$$

Chores

by Don Piele

EACH SPRING, THE USA COMPUTING OLYMPIAD staff is faced with the job of creating a challenging set of informatics problems for our national competition. These problems are similar to the programming challenges presented at the International Olympiad in Informatics. They focus on tasks that can be solved with an efficient algorithm that will quickly dispose of the 10 test data sets. An optimal program must find a solution for each data set within a few seconds to receive the maximum score. Less efficient programs receive partial credit, depending on how many test cases it can solve within the time limit. The scoring is done with an automated grading system.

Creating the problems is anything but automated. The entire staff participates by submitting new problem ideas to the head coach—Rob Kolstad. A list of approximately ten possible problems are considered, and five problems are selected for the competition. Then the fun begins: every problem gets "cowified."

Cowification is the process of transforming an ordinary looking programming task into a barnyard chore. It is also a license to work in some of our "dairy state" humor. Presented in this column is one of the easiest tasks from our recent US Open Competition.

The problem

Farmer John's family pitches in with the chores during milking time. They round up the cows, put them in the stalls, wash the cows' udders, and perform many other tasks. Organizing the chores and completing them as quickly as possible is always desirable, because it leaves more time for hang gliding with the cows. Of course, some chores cannot be started until others have been completed. For instance, it is impossible to wash a cow's udder until a cow is in the stall, and you wouldn't want to attach the milking machine until you have washed the cow's udder. Farmer John has created list of N chores that must be completed. Each chore requires an integer number of minutes to complete, and there might be other chores to be completed before this chore can be done (i.e., prerequisites). At least one chore has no prerequisite: the very first one, numbered 1. Farmer John's list is nicely ordered, and chore K ($K > 1$) can have only chores 1, ..., $K-1$ on its dependency list. Write a program that reads both a list of chores from 1 through N with associated times and a list of chore prerequisites. Calculate the shortest time it will take to complete all N chores. Of course, chores that don't depend on each other can be performed simultaneously in parallel. In fact, a large number of chores could be taking place simultaneously.

INPUT FORMAT:

Line 1: One integer, N, the number of chores ($3 \le N \le 10{,}000$)

Line 2, ..., $N+1$: N lines, each with several integers:

The chore number (1, ..., N, supplied in order in the input file).

The length of the chore in minutes ($1 \le \text{length} \le 100$).

A list of no more than 100 prerequisite chores, if any are needed.

SAMPLE INPUT (file CHORES.IN): 7

```
1 5
2 1 1
3 3 2
4 6 1
5 1 2 4
6 8 2 4
7 4 3 5 6
```

OUTPUT FORMAT: A single line with a single integer that is the least amount of time required to perform all the chores. SAMPLE OUTPUT (file CHORES.OUT): 23.

The sample INPUT file is represented graphically in figure 1. The chores are numbered 1 to 7 and highlighted

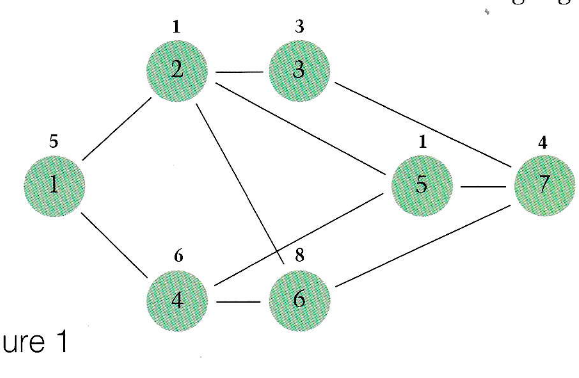


Figure 1

in green. The time required for the chore appears above the green dot. Lines between chores denote a dependency. Lower numbered chores must be completed before the higher numbered chores can be done.

It is clear from figure 1 that chores 1, 4, 6, and 7, which follow a line of dependency, will take 5 + 6 + 8 + 4 or 23 minutes to complete. This is the worst case and hence the shortest time that it will take to do all chores.

The solution

What we are looking for is an algorithm that guarantees a solution without summing up all chore paths. To sum up all chore paths would be, in the worst case, of order $(n-1)!$ for n chores if chore i depended on nearly all previous chores for $i = 1$ to n. A more efficient solution uses recursion and works as follows. If we have only one chore to complete, the solution is known: the time it takes to complete chore one. Now assume we know the minimum time it takes to complete chore $k-1$ and all dependent chores. The shortest time that chore k can be completed is equal to that maximum of all shortest times each dependent chore can be completed plus the time it takes to complete chore k. These times are indicated below in blue as they are generated for each chore from $k = 1$ to 7. Each frame in figure 2 shows one step in the process of finding the minimum time it would take to complete both a chore and all its dependent chores.

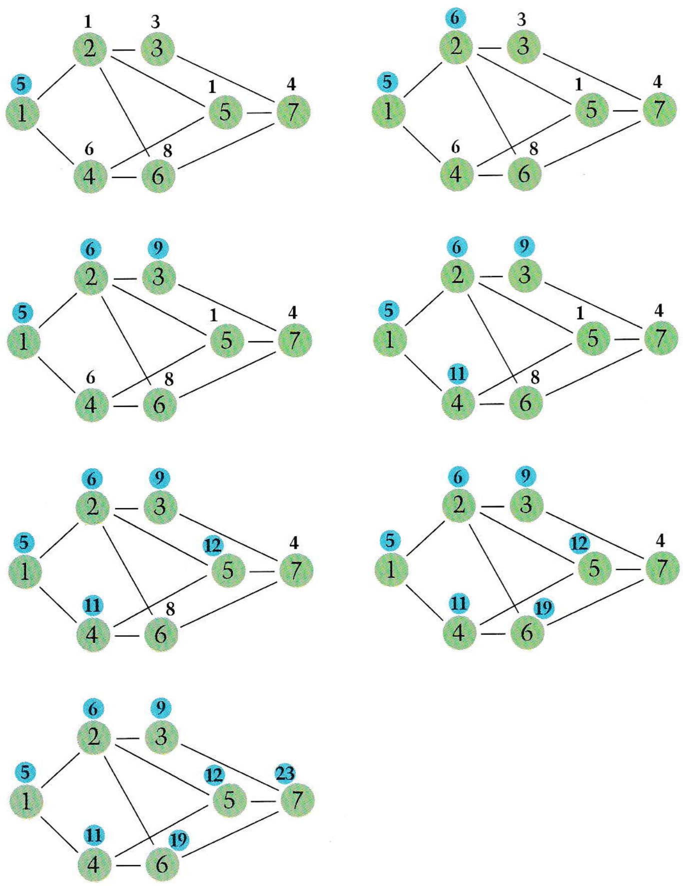


Figure 2

We now have the minimum length of time it would take to complete each chore and all dependent chores for all n chores. The solution to the problem is simply to take the maximum of all these values, which in this case is 23.

Pseudo code

Here is the pseudo code which encapsulated the algorithm described above.

1) Initialize

Chore[k] = Time to complete chore k.

DependentChores[k] = List of chores that must precede chore k.

2) Recursive algorithm for computing MinTime[k] = Minimum time needed to complete chore k and all dependent chores.

For $k = 1$ to n

MinTime[k] = Chore[k] + Max[MinTime[j], $j \in$ DependentChores[k]]

3) Solution = Max[MinTime[k], for $k = 1$ to n]

Mathematica code

Now for the actual code in *Mathematica* which mirrors the pseudo code:

```
(* INPUT file *)
CHORES = {{1, 5}, {2, 1, 1}, {3, 3, 2},
{4, 6, 1}, {5, 1, 2, 4}, {6, 8, 2, 4},
{7, 4, 3, 5, 6}};

(* n = number of chores *)
n = Length[CHORES];
(* pick off chore times *)

Chore[k_] := CHORES[[k, 2]]
(* pick off dependent chores *)

DependentChores[k_] := CHORES[[k,
Range[3, Lenght[CHORES[[k]]]]]]
(* compute MinTime to complete each
chore and all dependent chores *)

MinTime[1] = Chore[1];

AllMinTimes = Table[MinTime[k] =
Chore[k] + Max[MinTime /@
DependentChores[k]], {k, 2, n}]

{6, 9, 11, 12, 19, 23}

Max[AllMinTimes]

23
```

Your turn

Certain sequences of chores follow a path of dependency and take the full time needed to complete all chores. These sequences are called critical paths. There can be more than one critical path. Your chore is to modify the code ever so slightly to find a critical path. In our example, {1, 4, 6, 7} is a critical path because chore 7 depends on chore 6 which depends on chore 4 which

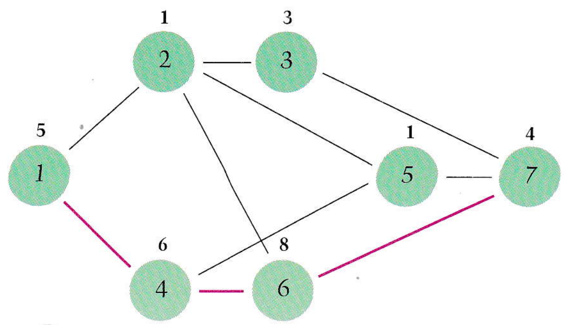


Figure 3

depends on chore 1, and the total time to complete all chores is 23: the minimum time needed to complete all chores. Your output should be a list of chores. Figure 3 shows graphically a critical path.

2000 US Open

The 2000 US Open, held in April of this year, attracted 378 entries from 34 countries, including 223 from the United States. For US students this is the final competition of the year leading up to the selection of the fifteen finalists. The finalists will spend an all-expenses-paid eight days at the University of Wisconsin-Parkside in June competing for one of four spots on the USA Computing Olympiad team. The finalists this year are:

Reid Barton, Arlington, MA; John Danaher, Springfield, VA; Vladimir Novakovski, Springfield, VA; Percy Liang, Phoenix, AZ; Yuran Lu, Presque Isle, ME; Jacob Burnim, Silver Spring, MD; Steven Sivek, Burke, VA; Jack Lindamood, Dallas, TX; George Lee, San Mateo, CA; Gary Sivek, Burke, VA; Richard Eager, Falls Church, VA, Tom Widland, Albuquerque, NM; Gregory Price, Falls Church, VA; Kevin Caffrey, Oakton, VA; Thuc Vu, Anaheim, CA.

The USACO team will have the opportunity to represent the United States in the Central European Olympiad in Informatics, August 24–31, in Romania (http://ceoi.ubbcluj.ro) and the International Olympiad in Informatics in Beijing, China, (http://www.ioi2000.org.cn), September 23–30, 2000.

The complete listing for all participants in the 2000 US Open can be found on the USACO website at www.usaco.org.

Finally

Waiting for two months to see a solution is not necessary, thanks to the Internet. All solutions to the problems presented in this column are available at the Informatics website:

http://www.uwp.edu/academic/mathematics/usaco/informatics/.

Index of Advertisers

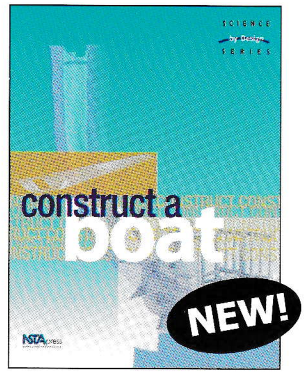
The new *Science by Design* series from NSTA

Construct-A-Glove

Think thermodynamics is beyond your students' grasp? Engage student interest in heat transfer and insulation with this volume. A challenging, hands-on opportunity for students to compare the function and design of many types of handwear and to design and test a glove to their own specifications. Students learn the basic principles of product design while exploring principles of physics and technology necessary to construct and test an insulated glove. **#PB152X1**

Construct-A-Boat

How do boats work? Why do they float? Explore principles of buoyancy, hull design, scale modeling, and seaworthiness. In this volume, students investigate the physics of boat performance and work with systems and modeling. Through research, design, testing, and evaluation of a model boat, students experience the practical application of mass, speed, and acceleration while applying the math and science necessary to build a scale model of a boat. **#PB152X2**



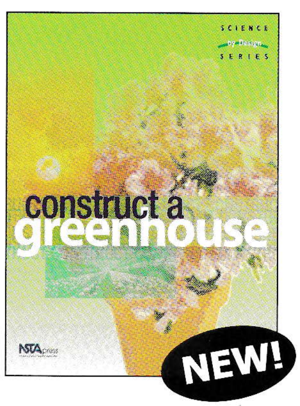
Science by Design Series

This series offers a hands-on approach for students to successfully develop and carry out product design. Each book includes teacher-tested units that introduce students to the design process and sharpen students' abilities to investigate, build, test, and evaluate familiar products. All four volumes are keyed to the *National Science Education Standards*, the *Benchmarks for Science Literacy*, and the *International Technology Education Standards*. Produced in cooperation with TERC. (Grades 9–12, 2000)

Each volume in series	**$19.95**	
Set of all four volumes	**$63.95**	***SAVE 20%!***

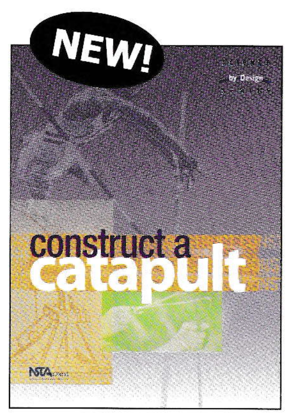
Construct-A-Greenhouse

How can physics help your garden grow? Engage your students in a problem-solving challenge to design and build a physical system that provides an optimal environment for plant growth. This volume helps you cultivate student interest in optics, energy transfer, and photosynthesis. In addition to learning and applying concepts in thermodynamics, light absorption, and plant biology, students must make a range of decisions as they encounter cost constraints, construction alternatives, and environmental changes while building a greenhouse model. **#PB152X3**



Construct-A-Catapult

How much can you launch? How far will it go? Catapult into physics and technology with the heavy weaponry of the Middle Ages. This volume integrates history, physics, mathematics, and technology in its challenge to students to design and build a working catapult system. Students investigate elasticity, projectile launching, and learn about frequency distribution while working through the process of product design. **#PB152X4**



P700Q1

To order, call 1–800–722–NSTA or order online www.nsta.org/scistore