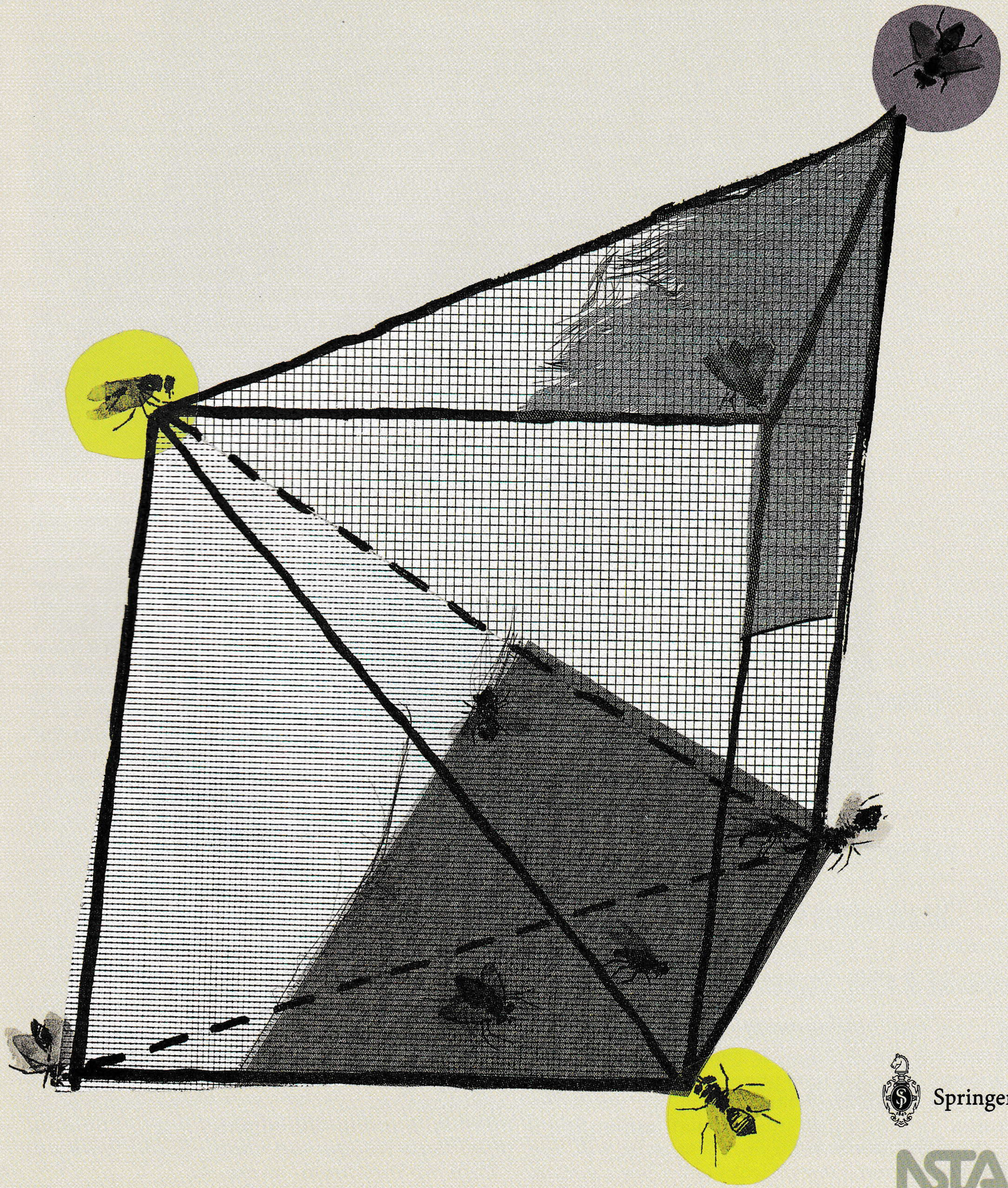
QUANTUM

JULY/AUGUST 1999

US $7.50/CAN $10.50



Springer

NSTA

Oil on canvas, 50 1/2 × 39. Samuel H. Kress Collection. © 1999 Board of Trustees, National Gallery of Art, Washington, D.C.

Interior of the Pantheon, Rome (c. 1734) by Giovanni Paolo Panini

AT THE TIME PANINI PAINTED THIS PICTURE, THE giant 8 m wide oculus in the middle of the Pantheon's immense dome had been peering up at the sky for 17 centuries. The dome, which is 43 m in diameter, remained the largest such structure until modern times.

With the bright disk of light shining on the wall, this grand architectural marvel looks rather like a pinhole camera, or camera obscura. Like the Pantheon, the camera obscura has been around for a long time. In fact, it is thought to have been used by the ancient Greeks. The camera obscura, as many people know, is quite useful in safely observing solar eclipses, but it also turns out to be handy for observing sunspots, as we will see in "Light in a Dark Room" on page 40.

QUANTUM

JULY/AUGUST 1999

VOLUME 9, NUMBER 6

Cover art by Yuri Vaschenko

If it takes you so long to solve a geometry problem that the problem starts attracting flies, then it's probably time to seek help. Fortunately, help is just a few pages away.

One useful method for avoiding geometrical flies in the ointment is to extend the sides of the given figure to create a new figure. For a full explanation of this method and more on dealing with pesky polyhedrons, turn to "Completing a Tetrahedron" on page 47.

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

NSTA *Quantum* (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 9 (6 issues) will be published in 1998–1999. *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. Printed in U.S.A.

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: $47. 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) Jay Feinman (212) 460-1682; 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.

Just for the fun of it!

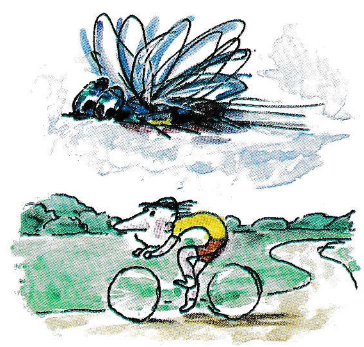
B266

Chess champs. Teresa, Robbie, and Alex played a chess tournament as follows. Two of them played a game, then the other played the winner, and so on. (If the game ended in a draw, the player who played white pieces is considered to have lost.) At the end of the tournament, Teresa had played 15 games, Robbie had played 9 games, and Alex had played 14 games. Who played in game number 13?

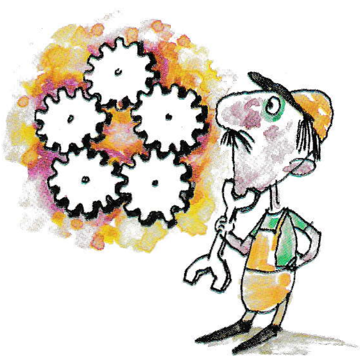
B267

Nesting triangles. The two legs of one isosceles triangle are equal to the two legs of another. Is it ever possible to place one of these triangles completely inside the other?

B268



Frequent flyer. Two cyclists, Kaitlin and Josh, simultaneously started toward each other from two towns 40 km apart. Josh rode at 23 km/h, and Kaitlin rode at 17 km/h. Before departure, a fly landed on Josh's nose. At the moment of departure, it started to fly toward Kaitlin at 40 km/h. When it reached Kaitlin, it immediately flew in the opposite direction at 30 km/h (the wind blew toward Kaitlin). As soon as the fly reached Josh, it turned back again, and so on. Find the total distance flown by the fly until the cyclists met (the speed of the fly was constant in each direction).



B269

Get it in gear. Five gears arranged as shown cannot rotate. Is it possible to arrange 101 gears in such a way that each of them meshes with two adjacent gears and that if one wheel rotates, then all the others rotate? (The axes of adjacent wheels do not have to be parallel.) (D. Anisov)

B270

Electrified sphere. Wires are connected to diametrically opposite points on a homogeneous metal ball. In what cross-section of the ball will the electric current produce the most heat?

ANSWERS, HINTS & SOLUTIONS ON PAGE 52

Art by Pavel Chernusky

Arithmetic obstacles

Can you get there from here?

by N. Vaguten

IN MANY MATHEMATICAL theories, applied problems, and puzzles, the following questions often arise: Is it possible to move from one position to another using certain "approved" operations? How can a desired sequence of moves be found if it exists, or how can it be proved that the desired transition does not exist? We will consider several problems of this type. These problems have another common feature: they involve integers, and the obstacles that inhibit certain transitions are usually of an arithmetic nature.

The examples that begin the discussion of each problem can be easily understood even by younger students. Exercises marked with an asterisk and proofs of the general results invite the reader to ponder. A number of difficult "olympiad-type" problems are given at the end of the article, and the last problem lies close to the theory of arithmetic groups, which is progressing rapidly at the present time.

Problem of a chess knight

Problem 1. The natural numbers m and n are given. A chess piece located on an infinite checkerboard can make L-shaped moves consisting of m squares in one direction and n squares in the perpendicular direction. We call this piece an $\{m, n\}$-knight. Which squares can this knight visit?

The common chess knight ($\{1, 2\}$-knight) can reach any square starting from an arbitrary square O. Indeed, it can reach any square adjacent to O in three moves, and it is clear that any other square can be reached by making a sequence of such elementary steps.

However, the $\{1, 3\}$-knight, whose moves are illustrated in figure 1, cannot reach the square adjacent (horizontally or vertically) to the starting square. This fact can be easily explained: the $\{1, 3\}$-knight always remains on squares of its starting color. On the other hand, it's easy to show that the $\{1, 3\}$-knight can reach any square of its starting color. Indeed, it requires three moves to get to the diagonally adjacent square (fig. 1), and any other square of the same color can be reached by making such elementary steps.

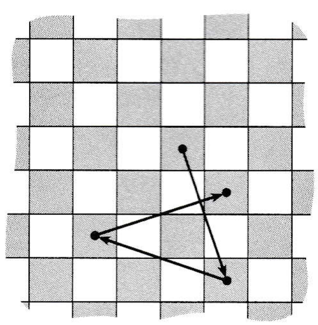


Figure 1. The $\{1, 3\}$-knight can reach any diagonally adjacent square: these elementary steps can be used to visit all squares of the same color.

Try to solve problem 1 for the following numbers m and n: (a) (2, 5), (b) (3, 7), (c) (10, 25), (d) (19, 79).

It turns out that the $\{m, n\}$-knight can reach any square if and only if m and n are of opposite parity and their greatest common divisor is 1.

The complete answer to problem 1 is given at the end of section 3 (exercise 10). Now we consider a simpler problem, the result of which is useful for the knight problem and for other, more serious, mathematical problems.

Representation of the GCD

We consider a one-dimensional analogue of our problem.

Problem 2. The natural numbers a and b are given. A move consists in adding or subtracting one of the numbers a or b to (or from) a certain integer. Given some number c, is it

Art by Sergey Ivanov

possible to obtain this number c from 0 using such moves?

In this problem, $\mathbf{Z}$ (the set of all integers) is the set of "positions."

First, consider a particular example. Suppose that a buyer and a cashier have an infinite number of bank notes of 10 and 25 dollars each (anything can happen in mathematical problems!). It is clear that the buyer can pay c dollars if and only if c is a multiple of 5.

Here is another example. Assume that moves ±19 and ±79 are allowed on the set $\mathbf{Z}$. Using such moves, we can obtain any integer: a combination of such moves makes it possible to change one's position by 3 and then by 1 (fig. 2a).

Here is a more complicated example: $a = 819$ and $b = 367$. In this case, the same method allows us to find the shortest displacement that results from a succession of moves. This shortest displacement equals 21 (fig. 3a). Thus, we can make any displacement that is a multiple of 21. On the other hand, both a and b are divisible by 21, so no other transitions are possible.

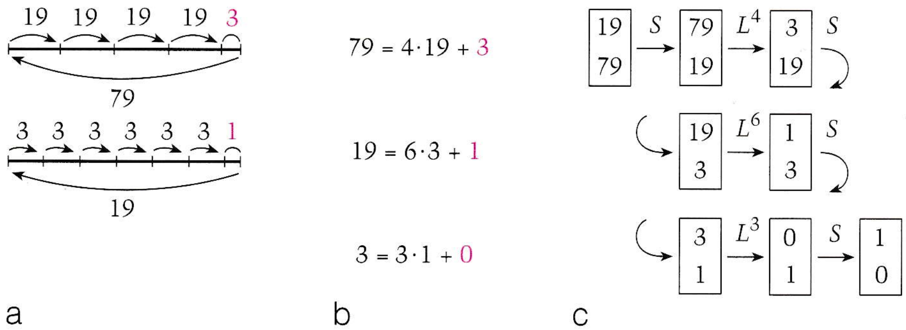


Figure 2. The numbers 19 and 79 are relatively prime (they are called coprimes); therefore, several divisions with remainder give the remainder 1 = GCD(19, 79).

Notice that 21 is the greatest common divisor of 819 and 367.

Exercise 1. (a) Prove that if a buyer and a cashier have an infinite number of bank notes of 3 and 5 dollars each, the buyer can pay any amount of dollars.

(b) Is it possible to move from 0 to 1000 if $a = 123$ and $b = 456$; and if

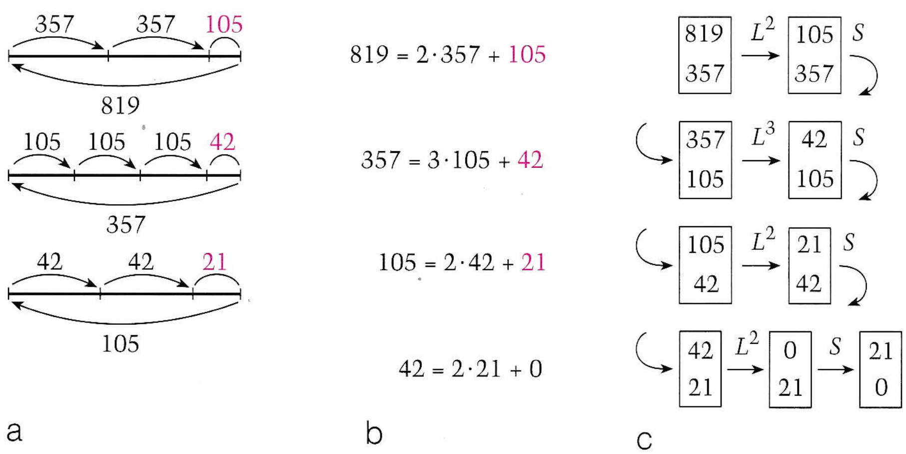


Figure 3. To find GCD(819, 357) = 21 using Euclid's algorithm, four steps are needed.

$a = 589$ and $b = 1984$?

(c) What transitions are possible if $a = 18$ and $b = 81$?

Now we formulate the answer to problem 2 in the general case. *Let the greatest common divisor of* a *and* b *(GCD) be* d. *Then, the transition from 0 to a number* c *is possible if and only if* c *is divisible by* d. Try to prove this fact.

This proposition will be obtained in another form in the next section.

Exercise 2. Prove that the answer to problem 2 does not change if we allow a only to be added and b only to be subtracted.

3. Is it possible to weigh the following amounts on a scale using weights of 36 and 60 grams: (a) 150 g, (b) 132 g?

Euclid's algorithm

In the following problem, the set of positions consists of all pairs of integers.

Problem 3. Three machines print pairs of integers on cards. Every machine reads a card (x, y) and prints a new card: the first one prints the card $(x - y, y)$, the second one prints the card $(x + y, y)$, and the third (y, x). Let the initial card be (1, 2). Is it possible to obtain the pairs (19, 79) or (819, 357) using the machines in any order desired? Which cards can be obtained if the initial card is (a, b)?

Denote the operations of the machines by L, R, and S, respectively. Once again, we begin with numerical examples. The pair (19, 79) can be obtained from the pair (1, 2). To find the desired sequence of machine operations, it would be more convenient to descend from (19, 79) to (1, 2) rather than to ascend in the opposite direction (fig. 4).

Then, write down the sequence of operations in reverse order (exchanging L with R) to obtain the desired ascent from (1, 2) to (19, 79).

In figure 4, the descent from (19, 79) to (1, 2) is continued to the pair (1, 0). The abridged notation of this process is shown in figure 2c, where L^k means that the operation L is performed k times. As a matter of fact, the same descent was performed in the example to problem 2 (fig. 2a).

An abridged notation of the descent from pair (819, 357) is shown in figure 3c (we invite the reader to write the full notation). In this example, the pair (1, 2) (and (1, 0)) cannot be obtained. This fact can be easily explained: there is an obstacle that prevents us from reaching this pair: all numbers obtained in the process are divisible by 21. No matter in what order we apply the operations L, R, and S, we cannot overcome this obstacle, since every operation used preserves the GCD:

$$\begin{aligned} GCD(x - y, x) &= GCD(x + y, x) \\ &= GCD(x, y). \end{aligned}$$

Therefore, we cannot reach the pair (1, 2) starting from (819, 357) and vice versa.

Exercise 4. Is it possible to perform the following transitions using the operations L, R, and S: (a) from (1, 10) to (5, 25), (b) from (18, 81) to (36, 63), (c) from (589, 1984) to (31, 1953)?

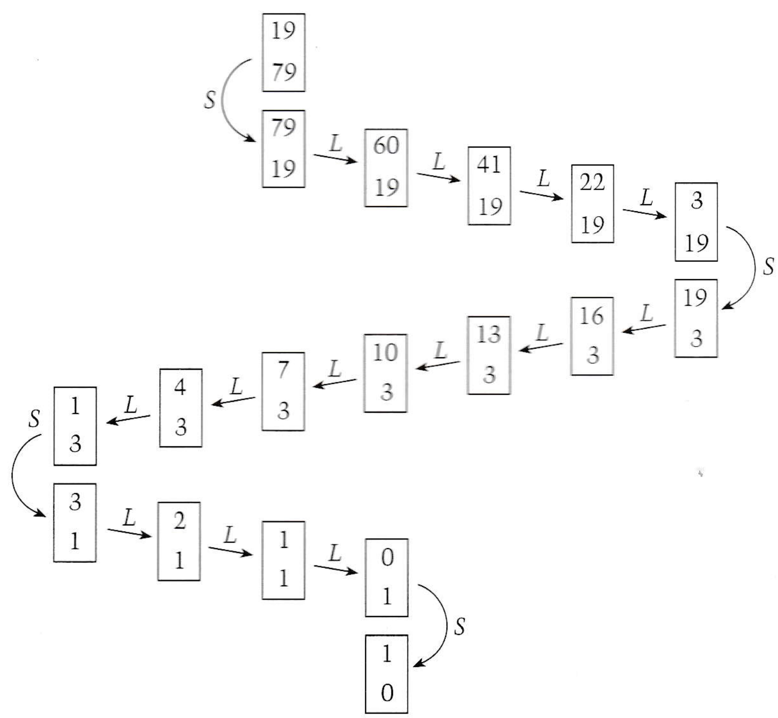


Figure 4. Every "flight of stairs" of this staircase represents one step of Euclid's algorithm.

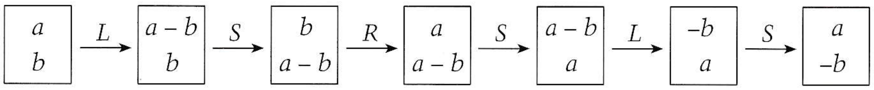


Figure 5. Operations *L*, *R*, and *S* can be used to change the sign of one of the numbers on the card.

Now we can answer the general question of problem 3: a pair (p, q) can be obtained from (a, b) if and only if $\text{GCD}(a, b) = \text{GCD}(p, q)$. This condition is necessary because the operations used preserve the GCD. It is also sufficient: if $\text{GCD}(a, b) = \text{GCD}(p, q) = d$, each of these pairs can be reduced to the pair $(d, 0)$ by a sequence of operations *L*, *R*, and *S*. Therefore, descending from (a, b) to $(d, 0)$ and then ascending from $(d, 0)$ to (p, q), we obtain the desired sequence of operations.

Let us prove that any pair (a, b) can be turned into $(d, 0)$. Notice that if one of the elements of the pair is negative, it can be easily made positive (fig. 5). Now, any pair (a, b) with natural numbers a and b can be turned into $(d, 0)$ using the same method that we used in the above examples: at each step (except for the operation *S*) the greater element of the pair is reduced until we reach the pair $(d, d) \to (0, d) \to (d, 0)$.

In solving problem 3, we have already found a convenient method for constructing the greatest common divisor of two numbers. First, from the pair (a, b), where $a > b > 0$, we move to the pair (b, r), where r is the remainder in the division of a by b. Then, we repeat the same operation until we reach a pair $(d, 0)$. The last nonzero remainder d is the desired $\text{GCD}(a, b)$ (fig. 2b and 3b). This method is called *Euclid's algorithm*.

Exercises

5. Prove that it is impossible to obtain the pair (1234, 5678) from (1357, 2468) and (7890, 1979) from (123, 457).

6. Find examples to demonstrate that *L* and *S*, as well as *R* and *S*, are not commutative: $LS \neq SL$ and $RS \neq SR$ (however, it is clear that $LR = RL$).

7. Using Euclid's algorithm, find GCD(589, 1984) and GCD(123456789, 987654321).

The set of all points on the plane with integer coordinates is called the *integer lattice*; it is denoted by $\mathbf{Z}^2$.

The following exercise and figure 6 illustrate the geometric sense of problem 3.

8. Let a segment OA, where O is the origin of coordinates and A is a point of the integer lattice $\mathbf{Z}^2$, be divided by other points of the lattice into d parts. Prove that point A can be moved to points $(d, 0)$ and $(-d, 0)$ by operations *L*, *R*, and *S*, but cannot be moved to any other point of axis Ox.

Exercises 9 and 10 generalize problems 1 and 2.

9*. Let n natural numbers $a_1, a_2, \ldots, a_n$ be given. Prove that an integer c can be obtained from 0 by moves $\pm a_1, \pm a_2, \ldots, \pm a_n$ if and only if c is divisible by $\text{GCD}(a_1, a_2, \ldots, a_n)$.

10*. (a) Let n vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ with integer coordinates be given in the plane. Prove that the set of all points D of the plane to which point O can be taken by moves $\pm\mathbf{v}_1, \pm\mathbf{v}_2, \ldots, \pm\mathbf{v}_n$ is the set of nodes of an oblique lattice (an oblique lattice is the set of the vertices of parallelograms formed by two equidistant families of parallel lines).

(b) Assume that with any vector $\mathbf{v}_i$, the set $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ contains a vector $\mathbf{v}_j$ perpendicular to $\mathbf{v}_i$ and equal to it in length. Then, the set of all accessible points D is the set of nodes of a square lattice.

Now, it is not difficult to find the complete answer to problem 1. Let $m = dm_1$ and $n = dn_1$, where $d = \text{GCD}(m, n)$. Then, if $m_1 + n_1$ is odd, all points (dx, dy), where x and y are any integers, are accessible (the lattice with step d); if $m_1 + n_1$ is even, all points (dx, dy) with $x \in \mathbf{Z}$, $y \in \mathbf{Z}$, and $x + y$ even are accessible (these points make up a lattice with step $d\sqrt{2}$ rotated 45° with respect to $\mathbf{Z}^2$).

Summing up

In the final part of the article, we present two more rather difficult

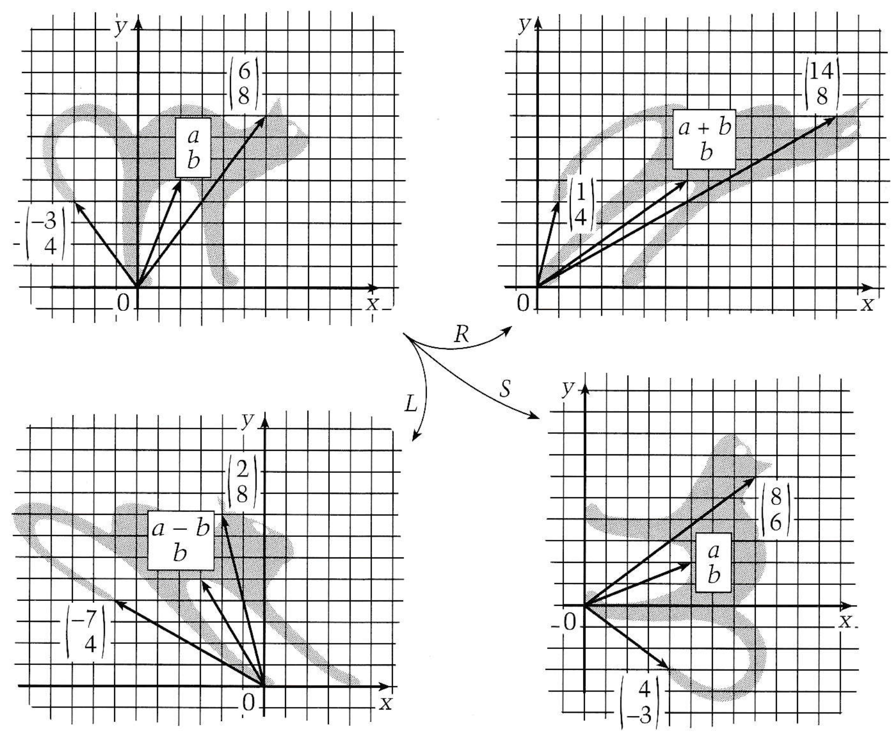


Figure 6. "Left skew" $L(x, b) \to (x - b, b)$, "right skew" $R(x, b)$ $(x + b, b)$, and "symmetry" $S(x, b) \to (b, x)$ are linear transformations that perform one-to-one mappings of the integer lattice Z^2 onto itself.

problems. However, we first review the preceding sections and try to formulate general rules that help us decide if a transition from one position to another is possible. We also present a number of mathematical terms for the notions we met in the problems.

1. To prove that a transition is impossible, we found a certain "obstacle," that is, a characteristic feature (called an *invariant*) of the position that remains the same under all admissible moves and is different for the initial and final positions. Thus, the proof of the impossibility of a transition was reduced to finding an appropriate invariant. For the {1, 3}-knight problem, we used the color of the square as such an invariant; for problem 2, it was the remainder in the division of the given number by the GCD(a, b); for problem 3, it was the GCD of the initial pair of numbers.

2. To find the desired sequence of moves on the lattice, it is often useful to find an elementary *key* move (or combination of moves) or reduce the problem to a simplest standard position and then formulate a general algorithm for finding the necessary moves. For example, in the {1, 3}-knight problem, it was sufficient to learn how to make diagonal moves, and in problem 2 it was sufficient to construct the "descent" to the standard position (d, 0).

3. In the problems that we have considered, all transitions were invertible: if position B could be reached from A, then A could be reached from B. In such problems, the entire set of positions can be partitioned into *equivalence classes:* every position of a class can be reached from any other position of the same class, but transitions between positions of different classes are impossible.

We now consider a problem that does not possess the property of reversibility. However, principles 1 and 2 do the job brilliantly.

Getting rid of twos

Problem 4. Three machines print pairs of natural numbers on cards as

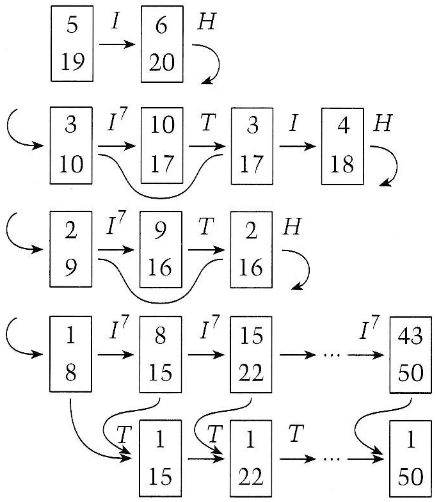


Figure 7. Using the operations I^q (increment by q), H (division by two), and T (transitive transformation), we can obtain card (1, 50) from (5, 19).

follows. Having read a card (a, b), the first machine prints card (a + 1, b + 1), the second one prints card (a/2, b/2) (it works only if both a and b are even), and the third machine reads two cards (a, b) and (b, c) and prints card (a, c). Assume that we begin with card (5, 19). Using the machines described above in any desirable order, is it possible to obtain card

(a) (1, 50) or
(b) (1,100)?

(c) Assume that we begin with card (a, b) ($a < b$). For which n can the card (1, n) be obtained?

Denote the operations that are performed by the machines by I, H, and T, respectively. Figure 7 shows how the "simplest" card (1, 8) can be obtained from (5, 19); then, the desired card (1, 50) can be obtained (as earlier, we write I^k when operation I is applied k times). Thus, the answer to problem 4a is affirmative.

However, the card (1, 100) cannot be obtained from (5, 19). We can find an obstacle by considering figure 7: *the difference of the numbers on every card is divisible by* 7. Regardless of the order in which the machines are used, this property remains true because it is preserved by all the operations I, H, and T. (This is obvious for I. For H, if a and b are even and $b - a$ is divisible by 7, then $b/2 - a/2$ is also divisible by 7. For T, if both differences $b - a$ and $c - b$ are divisible by 7, then $c - a = (c - b) + (b - a)$ is also divisible by 7.) However, the difference 100 – 1 = 99 is not divisible by 7. Thus, the answer to problem 4b is negative.

Exercise 11. Is it possible to use machines I, H, and T to obtain:

(i) cards (5, 29), (1, 101), and (1, 1978) from the card (3, 33)?

(ii) cards (3, 33), (1, 100), and (1,

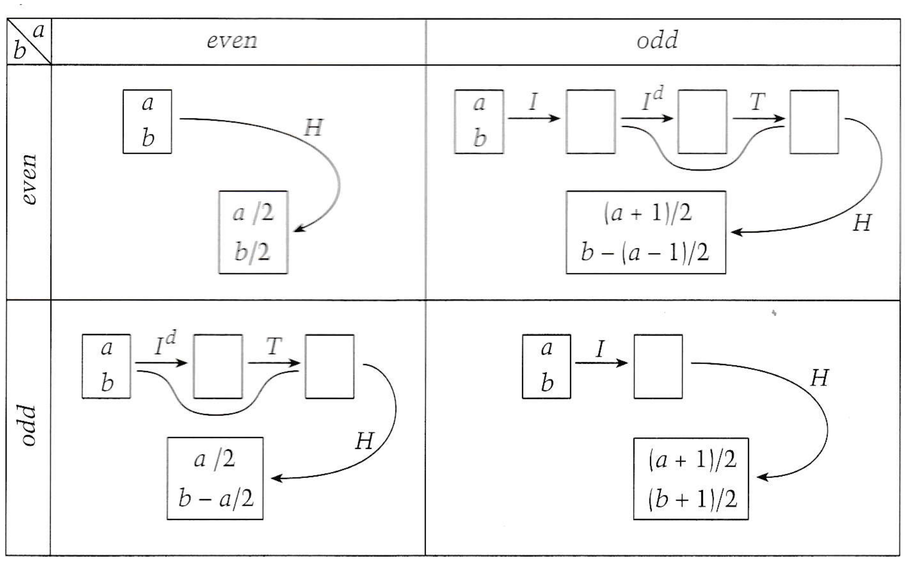


Figure 8. The greater number, b, on the card (a, b) can be decreased (here $d = b - a$, $a > 1$ or $a = 1$ and b is odd).

1979) from the card (5, 29)?

Now answer the general item (c) of problem 4: let the card (a, b) be such that $b - a = 2^m d$, where d is odd and greater than 0. Then only cards (p, q) in which the difference $p - q$ is divisible by d can be obtained from the card (a, b). Thus, we can eliminate all twos in the factorization of $b - a$ into prime numbers, but an odd divisor is an insurmountable obstacle.

Indeed, as we have already mentioned for $d = 7$, if the difference of the numbers in the card is divisible by an odd number d, operations I, H, and T yield cards that possess the same property. On the other hand, if $b - a = 2^m d$, where d is odd, we can obtain the card $(1, d + 1)$ from the card (a, b). One step of the sequence needed for such a transition is shown in figure 8. As soon as the card $(1, d + 1)$ is obtained, we can easily obtain any card of the form $(1, kd + 1)$ (as in problem 4a) from the card (1, 8) and then, any card of the form $(l, kd + l)$ with the difference that is a multiple of d.

Exercise 12. (a) Assume that the machine that performs operation T has broken down. Which cards can be obtained from (5, 19) and (5, 26)?

(b) Assume that the machine that performs operation H has broken down. Which cards can be obtained from (a, b)?

Exercise 13*. Which cards can be obtained by the operations I, H, and T from n given cards $(a_1, b_1), \ldots, (a_n, b_n)$?

Problem 4 looks rather artificial. Thus, it is interesting to note that it appeared as a lemma in a serious mathematical book (Ulam, S. *Unsolved Mathematical Problems*).

Pairs of vectors

The following problem is an extension of problem 3. The same three operations L, R, and S are used. However, in problem 3 they were applied to pairs of integers, and now we consider pairs of vectors (a, b) and (c, d) with integer coordinates as "positions" and apply these operations to both vectors of the pair simultaneously. It is convenient to write the coordinates of both vectors as two columns:

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

Thus, they make up a table of four numbers; in mathematics, such tables are called *matrices*.

Problem 5. The following operations may be performed on the matrix $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$:

$$L: \begin{pmatrix} a & c \\ b & d \end{pmatrix} \to \begin{pmatrix} a-b & c-d \\ b & d \end{pmatrix},$$

$$R: \begin{pmatrix} a & c \\ b & d \end{pmatrix} \to \begin{pmatrix} a+b & c+d \\ b & d \end{pmatrix},$$

$$S: \begin{pmatrix} a & c \\ b & d \end{pmatrix} \to \begin{pmatrix} b & d \\ a & c \end{pmatrix}.$$

Is it possible to obtain from the matrix $\begin{pmatrix} 1 & 5 \\ 2 & 7 \end{pmatrix}$ the following matrices: (a) $\begin{pmatrix} 1 & 3 \\ 2 & 9 \end{pmatrix}$, (b) $\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$, (c) $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, (d) $\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$, (e) $\begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}$?

Which matrices can be obtained from the matrix $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$?

Two matrices are called *equivalent* if one of them can be transformed into the other by operations L, R, and S (these operations are invertible, so all matrices are partitioned into equivalence classes).

We run into several difficulties when solving problem 5, and we will overcome them one by one.

(1) The matrices $\begin{pmatrix} 1 & 5 \\ 2 & 7 \end{pmatrix}$ and $\begin{pmatrix} 1 & 3 \\ 2 & 9 \end{pmatrix}$ are not equivalent: the second vector $\begin{pmatrix} 5 \\ 7 \end{pmatrix}$ cannot be transformed into $\begin{pmatrix} 3 \\ 9 \end{pmatrix}$, since GCD(5, 7) ≠ GCD(3, 9). In general, a condition necessary for the equivalence of two matrices $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ and $\begin{pmatrix} p & r \\ q & s \end{pmatrix}$ immediately follows from the solution to problem 3:

$$\begin{cases} GCD(a,b) = GCD(p,q), \\ GCD(c,d) = GCD(r,s). \end{cases} \qquad (*)$$

However, as we will see, this condition is not sufficient for two matrices to be equivalent. If this condition holds, we may divide each column of the matrix by its GCD and consider such *reduced* matrices (recall that the GCD of each column is preserved under our operations.)[1]

(2) Matrices $\begin{pmatrix} 1 & 5 \\ 2 & 7 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$ are not equivalent because any operation (L, R, and S) on the matrix $\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$ results in a matrix with equal columns: $\begin{pmatrix} p & p \\ q & q \end{pmatrix}$.

(3) Matrices $\begin{pmatrix} 1 & 5 \\ 2 & 7 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ are also not equivalent. Here we have another obstacle: the quantity

$$\Delta = \Delta\begin{pmatrix} a & c \\ b & d \end{pmatrix} = |ad - bc|$$

is preserved under all transformations L, R, and S. Let us check it for L:

$$(a - b)d - b(c - d) = ad - bc$$

(the reader is invited to check this property for R and S). Since $\Delta\begin{pmatrix} 1 & 5 \\ 2 & 7 \end{pmatrix} = 3$ and $\Delta\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1$, these matrices are not equivalent. Notice that

$$\Delta\begin{pmatrix} p & p \\ q & q \end{pmatrix} = 0.$$

The quantity $ad - bc$ occurs very often in various problems concerning matrices—it is called the *determinant* of the matrix $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$.

(4) The matrices $\begin{pmatrix} 1 & 5 \\ 2 & 7 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ are equivalent: the sequence of transformations that reduces the first vector $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ to the standard form $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and a number of additional tricks give the desired result (fig. 9). The invariant D is clearly seen in figure 9—it is the area of the parallelogram constructed on the vectors $\begin{pmatrix} a \\ b \end{pmatrix}$ and $\begin{pmatrix} c \\ d \end{pmatrix}$.

It remains for us to find out whether or not the matrices $\begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ are equivalent. It turns out that they are not equivalent, al-

[1]The term reduced appears quite natural if we consider the matrix $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ as a pair of fractions $(a/b, c/d)$.

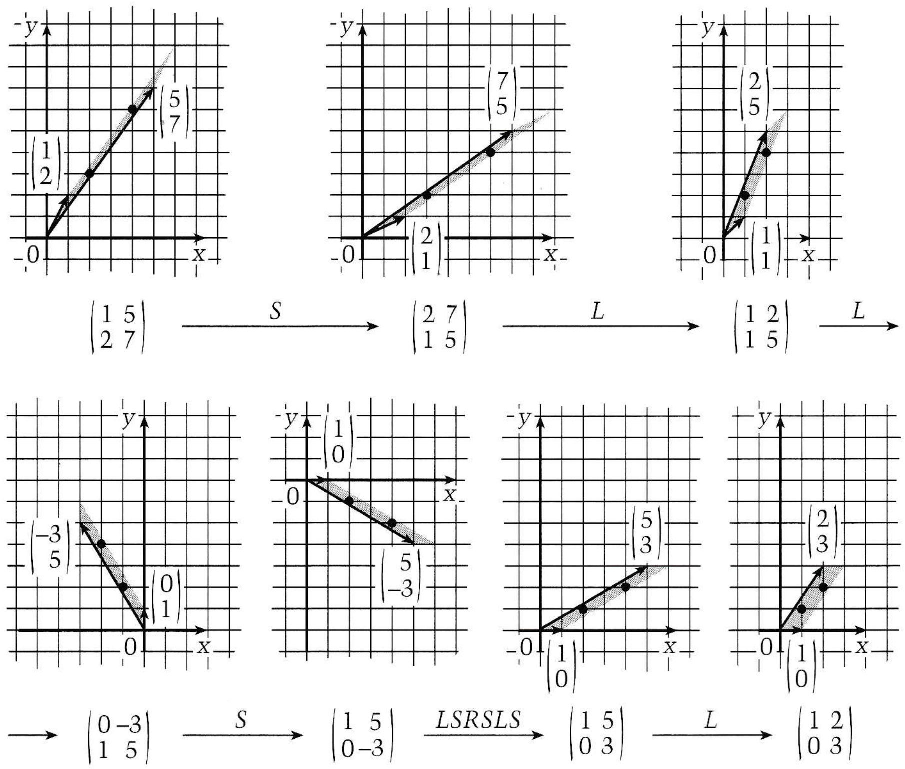


Figure 9. Reduction to the canonical form. Under all transformations L, R, and S, the area of the parallelogram and the location of the nodes of the integer lattice inside the parallelogram remain the same.

though it is not easy to find the corresponding obstacle in this case. Its geometric sense is clear from figures 9 and 10.

Now we are able to answer the general question of problem 5. Using operations L, R, and S, any reduced matrix can be transformed to a *canonical form*:

$$\begin{pmatrix} 1 & r \\ 0 & \Delta \end{pmatrix}, \text{ where } 0 \le r < \Delta, \quad \text{GCD}(r, \Delta) = 1 \qquad (1)$$

or

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \text{ if } \Delta = 0. \qquad (2)$$

Two matrices are equivalent if and only if conditions (*) are satisfied and the corresponding reduced matrices have identical canonical form. (The equivalence criterion is formulated in exercise 20 in a slightly different form.)

Indeed, any reduced matrix can be transformed to the canonical form in the same way as we transformed the matrix $\begin{pmatrix} 1 & 5 \\ 2 & 7 \end{pmatrix}$ (fig. 9). The fact that r is an invariant follows from exercises 14 and 15.

Exercises

14*. Let the canonical form of matrix $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ be (1). Then, parallelogram $OABC$ constructed on vectors $\mathbf{OA} = (a, b)$ and $\mathbf{OC} = (c, d)$ contains $\Delta - 1$ points with integer coordinates. All of these points M_1, M_2, . . ., $M_{\Delta-1}$, can be obtained using the following vector equalities:

$$\mathbf{OM}_j = \left\{\frac{j}{\Delta}\right\}\mathbf{OC} + \left\{\frac{j(1-r)}{\Delta}\right\}\mathbf{OA},$$

where $\{x\}$ denotes the fractional part of x.

15. Let GCD(a, b) = GCD(c, d) = 1 and $\Delta = |ad - bc| \ne 0$. Then, a unique r exists such that $0 \le r < \Delta$, GCD(r, Δ) = 1, both numbers $ra - c$ and $rb - d$ are divisible by Δ, and this number r is preserved under transformations L, R, and S of the matrix $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$.

The assertion of this exercise can be conveniently used to calculate r if Δ is relatively small.

16. Which of the following matrices are equivalent?

$$\begin{pmatrix} 5 & 1 \\ 7 & 2 \end{pmatrix}, \begin{pmatrix} 1 & -5 \\ 2 & 7 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 9 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 4 & 8 \end{pmatrix}, \begin{pmatrix} 4 & 7 \\ 5 & 8 \end{pmatrix},$$
$$\begin{pmatrix} 6 & 14 \\ 17 & 39 \end{pmatrix}, \begin{pmatrix} 5 & 13 \\ 19 & 50 \end{pmatrix}, \begin{pmatrix} 19 & 1 \\ 79 & 4 \end{pmatrix}, \begin{pmatrix} 39 & 60 \\ 50 & 77 \end{pmatrix}?$$

17. We have not considered matrices that have a zero column. When are such matrices equivalent?

18. Prove that any two matrices with $\Delta = 1$ are equivalent.

19*. How many classes of nonequivalent matrices with $\Delta = 3$, $\Delta = 4$, $\Delta = 5$, $\Delta = 10$, and $\Delta = 12$ exist? How many classes of reduced matrices are among them? For each class, show the location of integer nodes in the corresponding parallelogram (as in fig. 10).

20. Prove that for any matrix, a unique equivalent matrix of the form $\begin{pmatrix} k & l \\ 0 & m \end{pmatrix}$, where $k \le 0$, $m \le 0$, $l \le 0$, $l < m$, and $m \ne 0$ exists.

Another possible approach to problems 3 and 5 is to find out which transformations of the integer lattice can be obtained by the composition of operations L, R, and S (we solved a similar problem when we found out which translations could be obtained by the composition of the translations $\pm a$ and $\pm b$ in problem 2). It turns out that these transformations have the form $(x, y) \to (ax + by, cx + dy)$, where a, b, c, and d are integers and $|ad - bc| = 1$. However, this is a topic for another article dedicated to linear algebra.

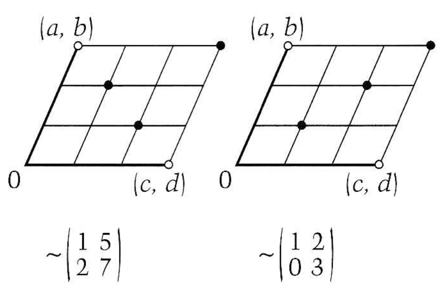


Figure 10. The location of the nodes of the integer lattice inside the parallelograms corresponding to the matrices $\begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ is different.

HOW DO YOU FIGURE?

Challenges

Math

M266

Tower of powers. Solve the equation $x^{x^8} = 2$. (M. Volchkevich)

M267

Triangular tribulation. In a triangle ABC, angle BAC is 60°. A point P is selected inside the triangle in such a way that angles APB, BPC, and CPA are 120°. Segment $AP = a$. Find the area of triangle BPC.

M268

Working the system. Solve the following system of equations (M. Volchkevich):

$$\begin{cases} \dfrac{x+y}{1+xy} = \dfrac{1-2y}{2-y}, \\ \dfrac{x-y}{1-xy} = \dfrac{1-3x}{3-x}. \end{cases}$$

M269

Integers, naturally. If x and y are natural numbers, and the sum

$$\frac{x^2-1}{y+1} + \frac{y^2-1}{x+1}$$

is an integer, prove that each of the fractions $(x^2 - 1)/(y + 1)$ and $(y^2 - 1)/(x + 1)$ are themselves integers.

M270

Angular features. Points D and F are chosen on the bisector of angle A of a triangle ABC in such a way that $\angle DBC = \angle FBA$. Prove that (a) $\angle DCB = \angle FCA$; (b) the circle that passes through D and F and is tangent to BC is also tangent to the circle circumscribed around triangle ABC.

Physics

P266

Particular trajectory. Two interacting particles with masses m_1 and m_2 compose a closed system. Figure 1 shows the trajectory of the first particle and the positions of both particles at the moment when the velocity of the first particle was **v** and the velocity of the second particle was $-3\mathbf{v}$. Plot the trajectory of the second particle for the case $m_1/m_2 = 3$.

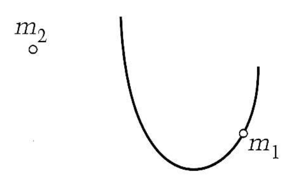


Figure 1

P267

Terrestrial and solar densities. The angle α at which the Sun is seen from Earth (the Sun's angular diameter) is about 10^{-2} rad. The radius of Earth $R_E = 6400$ km. Acceleration due to gravity on Earth's surface is $g \cong 10$ m/s^2. Using these data, find the ratio of mean densities of Earth and the Sun. Two valuable hints: 1 year $\cong 3 \cdot 10^7$ s, and the volume of a sphere is $V = (4/3)\pi R^3$, where R is the sphere's radius.

P268

Hot plate. A large, thin conducting plate with area S and thickness d is placed in a homogeneous electrical field **E**, which is perpendicular to the plate. How much heat is dissipated in the plate when the field is switched off? (P. Zubkov).

P269

Critical capacitance. A capacitor and a coil with inductance 1 H are coupled in series and connected to a power supply with an alternating voltage of 220 V and 50 Hz. A voltmeter with a very high internal resistance is connected in parallel to the capacitor. At what capacitance will the voltmeter read 220 V? What capacitance must never be used in such an experiment? (A. Zilberman)

P270

Optical illustration. Draw the image of a square formed by a converging lens (fig. 2). The midpoint of the square's side that lies on the principal axis coincides with the focal point of the lens. (B. Bukhovtsev)

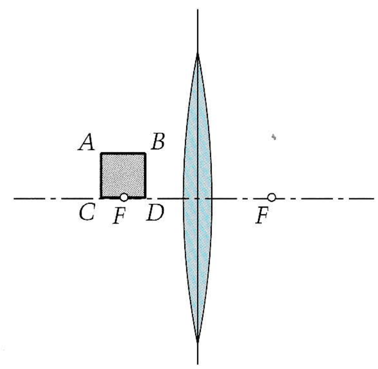


Figure 2

ANSWERS, HINTS & SOLUTIONS ON PAGE 50

Vacuum

Making something out of nothing

by A. Semenov

IT IS EVERYWHERE, BUT NObody feels it. It is invisible, impalpable, and inconspicuous, although it stores enough energy to produce a new universe. It is nothing, it is emptiness, but this emptiness can become the source of everything. It is the vacuum.

Many contemporary physicists consider the vacuum to be a major subject of twenty-first century science. However, until the end of the last century the vacuum was only a subject for the debates of philosophers. For example, in the seventeenth century René Descartes (1596–1650) used a long chain of logical conclusions to decide that the vacuum cannot exist: if "nothing" separates two particles, then they are separated by nothing (and therefore cannot be considered two individual objects). In the early days of science, logical arguments were often confused with linguistic ones.

However, the vacuum did exist, and quantum theory filled it with sense and content. In 1911 Max Planck (1858–1947) showed that a body retains energy even at absolute zero. What is the origin of this energy?

Physicists began to search for this mysterious vacuum energy, and in 1925 Robert Millikan (1868–1953) detected it for the first time. It was revealed in the emission spectra of boron monoxide. The frequency of radiation resulting from electrons jumping from one orbit to another was at odds with the theoretical estimates, as if an electron "collided" with something in its orbital flight.

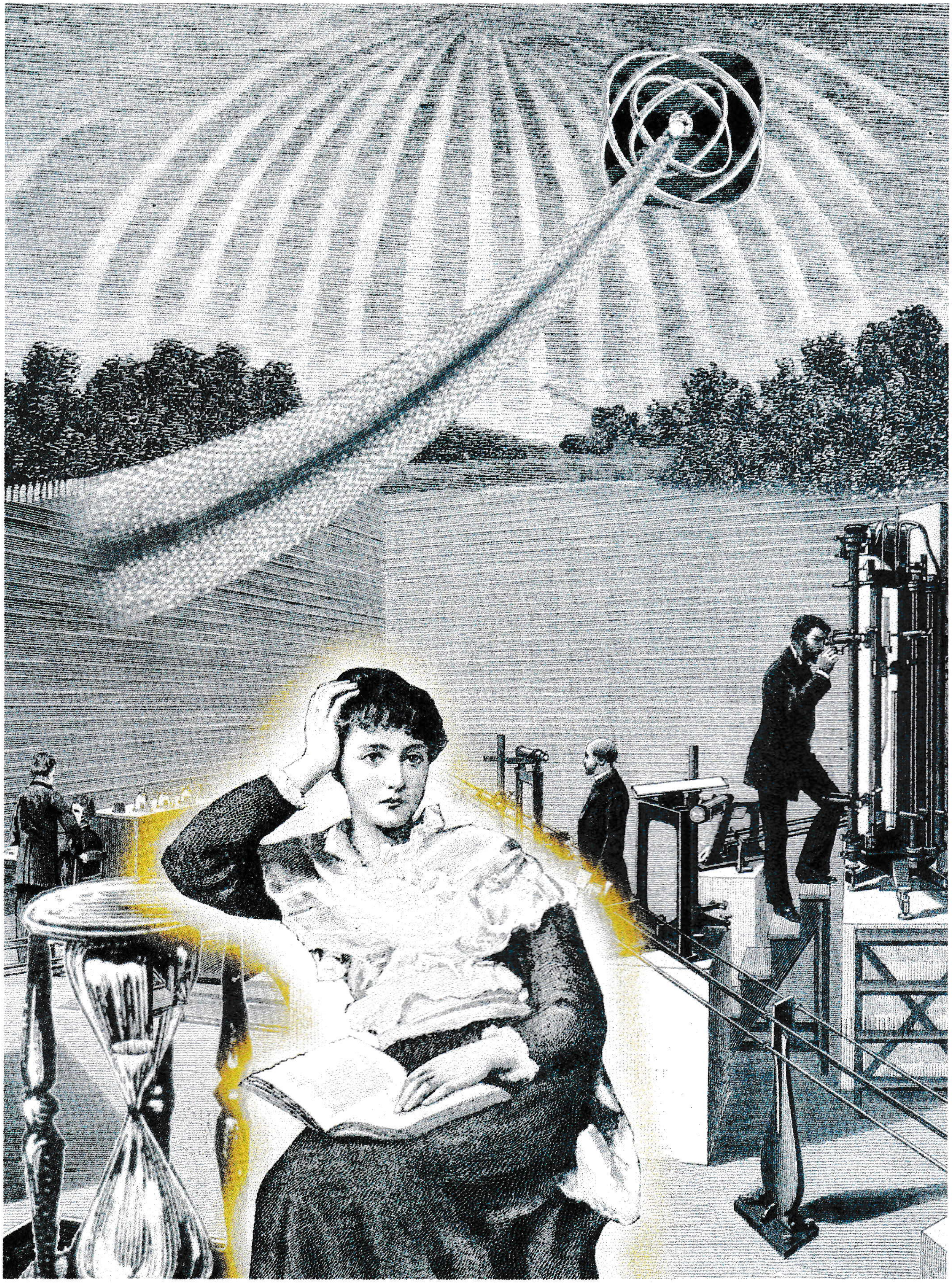
Just two years later, Werner Heisenberg (1901–1976) advanced the famous uncertainty principle and showed that a pair of virtual particles can appear and annihilate each other even in an absolute vacuum, although their lifetime is very short. The point is that energy fluctuations are possible in any system, but the larger the energy violation, the shorter period it exists. Mathematically, the product of the uncertainty of the energy and the uncertainty of its duration (that is, the value of the energy burst multiplied by its lifetime) can be no less than Planck's constant: $\Delta E \cdot \Delta t \gtrsim h$. So, these virtual pairs of particles impede the orbital motion of electrons.

Vacuum fluctuations are also manifest in the stochastic noise of electronic devices. They impose limitations on the gain of radio amplifiers. The van der Waals' forces acting between molecules also originate from the vacuum fluctuations of the molecular energy. The energy "hidden" in the vacuum doesn't permit liquid helium to change to the solid state at any low temperature. This same energy "turns on" the discharge in a mercury vapor lamp. When a mercury atom is excited by the electric discharge in such a lamp, vacuum fluctuations return the atom from the excited to the ground state. Thus, when you switch the lamp on, remember that you are starting the very process that once produced the Universe!

Born from nothing

For the last few decades the triumphant theory in astronomy has been the Big Bang model. It supposes that our Universe was created 20 billion years ago by an explosion of a hyperdense and superheated point. However, an accurate study of conclusions and predictions of this model revealed a number of difficulties it could not explain. Allan Guth of the Massachusetts Institute of Technology and former Moscow theoretician Andrei Linde (now a physics professor at Stanford University) proposed a modified explosion theory called the inflationary universe.

Art by Vera Khlebnikova



The new model says that in the first instants after the birth of the Universe, the vacuum was unstable and had a very large internal energy. One can compare it with a ball at the top of a hill. Such a state is unstable, so the "vacuum ball" rolled down and released a huge amount of energy. In a tiny fraction of a second, the Universe inflated itself by so many times that this number is expressed as one with one hundred zeros after it!

This inflation made our Universe very homogeneous and flat. The inflationary universe model removed many obstacles that were met on the road of the Big Bang hypothesis. For example, scientists have not detected magnetic monopoles[1], particles that have only one magnetic pole. This is strange, because a huge number of them should have been produced by the Big Bang. So where are the monopoles now? The answer: inflation scattered all the monopoles over such unimaginable distances that only a few of them remained in our tiny observable part of the Universe. Thus, all variety of quasars, pulsars, planets, rockets, and even human beings were born from this mighty "nothing" due to the relationship between mass and energy found by Einstein.

However, most cosmologists would be glad if the vacuum moderated its activity after it had finished its splendid endeavor of creating the Universe. No, life is not an easy road, so now and then the vacuum demonstrates its existence and stirs scientific minds. The problem is that the great internal energy of the vacuum complicates the equations of the general theory of relativity by adding some terms. Such theories are far from being completed, and who knows the true nature of things?

Is the vacuum full or empty?

In recent decades the physics community has been perturbed now and then by news from astronomers who reported that our Universe is younger than its own stars. The new estimates of the Universe's age are based on measurements of the Hubble constant. In addition to false rumors, the indeterminacy of these measurements motivated physicists to rethink their models of the development of the Universe.

One possible explanation of this contradiction is that the vacuum did not squander all its energy but instead secretly continues its subversive activity of expanding the Universe. This process increases the rate of recession of galaxies and misleads astronomers: the larger the rate of recession, the closer we are to the moment of the Big Bang.

However, the vacuum can do some other tricks. There is another problem in modern cosmology: the dark matter. Inflation predicts a certain density of matter in the Universe. By contrast, observations have yielded only 10–20% of this value. There are a variety of hypotheses that say where the dark matter should be located, but none of them have helped detect it.

George Efstathion of Oxford believes the dark matter is hidden as vacuum energy. This view is shared by Chris Coshaneck at Cambridge, who believes that the existence of gravitational lenses supports this hypothesis.

A gravitational lens is a very massive galaxy around which light coming from distant stars passes before reaching observers on Earth. Due to the curvature of the light rays, one can see multiple images of the same distant object. At present, a number of putative gravitational lenses have been found. Coshaneck considers that their number provides a key to estimating the energy hidden in the vacuum, which could be responsible for as much as half of the dark matter in the Universe.

It should be stressed here that we are discussing hypotheses rather than established facts. These are the working models of theoreticians, and they illustrate how these magicians of science work to find correct answers. However, the vacuum is not merely a favorite toy of astronomers; it is also involved in pure earthly matters.

Inertia

One of the most intriguing everyday manifestations of the hypothetical properties of the vacuum is inertia, the property of maintaining motion. Everybody is familiar with inertia: recall your experience of crashing into a snow drift while ice skating. The nature of inertia was enigmatic to such giants of science as Albert Einstein and Richard Feynman. Einstein assumed that the acceleration of one body somehow indirectly affected other bodies. He could not, however, explain the process.

A few years ago Bernard Hewish from Palo Alto and Shel Pythov from Texas tried to revive this idea of Einstein. They supposed that a body has inertia due to its interaction with vacuum fluctuations. They even modified Newton's law by replacing mass with a parameter that characterizes the interaction of a body with a vacuum.

In simpler terms, their hypothesis says that vacuum fluctuations generate a field similar to a magnetic field. The more atoms in a body, the stronger it interacts with this "vacuum" field, and the more difficult it is to accelerate the body.

However, this is just an idea, which at present is not substantiated by precise calculations. All estimates differ from the experimental data by a stunningly large factor: this number is one followed by a good hundred zeros. Steven Weinberg, a Nobel Prize winner, joked that this number is the most inexact prediction in physics.

Still, the authors of the vacuum hypothesis have not lost heart. In fact, they have even discussed ways of extracting energy from a vacuum. It seems like science fiction, but Hewish reminds us that only 100 years ago nobody knew about jet planes, radios, or televisions, to say nothing of the atomic bomb.

---

[1]Read about monopoles in J. Wylie, "Magnetic Monopoly," May/June 1995, pp. 4–9.

A lyrical digression

A vacuum is a very enigmatic thing. Many outstanding minds have guessed at its nature—from Aristotle to Feynman. "Nature does not permit emptiness"—this dictum designates the vacuum as an abstract, unreal conception, which predicts the failure of any scientific attempt to study it. Nevertheless, modern theoreticians cannot ignore the vacuum. Such notions as the "vacuum state" or "vacuum energy level" are used in virtually every physical model of the microcosm. It seems as if up until now scientists have disregarded the unbelievable global character of the vacuum problem. To spend time considering the vacuum was considered something of a scientific sacrilege.

The problem of the vacuum was circumstantially treated by science fiction authors. Even for them it was too fantastic to allow their characters to take energy from "nothing," so they took it from other unexpected sources: time, future, or hyperspace. Perhaps the vacuum will turn out to be the unexpected source of inexhaustible energy after all.

Consider a social phenomenon that has nothing to do with classical science: the growing activity of "specialists" in extrasensory perception and in other mystical events such as intercontinental non-electromagnetic information transmission, inner communication with space, and diagnostics of astral bodies. To tell the truth, it is a wonder that these supernatural people do not use the concepts of vacuum and vacuum fluctuations. Sometimes even the craziest-sounding hypotheses should be considered. What if all these astral fields and space fluxes are somehow related to the vacuum? Nothing is clear in both fields, so perhaps they should be studied. Of course, any study must be scientific and free from illiterate mystical gibberish, but these problems should not be simply brushed aside.

A common view of the vacuum sounds like King Lear's remark: "Nothing will come of nothing." The true interpretation of the vacuum as "nothing" can only mean one thing—we really know nothing about it. Recently, electromagnetic fields and X-rays were similar invisible and imperceptible notions.

All the life of our world is based on the interaction of particles and the fields that connect them. According to the common view, the vacuum is the negation of existence. Maybe it is just another form of existence?

What can be squeezed out of a vacuum?

The idea of obtaining something useful from the vacuum is rather old. In 1948 the Dutch scientist Hendrick Casimir proposed bringing two plates close to each other and measuring the tiny attraction that should arise between them. This attraction is caused by the external vacuum fluctuations that push the plates toward each other. However, the predicted value of this vacuum pressure is very small: Just one hundred millionth of an atmosphere for a separation of 1 μm. However, not the value but the fundamental principle is the major point here. If we take two plates of $1 m^2$, polish them, position them 1 μm apart, and release them, the attractive forces will develop a power of about 1 μW. This is not very much, but who can tell for sure?

I am not a novice in scientific work, so I meet such extravagant projects with a skeptical smile. However, my scientific time has gone, and my skepticism belongs to the twentieth century. Now the next century swiftly approaches. To revive the common belief in science, to make science flourish, it is necessary to discover something that cannot be found with a skeptical smile. Hopefully, the person who will make this step will not be embedded in modern pseudoscientific gibberish but will make a real breakthrough, taking the best from the receding century.

Well, the formation of the Universe, the nature of inertia, and inexhaustible sources of energy—these are the promising prospects glimpsed in the study of the vacuum. Much is still unclear, and estimations have not replaced speculations. Yet one thing is clear: it is time to start. The Century of the Vacuum is coming! ◘

Quantum on the Universe and cosmology:

Y. Zeldovich, "A Universe of Questions," January/February 1992, p. 6–11.

A. D. Chernin, "Grand Illusions," January/February 1992, p. 24–29.

Y. Solovyov, "The Universe Discovered," May/June 1992, p. 12–18.

W. A. Hiscock, "The Inevitability of Black Holes," March/April 1993, p. 26–29.

G. Myakishev, "The 'Most Inertial' Reference Frame," March/April 1995, p. 48.

S. Silich, "Interstellar Bubbles," November/December 1997, p. 14–19.

I. D. Novikov, "The Thermodynamic Universe," March/April 1998, p. 10–14.

The new Earth

We must approach global construction with the utmost care

by A. Stasenko

INITIATED BY A LOVE FOR humanity, someone suggested increasing the surface area of Earth by a factor of 100 by excavating its interior.

What a noble project this would be, offering solutions to all territorial disputes and providing space for many new summer houses and gardens! An additional benefit is that trips within Earth's empty shell would require almost no energy expenditures and people could jump far and high because of the decreased force of gravity on the surface!

But every new project, and a global one in particular, requires quantitative estimates of expenditures and consequences. So, let's begin.

First, we evaluate the size of New Earth (with radius R and shell thickness δ). According to the plan, the area of its surface $S = 4\pi r^2$ would be 100 times greater than that of Old Earth $S_0 = 4\pi R_0^{\,2}$, from which we obtain $R^2 = 100R_0^{\,2}$ and $R = 10R_0$. This means that the inner radius of the New Earth will be $R - \delta$.

From conservation of mass, we have

$$\frac{4}{3}\pi R_0^3 = \frac{4}{3}\pi\left(R^3 - (R-\delta)^3\right)$$
$$= \frac{4}{3}\pi\left(R^3 - R^3 + 3R^2\delta - 3R\delta^2 + \delta^3\right)$$
$$\cong 4\pi R^2\delta.$$

The first two items in the parentheses cancel out, and the last two are neglected, because they contain the square and the cube of a small value, the thickness of the shell. We assume it to be small, but we won't forget to check this point later.

We have assumed that Earth's density is homogeneous and doesn't vary during the construction of New Earth, so we immediately canceled it from both sides of the equality. Therefore,

$$\frac{\delta}{R} \cong \frac{1}{3}\left(\frac{R_0}{R}\right)^3 = \frac{1}{3000},$$

a rather small value. According to our project, the entire mass of Earth must be confined within a very thin shell (with a thickness of about 20 km). That is, all elements of that mass will be virtually at the same distance R from the center.

Let's start by calculating the gravitational work required to move the material of Old Earth to a distance R. Since the problem possesses spherical symmetry, we begin our consideration with an elementary layer of radius r and thickness dr inside Old Earth (figure 1a). Because we are concerned with this layer, we are assuming that the outer layers have already been moved to the required distance. As we know, these outer layers don't generate a gravitational field inside New Earth (readers unfamiliar with fields and potentials may want to read the articles referenced at the end of the article).

The selected layer, when located a distance r' from the center, will be attracted only by the remaining mass $m(r)$. According to Newton's law of gravitation, this force is

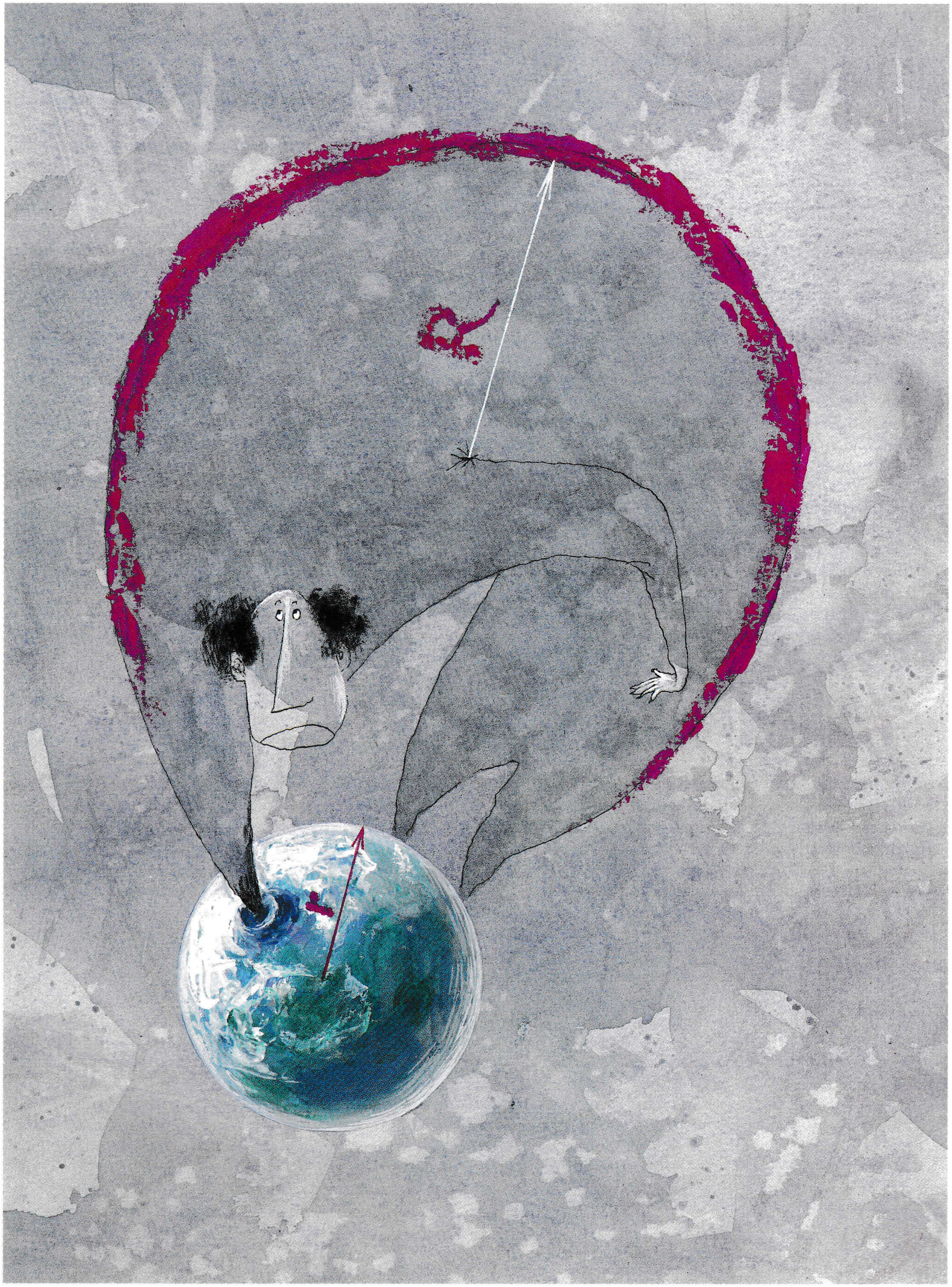
$$-G\frac{m(r)dm(r)}{r'^2}.$$

To move the layer a little distance dr', the additional work

$$G\frac{m(r)dm(r)}{r'^2}dr'.$$

is required. The total work performed by moving the layer for a

Art by Pavel Chernusky



distance from r to R is given by the integral

$$dW = \int_r^R Gm(r)dm(r)\frac{dr'}{r'^2} = Gm(r)dm(r)\left(\frac{1}{r} - \frac{1}{R}\right).$$

Here we should note an important property of the gravitational field. Figure 1a shows the trajectory of a part of an elementary layer as a curve with arrows on it. It's possible to draw it as curved as you like—the work required to transfer such a fragment will not vary, provided we do not change the initial and final distances from the center of gravity (in our case from the center of the remaining mass $m(r)$).

The reasoning is valid only for an elementary (dashed line) layer. When we take mass away layer by layer from the very beginning, we need to integrate the expression for the work with respect to r from 0 to R_0:

$$W = \int dW = G\int_0^{R_0} m(r)dm(r)\left(\frac{1}{r} - \frac{1}{R}\right).$$

Assuming Old Earth to be a homogeneous ball, and its density ρ_0 to remain constant during the transformation of the planet (we have already used this assumption in calculating the thickness), we have

$$m(r) = \frac{4}{3}\pi\rho_0 r^3$$

and

$$dm = 4\pi\rho_0 r^2 dr,$$

so

$$W = G\frac{4}{3}\pi\rho_0 \cdot 4\pi\rho_0\int_0^R \left(r^4 - \frac{r^5}{R}\right)dr = \frac{GM^2}{R_0}\cdot\frac{3}{5}\left(1 - \frac{5}{6}\frac{R_0}{R}\right),$$

where $M = (4/3)\pi\rho_0 R_0^3$ is the total mass of Earth.

Now all that's left are the calculations. First, we simplify the formula by combining the mass and the radius of Earth in an expression equal to the acceleration due to gravity g_0:

$$\frac{GM}{R_0^2} = g_0.$$

Then we have

$$W = Mg_0R_0\cdot\frac{3}{5}\left(1 - \frac{5}{6}\frac{R_0}{R}\right).$$

However, you may not know the mass of our planet offhand and may not have a reference book nearby. This is not an obstacle—we can find the answer without it. Most of us remember that g_0 is 9.8 m/s^2 and R_0 is 6400 km, and if we also remember that G is $6.6 \cdot 10^{-11}$ m^3/(s^2 · kg), it is easy to determine the mass of Earth from the formula for g_0: $M = 6 \cdot 10^{24}$ kg. (Once upon a time, Henry Cavendish (1731–1810) "weighed" Earth for the first time in history by this very method—that is, by determining the constant G in his famous experiment with a torsion balance).

Now everything is ready for the estimation of the minimum work needed to create New Earth:

$$W \cong 2 \cdot 10^{32}\text{ J}.$$

If we take into consideration that the

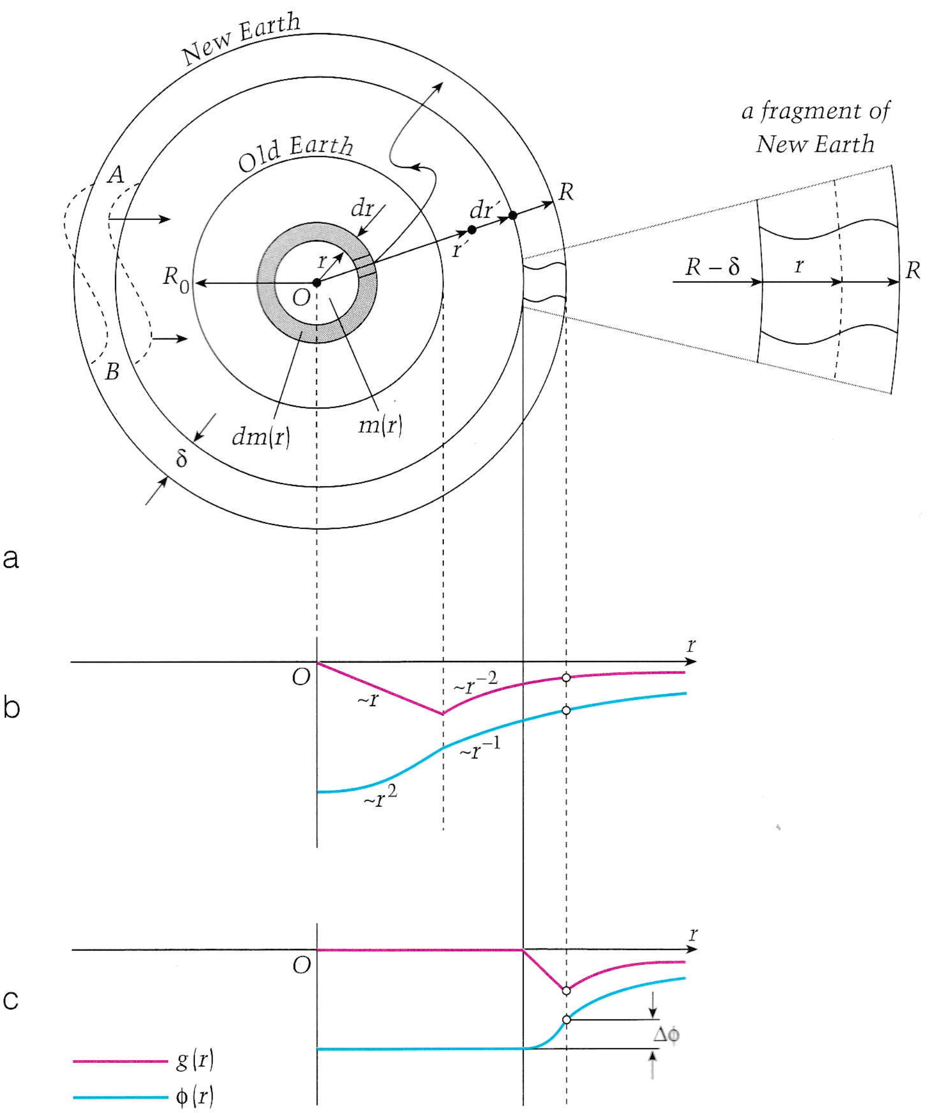


Figure 1

rate of energy consumption of humanity is currently about 10^{20} J/year, it's easy to estimate how many years it will take to realize our "beneficial" project.

Imagine that New Earth has been created. How will the new configuration affect your life? First of all, the acceleration due to gravity on its surface will be

$$\frac{g}{g_0} = \left(\frac{R_0}{R}\right)^2 = \frac{1}{100}$$

times the present value. Physical objects will be lighter by the same factor, and a stone thrown with the same velocity at the same angle to the horizon will fly farther by the same factor. A pendulum will oscillate one-tenth as fast ($\sqrt{g_0/g} = 10$), and inside New Earth it won't oscillate at all. The escape velocity will be changed by a factor of

$$\frac{\sqrt{gR}}{\sqrt{g_0 R_0}} = \sqrt{\frac{R_0}{R}} = \frac{1}{\sqrt{10}} \cong \frac{1}{3}.$$

As already mentioned, it will be possible to travel inside New Earth without energy expenditure, because there will be no gravity ($g = 0$). Therefore, no work is needed to move between any two points there. Or, as a physics teacher might say, the volume inside the spherical shell is an equipotential: ϕ = const when $r < R - \delta$. Look carefully at figures 1b and 1c and compare the plots of functions $g(r)$ and $\phi(r)$, drawn for Old Earth and New Earth, respectively.

By the way, what work must be performed to travel from the inside to the outside of New Earth? Set the spherical surface with radius r inside the shell (see figure 1a, where a magnified part of the shell is shown on the right). As we already know, the acceleration due to gravity on that surface is produced only by the part of the mass located inside the surface. This acceleration is equal to

$$\frac{4}{3}\pi\left(r^3 - (R-\delta)^3\right)\rho_0,$$

so

$$g(r) = \frac{4}{3}\pi G\frac{\left(r^3 - (R-\delta)^3\right)\rho_0}{r^2}.$$

In figure 1c this function is shown between the points $r = R - \delta$ and $r = R$.

To find the work W_1 needed to remove a body of mass 1 kg from the interior chamber, we must calculate the integral of $g(r)$ between the limits $R - \delta$ and R, which the reader is invited to try independently. Here we evaluate W_1 by approximating $g(r)$ with a linear function that varies from 0 (interior of the shell) to $g = g_0(R_0/R)^2$ (exterior space):

$$W_1 = g_{\text{mean}} \cdot \delta = \frac{g}{2}\delta$$
$$= \frac{g_0}{2}\left(\frac{R_0}{R}\right)^2 \cdot \delta \cong 10^3 \text{ J/kg}.$$

This work per unit mass is also referred to as the potential difference between the inner and outer surfaces. It is denoted by $\Delta\phi$ in figure 1c.

Would we be able to breathe on the new planet? We assume the mass of the atmosphere to be the same on Old Earth and New Earth. Let's evaluate it for Old Earth. If the temperature didn't vary with altitude, meaning T = const, its density varies according to Boltzmann's formula:

$$\rho = \rho(R_0)e^{-\frac{mg_0 y}{kT}},$$

where m is the mass of an air molecule and k is Boltzmann's constant. Note that the numerator of the exponent contains the difference of potentials at the altitude y and at the surface of Earth (where $y = 0$).

Taking $mg_0 y_*/kT = 1$, we determine the characteristic altitude where the atmospheric density is only $1/e$ of that at the surface, which is to say that it is almost one-third as small. This altitude is

$$y_* = \frac{kT}{mg_0} \cong 8 \text{ km}.$$

Now we can calculate the mass of the atmosphere to be

$$M_a \cong 4\pi R_0^2 y_* \rho(R_0) \cong 5 \cdot 10^{18} \text{ kg}.$$

On New Earth the atmosphere will spread evenly inside the spherical shell (because the difference in potential between any two points inside the shell is zero), and above the surface its density will decrease exponentially with altitude in the usual manner, although with a new characteristic altitude, which is $g_0/g = 100$ times larger than the old one. Therefore, the characteristic altitude of the new atmosphere will be 10^3 km, a value still negligible compared with the radius of New Earth $R = 10R_0 \cong 6 \cdot 10^4$ km.

The ratio of densities inside and outside will be

$$\frac{\rho^-}{\rho^+} = \frac{\rho(R-\delta)}{\rho(R)} = e^{\frac{W_1 m}{kT}},$$

where W_1 is the specific work needed to move a mass of 1 kg from the inside to the outside. Let's evaluate the power of the exponential function:

$$\frac{W_1 m}{kT} \cong 1.2 \cdot 10^{-2}.$$

This turned out to be negligibly small, so we can assume that the densities inside and outside New Earth are virtually identical.

Now let's formulate the conditions for conservation of mass of the atmosphere:

$$M_a = \frac{4}{3}\pi\rho^-(R-\delta)^3 + 4\pi R^2\rho^+ \cdot 100y_*$$
$$\cong \frac{4}{3}\pi\rho^- R^3\left(1 + 3\frac{100y_*}{R}\right) \cong \frac{4}{3}\pi\rho^- R^3$$

(since the second item in the parentheses is much less than 1, we can assume that almost all the atmosphere is inside), from which we obtain

$$\rho^+ \cong \rho^- \cong \frac{M_a}{\frac{4}{3}\pi R^3} \cong 5 \cdot 10^{-6} \text{ kg/m}^3.$$

It seems to be dangerous to breathe

inside or outside this new planet.

In addition, the spherical shell of New Earth will be unstable: any disturbance AB (figure 1a on the left) of its shape will increase with time, because there will be no restoring force. Thus, we must be careful to maintain the integrity of the New World.

New Earth still has many surprises to reveal. For example, with what angular velocity will New Earth rotate? There is a law of conservation of angular momentum in mechanics. Similar to conservation of linear momentum $m_0v_0 = mv$, it can be written easily by substituting angular velocities for the linear velocities. In doing so, we must also replace the masses by the moments of inertia.

However, the moments of inertia for a homogeneous ball and a spherical shell are known: $(2/5)MR_0^2$ and $(2/3)MR^2$, respectively. For our estimates we can drop the numerical coefficients in these formulas. After canceling the masses, we get

$$R_0^2\omega_0 \cong R^2\omega,$$

from which the new day is found to be longer than the old day by

$$\frac{T}{T_0} = \frac{\omega_0}{\omega} \cong \left(\frac{R}{R_0}\right)^2 = 100 \text{ times.}$$

So, the seven-day "week" on New Earth will last more than a year. That's quite a long wait for the weekend! That's too bad, indeed. Now we see why we must approach global reconstruction with the utmost care, analyzing all possible consequences and using them for a comprehensive study of physical laws.

Quantum on rotating bodies and rigid dynamics:

S. Krivoshlykov, "Head over Heels," May/June 1995, p. 62–65.

A. Eisenkraft and L. D. Kirkpatrick, "Pins and Spin," July/August 1995, p. 34–36.

L.Borovinsky, "Why Won't Weeble Wobbly Go to Bed?" May/June 1996, p. 64–65.

V.Surdin, "A Venusian Mystery," July/August 1996, p. 4–8.

M.Emelyanov, A.Zharkov, V. Zagainov, and V.Matochkin, "In Foucault's Footsteps," November/December 1996, p. 26–27.

L. D. Kirkpatrick and A. Eisenkraft, "Around and Around She Goes," March/April 1998, p. 30–33.

A. Stasenko, "Rivers, Typhoons, and Molecules," July/August 1998, p. 38–40.

Quantum on potential:

A. Stasenko, "From the edge of the Universe to Tartarus," March/April, 1996, p. 4–8.

A. Leonovich, "Do You Have Potential?" November/December, 1998, p. 28–29.

Bulletin Board

Duracell/NSTA Invention Challenge

By designing and building working devices powered by Duracell batteries, students participating in the Duracell/NSTA Invention Challenge can match wits with students from across the United States, develop the confidence to complete a complicated project, and use individual or team skills and interests to participate in science in a new way.

The Challenge also gives students the chance to win large monetary prizes in the form of U.S. savings bonds and an all-expenses paid trip to the NSTA national convention for an awards ceremony. The Duracell/NSTA Challenge has stimulated over 12,000 inventive ideas and recognized more than 1,000 student inventors and their sponsoring teachers since 1983. Some of the winning devices have been refined and commercially marketed.

The Challenge is open to students in grades 6–12 who are under 21 years of age and who are U.S. citizens and reside in the United States or U.S. Territories. Students may enter individually or in teams of two. Teachers are an integral part of the competition process. Sponsoring teachers sign entry forms attesting that each student or pair of students built each device.

To enter:

1. Obtain the Official Entry Form from Duracell/NSTA Invention Challenge, 1840 Wilson Blvd., Arlington, VA 22201–3000; or call toll-free 1-888-255-4242, or e-mail your mailing address to duracell@nsta.org. The entry form will be available beginning in August. Fill out the entry form completely, and obtain all necessary signatures.
2. Design and build a device that runs on Duracell batteries.
3. Write a two-page description of the device and its uses.
4. Draw a schematic (wiring diagram) of the device.
5. Photograph the device (use clear photos only).
6. Mail Official Entry Form, typewritten description, schematic, and photos (do not send the actual device at this time). Entry form must be RECEIVED by **January 12, 2000.**
7. The 100 top finalists (or pairs) will be notified where to send their actual devices for final judging.
8. The first- and second-place winners must attend awards events to receive their savings bonds.

For more information on the Duracell/NSTA Invention Challenge, point your web browser to http://www.nsta.org/programs/duracell/.

Frequent fly

This month's Cyberteaser winners didn't need to carry a big swatter to solve this problem (B268 in this issue). They followed a fly back and forth between two bicycle-riders to find out the total distance it flew before the riders bridged the gap between them. Here are the 10 fastest problem-solvers:

Bruno Konder (Rio de Janeiro, Brazil)
Christopher Franck (Redondo Beach, California)
Theo Koupelis (Wausau, Wisconsin)
Jerold Lewandowski (Troy, New York)
Rafael Shusterovich (Rishon-le-Zion, Israel)
Vincze Zsombor (Szeged, Hungary)
Sergio Moya (Culiacan, Mexico)
Anastasia Nikitina (Pasadena, California)
Melamed David (Kiryat Tiv'on, Israel)
T. Scott Frick (Dallas, Pennsylvania)

Congratulations! Each winner will receive a copy of the July/August issue and a *Quantum* button. Everyone who submitted a correct answer in the time allotted was entered in a drawing for a copy of *Quantum Quandaries,* a collection of the first 100 *Quantum* brainteasers. Try your luck at winning a prize of your own by visiting www.nsta.org/quantum and clicking the Contest button for the current Cyberteaser.

Talk back to us

We actually like it! We want to know what you think about *Quantum*. So,

drop us a line at *Quantum*, 1840 Wilson Blvd., Arlington VA 22201

or

zip us an electron or two at quantum@nsta.org

2-adic numbers

Employing Hensel's rational insights

by B. Becker, S. Vostokov, and Y. Ionin

IF YOU WERE ASKED TO DEFINE A "DISTANCE between two rational numbers," you would probably answer that this is the absolute value of their difference. This answer is quite reasonable: it satisfies all axioms of distance. However, it turns out that another distance between rational numbers can be defined that also satisfies all the axioms of distance. This was done by the German mathematician Kurt Hensel (1861–1941). He invented an entire class of such distances; here we discuss one of them.

Hensel's studies proved important for algebra and mathematics in general. We will use Hensel's distance to solve two problems that at first glance do not seem related to any distances.

2-adic distance

Let a and b be rational numbers. If $a \neq b$, we represent the difference $a - b$ as $a - b = 2^k(m/n)$, where m and n are odd integers and k is an integer (positive, negative, or zero). The 2-adic distance between numbers a and b ($a \neq b$) is defined as the number $\rho(a, b) = 1/2^k$. If $a = b$, we set $\rho(a, b) = 0$.

A *distance function* or *metric* is usually defined to be a function of two numbers that satisfies the following axioms:

A1. $\rho(a, b) > 0$ if $a \neq b$, and $\rho(a, b) = 0$ if $a = b$.

A2. $\rho(a, b) = \rho(b, a)$.

A3. $\rho(a, c) \leq \rho(a, b) + \rho(b, c)$.

It is evident that properties A1 and A2 are true for the 2-adic distance. Property A3 is also clearly true in the case when $a = b = c$.

Let us prove property A3 for the case of distinct rational numbers a, b, and c. Let $a - b = 2^{k_1}(m_1/n_1)$, $b - c = 2^{k_2}(m_2/n_2)$, and $a - c = 2^{k_3}(m_3/n_3)$, where all the m_i and n_i are odd integers. Since $a - c = (a - b) + (b - c)$, k_3 cannot be less than the smaller of the numbers k_1 and k_2. Then, $1/2^{k_3}$ does not exceed the greater of the numbers $1/2^{k_1}$ and $1/2^{k_2}$, so $1/2^{k_3} \leq 1/2^{k_1} + 1/2^{k_2}$.

Thus, we see that all the axioms are valid, and ρ can be called a distance.

What quantity does this distance measure? It turns out that it measures (roughly speaking) the degree of divisibility of a rational number by 2. The "better" that 2 divides a number (for example, the higher the power of 2 that divides it, if it is an integer), the closer it is to zero. For example, 8 is closer to zero than 1/2, 16 is closer to zero than 8, 480 is closer to zero than 16, and 384 is closer to zero than 480.

In fact, we have proved that 2-adic distance possesses a property A3′, which is stronger than A3:

A3′. The distance $\rho(a, c)$ does not exceed the greater of the distances $\rho(a, b)$ and $\rho(b, c)$.

Exercise 1. Prove that if $\rho(a, b) \neq \rho(b, c)$, then $\rho(a, c)$ equals the larger of the numbers $\rho(a, b)$ and $\rho(b, c)$, and if $\rho(a, b) = \rho(b, c) \neq 0$, then $\rho(a, c) < \rho(a, b)$.

Property A3′ has interesting consequences. We call the set of all rational numbers x such that $\rho(a, x) < r$ (where a is a rational number and r is a positive real number) the *2-adic circle* of radius r centered at point a.

Exercise 2. Prove that if two 2-adic circles have a nonempty intersection, one of them includes the other.

Exercise 3. Prove that the 2-adic circle of radius r includes infinitely many pairwise nonintersecting 2-adic circles of radius r.

By analogy with the usual absolute value (sometimes called the *modulus*), we define a 2-adic modulus a of a rational number $\|a\|$ as the 2-adic distance from this number to zero: if $a = 2^k(m/n)$, where m and n are odd

Art by Sergey Ivanov

numbers, then $\|a\| = \rho(0, a) = (1/2)^k$. The following properties of the 2-adic modulus can be easily established:

M1. $\|a\| > 0$ if $a \neq 0$ and $\|0\| = 0$.

M2. If $\|a\| > \|b\|$, then $\|a + b\| = \|a\|$; if $\|a\| = \|b\| \neq 0$, then $\|a + b\| < \|a\|$. (Thus, in any case $\|a + b\| \leq \|a\| + \|b\|$.)

M3. $\|ab\| = \|a\| \cdot \|b\|$.

Exercise 4. Derive the following properties of the 2-adic modulus from properties M1–M3:

(a) $\|-a\| = \|a\|$; (b) if $\|a\| \neq \|b\|$, then $\|a - b\| = \|a + b\|$.

Exercise 5. Prove that if $\|1 - x\| < 1$ and $\|1 - y\| < 1$, then $\|1 - xy\| < 1$.

Decomposition of a square

Figure 1 depicts a square that is decomposed into congruent triangles. The squares in figure 2 are decomposed into triangles of equal area. In each of these examples, the number of triangles is even.

Problem. Prove that the square cannot be decomposed into an odd number of triangles of equal area.

Choose a system of coordinates in the plane such that the vertices O, A, B, and C of the given square have the

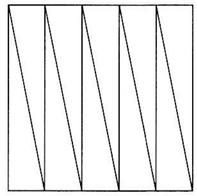


Figure 1

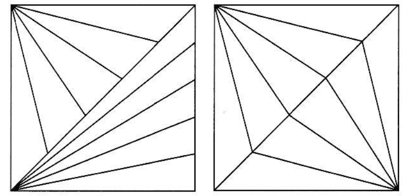
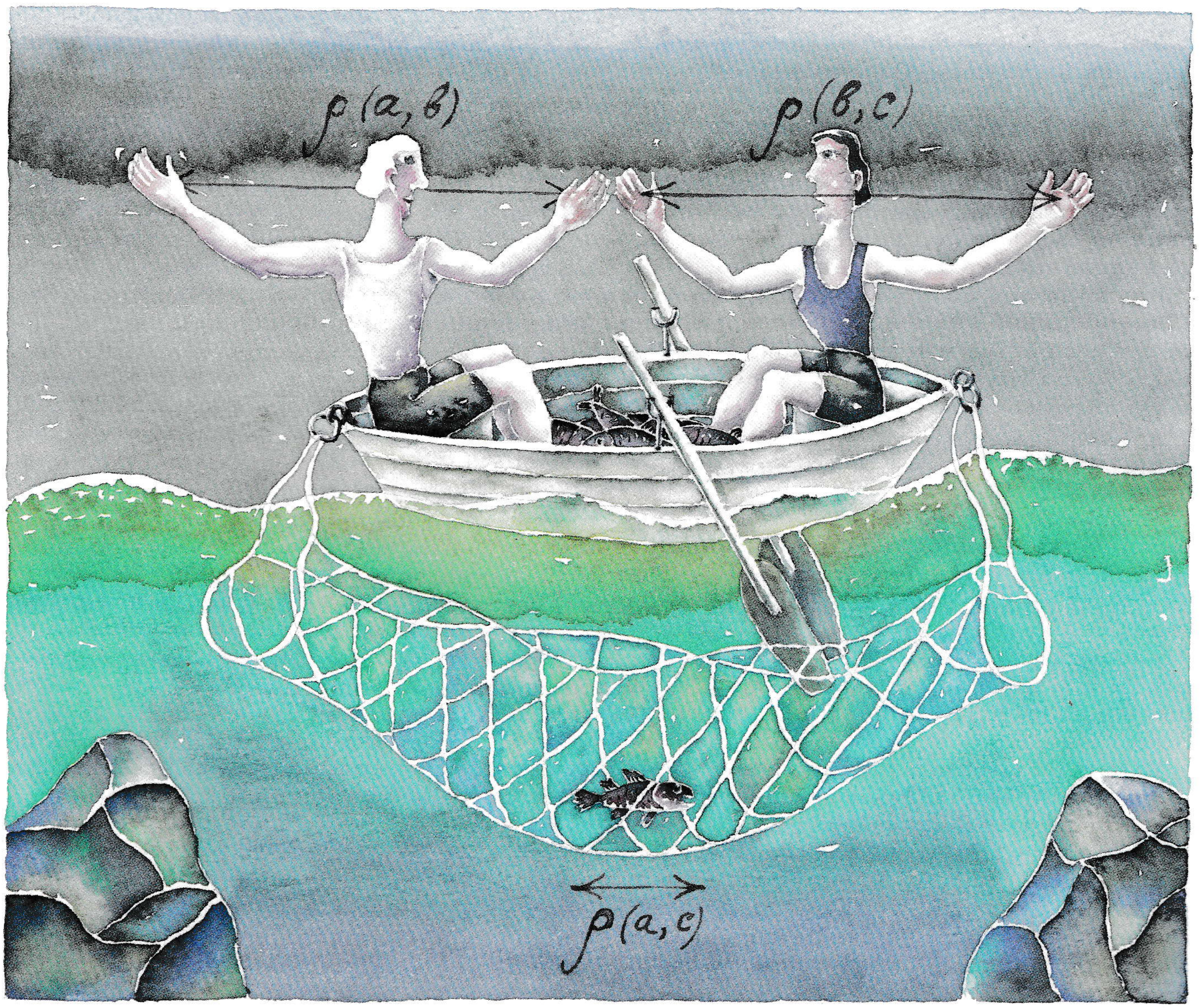


Figure 2

coordinates $O(0, 0)$, $A(1, 0)$, $B(1, 1)$, and $C(0, 1)$. Assume that the square is decomposed into n triangles of equal area. Then, this area is $1/n$. If n is odd, then $\|1/n\| = 1$; if it is even, $\|1/n\| \geq 2$.

Consider a particular case. Let the vertices of all the triangles of the decomposition be points with rational coordinates. In this case, we can color every vertex (x, y) green, red, or blue according to the following rule: if $\|x\| < 1$ and $\|y\| < 1$, the point is green; if $\|x\| \geq \|y\|$ and $\|x\| \geq 1$, the point is red; and if $\|x\| < \|y\|$ and $\|y\| \geq 1$, the point is blue (fig. 3). We will assume that all rational points of the plane are colored, not just the vertices.



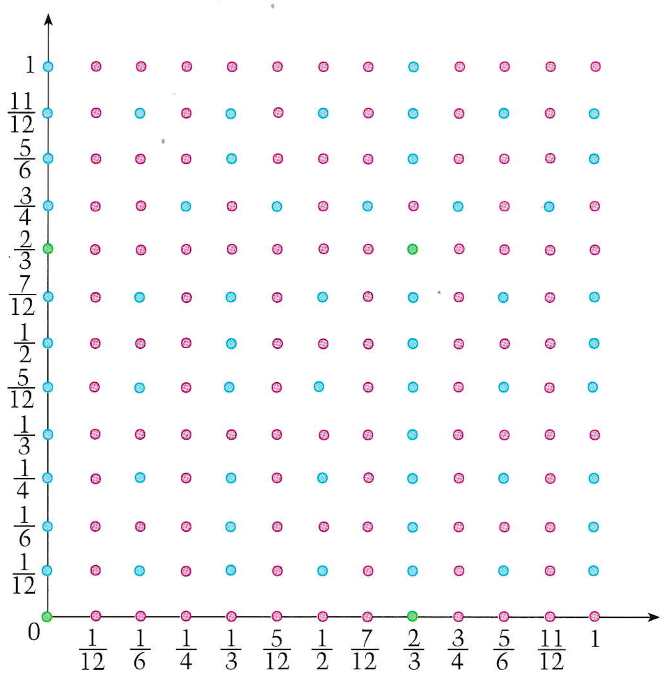


Figure 3

Exercise 6. (a) Prove that if P is a green point, then a translation by the vector **PO** preserves the color of the points.

(b) Prove that no line can include points of all three colors.

Suppose that the vertices of a triangle belonging to the decomposition are of three different colors (we prove in the Appendix that such a triangle does exist). Let K be the green vertex of this triangle.

By virtue of exercise 6a, a translation by the vector **KO** yields another triangle with vertices of all three colors. Denote by $L_1(x_1, y_1)$ the red vertex of this triangle and by $L_2(x_2, y_2)$ its blue vertex (the green vertex coincides with point O). Since the triangle OL_1L_2 is obtained by a translation of a decomposition triangle, its area is $1/n$. On the other hand, its area is $1/2|x_1y_2 - x_2y_1|$. (We invite the reader to prove this.)

Thus, we obtain the equation $1/n = 1/2|x_1y_2 - x_2y_1|$. Now it is not difficult to establish the inequality $\|1/n\| \geq 2$. Indeed, since L_1 is a red point and L_2 is blue, we have $\|x_1\| \geq \|y_1\|$ and $\|x_2\| < \|y_2\|$. Multiplying these two inequalities, we obtain $\|x_1\|\ \|y_2\| > \|x_2\|\ \|y_1\|$, and therefore, by property M3, $\|x_1y_2\| > \|x_2y_1\|$. By virtue of property M3 and exercise 4b, $\|x_1y_2 - x_2y_1\| = \|x_1y_2\|$. In addition, $\|x_1\| \geq 1$ and $\|y_2\| \geq 1$, so that $\|1/n\| = 2\|x_1\|\ \|y_2\| \geq 2$. Hence n is even.

To solve the problem in the general case, it is sufficient to prove that the 2-adic modulus can be extended to the set of all real numbers. That is, a function $x \to \|x\|$ exists that is defined on the set of all real numbers, satisfies properties M1–M3, and coincides with the 2-adic modulus on the set of all rational numbers. Such a function actually exists, but the proof of this fact requires tools that are far beyond the scope of this article.

2-adic expansion

It is well known that any natural number can be represented as a sum of powers of 2. For example, $1000 = 2^3 + 2^5 + 2^6 + 2^7 + 2^8 + 2^9$. We can obtain this expansion by using the 2-adic distance in the following way. First, find a power of 2 that is at the same 2-adic distance from 0 as the number 1000. Having found this number (2^3), subtract it from the given number 1000, and for the new number, 992, find a power of 2 that is at the same 2-adic distance from 0. Then, find a power of 2 that is at the same 2-adic distance from 0 as number $960 = 992 - 2^5$, and so on.

Using negative powers of 2, we can construct similar expansions for rational numbers of the form $m/2^k$, where m and k are natural numbers. For example,

$$\frac{1477}{256} = \frac{1 + 2^2 + 2^6 + 2^7 + 2^8 + 2^{10}}{2^8}$$
$$= 2^{-8} + 2^{-6} + 2^{-2} + 2^{-1} + 2^0 + 2^2.$$

Other rational numbers cannot be represented as a sum of powers of 2. However, any rational number can be approximated by such sums as accurately as desired. Indeed, let a be a rational number and $\|a\| = 1/2^{k_1}$. Define $a_1 = a - 2^{k_1}$. Then, a_1 is closer to zero than a (exercise 1). Therefore, either $a_1 = 0$ or $\|a_1\| = 1/2^{k_1}$, where $k_2 > k_1$. Setting $a_2 = a_1 - 2^{k_2}$, we again obtain either $a_2 = 0$ or $\|a_2\| = 1/2^{k_3}$, where $k_3 > k_2$, and so on. Thus, either a certain a_s is zero, and then a is a sum of powers of 2: (That is, $a = 2^{k_1} + 2^{k_2} + \ldots + 2^{k_s}$, or all a_s are not zero and then a is approximated by the sums $2^{k_1} + 2^{k_3} + \ldots + 2^{k_s}$ with any accuracy desired.) In this case, it is convenient to consider a as equal to the infinite sum $a = 2^{k_1} + 2^{k_2} + \ldots + 2^{k_s} + \ldots$

Definition. Let $x_1, x_2, \ldots, x_n$ be a sequence of rational numbers. We will say that a equals the infinite sum $x_1 + x_2 + \ldots + x_n + \ldots$ if the 2-adic distance between a and the sums $S_n = x_1 + x_2 + \ldots + x_n$ tends to zero as n tends to infinity. That is,

$$\lim_{n \to \infty} \rho(a, S_n) = 0.$$

Exercise 7. Prove that if $\|q\| < 1$, then the infinite sum $a + aq + aq^2 + \ldots + aq^n + \ldots$ exists and equals $a/(1 - q)$.

We have proved that any rational number can be represented either as a finite sum $2^{k_1} + 2^{k_2} + \ldots + 2^{k_n}$ ($k_1 < k_2 < \ldots < k_n$) or as an infinite sum $2^{k_1} + 2^{k_2} + \ldots + 2^{k_n} + \ldots$, where k_i are integers and $k_1 < k_2 < \ldots$ In both cases, this representation, which is called the *2-adic expansion* of number a, can be written as

$$\varepsilon_k 2^k + \varepsilon_{k+1} 2^{k+1} + \ldots + \varepsilon_{k+n} 2^{k+n} + \ldots,$$

where k is an integer, each of the numbers

$$\varepsilon_{k+1}, \varepsilon_{k+2}, \varepsilon_{k+n}, \ldots$$

is either 0 or 1, and $\varepsilon_k = 1$. The numbers $\varepsilon_k, \varepsilon_{k+1}, \ldots$ are called 2-adic digits of the number p/q. It is not difficult to prove that any rational number has one and only one 2-adic expansion.

Exercise 8. Prove that

$$-1 = 1 + 2 + 2^2 + 2^3 + \ldots + 2^n + \ldots$$

and

$$1/6 = 2^{-1} + 1 + 2 + 2^2 + 2^4 + \ldots + 2^{2n} + \ldots$$

By virtue of exercise 8, the 2-adic digits of number −1 form the sequence 1, 1, 1, . . ., and the 2-adic digits of number 1/6 form the sequence 1, 1, 0, 1, 0, 1, 0, 1, . . . Both these sequences are periodic. We can prove that the sequence of 2-adic digits of any rational number is periodic. (In the same sense in which a decimal fraction that represents a rational number is periodic beginning from a certain digit.)

To prove this, consider the sequence of integers $a_1, a_2, \ldots, a_n, \ldots$ defined by equations

$$\frac{p}{q} - \varepsilon_k 2^k = \frac{a_1}{q} 2^{k+1},$$

$$\frac{p}{q} - \left(\varepsilon_k 2^k + \varepsilon_{k+1} 2^{k+1}\right) = \frac{a_2}{q} 2^{k+2}, \ldots,$$

$$\frac{p}{q} - \left(\varepsilon_k 2^k + \varepsilon_{k+1} 2^{k+1} + \ldots + \varepsilon_{k+n-1} 2^{k+n-1}\right) = \frac{a_n}{q} 2^{k+n},$$

where $\varepsilon_k, \varepsilon_{k+1}, \ldots, \varepsilon_{k+n-1}, \ldots$ are the 2-adic digits of p/q (the reader can check that $a_1, a_2, \ldots$ are indeed integers).

Now we can notice that the 2-adic expansion of p/q can be obtained from the 2-adic expansion of a_n/q by multiplying each term of this latter expansion by 2^{k+n} and adding to the beginning of the sum obtained the following:

$$\varepsilon_k 2^k + \varepsilon_{k+1} 2^{k+1} + \ldots + \varepsilon_{k+n-1} 2^{k+n-1}.$$

Therefore, the periodicity of the sequence will be proved if we prove that a certain number occurs twice in the sequence $a_1, a_2, \ldots, a_n, \ldots$ In fact, this sequence is bounded, and since its members are integers, there will inevitably be equal numbers among them. The boundedness of this sequence can be proved by the following chain of inequalities:

$$|a_n| = \left| \frac{p}{2^{k+n}} - \left(\frac{\varepsilon_k}{2^n} + \frac{\varepsilon_{k+1}}{2^{n-1}} + \ldots + \frac{\varepsilon_{k+n-1}}{2} \right) q \right|$$

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

Exercise 9. Using the result of exercise 7, prove that if $\varepsilon_k, \varepsilon_{k+1}, \ldots, \varepsilon_{k+2}, \ldots$ is a periodic sequence of digits 0 and 1, then the infinite sum $\varepsilon_k 2^k + \varepsilon_{k+1} 2^{k+1} + \ldots + \varepsilon_{k+2} 2^{k+2} + \ldots$ is the 2-adic expansion of a certain rational number.

2-adic numbers

In addition to periodic expansions, which determine rational numbers, we can consider nonperiodic expansions and say that they define a new class of numbers. These new numbers together with the rational numbers form a set Q_2, and the elements of this set are called *2-adic numbers*.

The elements of Q_2 can be added and multiplied. This is done in the following way. Let $\alpha = \varepsilon_k 2^k + \varepsilon_{k+1} 2^{k+1} + \ldots$ Write α as an infinite fraction: $\alpha = \varepsilon_k \ldots \varepsilon_0 \varepsilon_1 \varepsilon_2 \ldots$ The sum (product) of two numbers written in this manner can be calculated in the same way as the sum (product) of two infinite fractions, but with the proviso that we carry digits from left to right (two examples are given in figure 4). We cannot consider this issue in more detail here, and we invite the reader to prove that the sum and product of two elements of Q_2 is also an element of Q_2 and that these operations obey the commutative, associative, and distributive laws.

```
  1011.0101¹...        110.11001¹...
+  101.1010...      ×    0.01101...
  ------------      ---------------
  1110.0000...           1.10110...
                         0.11011...
                         0.00110...
                         .......
                    ---------------
                         1.001....
```

Figure 4

The 2-adic modulus and distance can be naturally extended for Q_2: if $\alpha \in Q_2$ and $\alpha = \varepsilon_k 2^k + \varepsilon_{k+1} 2^{k+1} + \ldots$ (where $\varepsilon_k \neq 0$), then $\|\alpha\| = 1/2^k$; if $\alpha, \beta \in Q_2$ then $\rho(\alpha, \beta) = \|\alpha - \beta\|$.

Exercise 10. Given the expansion of a number $\alpha \in Q_2$, construct the expansion of $-\alpha$.

Exercise 11. Prove that the ordinary sum of two rational numbers and the sum of these numbers in Q_2 correspond to the same number. Prove the same for the product.

Exercise 12. Learn to perform division in Q_2.

We know that the sum of an infinite sequence $\{x_n\}$ can be defined under certain conditions. A necessary (however, not sufficient) condition is that $\lim_{n\to\infty} x_n = 0$. The situation is simpler for 2-adic numbers. The condition $\lim_{n\to\infty} \|x_n\| = 0$ is necessary and sufficient: if $\lim_{n\to\infty} \|x_n\| = 0$, then the limit

$$\lim_{n\to\infty} \sum_{k=1}^{n} x_k$$

exists.

Exercise 13. Prove the sufficiency of the condition $\lim_{n\to\infty} \|x_n\| = 0$.

Exercise 14. Prove that the sum $1 \cdot 1! + 2 \cdot 2! + \ldots + n \cdot n! + \ldots$ exists and find its value.

Exercise 15. Prove that if $x \in Q_2$ and $\|x\| \le 1/2$, the sum $x/1 + x^2/2 + \ldots + x^n/n + \ldots$ exists.

2-adic logarithm

In conclusion, consider the following problem: The number

$$\frac{2}{1}+\frac{2^2}{2}+\frac{2^3}{3}+\cdots+\frac{2^n}{n}$$

is represented as an irreducible fraction p_n/q_n.

(i) Prove that p_n is even.

(ii) Prove that if $n > 3$, then p_n is divisible by 8.

(iii) Prove that for any natural k, a number n can be found such that $p_n, p_{n+1}, p_{n+2}, \ldots$ are all divisible by 2^k.

Consider the function

$$L(x)=\frac{x}{1}+\frac{x^2}{2}+\cdots+\frac{x^n}{n}+\cdots$$

defined on the set of 2-adic numbers x that are not greater than 1/2 in modulus (see exercise 15). To prove item (iii) of the problem, it is sufficient to show that $L(2) = 0$. Define another function by the formula $\log x = -L(1-x)$. The equality $L(2) = 0$ means that $\log(-1) = 0$. The function log (it is called the *2-adic logarithm*) possesses the basic property of the logarithmic function: $\log(xy) = \log x + \log y$. This property immediately implies that $\log(-1) = 0$. Indeed, $\log 1 = \log(1 \cdot 1) = \log 1 + \log 1 = 2\log 1$, from which it follows that $\log 1 = 0$. On the other hand, $\log 1 = \log((-1)\cdot(-1)) = \log(-1) + \log(-1) = 2\log(-1)$. Therefore, $\log(-1) = 0$. The 2-adic logarithm plays the same role in the set of 2-adic numbers as the common logarithm does in the set of real numbers. The 2-adic logarithm is defined for x with modulus 1: $\|x\| = 1$, since it is exactly the case when $\|1 - x\| \le 1/2$; in particular, $\log(-1)$ is defined.

Besides the logarithm, other remarkable functions of a 2-adic variable exist. One of them is the *exponential function* that is defined as follows:

$$\exp x = 1 + x + \frac{x^2}{2}+\cdots+\frac{x^n}{n!}+\cdots$$

Exercise 16. Prove that the exponential function is defined on the set of 2-adic numbers x that satisfy the condition $\|x\| \le 1/4$.

The basic properties of the 2-adic exponent are similar to those of the usual exponent: $\exp(x + y) = \exp x \cdot \exp y$, $\exp(\log x) = x$, and $\log(\exp x) = x$. These identities hold for the values of the variables for which the corresponding functions are defined.

Just as the 2-adic distance has been defined, we can define a p-adic distance taking an arbitrary prime number p. It turns out that all distances defined for rational numbers are equivalent to the usual distance or to one of the p-adic distances. However, this is a topic for another article.

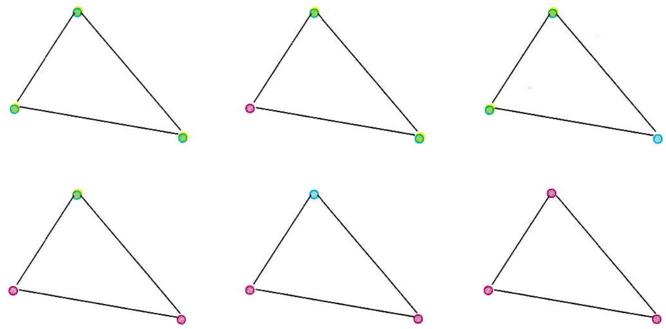


Figure 5

Appendix

It remains for us to prove that a triangle with vertices of the three different colors does exist. We formulate and prove a more general proposition: *Let a square* OABC *be decomposed into several triangles. Assume that each vertex of these triangles is colored green, red, or blue in such a way that no line contains points of all three colors. Let* O *be the green point,* A *and* B *be red, and* C *be blue. Then there is a triangle with vertices of three different colors among the triangles of the decomposition.*

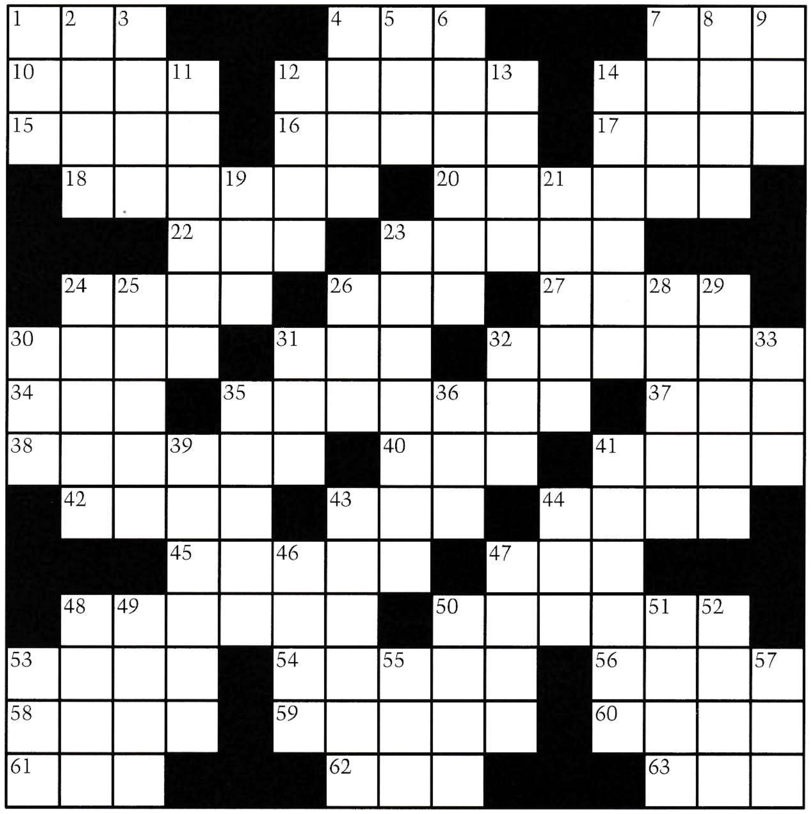
Proof. It is convenient to differentiate between the sides of triangles and segments that are parts of the triangles' sides formed by vertices of other triangles falling on these sides. If a side of a triangle contains no vertices of other triangles, it is considered as a segment itself.

We distinguish six types of segments and six types of sides, depending on the colors of their endpoints: *GG* (both endpoints are green), *GR* (one endpoint is green and the other is red), and so on (*GB*, *RR*, *RB*, and *BB*). We prove that the sides of a triangle that has two vertices of the same color contain an even number of *GR* segments. Indeed, since no lines contain points of all three colors, a *GR* segment can lie only on sides of types *GG*, *GR*, and *RR*; sides of type *GG* and *RR* contain an even number of such segments and sides of type *GR* an odd number. Therefore, the sides of each of the triangles depicted in figure 5 contain an even number of segments of type *GR* (triangles with two or three blue vertices do not contain such segments).

Now assume that none of the triangles of the decomposition has vertices of all three colors—that is, each of the triangles has two vertices of the same color. Every segment that lies on a side of the square *OABC* belongs to a side of exactly one decomposition triangle, and every segment inside the square belongs to sides of two decomposition triangles. Since any decomposition triangle has two different-colored vertices, the sides of the square contain an even number of segments of type *GR*. On the other hand, sides *OC* and *BC* contain no segments of this type, side *OA* contains an odd number of such segments, and side *AB* contains an even number, which gives an odd number in total. Thus, we have arrived at a contradiction. ◘

Criss×cross science

by David R. Martin



Across

1 ___ bolt (barbed anchor bolt)
4 ___ cycle (Krebs cycle)
7 Trig. function
10 44,523 (in base 16)
12 Sodium hypochlorite
14 1/760 atmosphere
15 10^{-9}: pref.
16 Astronomer ___ Cannon (1863–1941)
17 39D arsenide
18 Strong nuclear particles?
20 Type of engine
22 Type of parity
23 18A members
24 Swedish chemist ___ Bergstrom
26 Zirconium carbide
27 Baseball scoreboard item
30 Ionizing rad. units
31 *The ___ of Physics*
32 Soap ingredient
34 Sphere
35 Fermions
37 Rocky peak
38 Element 96
40 Mine output
41 South African writer Alex La ___
42 German novelist Thomas ___
43 Genetic material
44 ___ mater
45 Letterman and Roth
47 Unit of mass: abbr.
48 An asteroid
50 Two dimensional quasiparticles?
53 Boyfriend
54 Balanced constellation?
56 Lowest tide
58 Peaceful
59 Follow
60 Alone: comb. form
61 Calcium oxide
62 Liquid hydrocarbon
63 Nitrilotriacetic acid

Down

1 Astronomer ___ Oort
2 Swedish botanist ___ Afzelius (1750–1837)
3 Unit of heredity
4 Sunbathes
5 Pro and ___
6 Having a pH < 7
7 Retina cell
8 Type of exam
9 Strontium sulfide
11 Particles with integer spin
12 A logic circuit
13 Smooth: comb. form
14 Group of similar cells
19 Songlike poem
21 Compounds containing –CH=C(OH)–
23 Nuclear particles
24 Blood ___
25 Full shadow
26 Comic-strip word
28 1958 Physiology Nobelist Edward L. ___
29 Leaf pore
30 Electrical conductivity unit
31 Wave guide mode: abbr.
32 Identity element of a set
33 Paleozoic or Precambrian, e. g.
35 Type of eclipse
36 Lake Garda wind
39 Element 49
41 Hypothetical massless particles
43 Actor Robert ___
44 British aviator ___ Johnson (1903–1941)
46 Unit of electrical potential
47 ___ sphincter (certain muscle)
48 52,906 (in base 16)
49 Luminous ring
50 Seed covering
51 Element 10
52 Sodium chloride
53 Crystal structure: abbr.
55 Yellow mist in China
57 ___ mater (brain membrane)

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE MAY/JUNE PUZZLE

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

The pointed meeting of a

THERE IS MORE than one way to skin a cat, and the same can be said for proving a geometric theorem. This is particularly true for the well-known theorem stating that the altitudes of any triangle are concurrent—they all pass through the same point, which is called the orthocenter of the triangle. (To be more precise, we mean here that the lines containing these altitudes meet at a point.) We present here several ways of proving this important theorem.

But first some preliminaries. Let us note that it is sufficient to prove this theorem for acute triangles. Indeed, suppose we have done so, and let HBC be an obtuse triangle, with an obtuse angle at vertex H. Draw perpendiculars from points B and C to the opposite sides of the triangle and label the point of intersection of these lines point A (fig. 1). Triangle ABC is acute, and so, by our assumption, its altitudes intersect in a point. So if we draw perpendiculars for C to AB, and from B to AC, their point of intersection will lie on the perpendicular from A to BC. But their point of intersection is just H! So the line through A and H is perpendicular to BC and is the third altitude (both in triangle ABC and in HBC). In particular, the altitudes of HBC concur at A.

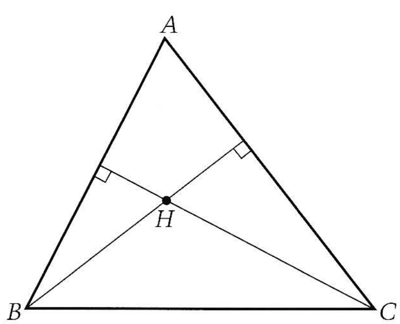


Figure 1

Right under our nose, we have proved a very interesting theorem: If H is the orthocenter of triangle ABC, then:

A is the orthocenter of triangle HBC,
B is the orthocenter of triangle HAC,
C is the orthocenter of triangle HAB.

Proof 1. *Auxiliary circle method.* Consider an acute triangle ABC (fig. 2). Let BB_1 and CC_1 be the altitudes

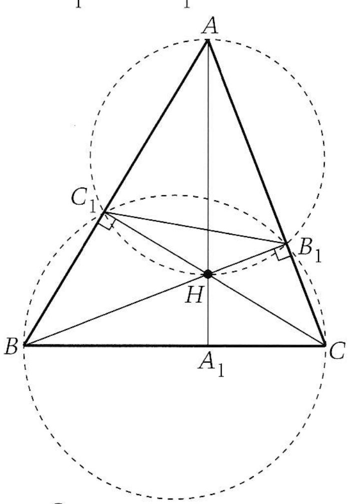


Figure 2

of this triangle and let H be their point of intersection. Draw the line AH and let A_1 be the point of its intersection with BC. Notice that points B, C, B_1, and C_1 lie on a circle (with diameter BC). Therefore, $\angle B_1C_1C = \angle B_1BC$ (these angles subtend the same arc of the auxiliary circle). Notice that points A, B_1, H, and C_1, also lie on a circle (with diameter AH). Therefore, $\angle B_1C_1C = \angle B_1C_1H = \angle B_1AH = \angle CAA_1$. Triangles CAA_1 and CBB_1 have a common angle and a pair of equal angles. Therefore, the remaining angles of these triangles are equal. Hence, $\angle AA_1C = \angle BB_1C$, and AA_1 is the the altitude of the triangle considered.

Proof 2. *Other auxiliary circles.*

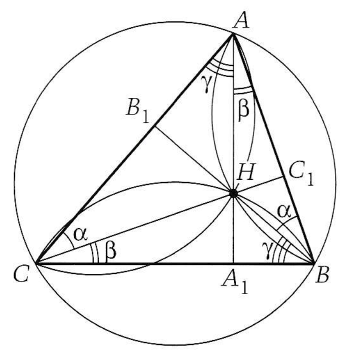
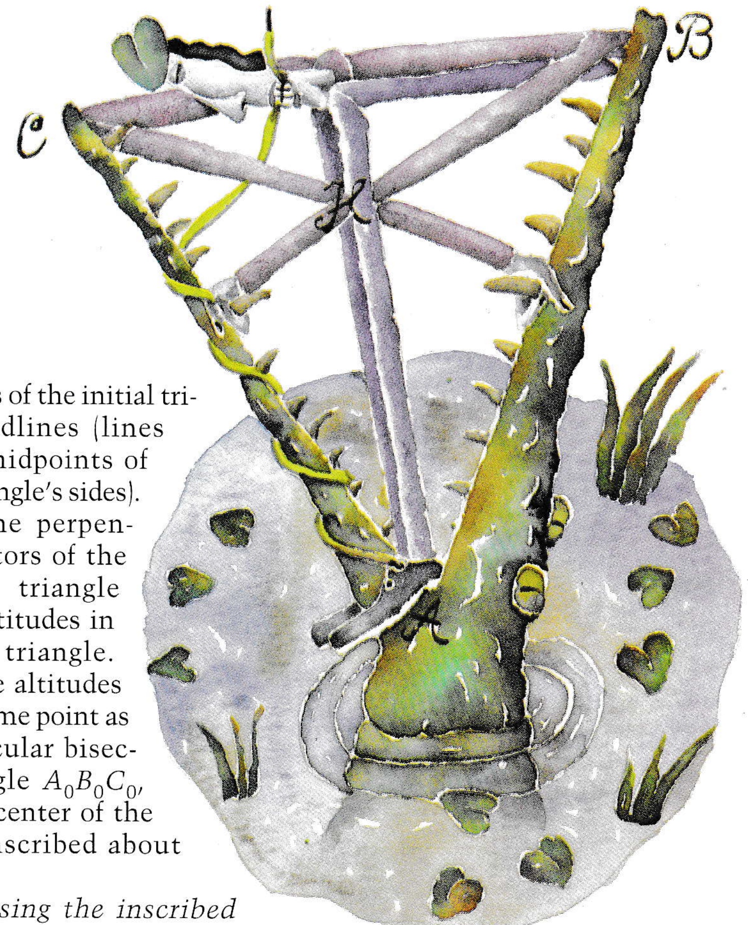


Figure 3

Consider a circle circumscribed around an acute triangle ABC. The sum of the arcs of this circle that are subtended by the triangle's sides is 360°. Therefore, the arcs that are symmetric to these arcs with respect to the triangle's sides meet at a point, which is labeled H in figure 3 (the proof of this interesting fact is left to the reader). We can see that $\angle ABH = \angle ACH$, since these angles are acute and subtended by the same chord of equal circles. Let each of these angles be α. Similarly, we have two other pairs of equal angles: $\angle BCH = \angle BAH = \beta$, and $\angle ACH = \angle ABH = \gamma$. Since two copies each of α, β, and γ exactly cover the three angles of triangle ABC, we have $\alpha + \beta + \gamma = 90°$.

Before we continue, we look a bit at the context of what we are doing. The concurrence of the three altitudes of a triangle is perhaps the most difficult of several concurrence theorems in elementary geometry.

f a triangle's altitudes



Let us review these.

Theorem: The three perpendicular bisectors of the sides of a triangle concur at the center of a circle through the triangle's three vertices.

This is easy to prove. We take a triangle ABC and draw the perpendicular bisectors of sides AB and AC. Their point of intersection (since they cannot be parallel!) is equidistant from A and B, and also from B and C. Thus it is equidistant from A and C, and so lies on the perpendicular bisector of BC.

Theorem: The three bisectors of the angles of a triangle concur at the center of a circle tangent to the triangle's three sides.

This is also easy to prove. We take a triangle ABC and draw the bisectors of angles A and B. Their point of intersection (since they cannot be parallel!) is equidistant from lines AC and AB, and also from lines BC and AB. Hence it is equidistant from lines AC and BC and so lies on the bisector of angle C.

We can sometimes use these theorems, which are so easy to prove, to prove the concurrence of the three altitudes.

Proof 3. *Using the circumscribed circle.* Through the vertices of ABC, draw the lines parallel to the opposite sides of the triangle (fig. 4) to obtain a triangle $A_0B_0C_0$. In this triangle the sides of the initial triangle are midlines (lines connecting midpoints of two of the triangle's sides). Therefore, the perpendicular bisectors of the sides of the triangle $A_0B_0C_0$ are altitudes in the initial triangle. Therefore, the altitudes meet at the same point as the perpendicular bisectors of triangle $A_0B_0C_0$, which is the center of the circle circumscribed about $A_0B_0C_0$.

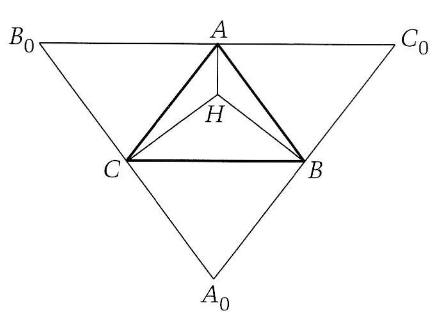


Figure 4

Proof 4. *Using the inscribed circle.* Let A', B', and C' be the points at which the altitudes of triangle ABC meet the circle circumscribed around it (fig. 5). We have $\angle ABB' = \angle ACC'$

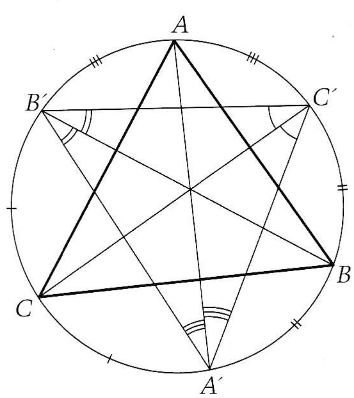


Figure 5

(they're both complementary to $\angle CAB$), so $AB' = AC'$. Now we can see that AA', BB', and CC' are angle bisectors of triangle $A'B'C'$. Therefore, they meet at the center of the circle inscribed in triangle $A'B'C'$.

Let us look back for a minute at the proofs of the concurrence of the perpendicular bisectors and the angle bisectors. What made these proofs so easy? We were able to describe each of the lines we were discussing as a locus, satisfying some condition of equality, and involving two of the vertices of the triangle. Then the transitive property of equality (if $x = y$ and $y = z$, then $x = z$) did the work for us. If we can describe the triangle's altitude as loci, we can get another simple proof of their concurrence.

But first let us practice on the medians of a triangle. There are many proofs that these concur (see "The Medians," by V. Dubrovsky in the November/December 1994 issue, page 32).

Theorem: Median AM of triangle ABC is the locus of point P such that the areas of triangles CPA and CPB are equal.

Proof: For any point P on AM, triangles APM, BPM are equal in area, since they have the same altitude from P, and equal bases $AM = BM$. The same thing is true of triangles ACM, BCM. So the differences in these two areas are equal: using ab-

CONTINUED ON PAGE 46

Image charge

by Larry D. Kirkpatrick and Arthur Eisenkraft

> *"We operate with nothing but things which do not exist, with lines, planes, bodies, atoms, divisible time, divisible space—how should explanation even be possible when we first make everything into an image, into our own image!"*
> —*Friedrich Nietzsche (1844–1900)*

WE ARE ALL FAMILIAR WITH images formed by mirrors and lens. In fact, we often pay to see images produced in special ways such as those in fun houses or at the Haunted Mansion in Disneyland.

But what is an "image charge?" Does it have anything to do with optical images? Let's investigate this by considering the following problem.

We are given a conducting plate that is so large that we can imagine that it reaches to infinity in the plane of the plate. Alternatively, we can work close enough to the plate and far enough from the edges that the plate might as well stretch to infinity.

Let's now place a charge q a distance d in front of the middle of the plate. If the plate is grounded, is the charge q attracted to or repelled by the plate? And what is the strength of the force acting on the charge?

Experimentally, we can demonstrate that there is an attractive force between the charge and the metal plate. Run a comb through your hair and use it to pick up small pieces of aluminum foil.

We can also see this qualitatively. If we bring a positive charge near a metal plate, the positive charge will attract the electrons in the plate, causing them to concentrate in the area nearest the charge. Because these electrons have moved closer to the charge, their attractive force is larger than the repulsive force of the positive ions left behind.

Quantitatively, this looks like a complicated problem, but it can be solved rather easily using one of the "tricks of the trade." This technique relies on a uniqueness theorem for the electrostatic potential. Remember that the electrostatic potential at a point in space is the amount of work required to bring a unit positive charge to the point from a place where the potential is zero. Suppose that we are given the value of the electrostatic potential at every location on the entire boundary (surface) of a volume of space. If by hook or crook, we can find a formula that gives the correct values at all points on the boundary, this formula also gives the values of the electrostatic potential throughout the volume. Furthermore, the formula is unique. No matter what other technique we may use, we will get the same potential. This means that we can search around for a simple way of finding the electrostatic potential.

One way is to use the method of images. We imagine that we can replace the metal plate with an image charge Q. This charge is located behind the original metal surface along the normal from the original charge to the surface as shown in figure 1. (Remember that the image of an object in front of a plane mirror is located along a similar line. In fact, we might guess by analogy that the image charge is located a distance d behind the original surface.) For the moment, let's leave the distance D of the image charge behind the plate as unknown.

If we choose the potential to be zero at infinity, the potential at a distance r from a point charge q is given by

$$V = \frac{kq}{r},$$

where k is Coulomb's constant $= 1/4\pi\varepsilon_0$. One of the reasons for in-

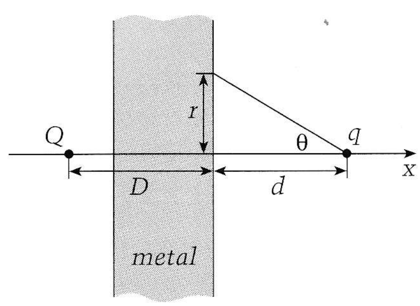
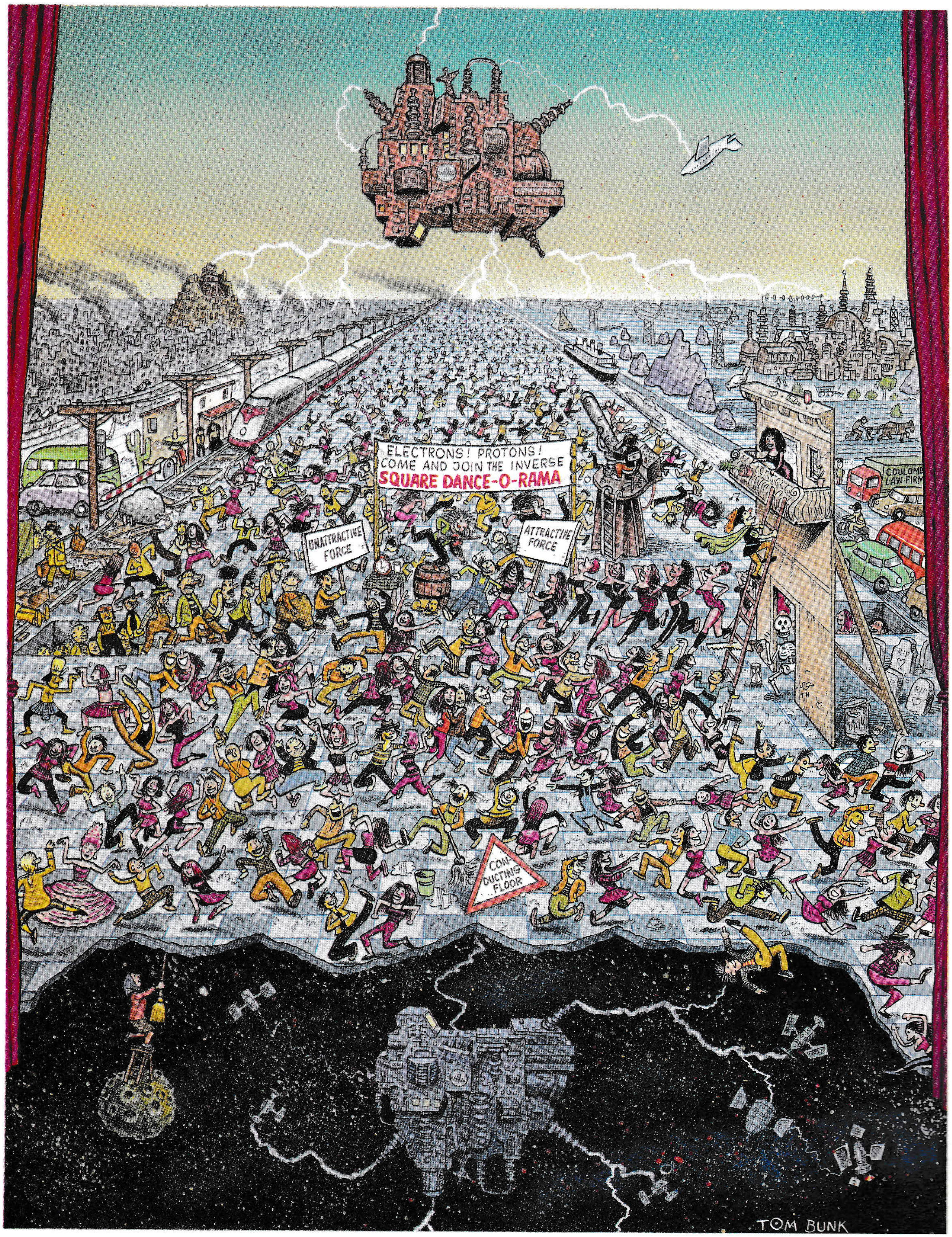


Figure 1

Art by Tomas Bunk


ELECTRONS! PROTONS! COME AND JOIN THE INVERSE
SQUARE DANCE-O-RAMA
UNATTRACTIVE FORCE
ATTRACTIVE FORCE
COULOM LAW FIRM
CON-DUCTING FLOOR
TOM BUNK

troducing the idea of potential is that potential is a scalar. The potential due to a collection of point charges is just the sum of the potentials due to each charge. In contrast, we must add the electric field due to each charge as vectors to get the total electric field.

Returning to our problem, let's choose the origin of our coordinate system to be on the surface along the normal to the charge. At any location along the original surface, the potential due to both charges is

$$V = \frac{kq}{\sqrt{d^2 + r^2}} + \frac{kQ}{\sqrt{D^2 + r^2}}.$$

where r is the distance along the plane measured from the normal. Because the plane is grounded, $V = 0$ and

$$\frac{q}{\sqrt{d^2 + r^2}} = \frac{-Q}{\sqrt{D^2 + r^2}}.$$

We need to find the values of two variables, Q and D, but we only have one equation. However, we also have the condition that this equation must be satisfied for all values of r. In this particular case, we can guess the solution, $Q = -q$ and $D = d$, but it is instructive to obtain the solution in a more formal manner.

Let's square the equation and multiply through by both denominators to obtain

$$\left(D^2 + r^2\right)q^2 = \left(d^2 + r^2\right)Q^2.$$

We now collect terms containing r^2 on one side of the equation and all other terms on the other side:

$$D^2q^2 - d^2Q^2 = r^2\left(Q^2 - q^2\right).$$

For this equation to be valid for all values of r, the coefficient of r^2 must be zero. Therefore, $Q = \pm q$. We choose the minus sign because this is the only way the potential can be zero at the surface. The left-hand side of the equation must also be zero, giving us $D = d$.

This tells us that the charge q can be imagined to induce a charge $-q$ a distance d behind the metal surface. According to the uniqueness theorem, we can now use this image charge to calculate results at all points in front of the metal.

Armed with this result we can answer our original questions. Because q and Q have opposite signs, the charge q is attracted to the metal plate. The strength of this attractive force is given by Coulomb's law:

$$F = k\frac{qQ}{(d + D)^2} = -k\frac{q^2}{4d^2},$$

where k is Coulomb's constant.

The electrostatic potential is obtained by substituting our conditions back into the formula for the potential:

$$V = kq\left[\frac{1}{\sqrt{(x-d)^2 + r^2}} - \frac{1}{\sqrt{(x+d)^2 + r^2}}\right].$$

We can also use the image charge to calculate the electric field at all points in the volume and on its boundary. In particular, let's do this along the metal surface. Symmetry tells us that

$$E_r(0, r) = 0.$$

We also know this because the electric field must be normal to all metallic surfaces. The two charges contribute equal amounts to the normal component:

$$E_x(0,r) = \frac{-2kq}{d^2 + r^2}\cos\theta = \frac{-2kqd}{\left(d^2 + r^2\right)^{3/2}}.$$

One reason for calculating this electric field is that it allows us to find the actual charge distribution induced on the metal surface by the charge q. The induced surface charge density is

$$\sigma(r) = \varepsilon_0 E_x(0,r) = \frac{-qd}{2\pi\left(d^2 + r^2\right)^{3/2}}.$$

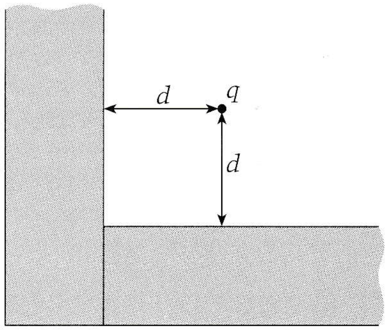


Figure 2

It is interesting to note that if you integrate this charge density over the entire surface, you obtain a total charge of $-q$, the same value we obtained for the image charge.

This brings us to this month's contest problems:

A. Two large metal plates form a right-angle corner. A charge q is placed within the corner, equal distances d from both plates, and far from the edges of the plates (see figure 2). What is the force acting on the charge?

B. The second of this month's contest problems is based on a problem given on the semifinal exam used to select members of the US Physics Team that will compete in Italy in July. A charge q is placed a distance d from the center of a grounded metal ball with radius $c < d$. The electrostatic potential is chosen to be zero at infinity. What is the force acting on charge q?

Sportin' life

In the January/February issue, we posed a series of problems that required an understanding of trajectories in sports events. Before beginning with the solutions, we must apologize for two errors that appeared in the article. In the first equation showing the dependence of the y-coordinate on the x-coordinate and the angle, the first term should have had an x^2 rather than an x in the numerator. The second error was the reversal of the axis labels on the graph.

The first problem asks for the locations where a soccer ball cannot land if a wall of defenders 1.8 m high is set up 15 m from the free kick.

The strategy for the solution is to find the two angles for which the ball can be kicked at a certain velocity that will take it to the point (1.8 m, 15 m). Knowing these two angles, we then find the corresponding x values for the ball landing on the ground ($y = 0$):

$$y = \frac{-gx^2}{2v_0^2}\tan^2\theta + x\tan\theta - \frac{gx^2}{2v_0^2}.$$

Substituting the values for x, y, and $v_0 = 35$ m/s and solving the quadratic equation yields values for θ of 86.5° and 10.4°. We use the range equation to find the points where these two trajectories hit the ground.

$$x = \frac{2v_0^2}{g}\cos\theta\sin\theta = \frac{v_0^2}{g}\sin 2\theta.$$

The corresponding distances are 15.11 m and 44.04 m. The shadow region lies behind the wall and extends for 0.11 m.

Part b asks what happens to this shadow region as the wall moves relative to the kicker. Calculating the shadow region on a spreadsheet, we determine that the shadow region increases as the distance from the kicker increases. This makes sense because the angle required to just clear the wall will decrease and therefore the ball will not be able to fall as sharply behind the wall.

Problem 2 shifted to trajectories in basketball. It required readers to find the relationship between the initial velocity v_0 and the initial angle θ_0 given a fixed shot position where the rim is h meters above the ball and L meters away horizontally.

Beginning with our trajectory equation

$$y = \frac{-gx^2}{2v_0^2\cos^2\theta} + \frac{x\sin\theta}{\cos\theta},$$

we solve for v_0:

$$v_0^2 = \frac{gL}{2\cos^2\theta}\,\frac{1}{\tan\theta - \dfrac{h}{L}}.$$

Part b of this problem asks for the constraints on the initial angle if the ball is to enter the basket during its descent. The angle of entry can be defined as the angle between the horizontal and the angle of the tangent to the ball's trajectory. We will assume that we can ignore the size of the ball—a terrible assumption, but one that makes the analysis simpler.

The initial constraints should depend only on h and L, where

$$h = (v_0\sin\theta_0)t - \frac{1}{2}gt^2,$$

$$L = (v_0\cos\theta_0)t,$$

$$\tan\theta = \frac{v_y}{v_x} = \frac{v_0\sin\theta_0 - gt}{v_0\cos\theta_0}$$

$$= \frac{(v_0\sin\theta_0)t - gt^2}{(v_0\cos\theta_0)t}$$

$$= \frac{2(v_0\sin\theta_0)t - gt^2 - (v_0\sin\theta_0)t}{(v_0\cos\theta_0)t}$$

$$= \frac{2h}{L} - \tan\theta_0.$$

Because the ball must be falling in order to make a basket, θ and $\tan\theta$ must both be negative. This requires that the right side of the equation also be negative. Therefore, the constraint on the initial angle is

$$\tan\theta_0 > \frac{2h}{L}.$$

Part c asks for the angle where a minimum speed is required to sink the shot. To solve this, we can calculate the speed required for many different angles using the equation from Part a and a spreadsheet, assuming values of $h = 1$ m and $L = 5$ m. The graph in figure 3 of v vs. θ indicates a minimum at approximately 50°. Alternatively, we can take the derivative of the equation from Part a relating v to θ and set this derivative equal to zero to obtain

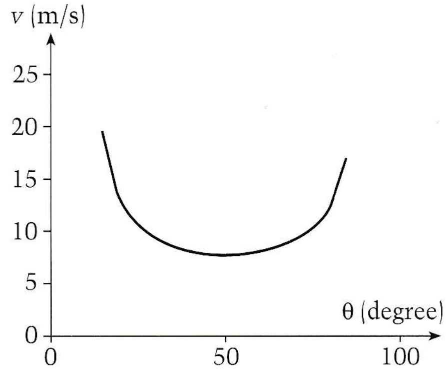


Figure 3

$$v_0 = \sqrt{\frac{gL}{2\cos^2\theta}\,\frac{1}{\tan\theta - \dfrac{h}{L}}},$$

$$\frac{dv}{d\theta} = 0,$$

$$\tan\theta = \frac{h}{L} \pm \sqrt{\frac{h^2}{L^2} + 1}.$$

Using the sample values of $h = 1$ m and $L = 5$ m, we arrive at an optimal shooting angle of 50.7°.

Peter Brancazio, Physics Professor Emeritus of Brooklyn College, took this analysis further and then compared theory and practice on the basketball court. We highly recommend Brancazio's 1981 article "Physics of Basketball" in the *American Journal of Physics* (49), 356–365.

AT THE BLACKBOARD I

A cardioid for a mushroom picker

by S. Bogdanov

ONE CLEAR SEPTEMBER morning my friends and I went to the forest to pick mushrooms. We went along a path in the forest to a well-marked pole, then decided to go west separately and meet 4 hours later at the pole. We didn't have compasses, but the landmarks were reliable: the path went strictly from north to south, the Sun was in the east, and two big lakes were on the edges of the forest (fig. 1).

So as to avoid distraction from picking mushrooms and to arrive on time at the meeting place, I decided to walk for half the time (2 hours) so that the Sun was shining on my back and then return in the opposite direction. I collected a full basket of mushrooms, then found that I had to make certain corrections to my route. As a result, I returned to the path about 1 km south of the meeting place and was 15 minutes late.

Then I decided to make sense of my trajectory and find out where I would find myself at the end of the walk if I had strictly adhered to my plan. After making certain reasonable assumptions, I formulated the problem as follows.

A point B (the Sun) moves in the plane XY with a constant angular speed ω along a large circle centered at the origin (fig. 2). Point A (a person) moves from the origin with a constant speed v in the direction "away from point B." In other words, at any moment in time, the velocity is oriented along the line "Sun–person." At a certain time t, the direction of the velocity is reversed, and point A begins to move in the direction "toward B." What are the coordinates of point A at time $2t$?

The solution proved to be quite simple and understandable, especially for those who have at least once walked "in circles" in the forest. Indeed, point A moves at a constant speed, but the velocity vector changes such that it is always directed "from point B," which describes a circle. Therefore, point A also moves along a circle with the same angular speed as point B. While point B moves to position B', the person moves along the circular arc AA' to position A'. It is clear that the return motion from the turning point A' to the final point A'' can be described similarly. The point moves along the circular arc $A'A''$ of the same radius, and its center, O', lies on the line that is perpendicular to $A'B'$ and passes through the point of tangency A'.

Using the well-known relation between linear and angular velocity, we obtain the following formula for the radius of the circle AA': $R = v/w$. In addition, it follows from the initial conditions that the center of this circle is at the point with coordinates $(v/w, 0)$. Simple geometrical considerations (for example, of trapezoid $AOO'A''$) make it possible to determine the location of point A'', which is the distance r from the initial point A at the angle $\theta = \angle A''AO$:

$$\theta = \pi - \omega t,\ r = AA'' = 2v\frac{1 + \cos\theta}{\omega}.$$

The quantities r and θ are called the polar coordinates of a point. Let's analyze this result for a characteristic interval of possible values of t: $0 \le t \le 6$. We use the fact that $\omega = 2\pi/24\ \text{h}^{-1}$. For small t, θ is a little less than π. That is, there is a considerable azimuthal deviation from the

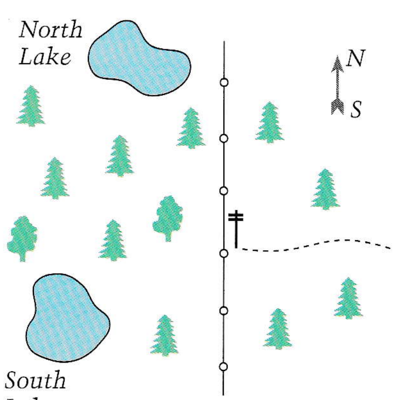


Figure 1

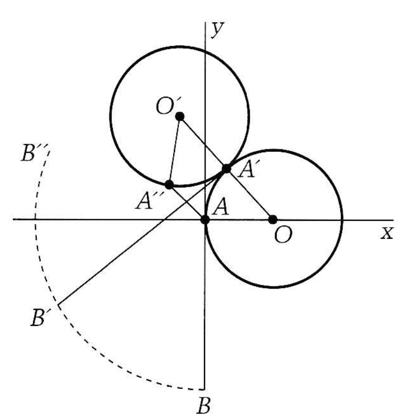
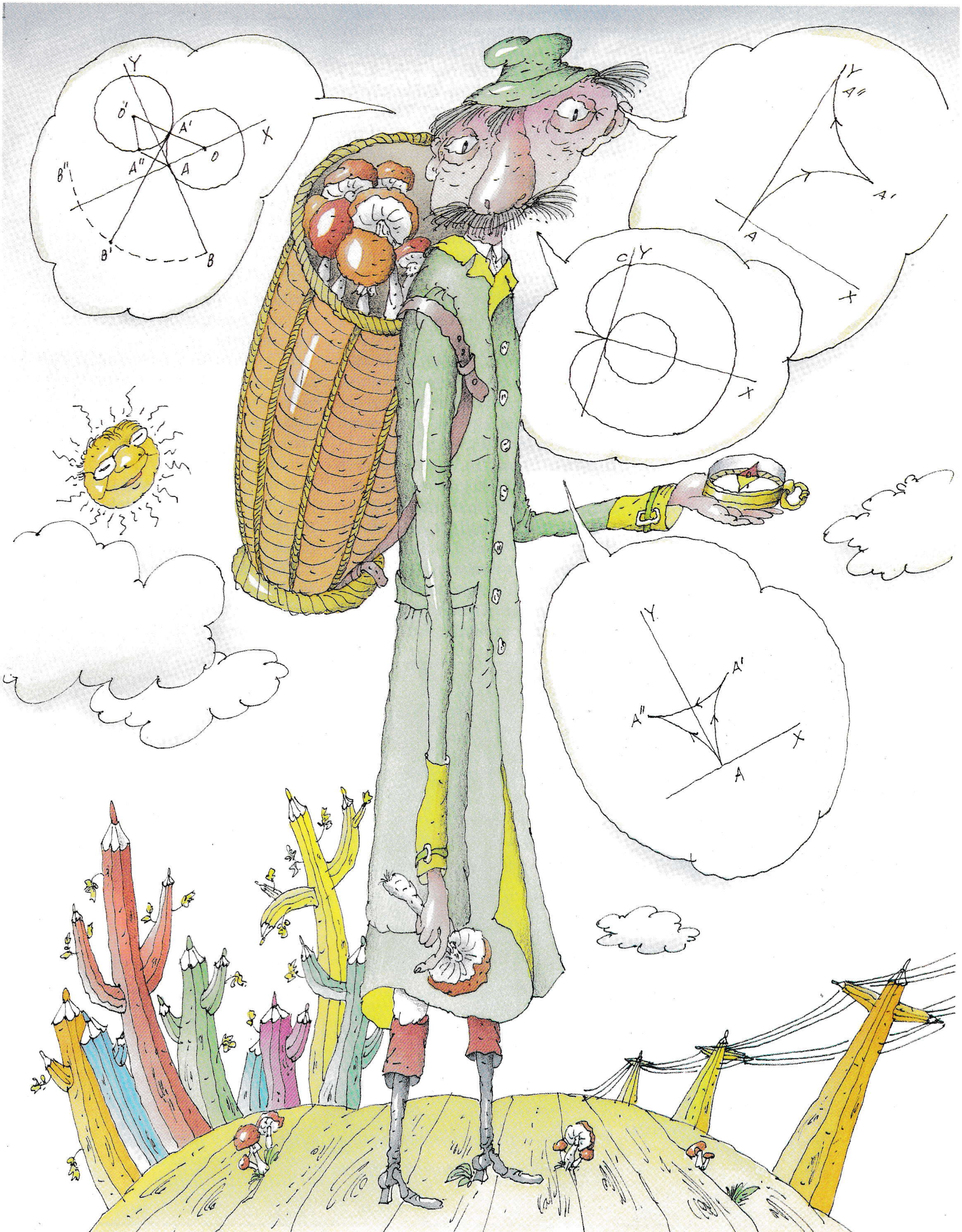


Figure 2

Art by V. Ivanyuk


Y
O'
A'
X
O
A''
A
B''
B'
B
Y
A''
A'
A
X
Y
X
Y
A'
A''
X
A

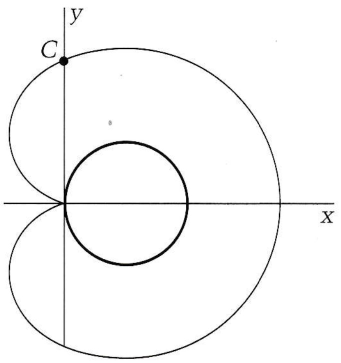


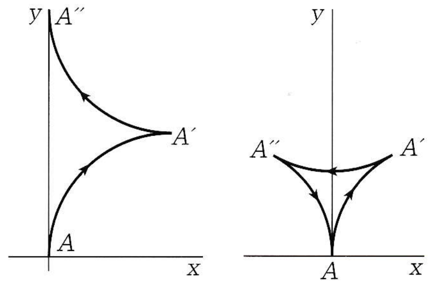
Figure 3

initial direction $\pi/2$, while the distance r is rather small. If t is about 6 h, the azimuth deviation is practically zero, but r becomes rather large.

For example, let us make a numerical estimate for $t = 2$ h and $v = 2$ km/h. In this case, $\theta = 5\pi/6$, and the radius R and distance r are about 8 km and 2.4 km, respectively.

Figure 2 also makes it possible graphically to demonstrate the location of the final point A''. Suppose that the circles O and O' touch each other at the initial moment at point A. Then, circle O' starts rolling without slipping along circle O. By the time t, when the circles touch each other at the turning point A', the initial point of tangency A moves to the position A''. Thus, the set of all possible final points of the route at a constant speed v and various t coincides with the trajectory of a point of circle O' that rolls along circle O. The corresponding circle is called a *cardioid*—its plot is shown in figure 3, and its equation was obtained above. This curve can be constructed using improvised tools, and the corresponding drawing may be useful in the forest if no map is available or as a complement to the map.

Point C in figure 3 is the final position at $t = 6$ h, and the corresponding trajectory is shown in fig. 4a. In this case, the azimuth deviation is zero, but $r \cong 16$ km for the characteristic values of the parameters. Thus, the whole-day stroll ($2t = 12$ h) with the strategy "from the Sun, to the Sun" is clearly unfortunate and even dangerous. However, an alternative three-segment trajectory (fig. 4b) can be suggested for such a stroll. This



a b

Figure 4

strategy keeps the main advantage, which is orientation with respect to the Sun only.

In conclusion, I would like to note that a more complex case, when the outbound and return motions have different speeds, can be analyzed in a similar way. In particular, it is not difficult to show that all final points will lie on the line $A'A''$. The increased speed at the way back only increases the distance from the initial point A.

We therefore can confidently conclude from our analysis that when you go into the forest, you should bring a compass!

Electric currents on coulomb hills

by E. Romishevsky

TODAY WE CONSIDER ELECtrical circuits carrying direct current that are composed of wires and batteries. First, however, we recall how a good old capacitor works.

When a voltmeter is connected to the plates of charged capacitor, a circuit is thereby closed. This means that the total work needed to transfer an electric charge along this circuit in the electric field is zero. We assume the voltmeter to be ideal (for example, of the electrostatic type), and thus it has infinite internal resistance and a capacitance of zero. Moving clockwise from positive plate 1 to negative plate 2 (fig. 1), we record the decrease in potential as $V_0 = Q_0/C$, where Q_0 is the charge and C is the capacitance of the capacitor. The potential doesn't vary along the connecting wires, and it will increase by V_0 in the voltmeter. This is the value that is recorded by an ideal voltmeter, according to the attractive force acting, for example, between its plates 3 and 4.

Connect plates 1 and 2 of the uncharged capacitor to a battery with an electromotive force (emf) $\mathscr{E}_b$. The capacitor will be charged to a voltage $V_0 = \mathscr{E}_b = Q_0/C$. Consider the closed circuit 1-2-3-4-1. We move along it in the clockwise direction (fig. 2). When we step from positive plate 1 with potential ϕ_1 to negative plate 2 with potential ϕ_2, the potential decreases. Therefore, the voltage difference $\phi_2 - \phi_1 = -V_0 = -Q_0/C$ is negative. When we step from the negative electrode (cathode) of the battery with potential ϕ_3 to the positive anode with potential ϕ_4, the same voltage difference in the battery's field is positive: $\phi_4 - \phi_3 = V_0$, so at plate 1 we arrive at the same potential:

$$\phi_2 - \phi_1 + \phi_4 - \phi_3 = -V_0 + V_0 = 0.$$

A question arises: what is a battery and what role does it play in physics? As we said, there is an electric field inside a battery, which generates (or is described by) a voltage difference $V_0 = \mathscr{E}_b$. This means that negative and positive charges are separated in the battery. When we connect the battery to an uncharged capacitor, an electric charge Q_0 passes in the battery in the direction opposite to the field. The energy of this charge increases (and so does the voltage, which is the potential energy of a unit charge). Thus, a battery increases the potential energy of the charges moving in an electric circuit. The only way to do this is to move charges against the electric field inside the battery. What forces perform such heroic work?

Recall that the electric field is similar to the gravitational field. Assume that we (the charges) come to an elevator at the ground floor (the negative electrode of the battery). The elevator is affected by the gravitational force (and the transferred charges are affected by the coulomb force). If the elevator is ascending uniformly, it is affected by the elastic tensile forces of the steel cables, which are equal to the force of gravity (to the coulomb force). What is analogous to the elastic tension of the steel cables, which perform work against gravity? These are the forces of a chemical nature that arise in batteries between the metal electrodes and the electrolyte. These

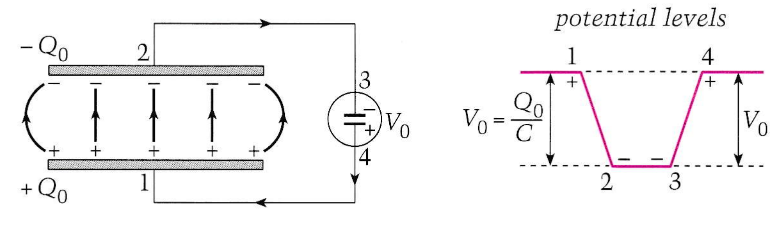


Figure 1

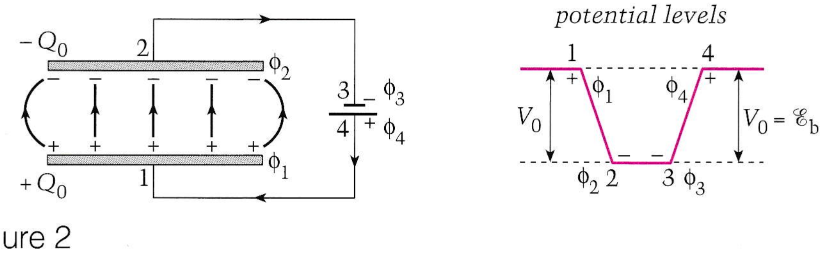


Figure 2

forces are called *extraneous* to stress their non-electrostatic nature. We can describe them with the parameter of *extraneous strength* $\mathbf{E}_{ext}$, which is equal to the force affecting a unit positive charge. For the charges traveling through the battery, the extraneous forces are equal and opposite in direction to the electric forces, so we have the following equation for the potential difference:

$$E_c l_b = V_b = -E_{ext} l_b = -\mathscr{E}_b,$$

where l_b is the distance over which both of these forces act inside the voltaic cell. Thus, $\mathscr{E}_b$ is the work performed in the battery by chemical forces to transfer a unit positive charge from the cathode to the anode against the force of an electric field with a potential difference equal to the emf of the battery $\mathscr{E}_b = V_0$.

Let's consider the distribution of fields and potentials inside some particular voltage source, for example a voltaic cell (fig. 3). Zinc and copper plates are immersed into a nonconducting jar filled with a water solution of sulfuric acid, which acts as the electrolyte. The chemical reaction of zinc with the electrolyte produces positive zinc ions, which diffuse into the electrolyte, so the surface layer of the metal electrode acquires a positive charge and the adjacent layer of electrolyte becomes negatively charged. The distance between these layers l_b is very small (on the order of the size of an atom), and the corresponding potential difference $\Delta\phi$ is about 1 V, so the strength of the electric field between the layers is comparable to that of

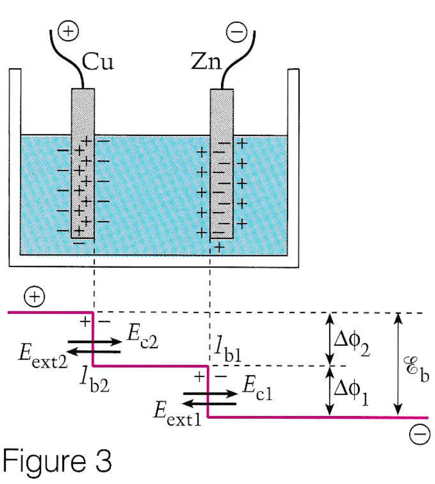


Figure 3

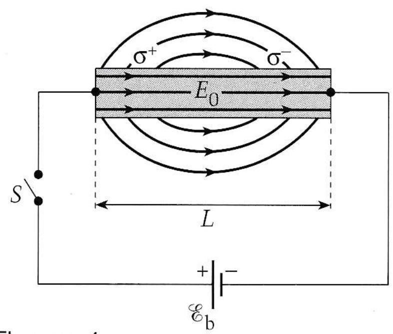


Figure 4

electric fields in atoms, and the surface density of these charges is also very high. The same values are characteristic of the strength of the extraneous forces in these layers, because $E_c = -E_{ext}$. Thus, the transfer of a unit positive charge from the negative plate into the electrolyte raises the potential to a higher level: $\Delta\phi_1 = E_{c1} l_{b1}$. In the bulk of the electrolyte solution the potential doesn't vary (the electric current doesn't flow yet). The second plate (anode) is made of copper, and its interaction with the electrolyte results in the accumulation of positive charges on its surface, while the adjacent layer of electrolyte becomes negative. Here the potential step $\Delta\phi_2 = E_{c2} l_{b2}$ has the same order of magnitude as $\Delta\phi_1$, and their sum is what we call the emf of the cell: $|\Delta\phi_1 + \Delta\phi_2| = \mathscr{E}_b$.

Let's connect the terminals of the battery with a thin, long, homogeneous cylindrical conductor, with resistance $R = \rho L/S$, where ρ is the resistivity, L is the length, and S is its cross-sectional area (fig. 4). If the battery has no internal resistance, this conductor will carry the direct current $I = \mathscr{E}_b/R$. Again questions arise: what role does the electric field play in this process, and what is its value inside and outside the conductor? When the switch S is closed, an electromagnetic impulse spreads along the circuit, which redistributes the free charges at the cylinder's surface in such a way that they generate an homogeneous electric field $\mathbf{E}_0$ inside the conductor and some nonhomogeneous field outside of it, whose lines of force emerge from the cylinder's surface and enter it at some angle (fig. 4). The distribution of surface charge density σ along the axis of a thin, long, homogeneous conductor turns out to be linear everywhere but at its ends.

Note that upon switching on a voltage source, the free electrons begin to move virtually simultaneously in all the parts of a conductor, similar to the way water starts to flow simultaneously through a pipe when we open the faucet. While electric current flows in a conductor, the numbers of positive and negative charges are strictly equal in any part of it.

It is known that electric current in a conductor is accompanied by the production of heat, which is dissipated in the conductor. What is the mechanism of conversion of the chemical energy in a battery into the thermal energy dissipated in the conductor?

In the absence of an electric source, the motion of free electrons in a conductor is stochastic. When the electric source is switched on, the electrons inside the conductor are affected by the electric field, which results in their ordered, directed flow. Simultaneously, opposing forces due to the crystal lattice of the conductor decelerate the electrons. These forces are similar to the frictional forces that act on a ball moving in a viscous medium. All of these forces produce a uniform flow of free electrons along the lines of force of the electric field with a very slow velocity in comparison with that of the chaotic thermal motion.

We can say that the production of heat by moving electrons in the conducting medium is analogous to the dissipation of heat caused by friction when a body slides uniformly down an incline under the influence of gravity. The electric current is in the direction of decreasing electric potential. We can show (although it is beyond the scope of this article) that the potential difference between the conductor's ends equals the product of the electric current in the conductor and its resistance: $V = IR$.

Let's consider a circuit consisting of two batteries connected in series that have emfs $\mathscr{E}_1$ and $\mathscr{E}_2$ and inter-

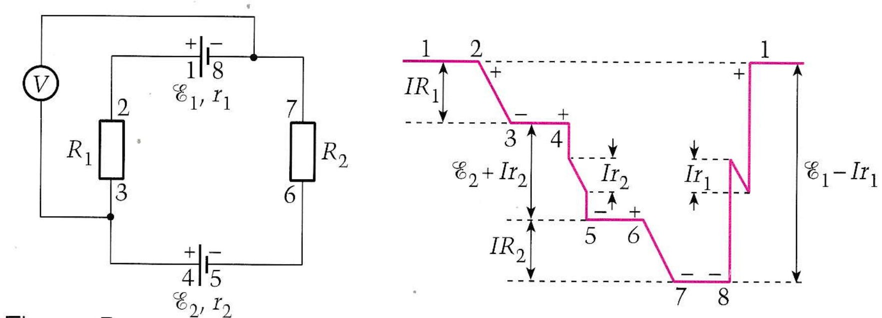


Figure 5

nal resistances r_1 and r_2. The batteries are connected with two external resistors R_1 and R_2 as shown in fig. 5. If in this circuit $\mathscr{E}_1 > \mathscr{E}_2$, the current is counterclockwise. Let's plot the changes in electric potential when moving along the closed circuit.

Point 1 has the highest potential and point 8 has the lowest. The conductor 1-2 has no resistance (no friction in the mechanical analogy), so the potential doesn't change in it. We assume that the conductor 2-3 with resistance R_1 is similar to our cylindrical conductor. In this part of the circuit, we "roll" downhill, and the potential decreases linearly between points 2 and 3. "The force of friction" is counterbalanced by the "rolling down" force (the coulomb force), so the voltmeter will read $V_1 = IR_1$. When passing from point 3 to point 4, the potential doesn't change. Point 4 corresponds to the positive plate of the second battery with emf $\mathscr{E}_2$. When passing from a positive plate into the electrolyte, we are in effect descending in an elevator, so the potential decreases, and the electric field performs work against the battery's extraneous forces. The same thing happens when we pass from the electrolyte to a negative plate. The transfer of electric charges inside the battery is accompanied by the overcoming of its internal resistance r_2, so the potential will further drop by Ir_2. The voltmeter connected to the terminals of this battery will read $V_{45} = \mathscr{E}_2 + Ir_2$. The potential will not vary between points 5 and 6, but it will drop across resistor R_2 by IR_2. Now we get on the negative plate (point 8) of the stronger battery with emf $\mathscr{E}_1$. First a powerful rise and then a fall across resistance r_1, then again a mighty rise when climbing from the electrolyte—and we are back on the positive plate 1.

After moving around the closed circuit, we returned to the initial potential. Therefore, the sum of potential "lifts" equals the sum of "falls." Therefore, the work along a closed circuit is zero:

$$-IR_1 - \mathscr{E}_2 - Ir_2 - IR_2 - Ir_1 + \mathscr{E}_1 = 0,$$

or

$$I(R_1 + R_2 + r_1 + r_2) = \mathscr{E}_1 - \mathscr{E}_2.$$

This equation is known as Ohm's law for a closed circuit or Kirchhoff's second law. Now we see that it is a direct consequence of the properties of electric fields.

Quantum on direct current:

I. Slobodetsky, "Direct current events," March/April 1992, p. 52–55.

A. Varlamov, "How does electric current flow in a metal?" September/October 1992, p. 49–50.

S. Murzin, M. Trunin, and D. Shovkun, "Beyond the reach of Ohm's law," November/December 1994, p. 24–29.

Light in a dark room

by V. Surdin and M. Kartashev

IN LATIN *CAMERA OBSCURA* means "dark room." The following trick is known from ancient times: in a dark room on a bright sunny day, make a small hole in a window blind. On the opposite wall you will see the inverted image of the street and passersby.

The pinhole camera was probably known to the ancient Greeks. It was also used by Arabian scientists, and at the end of the fifteenth century Leonardo da Vinci (1452–1519) gave the first detailed description of this wonderful device. However, the classical camera obscura was not widely used, because when the hole was rather large, the image was blurred, and when it was too narrow, the camera produced a clear but very dim image. In addition, the camera worked only in complete darkness and required the observer's eyes to adjust.

However, in the middle of the sixteenth century the pinhole camera was equipped with a mirror and an objective made of lenses. As a result, the image became large and bright, and the camera enjoyed wide popularity, in particular among amateur artists, who used it to sketch landscapes. There were large, human-sized cameras and small portable ones. Now we consider this simple optical device as a prototype of modern photographic cameras.

Unfortunately, the name of the camera obscura was not changed after it was outfitted with a lens objective. Thus, some historical records are controversial. For example, it is written that pinhole cameras were used in the first experiments on photography in the 1820s and 1830s. In this case, it is obvious lenses were used. However, other reports cannot be interpreted so unequivocally. For example, in 1611, independent of Galileo, the Dutch astronomer Johannes Fabricius (1587–c. 1615) discovered sunspots with the help of a telescope and a camera obscura. No question arises concerning the use of a telescope in such research, but how Fabricius could discern sunspots with a pinhole is an open question.

However, as early as 1609, Johannes Kepler (1571–1630) reported a small, dark spot that he observed on May 18, 1607 in the image of the solar disk obtained with a camera obscura. He erroneously assumed the spot to be Mercury. Such a mistake is justifiable: the diameter of the dark part of a typical sunspot is about 15,000 km, just a little bit larger than the diameter of Earth or Venus. Mercury is half the size of Earth, but when it passes between Earth and the Sun, it is two times closer to us than the Sun's surface. At this moment the angular size of Mercury is similar to that of a sunspot (about 0.3′). Is it possible to discern an object with so small an angular size using just a pinhole?

Of course, a simple camera obscura can be used to observe the phases of a solar eclipse. One of the authors once observed a solar eclipse in the morning using a small hole made by the tip of a pencil in the cover of a notebook. The quality of the image was excellent. However, a sunspot covers a rather small fraction of the solar surface. Fabricius most likely used a lens in addition to a pinhole, otherwise why weren't sunspots discovered long before telescopes were constructed?

The camera obscura with a lens is almost a telescope—it is a product of Renaissance high technology. It differs greatly from a simple classical camera obscura, which could be made by anyone in any century. Let's see what can be observed with such a simple camera obscura.

Practice

It is a very simple matter to make a camera obscura. Take any box with a length of 15–30 cm (a milk carton, for example). Using a pin, pierce a small hole in the bottom of the box. Close the upper opening of the box with oiled paper. A potato-chip box with a white frosted lid is ideal: you need only to eat the chips and pierce the bottom with a light stroke of a pin.

Note that observation requires bright light outside the box and pitch darkness on the screen's side. It is better to conduct experiments in a dark room with the screen carefully isolated from the surrounding light on a bright sunny day. To this end, you can use a tube made of thick paper with a length of 30–40 cm, attaching it firmly to the screen and pressing your face against its opposite end. If the shape of your

Art by Leonid Tishkov


Pinhole people walking

camera obscura is round, you may use a coat, throwing it over your head and passing the camera through a sleeve. To compare various "objectives," make several holes of different diameters and experiment with one hole while the others are covered by electrical tape.

After testing the camera obscura, one can see that a simple hole makes a serviceable objective: all objects produce similar sharp images independent of distance, the greater definition being produced by the smallest hole. However, in the latter case the images cannot be observed easily, because they are too dim. Still, modern photographic films are very sensitive, so they will certainly work even with the smallest hole.

So, why not build a photographic camera with a small hole instead of an expensive objective? By the way, some devices work on this principle. For example, astronomers make X-ray telescopes as a lead camera obscura, because there are no lenses that can focus hard X-rays. However, it turns out that the operation of a pinhole in the optical range is quite limited, as pointed out in the following discussion.

Theory

Every luminous point of a distant object sends a practically parallel beam of light to the pinhole. Having passed the hole of diameter D, the beam projects a circle of the same diameter onto the screen. Let F denote the distance between the hole and the screen. If the angular distance between two neighboring points in the object is less than D/F (measured in radians), the respective circles on the screen will overlap partially.

It is not a simple question to determine how much overlap the neighboring points of the objects can have and still be distinguished (resolved). The result depends on the contrast of the details in the original object, on the brightness of its image, and so on. It is possible to distinguish the details of an image with little contrast if they do not overlap

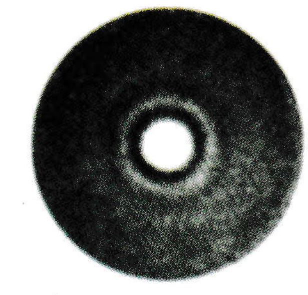


Figure 1

at all. Since sunspots produce images with high contrast, we can assume that they are resolved when their centers are separated by a distance equal to the radius of the circles. Now it is easy to find the minimum angular size of the distinguishable details of the object, or, as opticians say, the limiting *angle of resolution* determined by a certain finite size of the beam:

$$\alpha_1 = \frac{D}{2F}. \qquad (1)$$

Until now we have considered light as rays. This approach is characteristic of geometrical optics. However, light is an electromagnetic wave, and as a wave it is subject to diffraction and interference. If a plane wave (a beam of parallel rays) hits the opening of an optical device, the wave front will become slightly curved, so the beam will diverge. This phenomenon is called "diffraction." It is diffraction that limits the application of the laws of geometrical optics. After passing through a small hole of the camera obscura, the light beams diverge, and the pattern on the screen becomes blurred. To determine the degree of blurring, we must recall the property of interference, which is the addition of waves from different sources at the same point on the screen.

In our case, the independent sources of light are the infinite number of luminous points in the input aperture, and every point emits light in all directions due to diffraction (this is Huygens' principle). The incident waves superimpose on the screen according to their phases. At some points on the screen they add, and at other points they cancel. As a result, after passing through the hole, a parallel beam of rays produces a pattern on a screen of a bright spot surrounded by concentric dark and light rings of decreasing brightness (fig. 1). We can say that the camera obscura transforms any point of a luminous object into a bright spot surrounded by a "zebra" pattern of rings.

Usually it is assumed that the images of two neighboring points can be resolved on the screen if the centers of their bright spots are separated by a distance no less than the radius of the first dark ring (Rayleigh's criterion). The angular size α_2 of this radius as viewed from the opening can be evaluated knowing that the difference of the light paths from the nearest and farthest points of the objective to any point on the dark ring is approximately one wavelength λ. Thus, we get $\alpha_2 \cong \lambda/D$. Precise calculations yield the following value for the limiting angle of resolution due to diffraction:

$$\alpha_2 = 1.22\frac{\lambda}{D}. \qquad (2)$$

Since both effects (the geometric size of the beam and its diffraction) occur at the same time, we can suppose that the limiting angle of resolution of a camera obscura is $\alpha = \alpha_1 + \alpha_2$. Depending on the size of the hole, this angle varies as shown in figure 2. So, the best resolution of a camera with a given length F is achieved at some optimal diameter D_{opt}, corresponding to the mini-

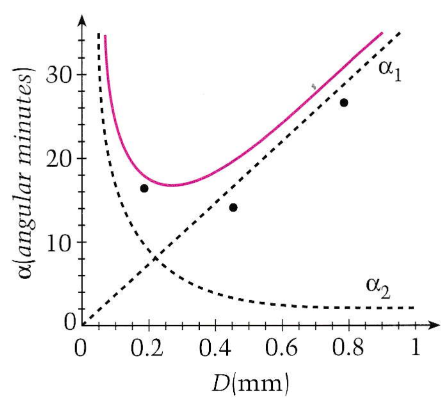


Figure 2

mum resolution angle α_{min}. It is not difficult to find the optimum diameter. Those who know elementary calculus and derivatives will guess that α reaches its minimum at the value of D where $d\alpha/dD = 0$. Readers can also see from figure 2 that the minimum is achieved at the point where $\alpha_1 = \alpha_2$. Both conventions are equivalent. They yield

$$D_{opt} = \sqrt{2.4\lambda F}$$

and

$$\alpha_{min} = \sqrt{\frac{2.4\lambda}{F}}. \qquad (3)$$

Now, what can the optimum classical camera obscura do? Assume that for visual observation we use light with wavelength $\lambda = 550$ nm. Equation (3) might be rewritten in the form ready for estimation:

$$D_{opt} = 1.2 \text{ mm} \cdot \sqrt{\frac{F}{1 \text{ m}}}$$

and

$$\alpha_{min} = 4' \cdot \sqrt{\frac{1 \text{ m}}{F}}, \qquad (4)$$

where F is measured in meters. A camera of a reasonable ("human") size (F = 2–5 m) has a limiting angle of resolution larger (worse) than that of a healthy human eye (about 1′).

This means that using such a device, we will not be able to discern the smaller details in comparison with what we can observe with an unaided eye (of course, protected by a dense filter). The role of such a filter can be played by clouds, smoke of a large fire, or the thick layer of air that protects our eyes when we watch the Sun at sunrise or sunset.

The Chronicles of some people report sunspots that were observed through the clouds and looked "just like nails." In principle, this is possible. Although a mean sunspot has an angular size of about 0.3′, sometimes very large spots or groups of spots appear on the Sun. For example, a group of sunspots 200,000 km in size was observed in March 1947. Similar sunspots appeared in 1957 and 1968. Due to their large angular size (4′), they could be easily seen by an eye, protected by a dense filter.

Caution! It is no mistake that we again mention the filter. *Never look at the Sun without proper protection!* The filter must be a very dark filter, not just sunglasses. One may use welder's glass or aluminum-coated cellophane used to wrap flowers.

Although even ancient people could observe the sunspots with unaided eyes, such episodic and anecdotal observation did not become scientifically established. There were no reliable and systematic observations of the Sun in ancient times, or perhaps the writings are lost forever. It is of principal interest: could sunspots be observed systematically with a classical camera obscura—say, by the ancient Greek astronomers? Formula (4) says that one needs a camera obscura 20–30 m in length to obtain a device better than the human eye. With a 100-meter pinhole camera, one can observe sunspots systematically. Did the Greeks overlook the opportunity?

Recall that the size of the image increases with distance from the hole. So, its brightness should decrease. The angular diameter of the Sun is about 0.5° (to be more precise, it is 32′), so the diameter of its image on the screen of a simple pinhole camera will be $F/107$. Thus, if the length of the camera is 100 m, the image of the Sun will be about 1 m across. The image is formed by the light that passes through a hole 1.2 cm in diameter, which means that illumination will be attenuated by a factor of 10,000. Wouldn't this be too dim?

The illumination of Earth's surface produced by the Sun is 10^5 lux, so the illumination of the Sun's image in a pinhole camera will be about 10 lux. This value seems to be small, but it is dozens of times larger than the illumination of Earth's surface produced by a full moon. One can discern the letters in a book illuminated by the full moon, so it would be much easier to observe the sunspots with our camera obscura, because they will be seen as a coin 1 cm in diameter on a 1-meter solar disk. It is not an easy matter to overlook such details! We must conclude, therefore, that theoretically, the ancient Greeks could have used a classical camera obscura to study the surface of the Sun!

Experiment

To check our theoretical considerations on the quality of images made by a pinhole camera, we carried out the following experiment: in a "Zenit" photographic camera, the objective was replaced by a piece of metal foil with a hole made by a pin. A specially prepared test pattern was photographed with the help of this camera obscura (figure 3). The

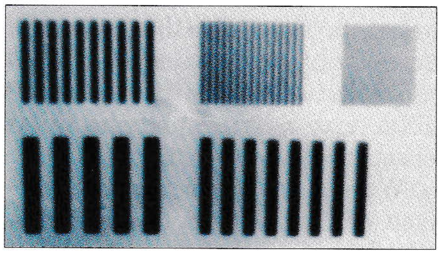
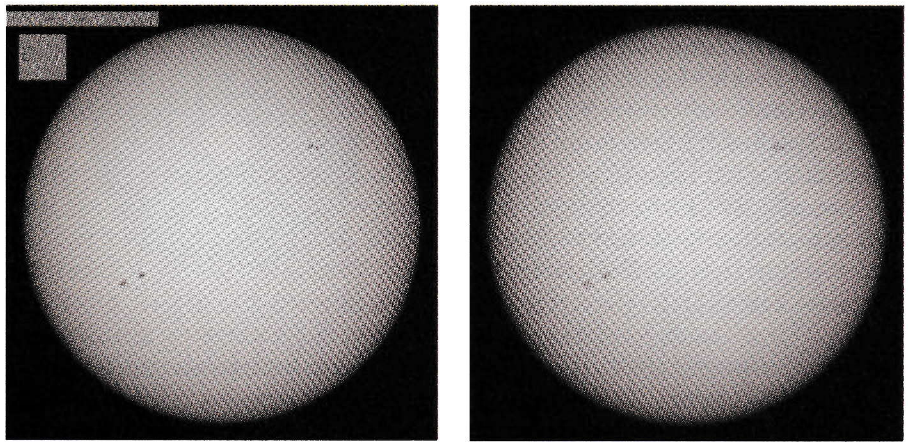


Figure 3

distance between the table and the hole was 30 cm, and that between the hole and the film 4.6 cm. We tested three holes with diameters 170, 420, and 840 μm. The pattern was illuminated by a table lamp, film sensitivity was 80 ASA, and the exposure time varied from several seconds to a few minutes depending on the diameter of the hole. After printing photos from the negatives, we determined the limiting angle of resolution from the visibility of the lines in the test pattern. The experimental angle turned out to be even smaller than the theoretical value, which was probably caused by the very high contrast of the original image and also by its linear appearance: straight lines are more easily perceived than points against a noisy background. By and large, our simple theory agrees with the experiment.

Having checked the theory on a simple test pattern, we decided to try detecting sunspots with a pinhole camera. The experiment was performed on May 19, 1998 at the P. K. Sternberg Astronomical Institute (an affiliate of Moscow State University) with the generous help of I. F. Nikulin, a senior researcher in the Department of Solar Investigations. Unfortunately, we couldn't construct a camera with a length of 100 or even 50 m. The instrument case of our improvised camera was the 17-meter long tube of a vertical solar telescope. The reflector-type objective was located at its base, so the tube was just a light-proof volume without any optical elements. We covered the opening of the tube with a thick lid, which had a small, round hole 5 mm in diameter. On a sheet of white paper placed at the lower end of the tube, we saw a bright image of the Sun with a diameter of 16 cm. There were two well-defined groups of sunspots in the image! This was a triumph: the solar obscura-telescope worked!

We also looked at the solar surface with modern optical devices, which showed that there were sunspots on the Sun that day. They were grouped in two clusters with angular sizes 15″ and 17″, separated by a distance of 1′. In addition, there were several small sunspots 3–5″ in diameter. We did not observe the small sunspots with our camera obscura, although two large spots (quite normal for the Sun) were clearly resolved and observed individually. We continued our observations for several days, noting the rotation of the Sun by the motion of the sunspots. The photographs in figure 4 show the Sun's surface on June 2, 1998. They were made with (a) a modern solar telescope and (b) our improvised obscura-telescope.



a b

Figure 4

Galileo and Fabricius discovered sunspots only after the invention of the optical telescope, although, as we have shown, the discovery could have been made as early as when the Egyptian pyramids were built. Maybe this thought will stimulate our readers to look for unrealized possibilities of our epoch. By the way, when Fabricius made his famous discovery, he was just a little bit older than 20. ◘

Quantum on the art of photography:

V. M. Bolotovsky, "What's That You See?" March/April 1993, p. 5–8.

M. L. Biermann, "Clarity, Reality, and the Art of Photography," September/October 1995, p. 26–31.

A. Leonovich, "How Enlightened Are You?" May/June 1996, p. 32–33.

A. Dozorov, "In Focus," September/October 1998, p. 48–49.

Square or not square?

by Mark Saul and Titu Andreescu

FINDING SQUARE ROOTS OF numbers, especially of natural numbers, is a common and useful mathematical task. But sometimes it is important also to recognize which numbers cannot be perfect squares. The problems below are of this form.

Problem 1: Prove that the number $N = 1234567891011121314151617181920212223$ is not a perfect square.

Solution: The basic fact we need here is that the last digit of a perfect square depends only on the last digit of its square root, and not on any other digits. So we can ask ourselves what the last digit of the square root of N could be. It cannot be 0, since then N would itself have to end in 0. Similarly, it cannot end in 1. In fact, if we examine each digit, we will find that its square never ends in 3:

1^2 ends in 1
2^2 ends in 4
3^2 ends in 9
4^2 ends in 6
5^2 ends in 5
6^2 ends in 6
7^2 ends in 9
8^2 ends in 4
9^2 ends in 1
0^2 ends in 0.

Therefore, we can see that $N = 1234567891011121314151617181920212223$ cannot be a perfect square. In fact, we have shown that a perfect square, written in decimal notation, can only end in the digits 0, 1, 4, 5, 6, and 9.

Problem 2: Show that the number $M = 12345678910111213141511171819202122$ cannot be a perfect square.

Problem 3: Show that no four-digit number can be formed from the digits 2, 3, 7, 8, which is a perfect square.

Problem 4: Show that neither of the numbers $5n + 2$ or $5n + 3$ can be a perfect square, for any natural number n.

Solution: The number $5n + 2$ ends in 2 or 7, and the number $5n + 3$ ends in 3 or 8. The solution to problem 1 shows that they cannot be perfect squares.

This solution to the problem relies on the technique we already have, using the final digits of the numbers' decimal representations. Let us look a bit deeper. What is it about the decimal notation system that allows for this?

Any number N, in decimal notation, can be written as $N = 10A + b$, where A is some natural number, and b is the last digit of N. Then $N^2 = (10A + b)^2 = 100A^2 + 20Ab + b^2$, and the first two terms, being multiples of 10, cannot affect the last digit. In fact, this expansion is what is behind the usual multiplication algorithm and explains why the square of a number ends in the same digit as the number itself.

We can reformulate the solution to problem 4 by generalizing this remark. If a number N has remainder r when divided by 5, then it can be written as $5A + b$, where A is some natural number. Then $N^2 = (5A + b)^2 = 25A^2 + 10Ab + b^2$, and the first two terms, being multiples of 5, do not affect the remainder when N^2 is divided by 5. It follows that N^2 has the same remainder as b^2, when divided by 5.

That is, if we want to find the remainder when N is divided by 5, we can do one of two things: (a) Square N, then divide N^2 by 5 and take the remainder; (b) Divide N itself by 5, take the remainder, then square this remainder (if the square of this remainder is greater than 5, we may have to divide by 5 once more).

Indeed, this observation holds if we are dividing by any number at all, and not just by 5. (Readers familiar with the idea of congruences in number theory will find this statement familiar.) In the present case, if the number $5n + 2$ is a perfect square, and its square root is the natural number N, then we can follow course (b) above, and take the remainder when N is divided by 5. It can only be one of the number 0, 1, 2, 3, or 4, and if we check the squares of these remainders, we find that none of them has remainder 2. But $5n + 2$ does have remainder 2, and so cannot be the square of N.

This argument is quite general, and can be applied in many circumstances.

Problem 5: Show that the numbers $4n + 2$ and $4n + 3$ cannot be perfect squares, for any integer n.

Problem 6: Show that the numbers $9n + 3$ and $9n + 6$ cannot be perfect squares.

Problem 7: The number N contains 1999 digits 1, one digit 2, and a number of digits 0. Show that it cannot be a perfect square.

Problem 8: Show that the numbers $9n + 2$, $9n + 5$, and $9n + 8$ cannot be perfect squares.

Problem 9: Show that the number

$$\underbrace{444\cdots4}_{1001\ 4\text{'s}}$$

cannot be a perfect square.

Problem 10: Show that the number $n! + 2000$ cannot be a perfect square for any natural number n.

Problem 11: Show that $n(n + 1)$ cannot be a perfect square for any positive integer n.

Solution: Note that $n^2 < n(n + 1) < (n + 1)^2$. This means that $n(n + 1)$ lies in between two consecutive perfect squares, and cannot itself be a perfect square.

Problem 12: Show that $n(n + 2)$ cannot be a perfect square for any positive integer n.

Solution: There is a solution analogous to that of problem 11, but we can also see that $n(n + 2) = n^2 + 2n = (n + 1)^2 - 1$. Being one less than a perfect square, the number itself cannot be a perfect square (unless it is 0, in which case n cannot be a positive integer).

Problem 13: Show that the number $2499 \cdot 2500 \cdot 2501$ is not a perfect square.

Problem 14: Show that the number $n(n + 1)(n + 2)(n + 3)$ cannot be a perfect square, for any positive integer.

Problem 15: Show that the number $(m + n)^2 + 3m + n + 1$ cannot be a perfect square for any distinct natural numbers m and n.

Problem 16: Show that the number $N = 1! + 2! + 3! + \ldots + n!$ cannot be a perfect square for any integer $n > 3$.

Problem 17: Show that $n!$ cannot be a perfect square for any integer $n > 1$. (You may want to use "Bertrand's postulate," actually a theorem, which says that for $n > 1$ there is always a prime number between n and $2n$.) ◙

ANSWERS, HINTS & SOLUTIONS ON PAGE 53

CONTINUED FROM PAGE 29

solute value for area, we can write $|ACM| - |APM| = |BCM| - |BPM|$, or simply $|CAP| = |CBP|$.

Conversely, if we have a point Q inside triangle ABC, such that simply $|CAQ| = |CBQ|$, we can show that it must lie on the median from C to AB. Indeed, extend CQ to meet AB at X. The equality of the two areas implies the equality of the perpendiculars from A and B to line CQ, which in turn implies the equality of the areas of triangles ACX, BCX. Since these two triangles also have the same altitude from C to AB, this means that their bases are equal, so that X is the midpoint of AB.

Now the stage is set for a simple proof of the following.

Theorem: The three medians of a triangle are concurrent.

Proof: Let medians AM, BN of triangle ABC intersect at point P. Then, by our locus theorem, $|CAP| = |BAP|$ and $|BAP| = |CBP|$. Hence $|CAP| = |CBP|$, and P is on the median from C as well.

Note that we have also shown that $|CPB|$ is 1/3 the area of $|ABC|$, so that the perpendicular from A to BC is three times the length of the perpendicular from P to BC. A simple argument from similar triangles (left to the reader) will show that $AP{:}PM = 2{:}1$, so we have:

Theorem: The centroid (intersection of the medians) of a triangle divides each median in the ratio 2:1.

So the method of loci turns out to be pretty powerful. We can apply it to prove the concurrency of the altitudes in two different ways.

Proof 5. *A method of locus.* Notice that for all points of any ray emanating from the vertex of an angle, the ratio of the distance to the angle sides is a constant. If the ray lies inside the angle, the ray is uniquely determined by this ratio.

Choose point H on altitude AA_1. From similar triangles we have $HC_1/HA = A_1B/AB = \cos B$, and $HB_1/HA = A_1C/AC = \cos C$, so $HC_1/HB_1 = \cos A/ \cos C$. This implies that for all points of the alti-

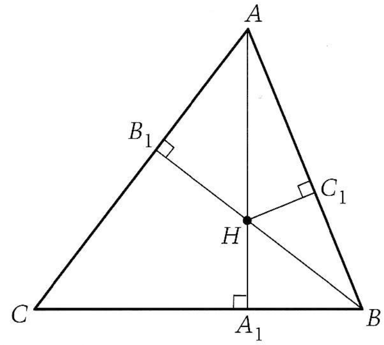


Figure 6

tude AA_1, the ratio of the distances to the sides AB and AC is $\cos B/\cos C$ (fig 6.). Similarly, for all points of the altitude BB_1, the ratio of the distances to the sides BA and BC is $\cos A/\cos C$. Let H be the point of intersection of AA_1 and BB_1. The ratio of the distances from this point to sides AC and BC is $\cos A/\cos B$. Therefore, point H lies on the altitude to side AB.

Proof 6. *Another locus method.* Let P be an arbitrary point on the line AA_1 (fig. 7, where AA_1 is the al-

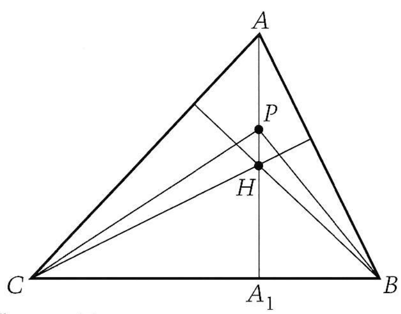


Figure 7

titude of the triangle ABC). Then,

$$BP^2 - CP^2 = \left(BP^2 - PA_1^2\right) - \left(CP^2 - PA_1^2\right) = BA_1^2 - CA_1^2.$$

Thus, $BP^2 - CP^2$ is a constant for all points P of this line. Also, we can verify that if $BP^2 - CP^2 = BA^2 - CA^2$, then P lies on the altitude of the triangle. Let H be the point of intersection of the altitudes to the sides BC and CA of the triangle. We have $BH^2 - CH^2 = BA^2 - CA^2$ and $CH^2 - AH^2 = CB^2 - AB^2$. Combining these equations, we obtain $BH^2 - AH^2 = CB^2 - CA^2$, which means that H lies on the altitude to side AB. ◙

—I. F. Sharygin

Completing a tetrahedron

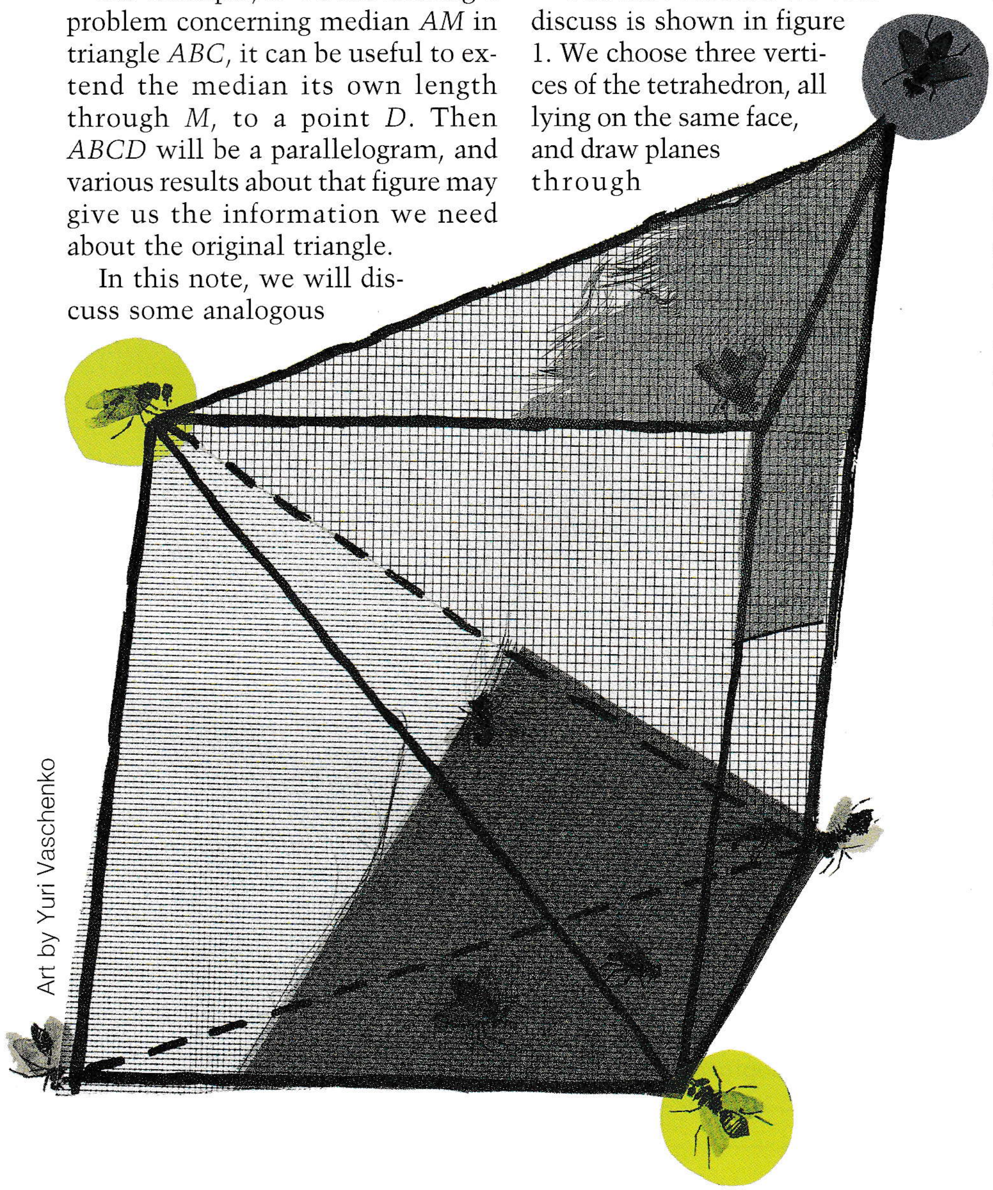
by I. F. Sharygin

AN ELEGANT TRICK THAT can sometimes be used to solve geometric problems is to change the figure under consideration into another figure that is in some way more convenient.

For example, if we are solving a problem concerning median AM in triangle ABC, it can be useful to extend the median its own length through M, to a point D. Then $ABCD$ will be a parallelogram, and various results about that figure may give us the information we need about the original triangle.

In this note, we will discuss some analogous way of solving problems about a tetrahedron (triangular pyramid). Many such problems can be solved by "completing" the given tetrahedron to obtain another polyhedron (usually a parallelepiped).

The first method we will discuss is shown in figure 1. We choose three vertices of the tetrahedron, all lying on the same face, and draw planes through each vertex, parallel to the opposite face. In figure 1, AA_1BD is the given tetrahedron, A_1BD the chosen face, and we have drawn planes CDD_1C_1, BCC_1B_1, and $A_1B_1C_1D_1$.



Art by Yuri Vaschenko

Problem 1. A triangular pyramid AA_1BD is given in which edges AA_1, AB, and AD are perpendicular to each other and have lengths of a, b, and c, respectively.

(a) Prove that vertex A of the pyramid, the point of intersection of the medians of face A_1BD, and the center of the sphere circumscribed about the given pyramid lie on a line.

(b) Find the radius of the sphere circumscribed about the given pyramid.

Let us complete the given pyramid to construct a (rectangular!) parallelepiped as shown in figure 1. Then, the sphere that is circumscribed about the given pyramid coincides with the sphere circumscribed about this parallelepiped. The radius of the sphere is half of the

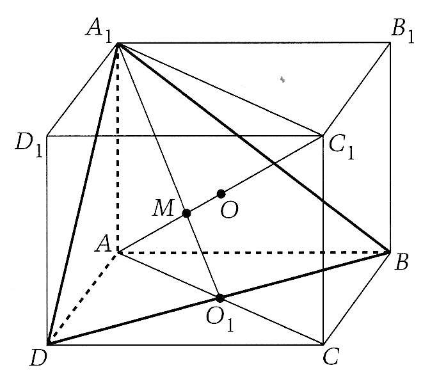


Figure 1

parallelepiped's diagonal:

$$\frac{1}{2}\left(a^2+b^2+c^2\right)^{1/2}.$$

This is the answer to part (b).

To prove the assertion of part (a), consider rectangle AA_1C_1C (the reader is invited to make a separate drawing of it). Center O of the sphere lies on the diagonal AC_1, and median A_1O_1 of triangle A_1BD intersects AC_1 at a point M. If we prove that $|A_1M|/|MO_1| = 2$, this will mean that M is the point of intersection of the medians of triangle A_1BD, which is the desired result. Indeed, it follows from the similarity of triangles A_1C_1M and AO_1M that

$$\frac{|A_1M|}{|MO_1|}=\frac{|A_1C_1|}{|AO_1|}=2.$$

Another frequently used method for completing a tetrahedron to create a parallelepiped is as follows. For every edge of the tetrahedron, construct the plane that contains it and is parallel to the opposite edge. These planes make up a parallelepiped (fig. 2) in which the edges of the initial tetrahedron are the diagonals of the faces. (In sketching the diagram in such situations, it is often convenient to begin the drawing with the parallelepiped rather than with the tetrahedron.)

Problem 2. Find the radius of the sphere that is tangent to all the edges of a regular tetrahedron with edges of length a.

As we can see from figure 2, the parallelepiped constructed as described above is a cube with edges of length $a/\sqrt{2}$. The sphere inscribed

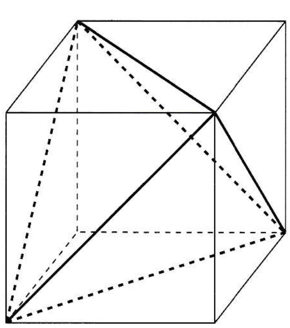


Figure 2

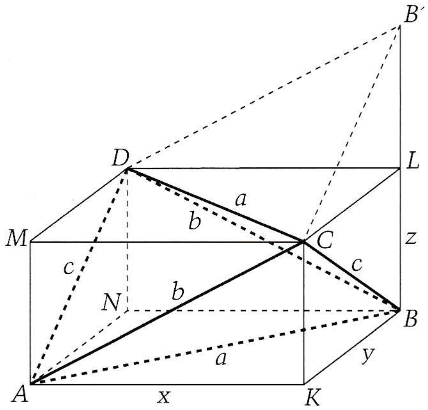


Figure 3

in this cube is the desired sphere, and its radius is $a/2\sqrt{2}$.

The first of the methods for completing a tetrahedron is more practical when plane angles at a vertex of the tetrahedron are given (especially if they all are right angles). The second method is often used when skew edges of the tetrahedron appear in the problem.

Problem 3. The lengths of two skew edges of a tetrahedron are a, two other skew edges are of length b, and two remaining edges are of length c. Find the distance between the center of the sphere inscribed in the given tetrahedron and the center of the sphere that is tangent to a face of the tetrahedron and extensions of its other faces.

In figure 3, $ABCD$ is the given tetrahedron, and the corresponding parallelepiped is bounded by the planes passing through each edge of the tetrahedron parallel to its opposite edge. The diagonals of every face of the parallelepiped are equal to the opposite edges of the tetrahedron, which are equal to each other by assumption. It follows that all the faces of the parallelepiped are rectangles and the parallelepiped itself is rectangular.

The center of the sphere inscribed in tetrahedron $ABCD$ coincides with the point of intersection of the parallelepiped's diagonals (the reader may enjoy proving this). Likewise, we may show that the center of the sphere that is tangent to face DCB of the tetrahedron and the extension of its other faces coincides with vertex L of the parallelepiped (this point is equidistant from faces DCB and ACD). A sketch of a proof follows. Consider pyramid $B'LCD$, where $|B'L| = |LB|$, which is congruent to pyramid $BLCD$: points A, D, B', and C lie in the same plane. Similarly, we can prove that L is equidistant from planes DCB and ACB as well as from planes DCB and ADB. We will see later that the result does not depend on the choice of face DCB.

Therefore, the distance sought is equal to the half of the parallelepiped's diagonal. Denote by x, y, and z the lengths of the edges of the parallelepiped (fig. 3). The Pythagorean theorem gives a system of three equations:

$$\begin{cases} x^2+y^2=a^2, \\ x^2+z^2=b^2, \\ y^2+z^2=c^2. \end{cases}$$

Combining these equations, we find that

$$\frac{|AL|}{2}=\frac{1}{2}\sqrt{x^2+y^2+z^2}$$
$$=\frac{1}{2}\sqrt{\frac{a^2+b^2+c^2}{2}}.$$

Problem 4. The area of the section of a tetrahedron by the plane parallel to and equidistant from its two skew edges is S. The distance between these two skew edges is h. Find the volume of the tetrahedron.

Let $ABCDA_1B_1C_1D_1$ be the parallelepiped bounded by the planes passing through each edge of the tetrahedron parallel to the opposite edges (fig. 4). Then, the volume of the tetrahedron A_1BC_1D is equal to the volume of the parallelepiped

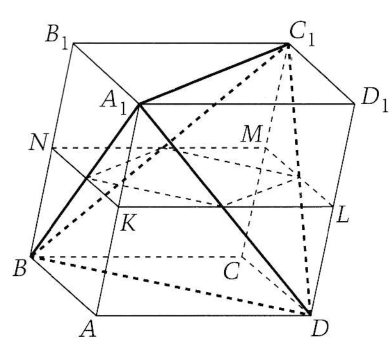


Figure 4

minus the volumes of four triangular pyramids (A_1ABD is one of them), and the volume of each of them is 1/6 of the volume of the parallelepiped (the proof of this fact is left to the reader). Therefore, $V_{tetr} = V_{par}/3$.

Let the skew edges given in the problem be A_1C_1 and BD, and let $KLMN$ be the section of the parallelepiped by the plane passing through the midpoints of AA_1, BB_1, CC_1, and DD_1. Then, the midpoints of the sides of parallelogram $KLMN$ are the vertices of the section given in the problem. Thus, the area of $KLMN$ is $2S$ and is equal to the area of the base $ABCD$ of the parallelepiped. Now, we can easily find the volumes of the parallelepiped and tetrahedron:

$$V_{tetr} = \frac{1}{3}V_{par} = \frac{2}{3}Sh.$$

Using this problem, it is easy to prove *Simpson's formula* for the volume of polyhedrons of a special kind: *Let a polyhedron be such that all its vertices lie in two parallel planes that are a distance* h *from each other. Let* S_1 *be the area of the face lying in the first plane,* S_2 *be the area of the face lying in the second plane, and* S_m *be the area of the section of the polyhedron by the plane parallel to the given planes and equidistant from them. Then, the area of the given polyhedron can be calculated by the formula*

$$V = \frac{h}{6}\left(S_1 + S_2 + 4S_m\right).$$

To prove this fact, check this formula for tetrahedrons. Then, partition the given polyhedron into tetrahedrons with the vertices in the given planes. The sum of the areas of the faces of these tetrahedrons that lie in the first plane is S_1, in the second plane S_2, and the sum of the areas of the midsections of the tetrahedrons is S_m.

We conclude with a problem in which a tetrahedron is completed to create a triangular prism rather than a parallelepiped.

Problem 5. Let the areas of two faces of a tetrahedron be S_1 and S_2,

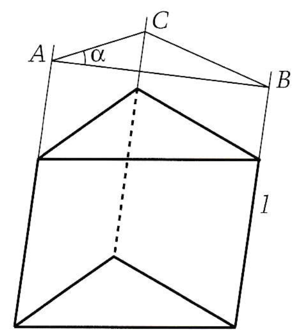


Figure 5

the dihedral angle between them be α, the areas of two other faces be Q_1 and Q_2, and the angle between them be β. Prove that

$$S_1^2 + S_2^2 - 2S_1S_2\cos\alpha = Q_1^2 + Q_2^2 - 2Q_1Q_2\cos\beta.$$

First we prove the following analogue to the law of cosines: **If the area of a lateral face of a prism is S, the areas of two other faces are S_1 and S_2, and the dihedral angle between them is α, then**

$$\mathbf{S_1^2 + S_2^2 - 2S_1S_2\cos\alpha = S^2.}$$

Indeed, let plane ABC (fig. 5) be perpendicular to the lateral faces of the prism and $\angle BAC = \alpha$. Write down the law of cosines for triangle ABC:

$$|BC|^2 = |AB|^2 + |AC|^2 - 2|AB| \cdot |AC| \cos\alpha.$$

Now multiply this equation by l^2, where l is the length of the lateral edge of the prism, and the result follows.

We now return to problem 5. Let $ABCD$ be the given tetrahedron, $S_{\triangle ABD} = S_1$, $S_{\triangle ADC} = S_2$, $S_{\triangle ABC} = Q_1$, $S_{\triangle DBC} = Q_2$, the dihedral angle at edge AD be α, and at edge BC be β.

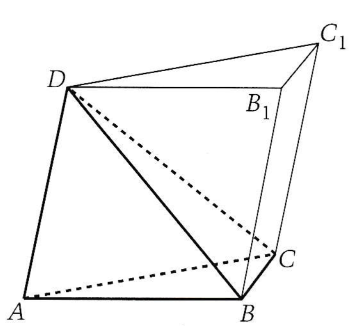


Figure 6

Consider a triangular prism with base ABC, and a lateral edge AD (fig. 6). Denote by S the area of parallelogram BB_1C_1C. Then we have, by the formula proved above,

$$4S_1^2 + 4S_2^2 - 8S_1S_2\cos\alpha = S^2.$$

We can represent S in terms of the lengths of edges BC, AD, and the angle θ between them: $S = |AD| \cdot |BC| \cdot \sin\theta$.

If we consider another prism with base ACD and lateral edge BC, we obtain

$$4Q_1^2 + 4Q_2^2 - 8Q_1Q_2\cos\beta = S^2.$$

This implies the assertion of the problem.

Exercises

1. Prove that the sum of the squares of the lengths of the tetrahedron's edges equals the quadruple sum of the squares of the distances between the midpoints of its skew edges.

2. Let a tetrahedron $ABCD$ be given. Prove that its edges AD and BC are perpendicular if and only if the following equation holds:

$$|AB|^2 + |DC|^2 = |AC|^2 + |DB|^2.$$

3. The lengths of two opposite edges of a tetrahedron are a, two other opposite edges are b, and two remaining edges are c. Find

(a) the volume of this tetrahedron,

(b) the radius of the sphere circumscribed about it.

4. The lengths of two opposite edges of a tetrahedron are a and a_1 and the angle between them is α; the lengths of two other opposite edges and the angle between them are b, b_1, and β, respectively; and the lengths of the two remaining edges and the angle between them are c, c_1, and γ, $(\alpha, \beta, \gamma \le \pi/2)$.

(a) Prove that one of the numbers $aa_1\cos\alpha$, $bb_1\cos\beta$, and $cc_1\cos\gamma$ equals the sum of two others.

(b) Find the angles α, β, and γ given a, a_1, b, b_1, c, and c_1.

ANSWERS, HINTS & SOLUTIONS ON PAGE 53

ANSWERS, HINTS & SOLUTIONS

Math

M266

Let $x^8 = y$. The equation becomes $y^{y/8} = 2$, or $y^y = 2^8$. We can see that $y > 1$. Let $f(x) = y^y$. We will show that *f increases monotonically* for $y > 1$. That is, if $a < b < 1$, then $a^a > b^b$. This is easy: we have $a^a > a^b > b^b$. Hence the equation $y^y = 1$ can have only one solution, and it is easy to guess that this unique solution is $y = 4$. Answer: $x = 2^{1/4}$.

M267

Triangles APB and CPA (fig. 1) are similar, since they both have a 120° angle at P and $\angle ABP = 60° - \angle BAP = \angle PAC$. Therefore, $BP/AP = AP/CP$,

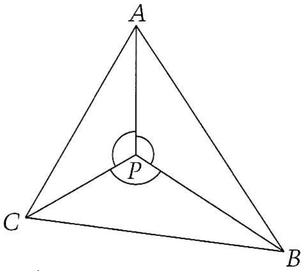


Figure 1

from which we obtain $BP \cdot CP = AP^2 = a^2$. Therefore, the area of triangle BPC is

$$\frac{1}{2} BP \cdot CP \sin 60° = a^2 \frac{\sqrt{3}}{4}.$$

Answer:

$$a^2 \frac{\sqrt{3}}{4}.$$

M268

Change the variables as follows:

$$x = \frac{u-1}{u+1},\ y = \frac{v-1}{v+1}.$$

Simple transformations yield the following system:

$$uv^2 = 3,\ u^2 = 2v.$$

Answer:

$$x = \frac{\sqrt[5]{12}-1}{\sqrt[5]{12}+1},\ y = \frac{\sqrt[5]{\frac{9}{2}}-1}{\sqrt[5]{\frac{9}{2}}+1}.$$

M269

Let $u = (x^2 - 1)/(y + 1)$ and $v = (y^2 - 1)/(x + 1)$. Both u and v are certainly rational numbers. The problem states that $u + v$ is an integer, and the product uv is also an integer (the denominators cancel out). So we can form a quadratic equation, with integer coefficients, whose two roots are u and v. Suppose this equation is $z^2 + mz + n = 0$ (so that $-m = u + v$ and $n = uv$). Let us examine the discriminant $m^2 - 4n$. It is certainly rational. We can show it has the same parity as m itself. Indeed, m^2 has the same parity as m, and $4n$ is even, so adding it to m^2 doesn't change the parity of the expression.

M270

Proof of item (a). Denote the distances from point D to AB, AC, and BC by x, x, and y, respectively, and the distances from point F to the same lines by m, m, and n. The equality of angles in the assumptions of the problem is equivalent to the equation $x/y = n/m$. We will show that the equality of angles in the assumption of the problem implies the equation $x/y = n/m$, which in turn implies the equality of the angles in the problem statement.

Indeed, let X be the foot of the perpendicular from D to AB. Let Y be the foot of the perpendicular from F to BC. The equality of the two given angles implies that $\angle FBY = \angle DBX$. Thus triangles FBY and DBX are similar, and $CX/FY = DB/FB$, or $x/n = DB/FB$. Drawing perpendiculars from D to BC and F to AB, we find two more similar triangles, which give us the equation $y/m = DB/FB$. It follows that $x/y = n/m$. The reader can prove that this last equation implies the statement of the problem.

Proof of item (b). Consider one of the two circles that pass through points D and F and are tangent to side BC (as shown in figure 2). Denote this point of tangency by P. Let M be the point of intersection of the bisector of angle A with the circle circumscribed around triangle ABC. Let K be the second point of intersection of line MP with the circle constructed (passing through points D and F). We prove that K lies on the circle circumscribed around triangle ABC. Notice that triangles DMB and FMB are similar, since they have a common angle M

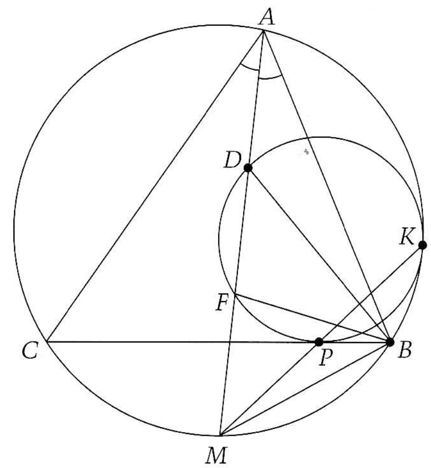


Figure 2

and

$$\angle BFM = \angle FAB + \angle FBA$$
$$= \angle MBC + \angle DBC = \angle DBM.$$

Therefore, $MF/MB = MB/MD$, from which we get $MB^2 = MF \cdot MD$. We know that if two secants are drawn to a circle from an outside point, the product of the secant and its external segment is constant. Thus we obtain $MF \cdot MD = MP \cdot MK$. Therefore, $MP \cdot MK = MB^2$, from which we get $MP/MB = MB/MK$. Therefore, triangles MPB and MKB are similar (they have a common angle M and the sides adjacent to this angle are proportional). Therefore, $\angle MKB = \angle MBP$. However, $\angle MBP = \angle MBC = \angle MCB$. Thus, $\angle MKB = \angle MCB$, which means that points B, C, M, and K lie on the same circle. Thus, we have proved that K lies on both circles whose tangency must be proved. The tangent to the circumscribed circle at point M is parallel to BC—that is, it is parallel to the tangent to the other circle (passing through points D and F) at point P.

Let KY be the tangent to the large circle (with points K and P on the same side of line MK). Then $\angle KMY = \angle KPY$, so $\overset{\frown}{MK}$ is equal in measure to $\overset{\frown}{PK}$. Hence $\angle PKY = (1/2)\overset{\frown}{MK} = (1/2)\overset{\frown}{PK}$, which means that KY is tangent to the small circle as well. So the two circles have a common tangent at K, and so are themselves tangent at K.

Physics

P266

From the problem statement, we can see that the momentum of the system is zero:

$$m_1\mathbf{v} + m_2(-3\mathbf{v}) = 0,$$

because $m_1 = 3m_2$. The absence of total momentum in a closed system means that its center of mass doesn't move.

In the case of two particles, the center of mass is located on the line connecting the particles. Moreover, it divides the line segment in the

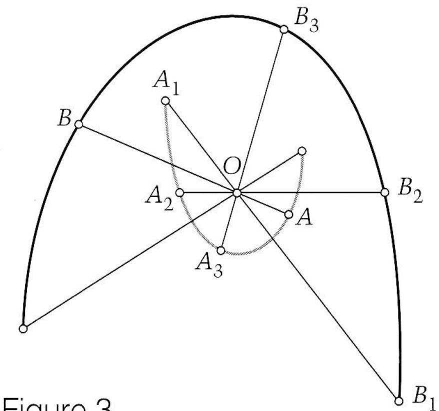


Figure 3

inverse ratio of the mass ratio. Thus, the second particle's trajectory is drawn as shown in figure 3. We draw a line AB connecting particles 1 and 2 at the given moment of time, when $\mathbf{v}_{m_2} = -3\mathbf{v}_{m_1}$. Then we subdivide this line segment into four equal parts and mark the point one-fourth of the way from the first particle. This determines the position of the motionless center of mass. Then we connect an arbitrary point of the first particle's trajectory (say, point A_1) with the center of mass by the line segment A_1O and extend it so that $OB_1 = 3A_1O$. Point B_1 will be the corresponding point on the trajectory of the second particle. By repeating this procedure for all points of the first particle's trajectory, we get the trajectory of the second particle.

P267

The force of gravitational attraction

$$F = G\frac{m_S m_E}{L^2}$$

of the Sun imparts the centripetal acceleration

$$a_c = \omega^2 L = \frac{4\pi^2}{T^2}L$$

to Earth, where ω is angular velocity; T = 1 year, the period of revolution of Earth around the Sun; m_S and m_E are the masses of the Sun and Earth, respectively; and L is the radius of Earth's orbit. According to Newton's second law,

$$G\frac{m_S m_E}{L^2} = m_E\frac{4\pi^2}{T^2}L.$$

However,

$$G\frac{m_E}{R_E^2} = g,$$

so

$$m_S g = \frac{R_E^2}{L^2} = m_E\frac{4\pi^2}{T^2}L,$$

or

$$\frac{m_E}{m_S} = \frac{gR_E^3T^2}{4\pi^2L^3}. \qquad (1)$$

Because

$$m_E = \frac{4}{3}\pi R_E^3\rho_E$$

and

$$m_S = \frac{4}{3}\pi R_S^3\rho_S,$$

equation 1 yields

$$\frac{\rho_E}{\rho_S} = \frac{gR_S^3T^2}{4\pi^2L^3R_E}.$$

Now we must only express R_S in terms of L and α. We can see that $R_S = L\alpha/2$ (fig. 4), so finally we have

$$\frac{\rho_E}{\rho_S} = \frac{g\alpha^3T^2}{32\pi^2R_E} \cong 4.4.$$

P268

When the conducting plate is placed in the electrical field, the free charges in the plate are rearranged

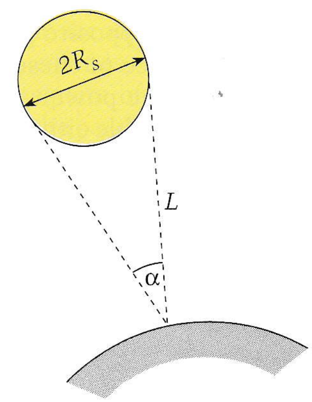


Figure 4

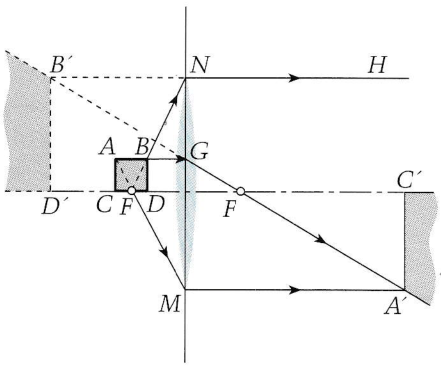


Figure 5

by the field. As a result, opposite charges accumulate at opposite faces of the plate. Inside the plate, the electric field is zero. This means that the free charges located on the plate's surface generate an electric field, whose intensity is –**E** inside the plate and zero outside of it.

Immediately after the external electric field is eliminated, only the field generated by surface charges remains inside the plate. The energy of this field is

$$U = \frac{E^2}{2}\varepsilon_0 Sd.$$

Under the effect of this field, the charges are spread out over the entire volume of the plate. During this process, the energy stored in the field is dissipated in the form of heat:

$$Q = \frac{E^2}{2}\varepsilon_0 Sd.$$

P269

The same current flows in the coil and capacitor, so the voltages across them have opposite phases. Thus, the difference of these voltages is equal to the power supply voltage. This is possible either when $V_C = 0$ (infinitely large capacitance) or when $V_C = 440$ V. In the latter case the capacitative impedance of the capacitor is twice the inductive impedance of the coil:

$$\frac{1}{\omega C} = 2\omega L,$$

from which we get

$$C = \frac{1}{2\omega^2 L} = \frac{1}{8\pi^2 f^2 L} \cong 5\ \mu\text{F}.$$

The dangerous value of capacitance corresponds to the case when the inductive and capacitative impedances are equal, which is when $C_{\text{forbid}} \cong 10\ \mu$F, because at this value, resonance occurs. At resonance, the resistance of the RC-circuit tends to zero, which means infinite current in the circuit.

P270

To locate the image of the square, it is most convenient to use the rays that pass through the focal point of the lens. The ray *ABGF* (fig. 5) is the most informative, because it travels along the upper side of the square, so the images of all the points of this side must lie either on ray *GF* or its extension. Let's trace ray *AM*, which travels through the left focal point *F* of the lens, and ray *BN*, which seems to emerge from this focal point. After passing through the lens, both rays travel parallel to the lens's principal optical axis.

The real image *A′* of point *A* is formed by the crossing of rays *GF* and *MA′*. Similarly, the virtual image *B′* of point *B* is at the intersection of the continuations of rays *GF* and *NH*. To find the positions of the images *C′* and *D′* of points *C* and *D*, it is sufficient to drop the corresponding perpendiculars to the principal axis. The images of sides *AC* and *BD*, which are perpendicular to the principal optical axis, must also be perpendicular to this axis.

To trace rays *AM* and *BN*, we must enlarge the lens on both sides. This is permissible, because any part of the lens produces the same image as the entire lens (they differ only in illumination).

Therefore, the image of the square consists of two parts: a real image (a part of the angle to the right of *A′C′*) and a virtual image (a part of the angle to the left of *B′D′*). The real image is produced by the half of the square that lies outside the focal plane, while the virtual image corresponds to the half of the square located nearer to the lens than the focal plane.

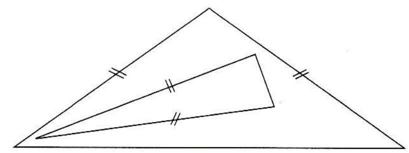


Figure 6

Brainteasers

B266

The total number of games was $(15 + 9 + 14)/2 = 19$. Each player missed no more than one game in succession. Robbie played 9 games. Therefore, he played only in even-numbered games. (Otherwise, the number of games he played would be greater.) Therefore, Teresa and Alex played in the 13th game.

B267

Yes, it is possible (see figure 6).

B268

The fly was airborne for 1 hour (the time that passed until the bikers met). The fly flew with the wind for 23 km more than it flew into the wind (this is the distance traveled by Josh). If we let t_1 and t_2 be the times during which the fly flew with the wind and into the wind, respectively, we can set up a system of equations: $t_1 + t_2 = 1$, $40t_1 - 30t_2 = 23$. We find $t_1 = 53/70$ and $t_2 = 17/70$. Therefore, the distance traveled by the fly is

$$40 \cdot \frac{53}{70} + 30 \cdot \frac{17}{70} = 37\frac{4}{7}\ \text{km}.$$

B269

Yes, it is possible. Mesh the gear wheels successively with each other such that they form a Möbius strip (see figure 7—the planes of two adjacent wheels are arranged at a small angle to each other, and the total angle between the planes of the first and last wheels is 180°). For a sufficiently large number of wheels, in particular, for 101 wheels, they can

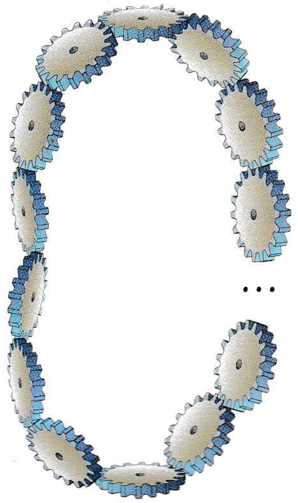


Figure 7

be arranged as described. An odd number of wheels thus arranged can rotate. However, if the number of wheels is even, they cannot rotate.

B270

Subdivide the ball into thin layers perpendicular to the diameter that connects the wires. Since each layer is pierced by the same current, and the resistance of a layer is inversely proportional to its area, more heat is dissipated in the layers with the smallest area—the polar layers. Thus, most heat is dissipated at the regions where the wires contact the ball.

Gradus

2. The number M ends in the digit 2. The argument in the solution of problem 1 shows that it cannot be a perfect square.

3. Any such four-digit number must end in 2, 3, 7, 8. But no perfect square ends in these digits. (This solution is correct whether you interpret the problem to mean that each digit is used exactly once, or whether you want to include multiple appearances of the same digit in your number.)

5. We look at remainders upon division by 4. These can be only 0, 1, 2, or 3, and their squares have remainders 0, 1, or 4. A perfect square cannot have a remainder of 2 or 3 when divided by 4.

6. The number $9n + 3 = 3(3n + 1)$ is a multiple of 3 but not of 9. Such a number cannot be a perfect square. A similar argument holds for $9n + 6 = 3(3n + 2)$. For a more general method, see the solution to problem 8.

7. The sum of the digits of such a number is 2001, which is a multiple of 3 but not of 9, so the number itself cannot be a perfect square (because the number is also a multiple of 3 but not of 9).

8. If we square all possible remainders when dividing by 9, we get the possible remainders of perfect squares when divided by 9: 0, 1, 4, 7. The given numbers have remainders 2, 5, and 8, respectively, and so they cannot be perfect squares.

9. If we divide by 3, the given number has the same remainder as $4 + 4 + 4 + \ldots + 4 = (1001)(4)$, which has remainder 2. But we can quickly check, using the methods we devised earlier, that a perfect square cannot have this remainder when divided by 3.

10. For $n = 1, 2, 3, 4, 5$, we can check directly that $n! + 2000$ is not a perfect square. For $n \geq 6$, $n!$ is a multiple of 9, and since 2000 has remainder 2 when divided by 9, $n! + 2000$ also has remainder 2. By the result of problem 7, it cannot be a perfect square.

13. We know that $2500 = 50 \cdot 50$ is a perfect square. If the given product is a perfect square, then so is $2499 \cdot 2501$. But this is impossible, by the result of problem 10.

14. We have

$$\begin{aligned} n(n+1)(n+2)(n+3) &= (n^2+3n)(n^2+3n+2) \\ &= (n^2+3n)^2 + 2(n^2+3n) \\ &= [(n^2+3n)+1]^2 - 1, \end{aligned}$$

which is one less than a perfect square, and so cannot itself be a perfect square.

15. The form of the given expression suggests we look at $(m + n + k)^2$, for small integer values of k. Indeed, we find that if $m > n$, the given number is strictly between $(m + n + 1)^2$ and $(m + n + 2)^2$. If $m < n$, our number is strictly between $(m + n)^2$ and $(m + n + 1)^2$. In either case, it cannot be a perfect square.

16. For $n = 4$, $N = 33$, and ends in 3 (so cannot be a perfect square, from the result of problem 1). For $n \geq 5$ the situation does not improve, because $k!$ ends in 0 for $k > 4$, so N still ends in 3, and still cannot be a perfect square.

17. Let p be the largest prime that divides $n!$. If we want $n!$ to be a square, it must contain at least one more multiple of p, namely $2p$. But according to Bertrand's postulate, between p and $2p$ there must be another prime. This contradicts our assumption that p is the largest prime that divides $n!$.

Tetrahedron

1. Complete the given tetrahedron to create a parallelepiped using the second method, then apply the formula that relates the sum of squared side lengths of a parallelogram to the sum of its squared diagonals.

2. Complete the given tetrahedron to create a parallelepiped using the second method. To satisfy the condition of the problem, it is necessary and sufficient that the corresponding face of the parallelepiped constructed be a rhombus.

3. (a)

$$\frac{\sqrt{2}}{12}\sqrt{\left(a^2+b^2-c^2\right)} \times \sqrt{\left(a^2-b^2+c^2\right)\left(-a^2+b^2+c^2\right)}$$

(b) $\frac{1}{2}\sqrt{\frac{a^2+b^2+c^2}{2}}$.

Hint: see problems 3 and 4.

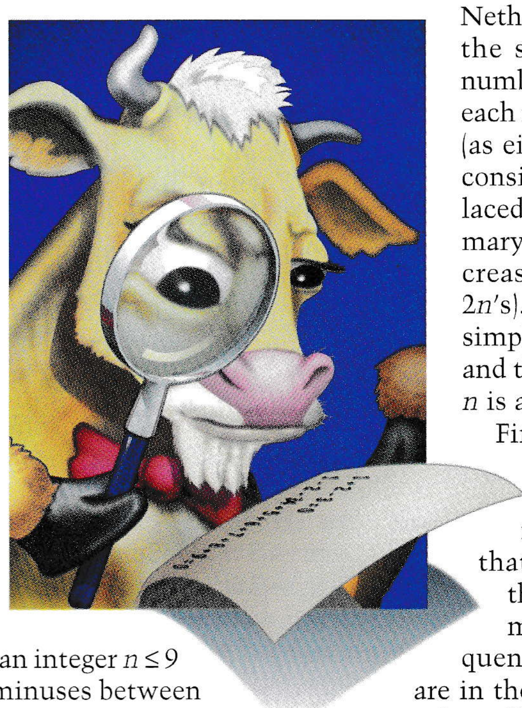
4. Complete the given tetrahedron to make up a parallelepiped using the second method. Using the law of cosines, represent the quadruple square of the length of each edge of the parallelepiped in terms of the lengths of the diagonals of the corresponding face and the angle between them (one face for each edge). Then, apply the theorem on the sum of squares of the diagonals' lengths of a parallelogram and the sum of the squares of its sides. Combine these equations to obtain the desired result.

Contented cows

by Dr. Mu

WELCOME BACK TO COWCULATIONS, THE column devoted to problems best solved with a computer algorithm. What does it take to produce grade AAA milk? You certainly need healthy cows, wholesome food, well ventilated and sanitary barns, clean udders, sanitary utensils, prompt cooling and protection against dust, flies, and other contamination. That will get you up to grade AA. But to get the triple A rating, you need a bit more. You need well adjusted, and most importantly, contented cows.

Can you buy contentedness in a feed bag? No, it only makes for fatter cows. If it were that easy, anyone could produce grade AAA milk. No, to have truly contented cows you must massage their brains. Here at Farmer Paul's Registered Holstein Dairy, we exercise the mind by feeding our herd a healthy diet of numbers to crunch on. Here's a typical tasty byte. Consider the digits 1234567 . . . n, where $n \leq 9$. Find a way to insert pluses or minuses between digits so the sum of the expression is zero. For example, with $n = 9$, $1 - 2 - 34 + 5 + 6 + 7 + 8 + 9 = 0$. Finding all such solutions can keep a herd contented all afternoon. This suggested the next problem, which is your next Challenge Outta Wisconsin.



Art by Mark Brenneman

COW 17

Write a program that takes as input an integer $n \leq 9$ and finds all ways to insert pluses or minuses between the digits 12345 . . . n so the resulting expression sums to zero. $1 - 2 - 34 + 5 + 6 + 7 + 8 + 9 = 0$ is one solution for $n = 9$. Find all of them.

Place the digits in a progression.
Insert signs to make an expression.
Find all those that sum to zero,
And you will be a barnyard hero.
Better yet, your grade will rise,
And your milk will be homogenized.

–Dr. Mu

COW 15

In Cow 15 you were asked to write a program that generates the terms of the sequence 1, 2, 3, 6, 4, 8, 5, 10, 7, 14, 9, 18, . . . Test your program by finding the 100,000th term. Speed and elegance count.

Solution

The most elegant solution was submitted by Russ Cox, a former gold medalist for the United States at the 1995 International Olympiad in Informatics in the Netherlands. Russ writes, "Basically, the sequence consists of pairs of numbers (n, $2n$) sorted by n, such that each number appears in the sequence (as either n or $2n$) exactly once. We consider the sequence as two interlaced sequences: one increasing 'primary sequence' (the n's) and one increasing 'secondary sequence' (the $2n$'s). To print the sequence we must simply devise a test for primary-ness and then print all pairs (n, $2n$) where n is a primary number.

First we note that if a number is not in the secondary sequence, it must be in the primary sequence. Next, we note that there are no odd numbers in the secondary sequence, so they must all be in the primary sequence. Since all the odd numbers are in the primary sequence, it must be true that all numbers of the form 2*odd, where odd is an odd number, must be in the secondary sequence. This means that all numbers of the form 4*odd are in the primary sequence. (If any number of the form 4*odd were in the secondary sequence, then 4*odd/2 = 2*odd would be a primary number, which it is not). Thus 8*odd must be a secondary number. So it turns out that numbers of the form 2^n odd, where n is even are primary numbers, and numbers of the form 2^n odd, where n is odd, are secondary numbers.

This leads to a very nice recursive definition in *Mathematica* of what it means to be a primary number.

```
primaryQ[n_] := If[OddQ[n], True, !
  primaryQ[n/2]]
```

The function **primaryQ** works as follows: If the number *n* is odd, then it is primary. If it is not odd, divide it by 2 and take the negative of the test applied to *n*/2. So if *n* has only one factor of 2, it will not be primary; if it has two, it will be primary; and so on. For example, 4, 12, 16, and 20 are all primary numbers, because they contain an even number of factors of 2.

```
Map[primaryQ, {4, 12, 16, 20}]
```

```
{True, True, True, True}
```

Now we can use **primaryQ** to select those numbers from 1 to *n* that are primary and put them together with the corresponding secondary number (2*primary number) and flatten them all into one list. This is done in *Mathematica* as follows:

```
n = 100;
  Flatten[Transpose[{primarySequence =
  Select[Range[n], primaryQ], 2
  primarySequence}]]
```

```
{1, 2, 3, 6, 4, 8, 5, 10, 7, 14, 9, 18,
11, 22, 12, 24, 13, 26, 15, 30, 16, 32,
17, 34, 19, 38, 20, 40, 21, 42, 23, 46,
25, 50, 27, 54, 28, 56, 29, 58, 31, 62,
33, 66, 35, 70, 36, 72, 37, 74, 39, 78,
41, 82, 43, 86, 44, 88, 45, 90, 47, 94,
48, 96, 49, 98, 51, 102, 52, 104, 53,
106, 55, 110, 57, 114, 59, 118, 60, 120,
61, 122, 63, 126, 64, 128, 65, 130, 67,
134, 68, 136, 69, 138, 71, 142, 73, 146,
75, 150, 76, 152, 77, 154, 79, 158, 80,
160, 81, 162, 83, 166, 84, 168, 85, 170,
87, 174, 89, 178, 91, 182, 92, 184, 93,
186, 95, 190, 97, 194, 99, 198, 100, 200}
```

The 100th number in this special sequence is 150.

```
%[[100]]
```

```
150
```

The only problem with the *Mathematica* solution above is that we don't know exactly when we have reached the 100,000th term and no more. This can be taken care of with a simple **While** loop that counts how many primary numbers have been found and stops at 50,000. Thus, the 100,000th term in the special sequence is twice this primary number.

```
n = 50000;
primary = 1; count = 1;
While[count < n, primary++;
  If[primaryQ[primary], count++]]
{primary, 2*primary}
```

```
{74999, 149998}
```

Here is another solution that I wrote in a procedural style. It is a bit harder to figure out how it works. Try it!

```
n = 50000;
primary[1] = 1; secondary[1] = 2;
i = 1; number = 1; index = 1;
Do[number++; i++;
  If[number == secondary[index], number++;
    index++];
  primary[i] = number; secondary[i] = 2
  number, {n}];
```

The answer to the 100,000th number in the sequence is the 50,000th secondary number.

```
secondary[n]
```

```
149998
```

A solution similar to Russ's was submitted by Mario Velucchi, from Piza, Italy.

And finally . . .

Send your solutions to drmu@cs.uwp.edu. Past solutions are available in *Mathematica* notebooks at http://usaco.uwp.edu/cowculations.

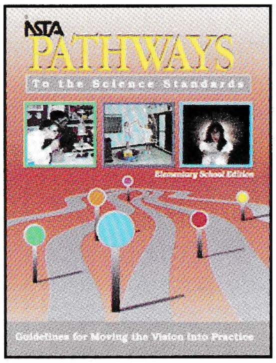
The USA Computing Olympiad has just selected the 15 finalists who will meet June 15–23, at the University of Wisconsin–Parkside in Kenosha, Wisconsin, to compete for one of four positions on the USA Computing Olympiad team. This team will represent the United States at the 11th International Olympiad in Informatics to be held in Antalya, Turkey, October 9–16, 1999. The finalists were selected by their rankings in the USACO National Competition and the three Internet Competitions held throughout the year.

The finalists are: David Cheng, Junior, Brandywine HS, Wilmington, Del.; John Danaher, Junior, Thomas Jefferson HS for Science and Technology, Alexandria, Va.; Gary Huang, Sophomore, Templeton West HS, Appleton, Wisc.; Bill Kinnersley, Junior, Lawrence HS, Lawrence, Kans.; Percy Liang, Junior, Mountain Pointe HS, Phoenix, Ariz.; Benjamin Mathews, Senior, St. Marks HS, Tex.; Jon McAlister, Senior, Langham Creek HS, Houston, Tex.; Ilia Mirkin, Sophomore, Thomas Jefferson HS for Science and Technology, Alexandria, Va.; Oaz Nir, Sophomore, Monta Vista HS, Saratoga, Cal.; Vladimir Novakovski, Freshman, Thomas Jefferson HS for Science and Technology, Alexandria, Va.; John O'Rorke, Junior, Centennial HS, Boise, Ida.; William Potscavage, Junior, Langham Creek HS, Houston, Tex.; Kaushik Roy, Senior, Montgomery Blair HS, Silver Spring, Md.; Daniel Wright, Senior, St. David's College (South Africa), now living in Lafayette, Colo.; Daniel Zaharopol, Junior, Vestal Senior HS, Vestal, N.Y.

If you are seeking contentedness, try crunching our brand of programming bytes at http://www.usaco.org.

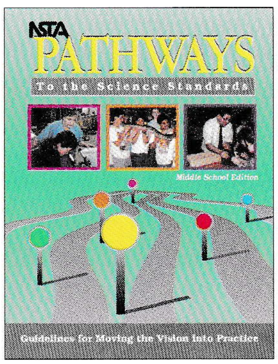
Volume 9 (1998–99)

Index of Advertisers



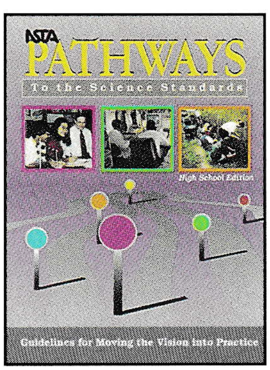
"Pathways is a roadmap, a model. It could get us closer to a strengthened science literate citizenry than any other guidelines currently available"

Dr. John R. Pancella
Science Education Consultant



"I am delighted to report that this book may be a means of rescuing many practitioners from the overwhelming nature of the task of reform before them."

Dr. Dennis W. Cheek
Rhode Island Department of Education, Director



Bring the *Science Standards* to life with the NSTA Pathways Series

Guidelines for Moving the Vision into Practice

Each practical guidebook demonstrates how you can carry the vision of the *Standards*—for teaching, professional development, assessment, content, program, and system—into the real world of the classroom and school. *Pathways* is also a tool for you to use in collaborating with administrators, school boards, and other stakeholders in science education. Filled with specific suggestions and clear examples that guide teachers on implementing each of the standards, *Pathways* is a valuable resource for everyone involved in science education. (Elementary School, Middle School, and High School Editions available)

EACH PATHWAYS BOOK ADDRESSES:

Science Teaching Standards
Professional Development Standards
Assessment Standards
Content Standards
Program Standards
System Standards

PRICING FOR NSTA PATHWAYS

Each Edition of NSTA Pathways
$34.95

▲ **Elementary School** **#PB124X**
(grades K–6, 1997, 152 pp.)

▲ **Middle School** **#PB125X**
(grades 5–9, 1998, 156 pp.)

▲ **High School** **#PB126X**
(grades 9–12, 1996, 196 pp.)

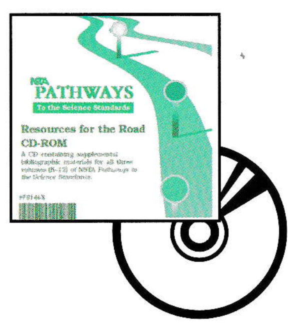
Resources for the Road CD-ROM (MAC/WIN)
#PB146X **$19.95**

Set of 3 Pathways & CD-ROM
(includes K–6, 5–9, and 9–12 books, and Resources CD-ROM)
#PB131X **$112.50**

To order, call 1–800–722–NSTA

Resources for the Road CD-ROM

Resources for the Road CD-ROM contains supplemental bibliographic materials for all three volumes (K–12) of NSTA Pathways. This CD has been designed with a list of contents arranged by science standard to agree with the table of contents appearing in the three Pathways books. (general, 1998)



P1/299Q2