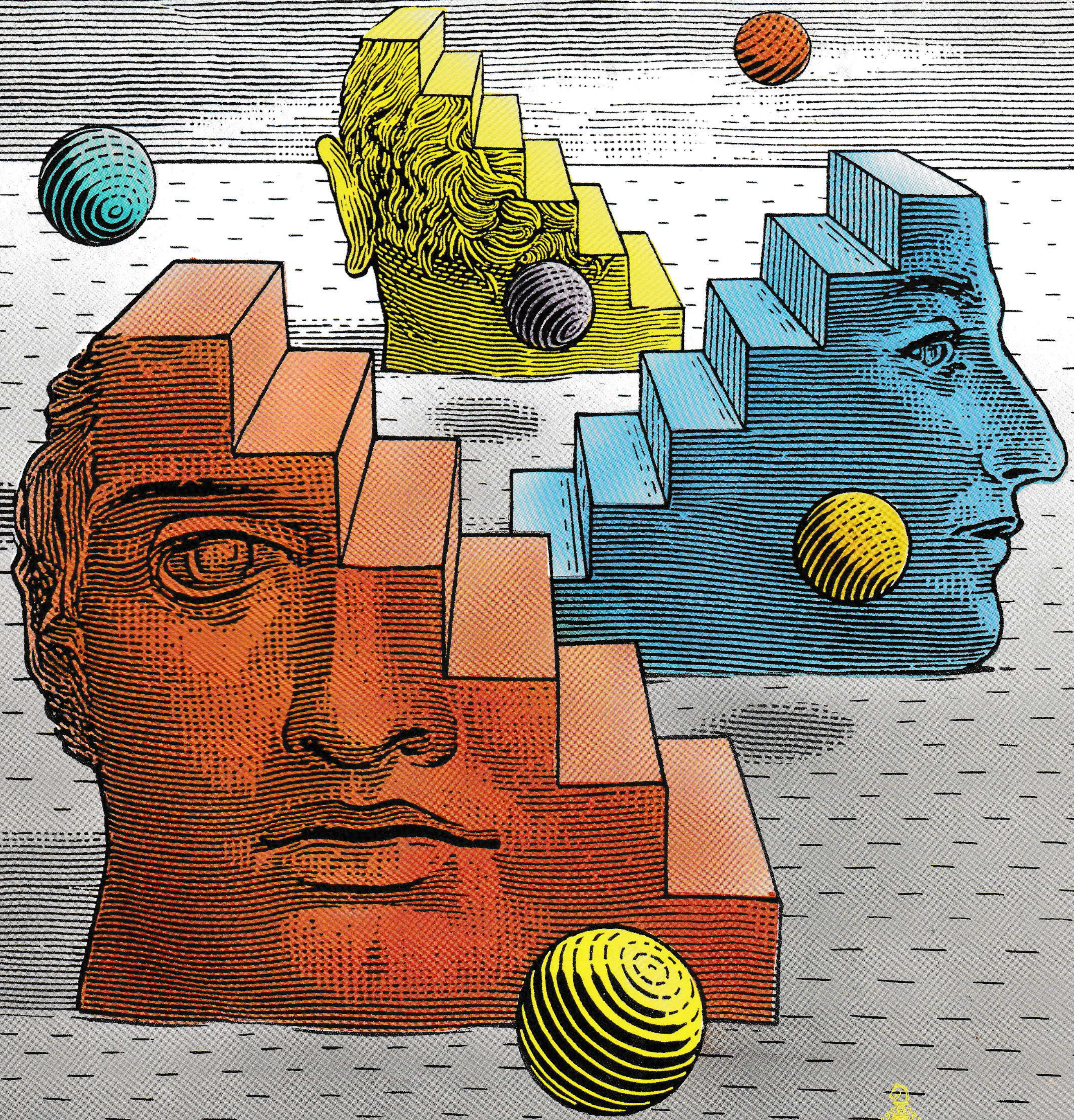
QUANTUM

MARCH/APRIL 1999

US $7.50/CAN $10.50



Springer

NSTA

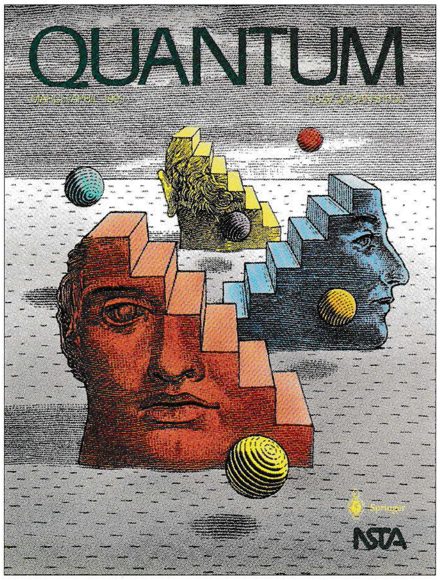
Velázquez Painting the Infanta Margarita with the Lights and Shadows of His Own Glory (1958) by Salvador Dali

HAVE YOU EVER WONDERED WHAT IT MIGHT BE LIKE to see the world through the eyes of an artist? Judging from this painting, one might conclude that Dali viewed the world through prisms and pinholes. We know that the range of patterns created by the interplay of light and objects can range from spectacular prismatic rainbows to more subdued black-and-white designs. Dali, however, sees no distinction when he combines the two in this portrait of Velázquez painting the Infanta Margarita. While Margarita has been exploded into her chromatic elements, Velázquez is left standing beneath an alternating overlay of shadow and light. To the trained eye, these patterns reveal something about the true nature of the elements involved, and maybe this was Dali's intention. Of course, we may be reading too much into the painting. To find out what truths can be discerned from similar patterns of light and dark, see "Diffraction in laser light" on page 33.

QUANTUM

MARCH/APRIL 1999

VOLUME 9, NUMBER 4



Cover art by István Orosz

Smooth curves can be approximated in stepwise fashion by aligning small squares along them. In the digital world of the pixel and its three-dimensional counterpart, the voxel, such approximations lead to some interesting mathematical challenges.

Prepare your palette and turn to page 10 to try your hand at "Painting the Digital World."

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

FEATURES

DEPARTMENTS

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 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

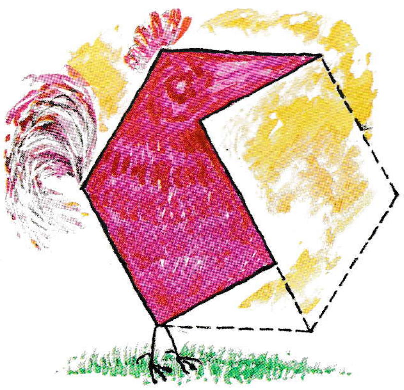
A publication of the National Science Teachers Association (NSTA) & Quantum Bureau of the Russian Academy of Sciences in conjunction with the American Association of Physics Teachers (AAPT) & the National Council of Teachers of Mathematics (NCTM)

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!

B256

Parenthetical puzzler. Place parentheses in the expression $2 \div 2 - 3 \div 3 - 4 \div 4 - 5 \div 5$ to obtain a number greater than 39.

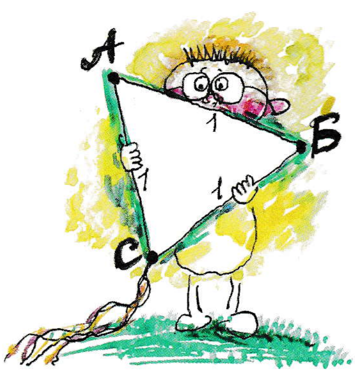


B257

Tripartite equality. The shaded figure shown is half of a regular hexagon. Cut it into three congruent parts.

B258

Rectangulation. A large rectangle is divided into several smaller rectangles. The areas of some of them are shown. Find the area of the rectangle labeled with the question mark.



B259

Triangulation. All sides of triangle ABC are 1 cm long. Point D lies 7 cm from point A. Find the distances from point D to points B and C if it is known that these distances, as expressed in centimeters, are integer values.

B260

Load within a lode. Imagine that you descend into an extremely deep mine with a friend. You carry a balance set at equilibrium by a 1 kg mass. Your friend holds a spring scale from which a 1 kg mass is suspended. Will the readings of these scales differ near the bottom of the mine?

ANSWERS, HINTS & SOLUTIONS ON PAGE 53

Art by Pavel Chernusky

Convection and displacement currents

Exploring the nature of electricity

by V. Dukov

WHAT IS ELECTRIC CURRENT? Usually this question is answered as follows: It is a directed motion of electric charges in a conductor. However, this is true only to a certain degree. To know the entire story, let's recall how this notion appeared in science.

In 1800 the Italian physicist Alessandro Volta (1745–1827) discovered a method to produce direct current using an emf source commonly referred to as a galvanic cell. There was a hypothesis at that time that treated electricity as a massless liquid capable of penetrating bodies via the tiniest pores. Naturally, electric current was considered to be the flow of this liquid along a circuit. Physics marched a long way before it was established that the real moving objects inside conductors are charged particles.

The first step in this direction was made by Michael Faraday (1791–1867), who showed that electric current in electrolytes is due to the motion of ions. An ion is an atom that has either a surplus or deficiency of electric charge. Therefore, every ion is characterized by two values: mass and electric charge. In electrolysis an ion transfers its charge to an electrode and adheres to the electrode, a process known as electroplating. Many experiments made by Faraday showed that ions of the same valence carry the same amount of electric charge. Many years later, this fact led Hermann von Helmholtz (1821–1894) to the conclusion that an elementary charge existed. The fact that this elementary charge is carried by the electron was discovered only at the end of the nineteenth century.

Note that if Faraday had postulated the existence of an elementary charge (electron) and two conservation laws (for energy and mass), then he could have obtained his two basic laws of electrolysis theoretically. In reality, he discovered them after many years of hard work.[1]

Imagine N ions moving in an electrolyte. Each of them carries a charge of Ze_0, where Z is the valence and e_0 is the elementary charge. Then the mass of the ions plated on the electrode will be

$$M = Nm_0, \qquad (1)$$

where m_0 is the ion's mass. These ions transfer the following amount of electric charge to the electrode:

$$q = NZe_0. \qquad (2)$$

Dividing equation (1) by equation (2) yields

$$M = \frac{m_0}{Ze_0}q. \qquad (3)$$

Substituting m_0/Ze_0 with k, we get Faraday's first law:

$$M = kq.$$

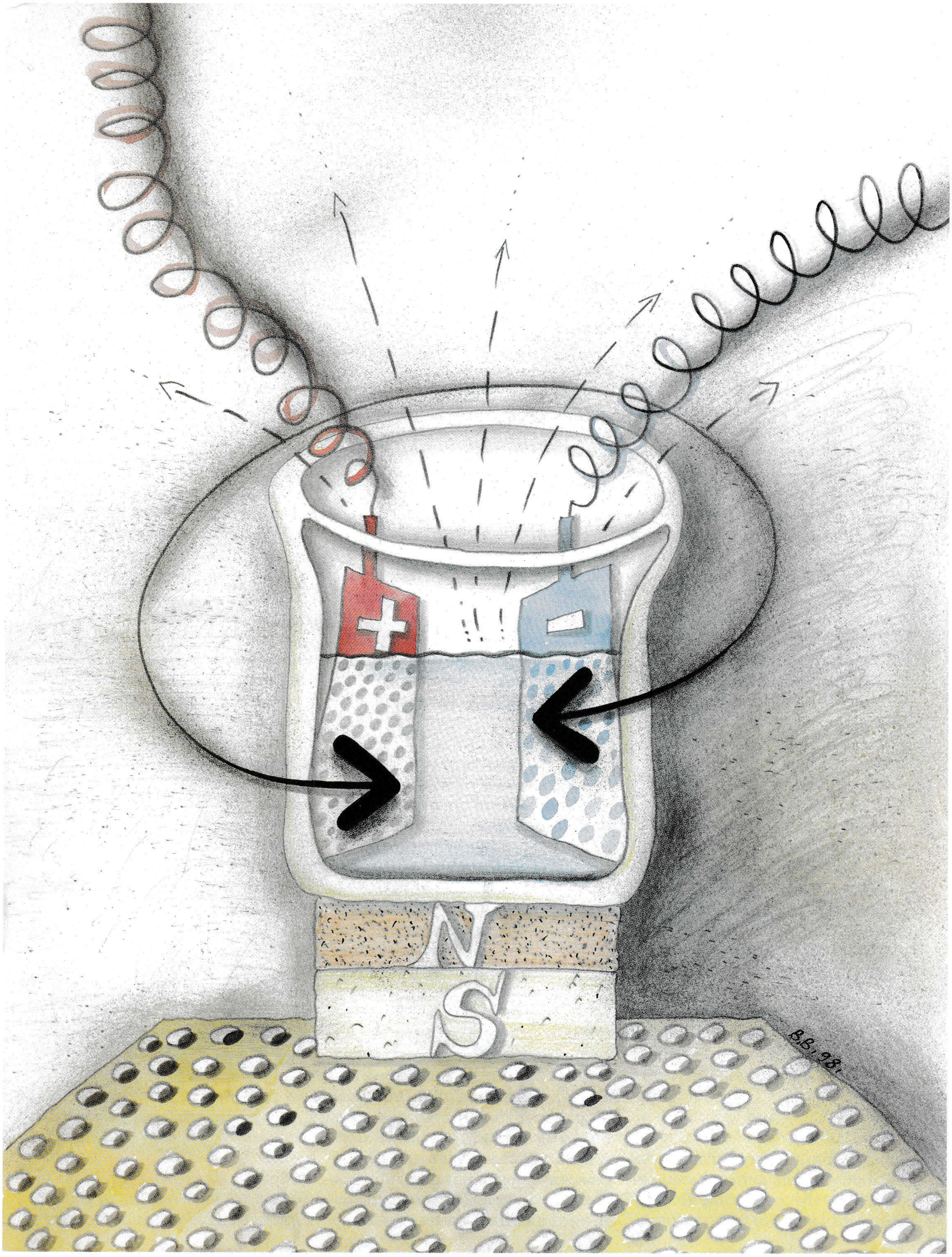
Now we solve this equation for k and multiply and divide the above formula for k by the same factor, Avogadro's number N_A:

$$k = \frac{N_A \cdot m_0}{Ze_0N_A} = \frac{\mu}{Ze_0N_A} = \frac{1}{F}\frac{\mu}{Z}. \qquad (4)$$

where μ is the molar mass of the material and $F = N_Ae_0$ is Faraday's number. Thus we obtained Faraday's second law. Therefore, Faraday's laws

[1] A postulate is a very efficient tool for learning a subject (such as geometry), but it is a dangerous instrument for discovering the laws of nature. Look what Bertrand Russell (1872–1970) wrote: "The method of postulating has many advantages. They are the same as the advantages of theft over honest toil."

Art by Vasily Vlasov


+
-
N
S
B.B. 98.

of electrolysis are underlain by the conservation laws and the existence of an elementary charge.

Equation (3) shows the physical meaning of the Faraday electrochemical equivalence of a substance: The mass of the plating depends on the charge-to-mass ratio of the ions. Equations (3) and (4) show that to obtain a larger mass one must take the substance with greater molar mass and smaller valence.

The ratio of a body's charge to its mass is called the specific charge—a very important value. Let's illustrate the tremendous difference in specific charges of elementary particles and macroscopic bodies. For a "light" particle such as an electron, the specific charge is

$$\frac{e_0}{m_e} = \frac{1.6 \cdot 10^{-19}\ \text{C}}{9.1 \cdot 10^{-31}\ \text{kg}} \cong 2 \cdot 10^{11}\ \text{C/kg},$$

while for the "heavy" proton it is

$$\frac{e_0}{m_p} \cong 10^8.$$

To appreciate the immense difference between these values, let's calculate the specific charge of an aluminum alloy ball 1 cm in diameter with a density of $\rho = 2 \cdot 10^3$ kg/m^3. Its mass is

$$\begin{aligned} m &= \rho V = 4/3\pi r^3 \cdot \rho \\ &\cong \frac{4 \cdot 3.14}{3}\left(10^{-2}\right)^3 \cdot 2 \cdot 10^3\ \text{kg} \\ &\cong 8 \cdot 10^{-3}\ \text{kg}. \end{aligned}$$

What charge can such a ball retain? It is known that intense leakage of electric charge from the surface of a body starts at an electric field $E_0 \cong 3 \cdot 10^6$ V/m. The electric field generated by a sphere of radius r that carries an electric charge q is

$$E = \frac{1}{4\pi\varepsilon_0}\frac{q}{r^2}.$$

The maximum charge that can be retained on this sphere is determined by the condition $E = E_0$. Therefore,

$$\begin{aligned} q_{\max} &= 4\pi\varepsilon_0 r^2 E_0 \\ &\cong 4 \cdot 3.14 \cdot 8.86 \cdot 10^{-12}\left(10^{-2}\right)^2 \cdot 3 \cdot 10^6\ \text{C} \\ &\cong 3 \cdot 10^{-8}\ \text{C} \end{aligned}$$

Accordingly, the maximum specific charge of the aluminum ball is

$$\frac{q_{\max}}{m} \cong \frac{3 \cdot 10^{-8}\ \text{C}}{8 \cdot 10^{-3}\ \text{kg}} \cong 4 \cdot 10^{-6}\ \text{C/kg}.$$

We see that the proton's specific charge is larger than that of the aluminum ball by a factor of 10^{13}!

Imagine that we try to generate an electric current of 1 A using the charged aluminum balls. By definition, $I = q/t$. The charge of a single ball is q_1, so $q = Nq_1$, where N is the number of balls. Thus, the electric current is $I = Nq_1/t$, where

$$N = \frac{It}{q_1} \cong \frac{1\ \text{A} \cdot 1\ \text{s}}{3 \cdot 10^{-8}\ \text{C}} \cong 3 \cdot 10^7.$$

The traffic of balls must be so dense that 30 billion of them cross a fixed area every second! If the balls "flow" in a tube, the tube's cross-sectional area should be

$$\begin{aligned} S &= N\pi r^2 = 3.14 \cdot 30 \cdot 10^6 \cdot \left(10^{-2}\right)^2\ \text{m}^2 \\ &\cong 10^4\ \text{m}^2. \end{aligned}$$

By contrast, a tiny wire a fraction of a millimeter in diameter can carry the same current, although it becomes red-hot. This phenomenon is possible because the elementary particles (electrons, for example) have a huge value of specific charge.

What is the nature of electric current? It is created by the motion of charged particles—that is, the particles move, not the massless charge! Although we write $I = qt$, we consider q to be the sum of the charges of the particles that have a certain mass. Therefore, electric current should have certain mechanical properties. These properties were first observed by Tolman and Stuart. A solenoid connected to a galvanometer was quickly rotated and then abruptly stopped. The galvanometer showed a current corresponding to the motion of negatively charged particles in the direction of rotation. Such experiments made it possible to measure the specific charge of the particles, whose inertia produced the electric current in the decelerating solenoid. This value coincided with the specific charge of the electron found by its deflection in electric and magnetic fields.

The experiments of Tolman and Stuart were the first to show the existence of free electrons in metals. Because of their weak linkage to the crystal lattice, the electrons continued to move due to their inertia after the solenoid was abruptly stopped, and thus a short-lived electric current was generated.

Electric current in electrolytes is the double flow of ions moving in opposite directions. The ions have the same size charges but their masses are different. For example, in a copper sulfate solution the current carriers are Cu^{+2} and SO_4^{-2} ions. The atomic mass of a copper ion is only 63.5, while the mass of a sulfate ion is appreciably greater: $32.1 + 4 \cdot 16 = 96.1$. This difference can be observed in a simple experiment, which can be carried out in a home lab.

A jar with two electrodes is filled with a copper (II) sulfate (also known as blue vitriol) solution, and some small pieces of paper are sprinkled on the liquid's surface. Then, a direct current source is applied across the electrodes. Finally, a permanent magnet is placed under the jar (fig. 1). Behold—the pieces of paper begin to rotate! This phenomenon is explained by the Lorentz force of the magnetic field, which deflects the

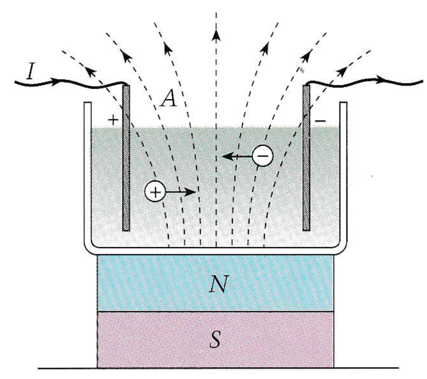


Figure 1

ions. In turn, the deflected ions push the water molecules, so the entire bulk of the solution starts to rotate. The pieces of paper on top of the solution make this effect visible. The Lorentz force deflects both the Cu^{+2} and SO_4^{-2} ions. However, the resulting rotation of the liquid is determined by the heavier particles—in this case, the SO_4^{-2} ions—because they more than cancel the counterrotation of the lighter Cu^{+2} ions. Changing the direction of the current or the polarity of the magnet causes the rotation of the solution to change direction.

A similar effect takes place in the case of leakage of electric charges from a spike. The electric charge density is known to increase drastically in the direction of a sharp point, so a strong electric field is generated in its vicinity. This field ionizes the neutral molecules in the surrounding air. If the sharp point is positively charged, then electrons in the air flow into the spike and positive ions move away from it. The flow of these positive ions carries the neutral air molecules with it—just as the stream of ions in the electrolyte moves the water molecules. The resulting phenomenon is known as "electric wind," which was demonstrated in the middle of the eighteenth century by Benjamin Franklin, who blew out a candle with such a wind.

Historically, Franklin's experiment with blowing out a candle led to the notion of convection currents. It was considered that in parallel with the ordinary convection (that is, the flow of neutral particles) there existed an electric convection, or a stream of charged particles. Indeed, further studies showed that electric currents in metals, electrolytes, and gases are flows of charged particles. Since these particles have masses, there is no principle difference between common convection and electrical convection.

In 1875 Hermann von Helmholtz wrote: "I use this word [*convection*] in the same sense as in thermodynamics, in order to denote the propagation of electricity by the motion of charged bodies." Helmholtz raised a very important question: Can the motion of charged macroscopic bodies be treated as a convection current? The answer to this question could be obtained in an experiment analogous to that performed by the Danish physicist Hans Christian Oersted, who in 1820 found that electric current carried by a conductor deflected a nearby magnetic needle. The needle was affected by a force proportional to the strength of the electric current. The direction of this force depends on the direction of the current.

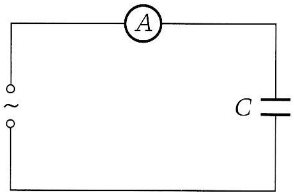


Figure 2

Helmholtz proposed to his pupil Henry Roulend to carry out the following experiment: charge a disk, place a magnetic needle near it, and then spin the disk. If the rotation of a charged disk is equivalent to a current (in a closed circuit), then it should generate a similar magnetic field.

Indeed, Roulend observed deflection of the magnetic needle, but the experiment was not easy. As we pointed out, macroscopic bodies cannot "transport" great charges. So Roulend had to find the magnetic effect of a very small current.

Thus, currents in any substance, fluid or solid, are convective in nature. Can an electric current "flow" in a vacuum where no particles of any kind are available? Consider an experiment with alternating current and a capacitor (fig. 2). When the circuit is closed, the ammeter shows a current. Experiments with various capacitors and different frequencies of alternating current showed that the strength of the current in this circuit is proportional to the frequency of oscillation generated by the emf source and to the capacitance. This experiment was known as early as the nineteenth century. It was explained as follows: The emf source (generator) forces the charged particles to oscillate in the conducting wires, so they "run" from one plate to another, while nothing occurs between the plates (in the vacuum). This explanation considered the current as the mechanical motion of the charged particles.

The outstanding English physicist James Clerk Maxwell (1831–1879) introduced a new concept. The moving charged particles are inherently coupled with electric and magnetic fields. Changes in electric current evoke changes in the fields. Faraday discovered the phenomenon of electromagnetic induction, which according to Maxwell is the generation of an electric field by a varying magnetic field. Being confident of the symmetry of electrical phenomena, Maxwell surmised that an alternating electric field generates a magnetic field.

In the circuit (fig. 2), the electric field changes between the capacitor's plates. According to Maxwell, this process generates a magnetic field. Since a magnetic field can also be produced by an electric current, the process that occurs between the capacitor's plates can be interpreted as a flow of a particular electric current. There are no "ends" of the current in this circuit! If a circuit is opened and the "gap" is filled with a dielectric or vacuum, the current will push its way on, but its nature will be different.

Maxwell baptized this type of electric motion "displacement current." Maxwell supposed that space, which we consider "empty," is in reality filled with a material medium of a particular kind—ether. This ether had a cellular structure (similar to a crystal lattice). The cells could be deformed under the action of an electric field—that is, they could be displaced just like the charged particles in a dielectric body.

According to Maxwell, electric current can be evoked both by convection and by displacement. In the case of convection current, its value is proportional to the velocity of the charged particles. The displacement current is determined by the rate of displacement, which is naturally proportional to the frequency of oscillation. The higher the capacitance, the greater the volume of ether (for a constant distance between the plates), which thus contains more cells and demonstrates a greater effect of displacement. Such was the physical model.

In mathematical terms Maxwell expressed his ideas as

$$i_{dis} = \varepsilon_0 S \frac{dE}{dt}.$$

This says that the displacement current is proportional to the area S at the opening of the circuit (that is, the area of the capacitor's plate) and to the rate of change of the electric field. The constant $\varepsilon_0 = 8.8 \cdot 10^{-12}$ F/m is known as the permittivity of free space.

Assume that the emf source is a generator that produces electric current at the industrial frequency of 60 Hz. Thus, the voltage across the capacitor's plates varies as $v = V_0 \sin \omega t$, where $V_0 = 120$ V. What should the area of the capacitor's plate be to obtain a current of 1 A in the circuit if the distance between the plates is $d = 1$ m?

As we remember, $E = v/d$. Therefore,

$$i_{dis} = \varepsilon_0 S \frac{1}{d}\frac{dv}{dt} = \frac{\varepsilon_0 S}{d}\omega V_0 \cos \omega t.$$

This yields

$$I_0 = \frac{\varepsilon_0 S \omega V_0}{d}$$

and

$$S = \frac{I_0 d}{\varepsilon_0 \omega v_0} \cong 3 \cdot 10^6 \text{ m}^2 = 3 \cdot 10^{12} \text{ mm}^2.$$

If the capacitor is taken away, the current between the ends of the wires (they have a cross-sectional area of 1 mm^2) will be only $3 \cdot 10^{-13}$ A. If the distance between the ends of the wires is decreased to 1 mm, the current will increase by 10^3 times and reach the value of a few ten-billionths of an ampere—virtually zero. In the same conditions we can get a current of 1 A by increasing the frequency of oscillation to 10^{11} Hz, which corresponds to radio frequencies.

These calculations show that the displacement currents become important only at very high frequencies of oscillation. Therefore, these currents are completely ignored in electrical engineering. By contrast, just the opposite takes place in radio engineering, where the displacement current plays the major role.

We always try to imagine a physical process clearly. What is electric current? The first association is this: The current is the displacement of charged particles along a conductor, or their directed motion. Yet, this mechanical model is only a rough approximation of reality. So, one should be reasonably cautious when using such a model.

Let's delineate the concept of electric current. The important thing is that moving particles have electric charge, so they are surrounded by an electromagnetic field. This field is described by two components, the vectors **E** and **B**. In the case of direct current, the electrical component **E** is not detected by devices. Indeed, in every segment of a current-carrying wire there is an equal number of positive and negative charges, and their total electric field is zero. It is only the magnetic component **B** that can be detected by devices in this case (Hans Christian Oersted was first to demonstrate this).

By contrast, in the case of alternating current both components of the electromagnetic field manifest themselves, and in this case induction plays a major role: Variation of the electric component generates the magnetic field and vice versa. Thus, the two fields generate each other, and this mutual generation makes it possible for the double field to live separately from its parent electric current in the conductor. Indeed, the alternating electromagnetic field breaks away from a current-carrying wire (called the transmitting antenna) and travels in space at the velocity of light. This process is known as electromagnetic radiation.

Let's write Maxwell's formula again, but without the proportionality factor $\varepsilon_0 S$:

$$I_{dis} \propto \frac{dE}{dt}.$$

The derivative dE/dt represents the rate of change of the electric field. The quicker **E** changes, the greater the displacement current. However, changes in **E** generate the magnetic field **B**. Therefore, the larger the displacement current, the greater the magnetic field.

Now, what takes place with the usual convection current? The larger the current, the greater the magnetic field. We see the same proportionality between the electric current and the magnetic field.

Therefore, when considering electric current of any kind, we must keep in mind the electromagnetic field generated by it. We must also remember what type of current flows in a particular case. If the current is constant, the key role is given to the mechanical motion of the charged particles in a conductor—that is, to the process of convection. When the current is alternating, the major actor is the electromagnetic field, and its role increases as the frequency of electromagnetic oscillation. ◘

Quantum on electromagnetic phenomena:

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

A. Byalko, "Backtracking to Faraday's law," January/February 1994, pp. 20–23.

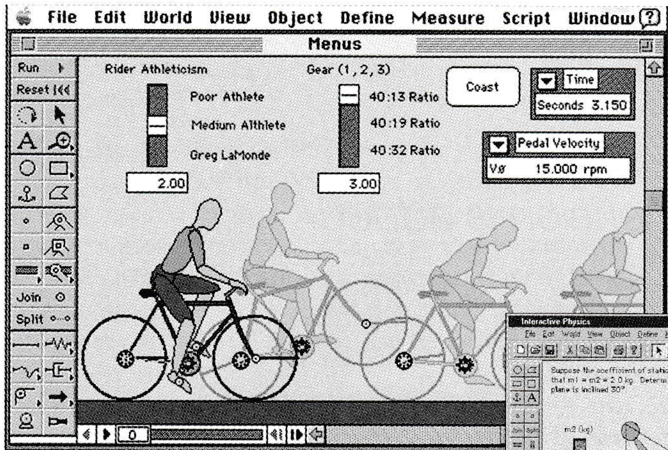
S. Murzin, M. Trumin, and D. Shovkin, "Beyond the reach of Ohm's law," November/December 1994, pp. 24–29.

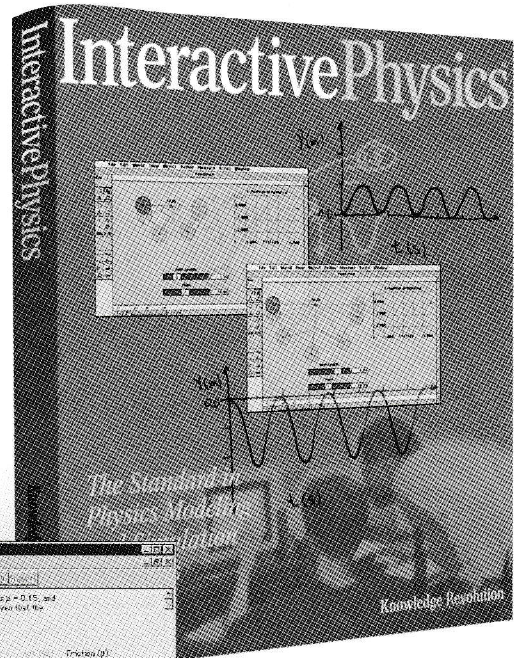
J. Wylie, "Magnetic monopoly," May/June 1995, pp. 4–9.

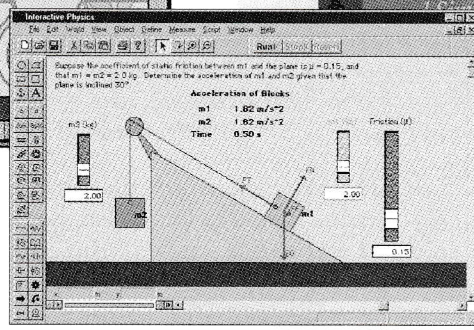
A. Stasenko, "Magnet, charges, and planets," May/June 1997, pp. 42–45.

A. Mitrofanov, "Can you see the magnetic field?" July/August 1997, pp. 18–22.

Revolutionize Your Curriculum!







Contact us for a free Demo CD today!

Create simulations by sketching out bodies and attachments with a simple point-and-click user interface.

Alter physical properties like friction and mass during the simulation to gain valuable insight into the system.

Our Integrated Mathematics and Science Solutions Redefine Your Educational Environment

Interactive Physics™ makes it easy to integrate modeling and simulation into your physics curriculum by offering a complete desktop motion lab.

Create models by drawing onscreen with a powerful and easy-to-use graphic interface. Add objects like springs, dampers, ropes and joints.

Measure physical quantities like velocity, acceleration, momentum and energy. Interact with your model in real time by changing properties as the simulation runs. Widely adopted by many textbook publishers and with more than half a dozen awards and thousands of educational users, Interactive Physics is the standard in physics modeling and simulation solutions.

TOUCHSTONE
1996 Northern California Technical Communication Competition

MARCH 1997
Macworld STAR Rating
★★★★ 8.3

Districts Choice 1997 Curriculum Administrator Award Top 100 Products

1997 EDDIE AWARDS
ComputEd Education Software Review

66 Bovet Road, Suite 200
San Mateo, California 94402
Fax 650 574 7541
Telephone 650 574 7777
Toll-Free 800 766 6615

Knowledge Revolution

www.krev.com

Circle No. 1 on Reader Service Card

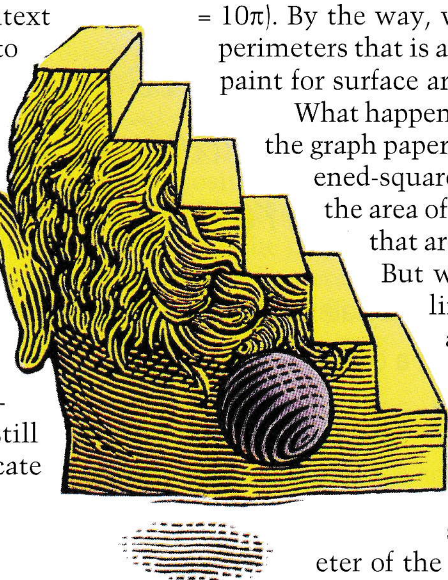
Painting the digital world

Pixels become voxels when a dimension is added

by Michael H. Brill

SOME FASCINATING RELATIONSHIPS CAN be found between areas and perimeters of figures in a plane, and also between volumes and surface areas of solid objects. The relationships are not always intuitively obvious, but can be appealing and simple. They can be visualized in the context of comparing the amount of paint needed to cover two shapes that appear much like each other.

This article introduces such relationships by looking at smooth objects and approximations to these smooth objects that are made out of a latticework of tiny identical cubes. (Maybe other such approximations will occur to the astute reader.) The subject we are about to discuss is perhaps not so exciting as fractals[1] but though tame, it is still interesting, much as the game of checkers is still interesting even though it is not so intricate as chess.



Painting circles and spheres

Let's start in two dimensions by drawing a circle on a piece of graph paper. We can imagine the graph paper as an array of pixels (picture elements) on a display screen, and thus the graph paper is a sort of "digital world." In figure 1, the circle is five box-widths in radius ($r = 5$). Within the circle, the graph-paper squares that lie strictly inside the circle are darkened.

We can see that the area of the circle is $\pi r^2 = 25\pi$ and that the area of the darkened squares is 60. It is no surprise that the area of the squares is smaller than the area of the circle. Somewhat more subtle is the comparison of perimeters. The perimeter of the darkened area is 32, which is larger than the perimeter of the circle ($2\pi r = 10\pi$). By the way, we can imagine a special paint for perimeters that is analogous in two dimensions to the paint for surface areas in three dimensions.

What happens to the areas and perimeters when the graph paper is made finer and finer? The darkened-square area clearly gets larger relative to the area of the circle and actually approaches that area (a behavior familiar in calculus). But what of the perimeter? What is its limiting value? As an aid to arriving at an answer, consider the following theorem.

Theorem: Given any region on the graph paper bounded by a smooth convex curve, no matter how irregular its darkened-square region may be, the perimeter of the darkened-square region is exactly twice the sum of its (maximum) length and its (maximum) width. In the case of darkened squares in the circle, this distance is just four times its width.

Proof: Question 1. Prove this theorem. (Answer on p. 52).

This remarkable relationship does not change in the limit as the squares get very small, in which case the width of the dark area is just the diameter of the circle, or $2r$ (in tiny-box-width units). So the perimeter of the dark area must be $8r$. Meanwhile, the perimeter of the circle (in the same units) is $2\pi r$. The ratio of these num-

[1]See B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, 1983.

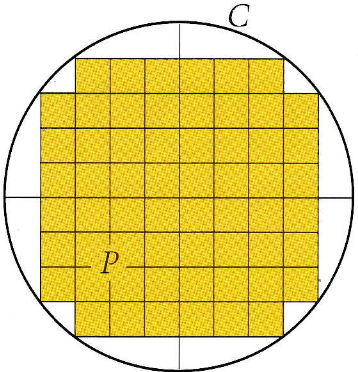


Figure 1

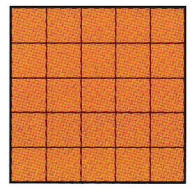


Figure 2

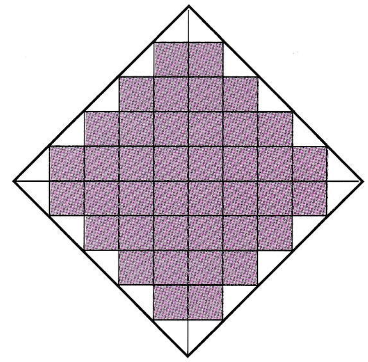


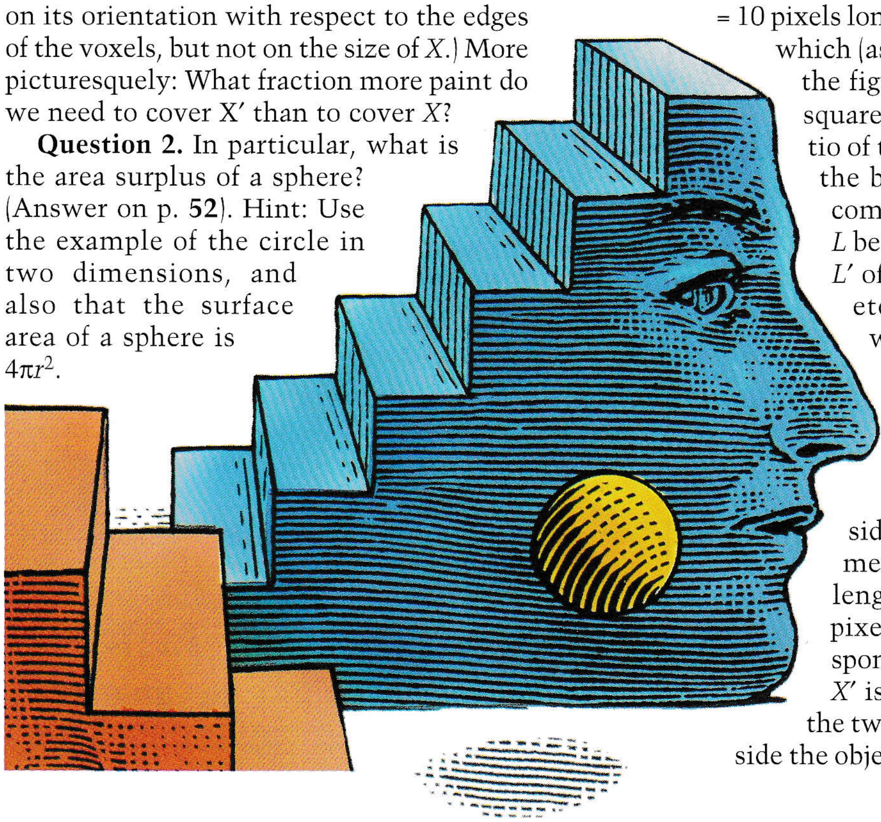
Figure 3

bers (dark-area perimeter/circle perimeter) is $8r/2\pi r = 4/\pi$. We call this number the *perimeter surplus*, since the number is greater than 1. As the squares got smaller, that ratio of areas went from $32/10\pi$ to this greater limiting value. Clearly, the unfilled region in the circle became less important as the squares became smaller, and the number of zigs and zags in the perimeter became more important.

Having gained some insight into perimeters in two dimensions, let's proceed to three dimensions. In this case the "real world" is a set of objects whose surfaces are smooth on a sufficiently small scale. The "digital world" is a three-dimensional mosaic of tiny identical cubes (called *voxels*) analogous to the square pixels we defined earlier for two dimensions. The voxels are tiny on any chosen scale. Each "digital-world" object X' is the maximal subset of these voxels that lies inside the corresponding "real-world" object X.

We want to compare the surface area of X' to that of X, using the ratio of these areas—call it the *area surplus*. (Clearly the area surplus depends on the shape of X and on its orientation with respect to the edges of the voxels, but not on the size of X.) More picturesquely: What fraction more paint do we need to cover X' than to cover X?

Question 2. In particular, what is the area surplus of a sphere? (Answer on p. **52**). Hint: Use the example of the circle in two dimensions, and also that the surface area of a sphere is $4\pi r^2$.



Art by István Orosz

Maximum and minimum paintfor the digital world

How can we find the three-dimensional shapes that have the minimum and maximum area surplus? The answer is not very difficult.

Let's go back to two dimensions for a moment and draw another figure on graph paper, this time a square (fig. 2). Let the square be aligned with the rulings of the graph paper, and let it have a width that is an integer multiple L of a graph-paper box. (In figure 2, $L = 5$.) Then the darkened area (made of graph-paper boxes) exactly matches the big square, and the perimeter is in each case $4L = 20$. The ratio between the perimeters is 1, and this number is the same as the number L gets very large. Therefore, the perimeter surplus is 1. We can see that the perimeter surplus cannot be less than 1, so the square aligned with the graph paper has the minimum perimeter surplus.

Now draw a square turned 45° to the graph paper, as in figure 3. In this case, the diagonals of the square are aligned with the graph paper, and the diagonals are $L' = 10$ pixels long. The darkened region has perimeter 32, which (as noted before) is four times the width of the figure ($4L$, where $L = 8$). However, the big square has perimeter $4L'/\sqrt{2} = 20\sqrt{2}$. The ratio of these perimeters is $(4/5)\sqrt{2}$. If the size of the big square increases indefinitely (L becomes large), then the darkened-area width L becomes fractionally closer to the diagonal L' of the big square. In the limit, the perimeter of the darkened region (4 times its width) becomes $4L'$, and the perimeter of the big square is $4L'/\sqrt{2}$. The perimeter surplus is the ratio of these numbers, or $\sqrt{2}$. This turns out to be the maximum possible perimeter surplus.

To see the maximality of $\sqrt{2}$, consider any almost-linear "side" of a two-dimensional object X with length s and unit-length perpendicular vector $\mathbf{n} = (n_1, n_2)$ in pixel coordinates. The length of the corresponding "side" of the pixel approximation X' is the sum of the projected lengths along the two pixel-axis viewing directions from outside the object X'. This sum is s times the sum of the components n_i in the pixel coordinate directions. A component n_i of a unit vector $\mathbf{n}$ is sometimes called a "direction cosine," because its value is the cosine of the angle between the i-th coordinate unit vector and the vector $\mathbf{n}$. Then the sum of the direction cosines is maximized when they are equal (and thus equal to $1/\sqrt{2}$). Therefore, the maximum perimeter surplus of the facet is $2/\sqrt{2}$, or $\sqrt{2}$. Note that this proof does not depend on the convexity of the object that is approximated by a square, because the argument applies to each local part of the object's boundary.

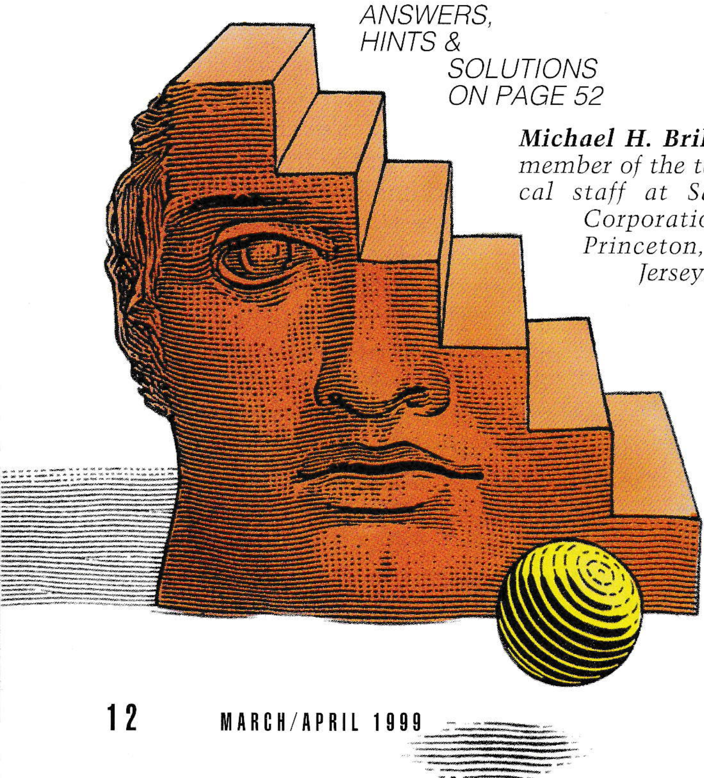
Now let's go through the exercises in three dimensions.

Question 3. What are the minimum and maximum possible area surpluses? Give examples of objects that produce these extremes.

Question 4. Generalize to $(N-1)$-dimensional boundaries of N-dimensional objects.

After you have gone through this exercise, it should be clear that the "digital-world" approximation to a smooth object is somewhere between a "flat-planar facet" approximation and a "fractal" approximation. For a "flat-planar facet" approximation, the facets are organized each with the average inclination of the piece of surface being approximated. As the facets get smaller on a smooth object, the area of facet approximation becomes as close as you like to the area of the original object. Thus the area surplus of the "flat-facet" approximation is 1.

On the other hand, suppose you approximate the original surface by a fractal—a surface made out of bumps, in which each bump has on it smaller bumps of the same shape. In that case, the area is infinite, and thus the area surplus is also infinite. The digital world, tricky as it is, is not nearly so weird as the world of fractals, which has been an object of fascination in mathematics during the past generation or so.



ANSWERS, HINTS & SOLUTIONS ON PAGE 52

***Michael H. Brill** is a member of the technical staff at Sarnoff Corporation in Princeton, New Jersey.*

HOW DO YOU FIGURE?

Challenges

Physics

P256

Gone with the wind. Measuring wind velocity in a sandstorm using conventional devices is difficult (and dangerous), because such storms are usually relatively small and continuously moving. Therefore, an enterprising student proposed measuring the wind velocity with a portable radar, which is possible because a sandstorm carries many small particles that reflect radio

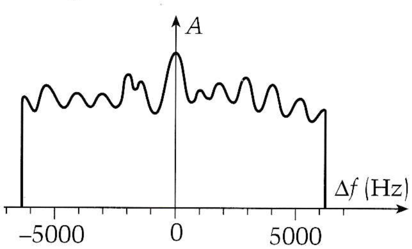


Figure 1

waves at a frequency of $f_0 = 10^{10}$ Hz. The spectrum of the signal after reflection from the sandstorm is given in fig. 1, where $\Delta f = f - f_0$. Find the maximum wind velocity in the sandstorm. (D. Kuptsov)

P257

Quicksilver plug. A small air column of height h = 76 cm is plugged by a mercury column in a vertical

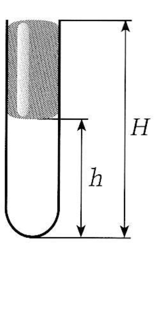


Figure 2

tube of height H = 152 cm (fig. 2). The atmospheric pressure is 10^5 Pa, and the temperature is T_0 = 17°C. To what temperature T_1 must the air in the tube be heated to drive all the mercury out of the tube? (E. Butikov, A. Bykov, and A. Kondratiev)

P258

Mote in a capacitor. A tuned circuit consists of an induction coil and a parallel-plate capacitor C with plate separation d. The natural frequency of oscillation of the circuit is ω_0. What will the value of the natural frequency be if a point-charge q with mass m is placed between the plates? Neglect gravity, edge effects, and electrostatic image forces. (A. Andrianov and D. Kuptsov)

P259

Rotating dipole. An electric dipole is made of two particles with the same mass m. The particles are attached to the ends of a rigid weightless rod of length l. They have electric charges $+q$ and $-q$. This di-

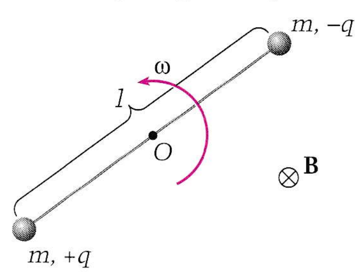


Figure 3

pole is rotating with an angular velocity in the horizontal plane around the vertical axis that passes through the dipole's center (fig. 3). At some time, a vertical magnetic field B is turned on. Describe the steady-state motion of the dipole. (S. Zdravkovich)

P260

Refraction in a ball. A narrow beam of light passing through the center of a glass ball of radius R is focused at a distance $2R$ from its center. Find the refractive index of the glass. (S. Gordyunin and P. Gorkov)

Math

M256

Natural powers. Find a natural number a such that $2a$ is a perfect square, $3a$ is a perfect cube, and $5a$ is a perfect fifth power.

M257

Rooting around. Find the sum of all the real roots of the two following equations:

$$x^3 + 6x^2 + 10x - 15 = 0$$

and

$$x^3 + 6x^2 + 10x + 23 = 0.$$

M258

Nested radicals. Solve the equation $\sqrt{2+\sqrt{2-\sqrt{2+x}}} = x$.

M259

Pyramid ensphered. A triangular pyramid is given such that all plane angles at one of the vertices are right. It is known that a point exists such that its distance from the given vertex is 3, and the distances from the other vertices are $\sqrt{5}$, $\sqrt{6}$, and $\sqrt{7}$, respectively. Find the radius of the sphere circumscribed around this pyramid.

CONTINUED ON PAGE 23

Core dynamics

Why does a transformer need a core?

by A. Dozorov

THE SIMPLEST TRANSFORMER (fig. 1) has two coils wound around an iron core. The primary coil is connected to a source of alternating voltage. A resistor (the load) is connected to the secondary coil. Both coils are threaded by the same alternating magnetic flux generated by the alternating current of the source. An induced emf

$$e_1 = -n_1 \frac{\Delta\Phi}{\Delta t}$$

arises in the primary coil that has n_1 turns. Here $\Delta\Phi$ is the change in the magnetic flux threading a turn during period Δt. An emf is also generated in the secondary coil, and it equals

$$e_2 = -n_2 \frac{\Delta\Phi}{\Delta t}.$$

Let's assume that the secondary coil is not connected to any load. In other words, we consider the idle

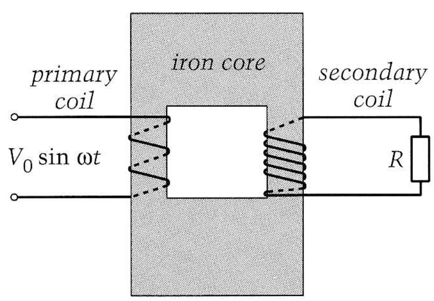


Figure 1

running of a transformer. We also assume that the ohmic resistance of the primary coil is very small compared to its inductive impedance. According to Kirchhoff's second law, the algebraic sum of all emfs in a closed circuit equals the sum of the potential drops in various sections of this circuit:

$$v_1 + e_1 = i_1 R_1.$$

Here v_1 is the emf of the voltage source, i_1 is the current, and R_1 is the ohmic resistance of the primary coil. Since R_1 is very small ($R_1 \to 0$), $v_1 + e_1 = 0$, or

$$v_1 = -e_1.$$

When the secondary coil is disconnected ($i_2 = 0$), the potential drop across its leads will be

$$v_2 = -e_2.$$

Thus, the voltage ratio is $v_2/v_1 = e_2/e_1$, which in the terms of effective (root mean square) values looks like

$$\frac{V_2}{V_1} = \frac{\mathscr{E}_2}{\mathscr{E}_1} = \frac{n_2}{n_1}. \qquad (1)$$

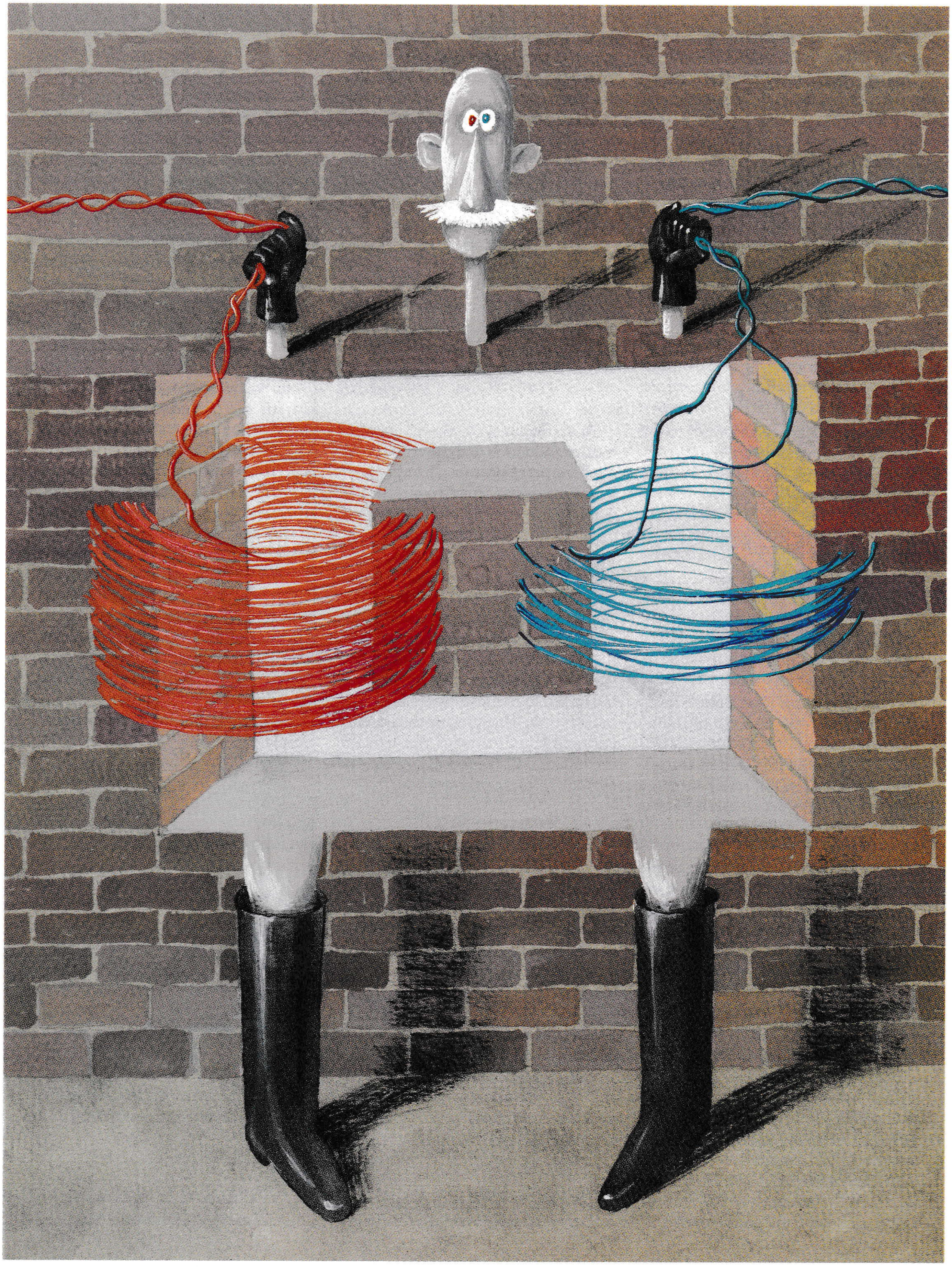
This equation poses the question of why a transformer needs a core. Really, the voltage ratio depends only on the number of turns in each coil. Paradoxically, the parameters of an iron core are absent from equation (1). Can these parameters be chosen arbitrarily? Perhaps a transformer needs no core at all? Under what conditions can this be true?

To crack this puzzle we must look more attentively at our mathematical manipulations. In deriving equation (1) we tacitly supposed that the magnetic fluxes threading through the primary and secondary coils are equal. However, this assumption may not be true: Some portion of the magnetic flux generated by the primary coil may not pass through the secondary coil, which should degrade the efficiency, performance, and technical quality of the transformer.

Perhaps the main role of the core is to decrease the scattering of the magnetic field? However, there are many other (and much easier) ways to keep the magnetic field inside a transformer. For example, one can attach the secondary coil just above the primary coil, or attach both of them to a toroidal (donut-shaped) coil. The same question arises: Why does a transformer need a rather heavy iron core, in which, by the way, some energy losses are unavoidable, because of eddy currents and hysteresis? In short, a core has many deficiencies. Why do we need it, after all?

As a rule, the characteristics of real devices are inferior to those of the ideal devices described by mathematical models (particularly by simplified ones). This is also true in

Art by Pavel Chernusky



the case of transformers. When can a transformer reasonably be considered an ideal one? First, let's find a qualitative answer. In a transformer the electromagnetic energy passes from the primary to the secondary coil. In an ideal device, the energy transferred to the load (connected to the secondary coil) equals the energy taken by the primary coil from the source. We can speak here of power instead of energy. It is also desirable to have the maximum power, which corresponds to the power coefficient $\cos\Phi \cong 1$. Mathematically these requirements are written as

$$V_{0_1} I_{0_1} = V_{0_2} I_{0_2},$$

where the index 0 corresponds to the amplitudes of the currents and voltages. This condition modifies equation (1) in the following way:

$$\frac{V_{0_2}}{V_{0_1}} = \frac{I_{0_1}}{I_{0_2}} = \frac{n_2}{n_1}. \qquad (2)$$

Note that in contrast to equation (1), equation (2) accounts for current flowing in the transformer's coils. Now let's calculate the currents I_{0_1} and I_{0_2} in the case when the transformer is running with a load. To do this, we need the notions of induction and mutual induction.

The magnetic flux Φ that threads through the circuit of area S oriented perpendicular to the circuit's plane is

$$\Phi = BS, \qquad (3)$$

where B is the magnetic field strength. When the magnetic flux is generated by an electric current, B is proportional to the strength of this current I:

$$\Phi \propto B \propto I, \text{ or } \Phi = LI. \qquad (4)$$

The proportionality coefficient L is called the inductance of the circuit. What is this parameter determined by?

Consider a long coil with a large number of turns (a solenoid). The magnetic field generated by a current-carrying wire is always proportional to the strength of the current, but it also depends on the wire configuration, the location of the point at which the field is measured, and the magnetic properties of the surrounding medium. Inside a solenoid the magnetic field is homogeneous. We can deduce that $B = \mu In/l$, where μ is the magnetic permeability, n is the number of turns in the solenoid, and l is its length. Plugging this formula into equation (3) yields

$$\Phi = BSn = \frac{\mu I n^2 S}{l}.$$

By comparing this equation with equation (4) we get the inductance of a solenoid:

$$L = \frac{\mu n^2 S}{l}. \qquad (5)$$

In a similar way, the concept of mutual induction and the corresponding parameter of mutual inductance M are introduced for two circuits. In our case it is mutual induction of the primary and secondary coils of the transformer. When the secondary coil is connected to a load, an electric current flows in the secondary circuit. We note its effective value as I_2. Simultaneously, the current in the primary coil changes. Now its effective value I_1 differs from the current flowing in the idle transformer. The magnetic flux Φ_2 generated by the primary coil and which threads through the secondary coil is proportional to the current I_1 in the primary coil:

$$\Phi_2 \propto I_1, \text{ or } \Phi_2 = MI_1.$$

Similarly, the magnetic flux Φ_1 threading through the primary coil and generated by the secondary coil is proportional to the current I_2 in the latter coil:

$$\Phi_1 \propto I_2, \text{ or } \Phi_1 = MI_2.$$

The coefficient of proportionality M (mutual inductance) is the same in both equations, and it can be obtained in the same way as the inductance of a solenoid:

$$M = \frac{\mu n_1 n_2 S}{l}. \qquad (6)$$

Expressions (5) and (6) result in

$$M = \frac{n_2}{n_1} L_1 = \frac{n_1}{n_2} L_2. \qquad (7)$$

Note that there is only one magnetic field in the transformer core, but mathematically it can be subdivided into two parts. The total magnetic field is almost entirely determined by the primary coil and the voltage of the source to which it is connected. However, it is convenient to consider the individual components of the magnetic fields generated by the currents I_1 and I_2.

Let's continue our story about a transformer running with a load. First, we write Kirchhoff's second law for the closed primary circuit, and then we do the same for the secondary circuit. In the primary circuit (fig. 1) the resistance is extremely low ($R_1 \to 0$), so the algebraic sum of all emfs in this circuit is zero. One emf in this circuit is the applied voltage $v_1 = V_0 \sin\omega t$, and the other two emfs result from induction. One of them is the self-induction emf e_1' generated by the alternating current i_1. Another emf (e_1'') results from mutual induction. It is generated by the current i_2, which produces an alternating magnetic flux in the primary coil. Bearing in mind equation (4), we get

$$e_1' = -L_1 \frac{\Delta i_1}{\Delta t} = -pM \frac{\Delta i_1}{\Delta t},$$

where

$$p = \frac{n_1}{n_2}$$

and

$$e_1'' = -M \frac{\Delta i_2}{\Delta t}.$$

Thus, for the primary circuit we have:

$$V_0 \sin\omega t - pM \frac{\Delta i_1}{\Delta t} - M \frac{\Delta i_2}{\Delta t} = 0. \qquad (8)$$

There are two emfs in the secondary coil: the self-induction emf

$$e_2' = -L_2 \frac{\Delta i_2}{\Delta t} = -\frac{1}{p} M \frac{\Delta i_2}{\Delta i}$$

and the mutual induction emf

$$e_2'' = -M\frac{\Delta i_1}{\Delta t}.$$

For the sake of simplicity, the secondary coil is connected to an ohmic load with resistance R. Therefore,

$$-\frac{1}{p}M\frac{\Delta i_2}{\Delta t} - M\frac{\Delta i_1}{\Delta t} = Ri_2. \qquad (9)$$

Solving the system of equations (8) and (9) is the final answer to the problem of how a transformer works. However, these equations are not very simple (they are called "differential"), because they contain the rates of changes of the unknown values ($\Delta i_1/\Delta t$ and $\Delta i_2/\Delta t$). Still, let's try to solve them.

Because the voltage applied to the primary coil is a sinusoid function of time, it is natural to suppose that both currents in the transformer vary according to the same sinusoid law, although with various phases and amplitudes. So, we seek the currents in the form $i_1 = A \sin(\omega t - \alpha)$ and $i_2 = B\sin(\omega t - \beta)$, where A, B, α, and β are constants. The rate of change of the first current is given by

$$\frac{\Delta i_1}{\Delta t} = \frac{i_1(t+\Delta t) - i_1(t)}{\Delta t}$$
$$= A\frac{\sin[\omega(t+\Delta t)-\alpha] - \sin(\omega t - \alpha)}{\Delta t}$$
$$= 2\frac{A}{\Delta t}\sin\frac{\omega\Delta t}{2}\cos\left(\omega t - \alpha + \frac{\Delta t}{2}\right).$$

Making Δt infinitesimally small, we may use the approximate formulas

$$\sin\frac{\omega\Delta t}{2} \cong \frac{\omega\Delta t}{2}$$

and

$$\cos\left(\omega t - \alpha + \frac{\Delta t}{2}\right) \cong \cos(\omega t - \alpha).$$

Thus,

$$\lim_{\Delta t\to 0}\frac{\Delta i_1}{\Delta t} = A\omega\cos(\omega t - \alpha).$$

Similarly,

$$\lim_{\Delta t\to 0}\frac{\Delta i_2}{\Delta t} = B\omega\cos(\omega t - \beta).$$

Now equations (8) and (9) become the trigonometric equations

$$V_0\sin\omega t - pMA\omega\cos(\omega t - \alpha) - MB\omega\cos(\omega t - \beta) = 0, \qquad (8a)$$

$$-\frac{1}{p}MB\omega\cos(\omega t - \beta) - MA\omega\cos(\omega t - \alpha) = RB\sin(\omega t - \beta). \qquad (9a)$$

Each of these equations can be rewritten in the form

$$a\sin\omega t + b\cos\omega t = 0, \qquad (10)$$

with the time-independent coefficients a and b. Equation (10) is true for any moment of time if simultaneously $a = 0$ and $b = 0$. Thus, two equations make four:

$$\begin{cases} V_0 - pMA\omega\sin\alpha - MB\omega\sin\beta = 0, \\ pA\cos\alpha + B\cos\beta = 0, \\ -\frac{1}{p}MB\omega\sin\beta - MA\omega\sin\alpha = RB\cos\beta, \\ \frac{1}{p}MB\omega\cos\beta + MA\omega\cos\alpha = RB\sin\beta. \end{cases}$$

After solving this system, we get the amplitudes of the currents in the primary and secondary circuits as well as the phase shifts of the currents relative to the applied sinusoid voltage.

Finally we get

$$I_{0_1} = A = \frac{U_0}{p^2R}\sqrt{1+\left(\frac{pR}{M\omega}\right)^2},$$
$$I_{0_2} = B = \frac{V_0}{pR},$$
$$\alpha = \arctan\frac{pR}{M\omega}, \quad \beta = \pi.$$

Does this solution meet the requirements of the ideal properties of the transformer (equation (2))? Let's check it. The ratio of amplitudes of the currents is

$$\frac{I_{0_1}}{I_{0_2}} = \frac{A}{B} = \frac{1}{p}\sqrt{1+\left(\frac{pR}{M\omega}\right)^2}.$$

The latter expression coincides with equation (2) only if the second term under the radical tends to zero:

$$\frac{R}{M\omega} = \frac{Rl}{\mu S n_1 n_2\omega} \to 0$$

(note that the turns ratio p must be constant). Therefore, a transformer can be considered as ideal if

1. the magnetic permeability of the core is large;
2. the frequency of the alternating current ω is rather high;
3. the numbers of turns in both coils are large;
4. the resistance of the secondary coil is low;
5. the length of any coil is small—that is, they are wound tightly.

Each of these conditions should be read as "sufficiently small" (large, high, low, and so on). To make a transformer similar to the ideal one, the most practical requirement must be met first. It is the choice of core, which should be made of a material with high magnetic permeability. For a vacuum, $\mu = \mu_0$, the permeability of free space, while the ferromagnetic materials are characterized by $\mu \cong 10{,}000\mu_0$. The other way is to increase the number of turns, but this is not practical, because it leads to a drastic increase in the size and price of the transformer. An increase in the frequency of the alternating current should be very large in comparison with the industrial frequency of 60 Hz (it must be increased by several thousand times for the efficient performance of coreless transformers). Such an increase creates a number of huge technical problems. However, high frequencies are widely used in electronic devices. You may guess that there are plenty of coreless transformers that are very close to the ideal model transformer. ◘

Quantum on electromagnetic phenomena:

Kaleidoscope: "Electromagnetic induction", March/April 1991, pp. 32–33.

J. Wylie, "Magnetic monopoly", May/June 1995, pp. 4–9.

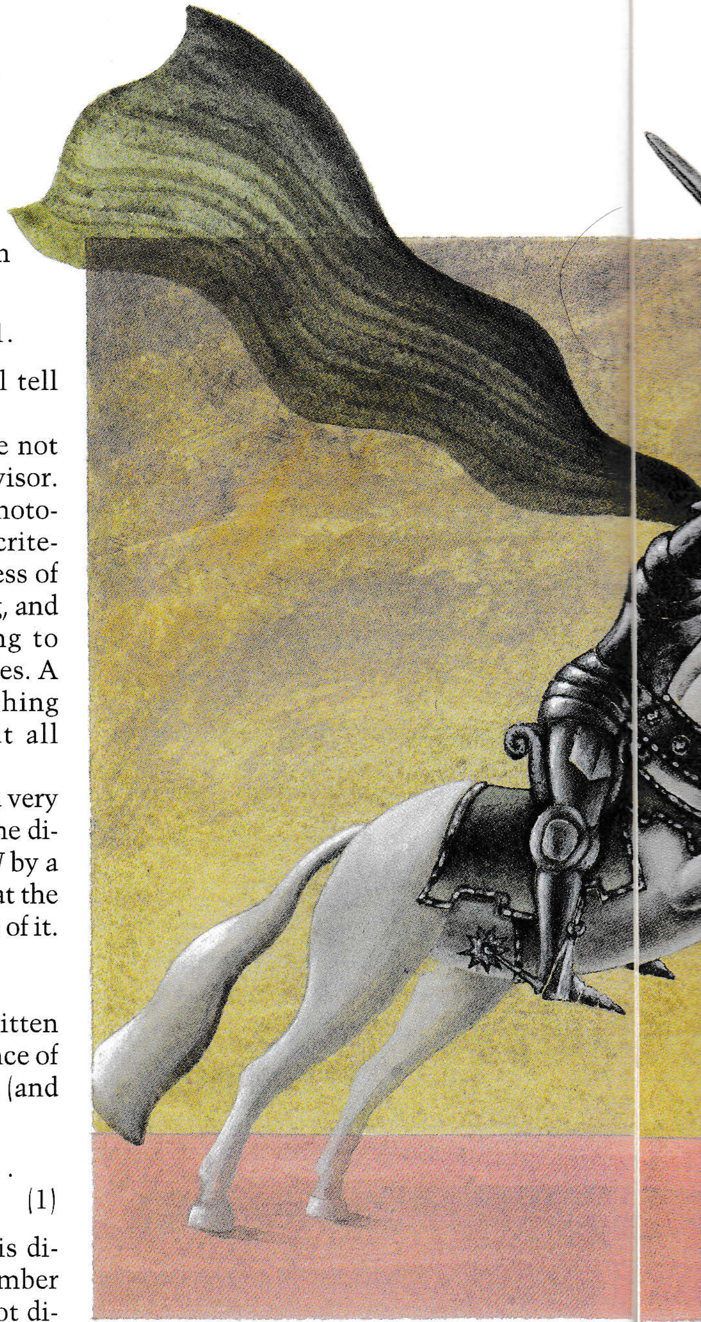
A. Stasenko, "Magnet, charges, and planets", May/June 1997, pp. 42–45.

A. Mitrofanov, "Can you see the magnetic field"? July/August 1997, pp. 18–22.

Divide and conquer!

How to master divisibility

by Ruma Falk and Eyal Oshry



HOW DO YOU TELL WHETHER a given natural number is divisible by, say, 9 (that is, without leaving a remainder)? Although your calculator may give you an instant answer, there is still a conceptual and mathematical challenge in being able to predict what it will tell you.

Many of us have encountered criteria for divisibility in the course of our school years. For example, we know that a number is divisible by 5 if it ends in 0 or 5, and by 2 if the last digit is even. A number is divisible by 4 if the number consisting of the last two digits on the right is divisible by 4, and it is divisible by 8 if the number formed by the last three digits is divisible by 8. The criterion for divisibility by 3 and 9 is also well-known. A number is divisible by 3 (or 9) if and only if the sum of its digits is divisible by 3 (or 9).

Predicting divisibility by 11 is more of a puzzle. The answer is fun: You have to calculate a "zigzag sum" of the digits, by alternately adding and subtracting successive digits of the number. The number is divisible by 11 if and only if that zigzag sum (whether it is positive, zero, or negative) is an integer multiple of 11. For example, consider the number 7,031,673. The zigzag sum is

$$7 - 0 + 3 - 1 + 6 - 7 + 3 = 11.$$

Indeed, as your calculator will tell you, $7{,}031{,}673 = 11 \cdot 639{,}243$.

However, simple criteria are not available for every possible divisor. The number 7, for example, is notorious for evading an efficient criterion. The apparent unrelatedness of the diverse criteria is intriguing, and we may rightly resent having to memorize a list of arbitrary rules. A general principle for establishing divisibility by any divisor at all would be much nicer.

In the following, we present a very basic and general criterion for the divisibility of a natural number N by a potential divisor d, and show that the familiar criteria are special cases of it.

A general criterion for divisibility

If N is a positive integer written in decimal notation as a sequence of $n + 1$ digits $a_n a_{n-1} \ldots a_2 a_1 a_0$ (and $a_n \neq 0$), then

$$N = a_n 10^n + a_{n-1} 10^{n-1} + \ldots + a_1 10 + a_0. \qquad (1)$$

We wish to know whether N is divisible by a given (natural) number d. Clearly, if $d > N/2$, it will not di-

Art by Jose Garcia

vide N: It will be too big.

Take any number c that has the same remainder as 10 when divided by d. We then say that 10 and c are congruent modulo d, and write $10 \equiv c \pmod{d}$. Obviously, if $d \leq 10$, then $c = 10 - d$ is congruent to 10 modulo d. It was the great discovery of Carl Friedrich Gauss, the "Prince of Mathematics," that these congruences for the most part follow the rules of usual arithmetic. In particular, if we replace 10 by c in any polynomial expression, the new value of the expression will be congruent to the old one, modulo d (we will not have changed the remainder when we divide by d).

Let us do this for expression (1). We obtain

$$N \equiv a_n c^n + a_{n-1} c^{n-1} + \ldots + a_1 c + a_0 \pmod{d}. \qquad (2)$$

Now the statement that N is divisible by d is just the statement that N has remainder 0 when divided by d, and this gives us our general criterion:

N is divisible by d if and only if $a_n c^n + a_{n-1} c^{n-1} + \ldots + a_1 c + a_0$ is divisible by d—that is, when $10 \equiv c \pmod{d}$.

(In fact, relationship (2) says a bit more. It lets us know how to compute the remainder when N is divided by d, without actually dividing. We will use this information later).

Examples and exercises

Problem 1. Show that the criterion for divisibility by 9 follows from relationship (2).

Solution. Note that $10 \equiv 1 \pmod{9}$. Thus we can let $c = 1$ in (2), to find that

$$N \equiv a_n + a_{n-1} + \ldots + a_1 + a_0 \pmod{9}.$$

Problem 2. Show that the number 8,333,557,778,844,466,686 is divisible by 9.

Solution. We find that the sum of the digits of this huge number is 108. We could check directly that 108 is divisible by 9, but we can also apply our trick a second time. That is, 108 is divisible by 9 if and only if $1 + 0 + 8 = 9$ is divisible by 9, and this is surely true.

The technique of applying a divisibility criterion repeatedly is an important one.

Problem 3. Find the digit x if the number $1473x94$ has remainder 5 when divided by 9.

Solution. The number

$$\begin{aligned} 1473x94 &\equiv 1 + 4 + 7 + 3 + x + 9 + 4 \pmod{9} \\ &\equiv 28 + x \pmod{9} \\ &\equiv 1 + x \pmod{9}. \end{aligned}$$

Since this must be congruent to 5 modulo 9, x can only be the digit 4.

Problem 4. Show that the criterion for divisibility for 3 follows from relationship (2). Hint: Show that you may take c to be 1.

Problems 5–8: (There may be more than one answer to some of these questions. If this is the case, find all possible answers.) Find the missing digit x if

5. $12345x9$ is divisible by 9.
6. $12345x9$ is divisible by 3.
7. $12345x9$ is congruent to 4 mod 9.
8. $12345x9$ is congruent to 2 mod 3.

Problem 9. We know that a number is even (divisible by 2) if and only if its rightmost digit is even. How does this (very easy) divisibility criterion follow from (2)?

Problem 10. Show that the criterion for divisibility by 4 follows from (2).

Solution. Note that $10 \equiv 2 \pmod{4}$. We let $c = 2$ in all but one of the terms in relationship (2):

$$N \equiv a_n 2^n + a_{n-1} 2^{n-1} + \ldots + a_1 10 + a_0 \pmod{4}.$$

(Can you see which term is treated differently? How can we justify this treatment?) Then we note that $2^n \equiv 0 \pmod{4}$ for $n > 1$. So $N \equiv a_1 10 + a_0 \pmod{4}$, which is exactly our criterion for divisibility by 4.

Problem 11. Show that the criterion for divisibility by 8 follows from relationship (2).

Problem 12. State and prove a criterion for divisibility by 16.

Solution. A number is divisible by 16 if the number formed by its last *four* digits is divisible by 16.

Problems 13–24. Find all possible values for the missing digits x and y if

13. $3578x8$ is divisible by 4.
14. $3578x8$ is divisible by 8.
15. $945x34$ is congruent to 2 mod 8.
16. $1435x9$ is congruent to 5 mod 16.
17. $23y579x$ is divisible by 6.
18. $x003561$ is divisible by 18.
19. $345xy$ is divisible by 24.
20. $456x7y2$ is divisible by 12.
21. $45x83y$ is divisible by 15.
22. $3014x5y$ is divisible by 12.
23. $9x2x1$ has remainder 3 when divided by 12.
24. $42673xy$ is divisible by 60.

To discuss divisibility by 11, we must talk about remainders of negative dividends. While our remarks are perfectly general, we will take the divisor 11 as an example.

We know that 47 has remainder 3 when divided by 11—that is, $47 \equiv 3 \pmod{11}$. If we add a multiple of 11 to 47, it is not hard to see that the result will still be congruent to 3, modulo 11 (try it). But we can also subtract multiples of 11 from 47 without changing this remainder. Indeed,

$$47 - 11 = 36,$$
$$47 - 22 = 25,$$
$$47 - 33 = 14,$$

and

$$47 - 44 = 3$$

are all congruent to 3 modulo 11. Let's continue our subtractions:

$$47 - 55 = -8,$$
$$47 - 66 = -19,$$
$$47 - 77 = -30.$$

We will agree to say that the numbers −8, −19, and −30 are also congruent to 3, modulo 11. It turns out that this is a very convenient way to speak of remainders of negative numbers. Indeed, in advanced work, one *defines* the statement $a \equiv b \pmod{11}$ to say that $a - b$ is divisible by 11.

And now, as mathematicians often do, we generalize the notion of congruence to negative numbers by making the above observation into a *definition*. That is, we say that $a \equiv b \pmod{d}$ for any two numbers a and b if and only if $a - b$ is a multiple of d.

Using this idea, we can say that $10 \equiv -1 \pmod{11}$, since $10 - (-1) = 11$.

Problem 25. Prove the "zigzag sum" criterion for divisibility by 11.

Problems 26–28. Find all possible values for the missing digits x and y if:

26. $143x59$ is divisible by 11.
27. $143x59$ is congruent to 5 mod 11.
28. $3074x8y$ is divisible by 33.

Problem 29. For any three digits A, B, and C, show that the number $ABCABC$ is always divisible by 7.

Solution. We have

$$ABCABC = 10^5A + 10^4B + 10^3C + 10^2A + 10B + C.$$

Using $d = 7$, $c = 3$ (since $10 \equiv 3 \pmod 7$), and relationship (2), we find

$$ABCABC \equiv 3^5A + 3^4B + 3^3C + 3^2A + 3B + C \pmod 7.$$

Therefore,

$$ABCABC \equiv (3^3 + 1)(3^2A + 3B + C) \pmod 7,$$

but $3^3 + 1 = 28$ is a multiple of 7. This proves that $ABCABC$ is divisible by 7. Alternatively, we can note that $ABCABC = ABC(1001)$, and $1001 = 7x11x13$. So the number $ABCABC$ is divisible by 11 (as can be verified by applying the zigzag-sum criterion) and 13 (try our general criterion with $c = -3$) as well.

Generalizing the divisibility criterion

The above criterion for divisibility depends on the decimal system: The number c has to satisfy $10 \equiv c \pmod d$. However, we can easily extend the criterion to any base.

Let b denote the base of our system of numbers. If N is written as a sequence of digits $a_na_{n-1} \dots a_1a_0$ (and $a_n \neq 0$) in the b-system, then

$$N = a_nb^n + a_{n-1}b^{n-1} + \dots + a_1b + a_0.$$

Clearly, if d is our potential divisor ($d \le N/2$), and we take any number c such that $b \equiv c \pmod d$, we obtain the same congruence as in relationship (2). Our extended general criterion is now as follows:

N is divisible by d if and only if $a_nc^n + a_{n-1}c^{n-1} + \dots + a_1c + a_0$ is divisible by d, where $a_na_{n-1} \dots a_1a_0$ are the digits of the number N, written in a system of base b, and $b \equiv c \pmod d$.

In particular, when $d = b - 1$, we get $b \equiv 1 \pmod d$. This means that a number is always divisible by $b - 1$ if its sum of digits is divisible by $b - 1$. Furthermore, this criterion applies to any divisor of $b - 1$, as in the case of the number 3 in the decimal system.

The difficulty of testing for divisibility by 7 now disappears. As Ralph P. Boas suggests in "The Lighter Side" of the *Two-Year College Mathematics Journal* (1979, vol. 10, p. 28): "You just express the number in base 8 and see if the sum of digits is divisible by 7."

Problem 30. For which base(s) is it true that a number is divisible by 2 if its sum of digits is even?

Solution. This is true for any odd base (b) that is no less than 3. This is because $b \equiv 1 \pmod 2$, and we may use $c = 1$.

Problem 31. Is the number 100111, as written in the binary system, divisible by 3?

Solution. For $b = 2$ and $d = 3$, we may take $c = -1$, because $2 \equiv -1 \pmod 3$. Therefore, we should check the zigzag sum $1 - 0 + 0 - 1 + 1 - 1$. The sum is zero, so the number is divisible by 3 (the number is 39 in decimal notation).

***Ruma Falk** is a professor of psychology and education at the Hebrew University of Jerusalem. **Eyal Oshry** is a graduate of the Hebrew University in statistics and psychology.*

Acknowledgments:
This study was supported by the Sturman Center for Human Development, the Hebrew University of Jerusalem. We thank Raphael Falk for his helpful comments and Mark Saul for his editorial contributions.

AT THE
BLACKBOARD I

Gliding home

by Albert Stasenko

A BRIGHT STUDENT OF AERO-dynamics was daydreaming about the upcoming semester break. He wanted dearly to spend his vacation at home. However, airline tickets were expensive and he had no desire to earn the money working as a retail salesman or a night-shift laborer.

Well, knowledge really does mean power! Suddenly an idea struck him. Why not construct a glider and accelerate it on the slippery ice-covered roof of his dormitory? Such an acceleration could be performed with the help of his friends by means of a weightless unstretchable cord threaded through a frictionless pulley (fig. 1). Were it also possible to disregard friction on the roof and air resistance . . .

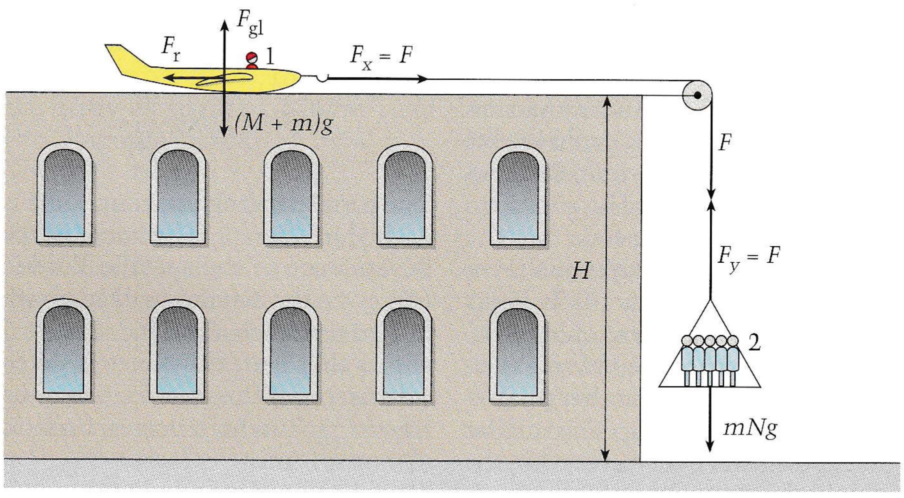


Figure 1. *A historic flight. Key: 1. Student of mass* m. *2. Friends of total mass* mN.

Well, let the glider's mass be M, let the mass of each friend and the student himself be m, and let the force applied by the cord be F. Now we can write Newton's second law for the horizontal motion of the glider with a pilot (mass $M + m$) and for the vertical motion of the platform with N friends (mass mN, the platform being weightless):

$$(M+m)a_x = F_x,$$
$$mNa_y = mNg - F_y.$$

Since the cord is unstretchable, the values of the accelerations of its ends are equal. That is, $a_x = a_y = a$, and $F_x = F_y = F$. So, these equations can be combined to form a single equation by canceling out F:

$$[M + m(N+1)]a = mNg. \qquad (1)$$

The meaning of this equation is clear: Its right-hand side is the force of gravity provided by the friends, and the left-hand side is the mass of the entire system $M_0 = M + m(N + 1)$ (including the mass of the student and the glider), which is accelerated by this force.

According to equation (1), the acceleration is constant, so the velocity will increase linearly with time, and the distance covered will increase as the square of the time:

$$v = at,$$
$$s = y = a\frac{t^2}{2},$$

where

$$a = \frac{mN}{M_0}g.$$

However, the time variable t is not important in this case. It is critical that the dormitory height H and the number of friends N are sufficient to impart to the glider some minimum velocity v_{min} for its successful take-off. Therefore, it is better to write how the velocity depends on the distance traveled. This is done by canceling t from the above equations:

$$v = \sqrt{2ay} \text{ or } \frac{v^2}{2} = ay. \qquad (2)$$

This equation looks very familiar to us. Inserting a into it, we get:

$$M_0\frac{v^2}{2} = (mNg)y. \qquad (3)$$

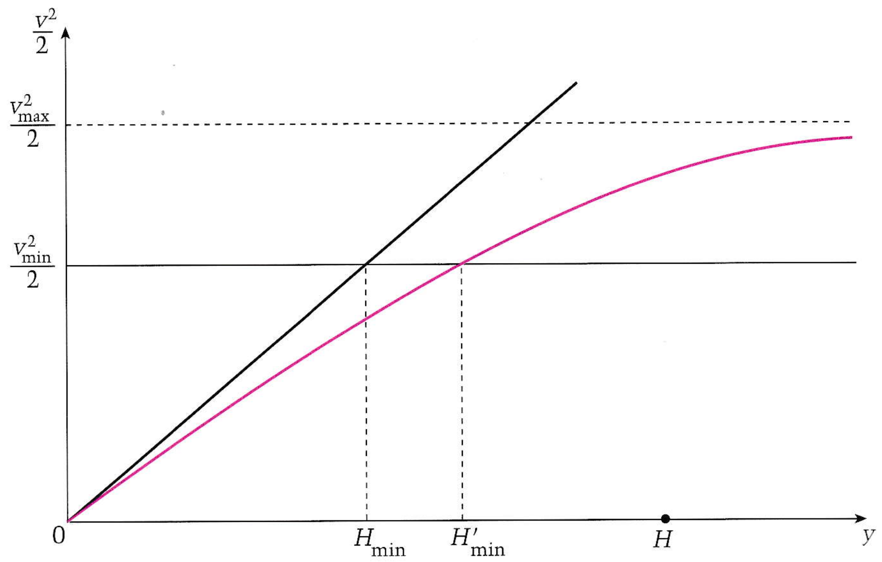


Figure 2. *Dependence of the glider's velocity on the acceleration path.*

Really, this is an old friend: the conservation of the mechanical energy. It says that the kinetic energy is acquired at the expense of the potential energy of the friends due to their descent from zero to $-y$, or in other words, due to the work performed by the constant force of gravity Nmg along the path $s = y$. In accordance with our intuition, in the absence of energy dissipation (we are neglecting friction), the "energy" equation (3) is equivalent to the "force" equation (1).

Thus, within the framework of the assumed simplifications, the specific kinetic energy of the system $v^2/2$ (that is, the energy per unit mass) is proportional to the distance y (see fig. 2 and equation (2)).

Can we decide what critical (minimum) velocity v_{min} must be developed to successfully launch the glider? The question is quite important, because although the dormitory is tall, its height is limited. Will it be enough to launch the glider? We can see that at this minimum velocity the glider's lift F_{gl} will precisely counterbalance its weight $(M + m)g$. From dimensional analysis, it is clear that the lift is proportional to the square of the speed, the wing's area S, and the density of the air ρ:

$$F_{gl} \propto \frac{v^2}{2} S\rho. \qquad (4)$$

This estimate differs from the precise formula only by a dimensionless proportionality coefficient C, which depends on the glider's design and can't be deduced with all the might of dimensional analysis (see "The Power of Dimensional Thinking" by Y. Bruk and A. Stasenko, in *Quantum*, May/June 1992, p. 34, for more details on this method). Another theory or additional experimental data are necessary to obtain this factor, which is known as the *drag coefficient*. Thus, to launch the glider, the following requirement must be met:

$$C\frac{v^2_{min}}{2} S\rho = (M + m)g. \qquad (5)$$

The horizontal line in fig. 2 marks this minimum velocity v_{min}.

At this point the student had another brainstorm: How could the lift exist without air resistance? Something was wrong in the previous analysis. We must insert either the air resistance F_r into equation (1) or its work into equation (3). It looks like the air resistance should be determined by the same physical values but taken with another proportionality coefficient C_x. The smaller the air resistance for a given lift, the more perfect the design of the aircraft. There is a concept of the aerodynamic quality of an aircraft $K = C/C_x$, which makes it possible to write the air resistance as

$$F_r = -\frac{F_{gl}}{K} = -\frac{C}{K}\frac{v^2}{2} S\rho.$$

Now let's add the work performed by this force to equation (3). Since this work varies during the acceleration of the glider, this equation should be applied not to some long distance y, but only to small increments Δy of the trajectory, where the variable force can be considered constant:

$$M_0\Delta\left(\frac{v^2}{2}\right) = mNg\Delta y - \frac{C}{K} S\rho\left(\frac{v^2}{2}\right)\Delta y. \qquad (6)$$

So now we have a differential equation for $v^2/2$ as a function of y. These magic words did not confuse the student. This equation was certainly not particularly difficult, and he surely could solve it. However, much can be seen even without solving this equation.

For example, we can see that at any instant (or at any coordinate y) the glider's speed with air resistance will be less than in the ideal case of no air resistance. Thus, the corresponding curve in figure 2 will be lower than the line ay. When the air resistance (which is persistently rising during the acceleration of the glider) becomes equal to the force of gravity mNg, the glider's speed will not increase any more, and the glider will move with the maximum speed v_{max}. Substituting zero for the left-hand side of equation (6) we get

$$\frac{v^2_{max}}{2} = mNg\frac{K}{CS\rho}. \qquad (7)$$

The corresponding horizontal line is dashed in figure 2. The curve of the dependence of the specific kinetic energy on the distance will approach but never reach this line. Figure 2 shows that if air resistance is taken into account, this curve crosses the line $v^2_{min}/2$ (which corresponds to the minimum velocity v_{min}) at $H'_{min} > H_{min}$. However, it may never cross this limiting line, because to make it possible, the requirement

$v^2_{max}/v^2_{min} > 1$ must be met. Dividing equation (7) by equation (5), we get another form of this condition:

$$\frac{mNK}{M+m} > 1.$$

Now the student knew what to do for a successful flight: invite more friends (increase N), design a lighter glider (decrease M), or make the glider more aerodynamic (increase K). He had guessed these solutions from the very beginning, but now he knew the precise answers to the "how much" questions.

We can see another important aspect of this equation (6) without solving it, but rather by transforming it. Let's use the characteristic velocity v_{max} as a scaling factor. To this end we divide equation (6) by equation (7) and then by the total mass M_0:

$$\Delta\left(\frac{v}{v_{max}}\right)^2 = \left(1-\left(\frac{v}{v_{max}}\right)^2\right)\left(\frac{CS\rho}{KM_0}\right)\Delta y. \qquad (8)$$

Now we see that the distance increment Δy "wants" to be compared with some characteristic value, which must also have the dimension of length. This value awaits us in the last brackets on the right-hand side of the equation. Let's denote it by h:

$$h = \frac{KM_0}{CS\rho}.$$

Assume the aerodynamic quality of the glider to be $K \cong 10$, the area of its wings to be $S \cong 10\ \text{m}^2$, the density of the frosty air to be $\rho \cong 1\ \text{kg/m}^3$, and the drag coefficient to be $C \cong 1$. Then, $h \cong 10$ m.

The parameter h is not just a haphazard combination of a set of physical values. It really characterizes the distance at which the glider's velocity "almost" reaches its maximum, or as physicists say, relaxes to the steady-state value. Thus h can be referred to as the relaxation length.

It seems we have now covered this problem in enough detail. Still, an inquisitive reader might want to know the precise dependence of the glider's velocity on distance. Why not? We must only solve equation (8) with the following initial condition: The glider's speed is zero at $y = 0$.

Notice that we can write

$$\Delta\left(\frac{v}{v_{max}}\right)^2 = -\Delta\left[1-\left(\frac{v}{v_{max}}\right)^2\right],$$

because the increment symbol Δ "annihilates" any constant value (and even the number 1 in the last equation). By introducing a new variable

$$\beta = 1-\left(\frac{v}{v_{max}}\right)^2,$$

we can rewrite equation(8) in the form

$$d\beta = -\beta\frac{dy}{h}.$$

This equation is a backbone of science. It is extremely widespread. For example, it describes the decay of radioactive atoms (in which case y is time and h is the half-life). It also describes the growth of microbes in a Petri dish (h is negative in that case). In our case the solution of this equation is

$$\left(\frac{v}{v_{max}}\right)^2 = 1-e^{-\frac{y}{h}}$$

(if you don't believe it, ask a passerby).

Now the student, having measured the height of his dormitory, could calculate how many friends he must invite to give the glider the necessary velocity.

Alas, after calculating all the expenditures; anticipating the objections of the dormitory supervisor, his friends, and the police; and considering the high probability of the coming thaw (in which case friction with the roof must be taken into consideration), the ambitious student decided that it would be cheaper to buy an airline ticket after all. Bon voyage, young inventor! ◘

CONTINUED FROM PAGE 13

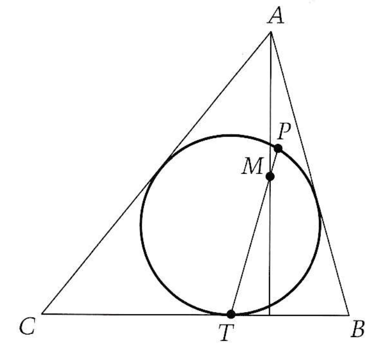


Figure 4

M260

Circular inscription. A circle inscribed in triangle ABC touches BC at point T, and M is the midpoint of the altitude drawn to BC. Point P is the second point of intersection of line TM with the inscribed circle. Prove that the circle that passes through points B, C, and P touches the circle inscribed in triangle ABC (fig. 4).

ANSWERS, HINTS & SOLUTIONS ON PAGE 48

One's best approach

O. T. Izhboldin and L. D. Kurlyandchik

Art by Michael Jones

ONE QUICKLY learns that simple mathematical questions often have complex and profound answers. Several such questions concern representing or approximating integers as sums of reciprocals of natural numbers. Here is one of them.

Problem: Take n natural numbers such that the sum of their reciprocals is less than 1. How large can their sum be?

This problem was first formulated by the American mathematician Kellogg. In 1915 Carmichael presented the problem in his book *Diophantine Analysis.* In 1922 the American mathematician Kersetz gave a solution. We will present here another, much shorter, proof.

We creep up on the problem itself in several steps. Step 1 introduces a sequence that will prove to be the hero of our story.

Step 1: The sequence $r_1, r_2, r_3, \dots$ is defined by the conditions:

$$r_1 = 2,\ r_{n+1} = r_1 r_2 \dots r_{n+1}.$$

The first few terms of this sequence are 2, 3, 7, 43, 1807, 3263443, . . .

Show that for any n,

$$\frac{1}{r_1} + \frac{1}{r_2} + \dots + \frac{1}{r_n} < 1.$$

Solution: The equation that determines r_{n+1} can be written as

$$r_{n+1} = (r_n - 1)r_n + 1,$$

or

$$\frac{1}{r_{n+1} - 1} = \frac{1}{(r_n - 1)r_n} = \frac{1}{r_n - 1} - \frac{1}{r_n}.$$

Therefore,

$$\frac{1}{r_1} + \frac{1}{r_2} + \dots + \frac{1}{r_n} = \left(\frac{1}{r_1 - 1} - \frac{1}{r_2 - 1}\right) + \left(\frac{1}{r_2 - 1} - \frac{1}{r_3 - 1}\right) + \dots + \left(\frac{1}{r_n - 1} - \frac{1}{r_{n+1} - 1}\right) = 1 - \frac{1}{r_{n+1} - 1} < 1,$$

because $r_1 - 1 = 2 - 1 = 1$. This is what we wanted to prove.

The assertion of step 1 shows that the sequence $1/r_n$ "creeps up on" the number 1. So it is a candidate for approximating the number 1, which is what we are asking about in our problem. But is it the best candidate?

Step 2: Suppose we have the sum of the n numbers

$$\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} + \dots + \frac{1}{r_n}.$$

We already know that this sum is less than 1. What is the smallest number a we can choose so that the sum

$$\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} + \dots + \frac{1}{r_n} + \frac{1}{a}$$

is still less than 1?

Solution: We must choose $a = r_{n+1}$. In fact, we've already proven this. Indeed, suppose we had tried to make it even a bit smaller, by choosing $a = r_{n+1} - 1$. Then the above calculation shows that our sum

$$\frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_3} + \dots + \frac{1}{r_n} + \frac{1}{r_{n+1} - 1}$$

would equal 1.

Thus, if we have already chosen

$$\frac{1}{r_1}+\frac{1}{r_2}+\frac{1}{r_3}+\ldots+\frac{1}{r_n}$$

to approximate 1, and we want to add one more term, the biggest step that we can take toward 1 (without actually touching the number 1) is $1/r_{n+1}$. But we have not yet shown that the sequence $\{r_i\}$ is the solution to problem 1. Perhaps if we had taken n different numbers, we could have gotten closer still with one more step.

What is this mystery sequence that solves our problem? We now give three criteria that this sequence must satisfy. Later on, we will see that (a) the sequence $\{r_i\}$ satisfies these criteria, and (b) $\{r_i\}$ in fact solves the problem.

Step 3: Suppose we have a sequence $\{a_i\}$ that gives our answer to problem 1. That is, suppose $a_1 \le a_2 \le a_3 \ldots \le a_n$ of n natural numbers, such that the sum

$$\frac{1}{a_1}+\frac{1}{a_2}+\frac{1}{a_3}+\ldots+\frac{1}{a_n}<1,$$

and such that this sum is the closest approximation to 1 by reciprocals of natural numbers. Let $b_k = 1/a_k$ for $k = 1, 2, \ldots, n-1$, and let $c = 1-(b_1 + b_2 + b_3 + \ldots + b_{n-1})$. (Thus c is the error committed in approximating 1 with the sum

$$\frac{1}{a_1}+\frac{1}{a_2}+\frac{1}{a_3}+\ldots+\frac{1}{a_{n-1}}.$$

Then the set $\{b_1, b_2, \ldots b_{n-1}, c\}$ satisfies the following three conditions:

(i) $b_1 + b_2 + b_3 + \ldots + b_{n-1} + c = 1$;
(ii) $b_1 \ge b_2 \ge \ldots \ge b_{n-1} \ge c \ge 0$;
(iii) $b_1 b_2 \ldots b_k \le b_{k+1} + b_{k+2} + \ldots + b_{n-1} + c$, for $k = 1, 2, \ldots n-1$.

After we take step 3, our strategy will be to show that c, the error of approximation, cannot be less than $1/(r_n - 1)$, and that if $c = 1/(r_n - 1)$, then $a_k = r_k$ for $k < n$ as well. This will solve our problem—well, almost.

But let us return to the solution to step 3. Condition (i) is merely a restatement of the definition of c (we will see later why we include it here).

Most of condition (ii) is almost as easy. We know from the definition of the sequence $\{a_i\}$ that

$$b_1 \ge b_2 \ge \ldots \ge b_{n-1} \ge 0,$$

and that $c > 0$. We must prove only that $b_{n-1} \ge c$. Since $\{a_i\}$ is the best approximation of 1, if we increase the we have

$$\frac{1}{a_1}+\frac{1}{a_2}+\ldots+\frac{1}{a_{n-2}}+\frac{1}{a_{n-1}-1}\ge 1,$$

or

$$1-\frac{1}{a_1}-\frac{1}{a_2}-\ldots-\frac{1}{a_{n-2}}\le\frac{1}{a_{n-1}-1}.$$

Therefore

$$\begin{aligned} c &= 1-\frac{1}{a_1}-\ldots-\frac{1}{a_{n-2}}-\frac{1}{a_{n-1}} \\ &\le \frac{1}{a_{n-1}-1}-\frac{1}{a_{n-1}}=\frac{1}{a_{n-1}(a_{n-1}-1)} \\ &\le \frac{1}{a_{n-1}}=\alpha_{n-1}. \end{aligned}$$

This proves that condition (ii) holds. It remains to prove the set of inequalities that constitute condition (iii). These follow from the equations

$$\begin{aligned} & b_{k+1}+b_{k+2}+\ldots+c=1-b_1-b_2-\ldots-b_k \\ &=1-\frac{1}{a_1}-\frac{1}{a_2}-\ldots-\frac{1}{a_k}=\frac{N}{a_1a_2\ldots a_k} \\ &= Nb_1b_2\ldots b_k, \end{aligned}$$

where N is a certain natural number. Therefore, the last expression in this equation is not less than $b_1 b_2 \ldots b_k$. This concludes the solution to step 3.

Step 4: The set $\{b_1, b_2, \ldots, b_{n-1}, c\}$ satisfying conditions (i) through (iii), and for which c is minimal, coincides with the set $\{1/r_1, 1/r_2, \ldots, 1/r_{n-1}, 1/(r_n - 1)\}$.

We prove this proposition by starting with the set $\{b_1, b_2, \ldots, b_{n-1}, c\}$ such that the product $b_1 b_2 \ldots b_{n-1} c$ is minimal. We then show that this implies that c is also as small as possible. Finally, we show that $b_i = 1/r_i$.

Now we can prove that for any set β, all the inequalities (iii) reduce to equalities.

For $n = 2$, relationships (i)–(iii) immediately imply that $\beta_1 \ge \beta_2 \ge 0$, $\beta_1 + \beta_2 = 1$, and $\beta_1 \le \beta_2$. Therefore, $\beta_1 = \beta_2 = 1/2$. Let $n > 2$. Suppose that there is at least one strict inequality among (iii). Then, we can find i and j such that $1 \le i \le j \le n$, inequalities (iii) are strict for $i \le k < j$, and such that these inequalities reduce to equalities for $k = i - 1$ and $k = j$ (obviously, we can consider the cases $k = i - 1$ and $k = j$ only for $i > 1$ and $j < n$, respectively). Let's prove that we can "jiggle" the numbers β_i and β_j in such a way that conditions (i)–(iii) remain true and the product $\beta_1 \ldots \beta_n$ decreases. Since this fact contradicts the choice of β as the set with the minimal product of elements, the supposition that there is a strict inequality among (iii) will be disproved.

Let us substitute β_i for $\beta_i' = \beta_i + \varepsilon$ and β_j for $\beta_j' = \beta_j - \varepsilon$, where $\varepsilon > 0$. The product of the elements of the new set is less than for set β, because

$$\beta_i'\beta_j'=\beta_i\beta_j-\varepsilon(\beta_i-\beta_j)-\varepsilon^2<\beta_i\beta_j.$$

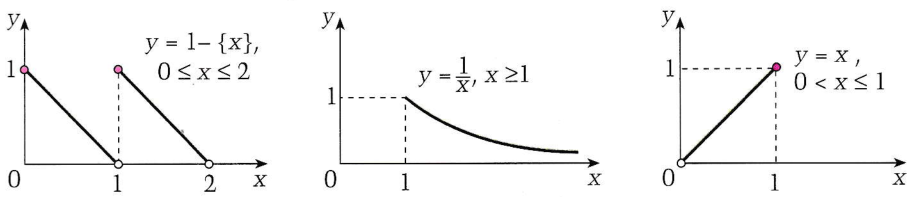
We can see that set β' satisfies condition (ii). The inequalities (iii) also hold for $k < i$ and $k \ge j$ for any $\varepsilon > 0$. For $i \le k < j$, the left-hand side of each of these inequalities increases by a factor of $(\beta_i + \varepsilon)/\beta_i$, and the right-hand side decreases by ε. Therefore, since these inequalities were strict for set β, they remain valid for sufficiently small ε.

Considering inequalities (i), we see that only inequalities $\beta_{i-1} \ge \beta_i$ and $\beta_j \ge \beta_{j+1}$ can be violated as a result of changing the set β (or inequality $\beta_n \ge 0$ for $j = n$ can be violated). However, they are not violated for sufficiently small ε, because the inequalities in question are strict. For example,

$$\beta_{i-1}\ge\beta_1\cdot\ldots\cdot\beta_{i-1}=\beta_i+\ldots+\beta_n\ge\beta_i$$

and for $n > 2$, at least one of these inequalities is strict, because $\beta_1 < 1$ and $\beta_n > 0$. Similarly, $\beta_j > \beta_{j+1}$ for $j < n$.

Let's sum up the results obtained. We now know that the sequence β_1,



a b c

Figure 1

Here are some examples of functions that do not take the minimal value on a given interval: (a) y = 1 – {x} *({x} denotes the fractional part of* x*); this function is not continuous on the interval 0 ≤* x *≤ 2; (b)* y = 1/x *for* x *≥ 1 (the domain of definition is unbounded); (c)* y = x *for 0 <* x *£ 1 (the domain of definition is not closed).*

. . ., β_n satisfies the system of equations

$$\begin{cases} 1 = \beta_1 + \ldots + \beta_n, \\ \beta_1 = \beta_2 + \ldots + \beta_n, \\ \beta_1\beta_2 = \beta_3 + \ldots + \beta_n, \\ \ldots \\ \beta_1 \ldots \beta_{n-1} = \beta_n. \end{cases}$$

Subtract the second equation from the first, the third equation from the second, and so on to obtain the following system of $n - 1$ equations:

$$\begin{cases} 1 - \beta_1 = \beta_1, \\ \beta_1(1 - \beta_2) = \beta_2, \\ \beta_1\beta_2(1 - \beta_3) = \beta_3, \\ \ldots \\ \beta_1\beta_2 \ldots \beta_{n-2}(1 - \beta_{n-1}) = \beta_{n-1}. \end{cases}$$

The first equation gives $\beta_1 = 1/2$. Then, we can rewrite the kth equation $(2 \le k \le n - 1)$ as

$$\frac{1}{\beta_1\beta_2 \ldots \beta_{k-1}} = \frac{1 - \beta_k}{\beta_k}$$

or

$$\frac{1}{\beta_k} = \frac{1}{\beta_1} \cdot \ldots \cdot \frac{1}{\beta_{k-1}} + 1.$$

Thus, $1/\beta_1 = 2 = r_1$. This implies that $1/\beta_k = r_k$ for all $k = 1, \ldots, n - 1$, and

$$\frac{1}{\beta_n} = \frac{1}{\beta_1} \cdot \ \ldots \ \cdot \frac{1}{\beta_{n-1}} + 1.$$

It remains to be proved that for any set $\alpha = (\alpha_1, \ldots, \alpha_n)$ satisfying conditions (i)–(iii) and different from β, $\alpha_n > \beta_n$. We have

$$\alpha_n^2 \ge \alpha_1 \ldots \alpha_{n-1}\alpha_n \ge \beta_1 \ldots \beta_{n-1} = \beta_n^2.$$

(the first inequality follows from condition (iii) for α for $k = n - 1$, and the second inequality follows from the choice of β. Thus, $\alpha_n \ge \beta_n$. If $\alpha_n = \beta_n$, then $\alpha_1 \ldots \alpha_n = \beta_1 \ldots \beta_n$. However, in this case, the set α satisfies conditions (i)–(iii) and has the minimal product of its elements, and thus is identical to β.

But we are not really done. So far, we have proceeded innocently under the assumption that there exists a set with the minimal product of elements. That is, we must prove the following.

Step 5: The function $f(\alpha_1, \ldots, \alpha_n) = \alpha_1 \ldots \alpha_n$ takes on a minimal value on the set A of all sequences $\alpha = (\alpha_1, \ldots, \alpha_n)$ that satisfy conditions (i)–(iii) (this set will be called A).

If it is not clear why we must prove this assertion, look at figure 1. It is not true that any function, defined on any set, must take on a minimal value on that set. Figure 1 shows several examples of functions defined on intervals, whose values are real numbers and which do not take on minimal values on the intervals we have chosen for their domains of definitions. The function in step 5 is a bit more complicated. It is a function of several variables, and its domain of definition is not just a simple interval. The statement in step 5 asserts that it doesn't act, on its domain, like the functions in figure 1 do on theirs: Step 5 asserts that this function does take on a minimal value.

It turns out that assertion 5 rests on some results from the calculus, and we will give a sketch of how to prove it. The following theorem is proven in advanced courses in calculus: *Any continuous function (of one or several variables) defined on a closed and bounded set must take on a minimal value on that set.*

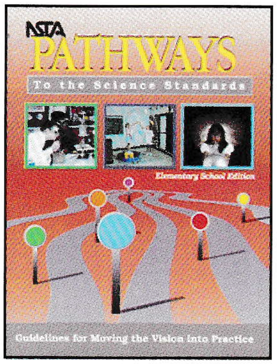
We will apply this theorem to the function in step 5.

First we argue that f is continuous. This means that its values on two different sequences whose corresponding elements are sufficiently close to each other can be made as close as desired. Second, the set A is bounded. This means that there exists some number N such that the elements a_i of any sequence from A do not exceed N (in absolute value). In our case, we can take $N = 1$.

Finally, set A is closed. This condition is a bit more complicated to describe, but not really difficult to understand. To say that A is closed means that a sequence of its elements converges to none of its own elements. This translates into the assertion about a sequence of sequences. Suppose we imagine a first sequence $\{\alpha_i\}$, then a second sequence $\{\beta_i\}$, then a third sequence $\{\gamma_i\}$, and so on. We then form the sequence of first elements $\alpha_1, \beta_1, \gamma_1, \ldots$, and suppose that it converges to a number p_1. Continuing in the same way, we get the sequence $p_i = (p_1, p_2, p_3, \ldots)$. The statement that set A is closed means that this sequence $\{p_i\}$ is an element of A, and this statement follows from the fact that none of the inequalities in conditions (i)–(iii) are strict inequalities.

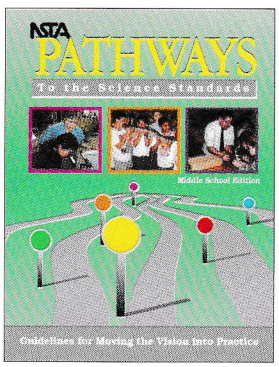
Having shown that the function f of assertion 5 satisfies the conditions of the theorem from calculus quoted above, we conclude that assertion 5 is correct, and that we haven't erred in assuming the existence of a set with a minimal product.

It is interesting that our problem, which was stated purely as a problem in arithmetic, required the use of a deep result in calculus. And yet it is typical of the flavor of mathematics as a discipline that a simply posed problem has a solution whose roots reach deep.



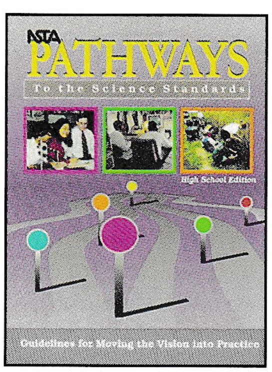
"Pathways is a roadmap, a model. It could get us closer to a strengthened science literate citizenry than any other guidelines currently available"

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

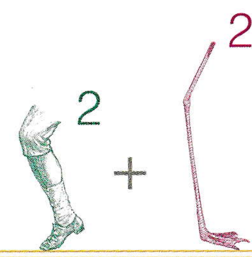
Set of 3 Pathways & CD-ROM
(includes K–6, 5–9, and 9–12 books, and Resources CD-ROM)
#PB131X $112.50

To order, call 1–800–722–NSTA

Resources for the Road CD-ROM

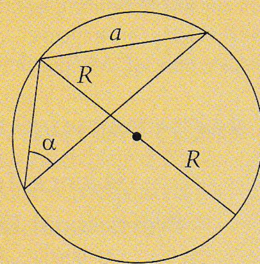
Resources for the Road CD-ROM contains supplemental bibliographic materials for all three volumes (K–12) of NSTA Pathways. This CD has been designed with a list of contents arranged by science standard to agree with the table of contents appearing in the three Pathways books. (general, 1998)

P1/299Q2

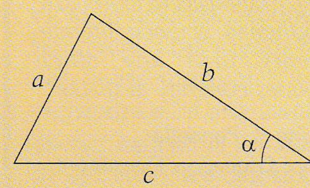


Remarkable geome

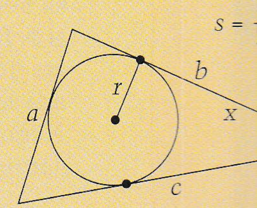
$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\gamma} = 2R,$$



$$a^2 = b^2 + c^2 - bc\cos\alpha,$$



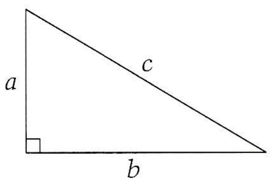
$$x = s - a,$$
$$K = rs,$$



THE BEAUTY OF GEOMETRY is also the beauty of geometric formulas. Here we have gathered some of the most striking algebraic relations in a discipline dominated by figures.

1. The equation

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

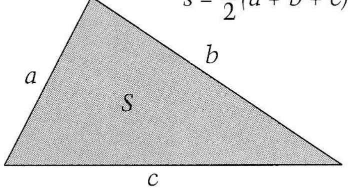


represents the Pythagorean theorem: The square of the hypotenuse of a right triangle equals the sum of the squares of two legs. The Pythagorean theorem is one of the oldest theorems in geometry. The algebraic equation that represents this theorem provides the basis for the metric theory of Euclidean geometry and trigonometry.

One of the oldest and most elegant formulas in geometry is Hero's formula, which represents the area of a triangle in terms of its sides:

$$K = \sqrt{s(s-a)(s-b)(s-c)}.$$

$$s = \frac{1}{2}(a + b + c)$$

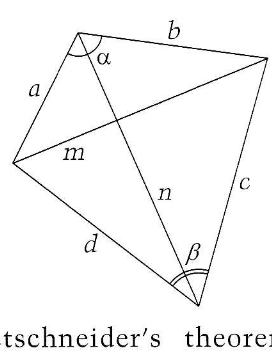


Not pretending to be complete, we present several more well-known formulas of elementary geometry (see box above). Some of them are included in the high school curriculum.

The first of these formulas is the "extended" law of sines. The second relation is the law of cosines, which is a generalization of the Pythagorean theorem.

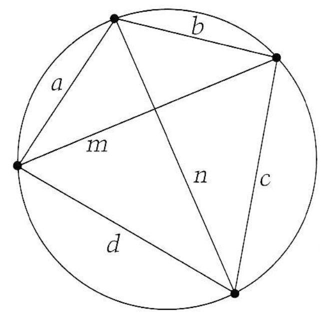
2. The law of cosines is the first representative of a broad class of cosine laws. Similar relations can be obtained for triangles in spherical and non-Euclidean geometry and for polygons, tetrahedrons, and other geometric figures. Rather often, several different relations can be obtained for the same geometric figure—for example, for a quadrilateral—that can be called a law of cosines. Bretschneider's theorem encompasses one such set of relations:

$$m^2n^2 = a^2c^2 + b^2d^2 - 2abcd\cos(\alpha + \beta).$$



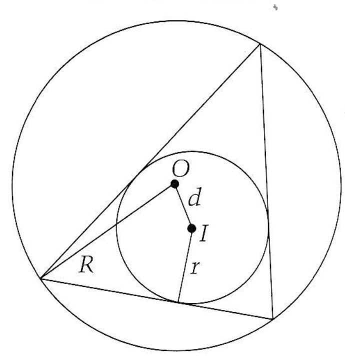
Bretschneider's theorem has many consequences. In particular, it implies Ptolemy's theorem for inscribed quadrilaterals:

$$mn = ac + bd.$$

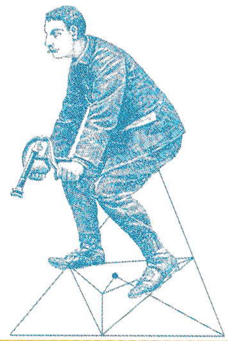


3. A triangle is usually specified by its three elements. For this reason, equations that relate three elements of the triangle are of particular interest. We have already listed several such relations in the box above. Let's give another remarkable equation called Euler's formula:

$$d^2 = R^2 - 2Rr.$$

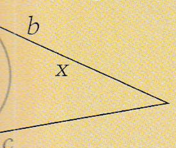


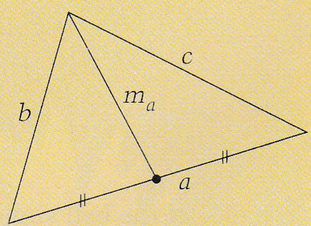
metric formulas

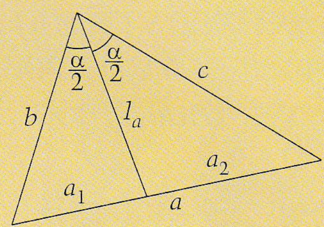


$$m_a^2 = \frac{1}{4}\left(2b^2 + 2c^2 - a^2\right),\qquad l_a^2 = bc - a_1a_2,\ l_a = \frac{2bc\cos\frac{\alpha}{2}}{b+c},$$

$s = \frac{1}{2}(a + b + c)$





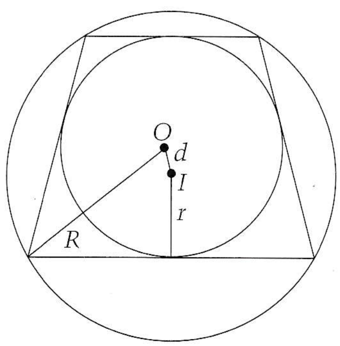


This formula expresses the distance between the centers of the inscribed and circumscribed circles of a triangle in terms of the radii of these circles. If these radii are known, the distance between the centers of the corresponding circles can be calculated. An infinite number of triangles exist for which these circles are inscribed and circumscribed circles, respectively. Any point on the larger circle can be taken as one of the vertices of such a triangle.

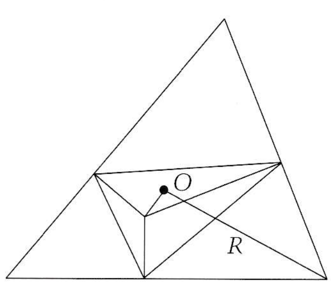
There is no similar formula for tetrahedrons, but there is a corresponding inequality. However, a corresponding equation, depending on the number of sides, can be written for any polygon that is circumscribed around a circle and inscribed in another.

A beautiful theorem of Poncelet (in projective geometry) implies that if an n-gon exists that is circumscribed around a circle and inscribed in another, then there is an infinite number of such n-gons, and that any point of the circumscribed circle can be taken as one of the vertices of such an n-gon. Here we write the equation for the inscribed-circumscribed quadrilateral:

$$\frac{1}{(R+d)^2} + \frac{1}{(R-d)^2} = \frac{1}{r^2}.$$



4. Many interesting geometrical figures are connected with the so-called *pedal triangle*. For any point chosen inside a given triangle, the pedal triangle of this point is a triangle whose vertices are the feet of the perpendiculars from the point to the sides of the given triangle (see figure below). Here is a remarkable



formula for the area of the pedal triangle (some authors give Euler credit for this formula):

$$Q = \frac{K}{4}\left|1 - \frac{d^2}{R^2}\right|.$$

Here Q, K, R, and d are, respectively, the area of the pedal triangle, the area of the given triangle, the radius of the circle circumscribed around the given triangle, and the distance from the given point to the center of the circumscribed circle. This formula has many interesting consequences. In particular, if the given point lies on the circumscribed circle, then the area of the pedal triangle is zero, and thus the corresponding points lie on the same line (the *Simpson line*). If we substitute d^2 in this equation from Euler's formula, we obtain a formula for the area of the triangle with vertices at the points where the inscribed circle touches the sides of the given triangle: $Sr/2R$.

5. Some formulas of plane geometry have analogues in space; others do not. However, there are remarkable relations for three-dimensional figures for which it would be hard to find plane analogues. Here is one such relation that is really amazing. Consider an arbitrary tetrahedron. Let its volume be V and let the radius of the circumscribed sphere be R. Consider three numbers that are equal to pairwise products of the opposite edges of the tetrahedron. It turns out that a triangle exists such that its sides are equal to these numbers, and the area of this triangle is

$$S = 6VR.$$

—I. F. Sharygin

PHYSICS CONTEST

Elevator physics

by Larry D. Kirkpatrick and Arthur Eisenkraft

One man's ceiling is another man's floor.
—Paul Simon

HAVE YOU TAKEN YOUR bathroom scales to an elevator for a ride as we suggested in the January/February 1998 issue of *Quantum*? This is guaranteed to start an interesting conversation, and you will be advancing the cause of physics at the same time.

Let's continue our exploration of physics in a uniformly accelerated reference frame by considering a ball of mass m dropped from a height h while the elevator has an acceleration a in the upward direction. Let's choose the upward direction as positive, start our stopwatch at the instant we drop the ball, and assume that the elevator has a speed v_0 at this time. Using the subscripts b and f for the ball and floor, respectively, we have the following equations for their positions in the laboratory frame of reference:

$$y_b = h + v_0 t - \frac{1}{2} g t^2$$

$$y_f = 0 + v_0 t + \frac{1}{2} a t^2.$$

Notice that even though we simply drop the ball in the elevator's reference frame, the ball has an upward velocity v_0 at this time. The ball is stationary relative to the floor.

We can find the length of time t_d for the ball to reach the floor by setting $y_b = y_f$. The time is

$$t_d = \sqrt{\frac{2h}{g+a}} \equiv \sqrt{\frac{2h}{g'}},$$

where we have defined $g' = g + a$. Notice that this is exactly the same formula that we would get if we dropped a ball while standing on the ground except that the normal acceleration due to gravity has been replaced by an "effective" acceleration due to gravity g'.

Let's now calculate the velocity of the ball relative to the floor just before the ball hits the floor. The equations for the velocities are

$$v_b = v_0 - gt,$$
$$v_f = v_0 + at.$$

The velocity of the ball relative to the floor is equal to the velocity of the ball minus the velocity of the floor. At the time the ball hits the floor, the relative velocity v_r is

$$v_r = v_b - v_f = (g + a)t_d = g't_d.$$

Once again this is the same result we would get on the ground using the effective acceleration due to gravity.

If we assume we are standing on the ground and drop an ideal ball on an ideal floor, the ball undergoes a completely elastic collision with the floor and returns to the height from which it was dropped. Conservation of kinetic energy tells us that the velocity of the ball simply reverses direction as it rebounds from the floor. In the more general case of a two-body elastic collision, it is the relative velocity of the two bodies that gets reversed.

Let's assume that we now return to the elevator and examine an elastic collision with the floor. We assume that we reset our stopwatch and choose the height of the floor to be zero. The velocities of the ball and floor (in the ground frame of reference) are now

$$v_b = v_h + v_r - gt = v_h + g't_d - gt,$$
$$v_f = v_h + at,$$

where v_h is the velocity of the floor at the time of the collision. We can find the time t_u when the ball reaches its maximum height above the elevator floor by setting the two velocities equal to each other. (After this time the floor is moving faster than the ball, and the ball approaches the floor again. It is falling.) Therefore,

$$g't_d - gt_u = at_u,$$

or

$$g't_d = (g + a)t_u = g't_u.$$

Once again we get the same result we would get on the ground—the time to fall is the same as the time to rise.

This analysis of the ball in free

Art by Tomas Bunk


...One Man's Ceiling Is Another Man's Floor
DOWN
UP
TOM BUNK
One Man's Floor Is Another Man's Ceiling

fall shows that we can simplify the calculations by assuming that we are standing on the ground if we replace the normal acceleration due to gravity by the effective acceleration. Although we have only shown that this is true for one problem, this is a general statement. In fact, one of the tenets of general relativity is the equivalence principle: *A constant acceleration is equivalent to a uniform gravitational field.* In general, we find the effective acceleration by taking the vector difference of the acceleration due to gravity and the acceleration of the system.

This leads us to our contest problems:

A. Show that the ball rises to the height from which it was dropped (relative to the floor of the elevator).

B. Argue that a ball dropped in a train moving with a constant horizontal acceleration a follows a straight line tilted at an angle with the vertical given by $\tan\theta = a/g$.

Leaf Turner developed our last problem for the first exam used to select the members of the 1999 U.S. Physics Team that will compete in Italy this summer.

C. A car accelerates uniformly from rest. Initially, its door is slightly ajar. Calculate how far the car travels before the door slams shut. Assume the door has a frictionless hinge, a uniform mass distribution, and a length L from front to back and that air resistance can be neglected.

If you are not convinced that using an equivalent acceleration due to gravity is the easiest way of solving this problem, try solving it in the inertial frame of reference attached to the ground.

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

Up, up and away

One solution to our September/October *Up, Up and Away* contest problem soared to the top of the pile. Congratulations to Dr. H. Leverett (a medical doctor from Texas) who loves physics and *Quantum*!

In part A of the problem, we ask to what temperature must a balloon's air be heated for the balloon to begin to float. Archimedes' principle informs us that the weight of the displaced air must be equal to the weight of the balloon for floating to occur:

$$m_1 g = m_a g + m_b g,$$

where m_1 is the mass of the displaced air, m_a is the mass of the air in the balloon, and m_b is the mass of the balloon material. Converting from masses to densities, we obtain

$$m_1 = \rho_1 V_1$$
$$m_a = \rho_a V_1$$
$$\rho_1 V_1 g = \rho_a V_1 g + m_b g$$
$$\rho_a = \rho_1 - \frac{m_b}{V_1}$$
$$\rho_a = 1.20\ \text{kg/m}^3 - \frac{0.187\ \text{kg}}{1.10\ \text{m}^3}$$
$$= 1.03\ \text{kg/m}^3.$$

This is the required density of air in the balloon for the balloon to float. Using the ideal gas law, we can find the corresponding temperature.

$$\rho_a T_a = \rho_1 T_1$$
$$T_a = \frac{\rho_1 T_1}{\rho_a} = \frac{\left(1.20\ \text{kg/m}^3\right)(293\ \text{K})}{1.03\ \text{kg/m}^3}$$
$$= 341\ \text{K}.$$

In part B, the air is heated to 110°C (383 K), and the balloon rises isothermally into the atmosphere, which has a constant temperature of 20°C (293 K). The net force acting on the balloon F_u is equal to the difference between the buoyant force F_B and the weight of the balloon F_W:

$$F_U = F_b - F_W$$
$$F_b = \rho_1 V_1 g$$
$$F_W = m_3 g + m_b g = \rho_3 V_1 g + m_b g$$
$$F_U = \left[(\rho_1 - \rho_3)V_1 - m_b\right]g,$$

where m_3 and ρ_3 are the new mass and density at this temperature.

The density of the air within the heated balloon can be found, once again, from the ideal gas law:

$$\rho_3 T_3 = \rho_1 T_1$$
$$\rho_3 = \frac{\rho_1 T_1}{T_3} = \frac{\left(1.20\ \text{kg/m}^3\right)(293\ \text{K})}{383\ \text{K}}$$
$$= 0.918\ \text{kg/m}^3$$
$$F_U = \left[\left(1.20\ \text{kg/m}^3 - 0.918\ \text{kg/m}^3\right)\right.$$
$$\left.\times\left(1.10\ \text{m}^3\right) - 0.187\ \text{kg}\right]9.80\ \text{m/s}^2$$
$$F_U = 1.21\ \text{N}.$$

Part C requires us to find the height to which the balloon rises. The balloon continues to rise until its buoyancy is equal to its weight:

$$F_B = \rho_y V_y g,$$

where ρ_y is the density of the displaced air at the balloon's height y. We have

$$\rho_y V_y g = \rho_3 V_3 g + m_B g$$
$$\rho_y = \rho_3 + \frac{m_B}{V_y}$$
$$\rho_y = 0.918\ \text{kg/m}^3 + \frac{0.187\ \text{kg}}{1.10\ \text{m}^3}$$
$$= 1.088\ \text{kg/m}^3.$$

Given the relationship between the air density and the height, we can now find how high the balloon will go:

$$\rho_y = \rho_1 e^{-\frac{g\rho_1}{P_0}y}$$
$$y = -\frac{P_0}{\rho_1 g}\ln\frac{\rho_y}{\rho_1}$$
$$y = \frac{-1.013\cdot 10^5\ \text{N/m}^3}{\left(1.20\ \text{kg/m}^3\right)\left(9.80\ \text{m/s}^2\right)}$$
$$\times \ln\frac{1.088\ \text{kg/m}^3}{1.20\ \text{kg/m}^3} = 844\ \text{m}.$$

In part D, we are asked to describe the subsequent motion if the balloon were pulled from its equilibrium position by 10 m and then released. We can see that if the balloon is below the position calculated in part C, then the balloon will rise. If it goes above this position, then it will sink. The balloon will therefore undergo harmonic motion. If air resistance is taken into account, the motion will be damped harmonic motion.

Diffraction in laser light

by D. Panenko

LASERS ARE SOURCES OF LIGHT with wonderfully high coherence. This is why they are used to obtain stable interference patterns and to observe the fine diffraction phenomena that cannot be produced by nonlaser light sources.

The following explains how to carry out diffraction experiments with cylinders, balls, or other bodies using a laser light source. We will begin by outlining the experimental setup.

Figure 1 shows the layout of the setup to project enlarged Fresnel diffraction patterns of various obstacles. It consists of a laser, a microscope objective M, a pinhole diaphragm S, and an object O that produces the diffraction pattern observed at the plane P_0P. At this plane the experimenters can set up a white screen for observing the pattern with the unaided eye, a camera without an objective lens to photograph the pattern, or a photodetector to record the interference bands.

The photodetector is shielded from the light source by a narrow slit or a pinhole diaphragm with a round orifice. It is mounted on a movable carriage, which shifts it uniformly together with the diaphragm along the y-axis in the observation plane. The signal of the photodetector is proportional to the intensity of the light that passes through the diaphragm. It is recorded by a plotter. The plotter has a carriage that moves with constant velocity along the x-axis and is deflected along the y-axis by the signal from the photodetector. Thus, the device plots (in the chosen scale) the distribution of illumination in the diffraction pattern as a function of the distance P_0P. The length of the recorded pattern is 35 mm.

In the first experiment we obtained a diffraction pattern with a polished metal cylinder 2 mm in diameter. The axis of this cylinder was perpendicular to the plane xy, and the distances were SO = 1 mm, OP_0 = 2.5 mm, and P_0P = 35 mm (fig. 1). In front of the photodetector was a slit of width 0.05 mm parallel to the cylinder's axis.

Figures 2 and 3a show the results of the experiment. The symmetrical diffraction pattern is not uniform, and we can distinguish two kinds of interference bands of different origin. At the edge and away from the geometric shadow, there is a clear-cut conventional edge diffraction pattern—that is, the alternating dark and bright bands. A specific feature of this pattern is decreasing space between the bands away from the axis of symmetry. In addition to this pattern, there are equally spaced bands (also parallel to the cylinder's axis) located within the geometric shadow and just outside of it. Both kinds of diffraction can be observed separately and analyzed individually.

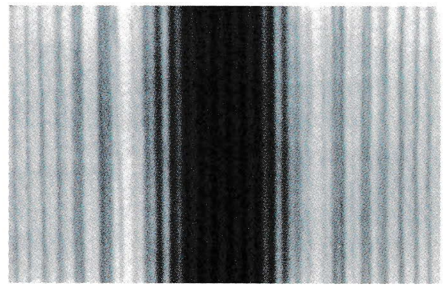


Figure 2. *Diffraction by a cylinder.*

If we cover the cylinder on one side with an opaque screen, we obtain diffraction only from the uncovered side of the cylinder (fig. 4). The resulting distribution of illumination is shown in figure 3b. This is a typical pattern of edge diffraction, which is a set of bands whose width and contrast decrease with distance from the geometric shadow of the edge.

Now we consider the second diffraction pattern composed of equidistant bands parallel to the cylinder's axis and located in the

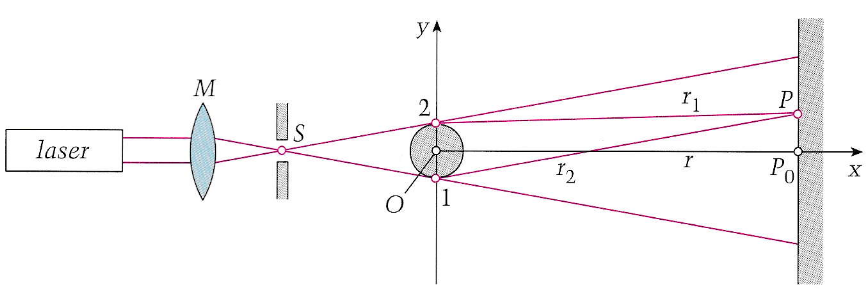


Figure 1

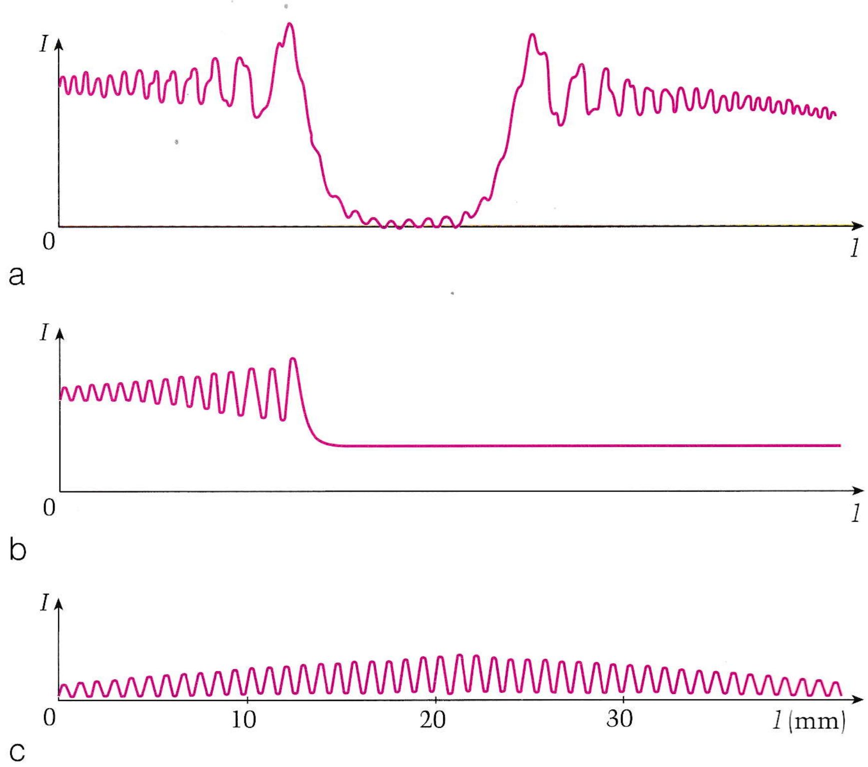


Figure 3. *Distribution of illumination in diffraction by a cylinder (a), on the edge of a cylinder (b), and on two slits at the edge of a cylinder (c).*

region of the geometric shadow. The nature of these bands can be explained by Thomas Young's (1773–1829) double slit interference. The regions of the wave front very near the cylinder's surface are the sources of the secondary waves. The interference of these waves produces the diffraction pattern in the geometric shadow region behind the cylinder.

To carry out such an experiment (which is similar to Young's double slit interference experiment), the rod was placed between the blades forming slit *S*. The width of the slit was chosen such that two openings 0.1 mm in width were formed between the cylinder's surface and the blades of the slit. When this arrangement was illuminated with laser light, two bright lines flashed at the cylinder's sides, which produced the Young's diffraction pattern. Figure 5 shows that this pattern is composed of equidistant bands. The distribution of illumination in this pattern is shown in figure 3c.

Thus, diffraction by a cylinder can be considered to be the addition of two elementary patterns formed by diffraction at both edges of the cylinder and by Young's double slit diffraction. (When comparing the plots in figure 3, we must note that the curves (a) and (b) were recorded at the same sensitivity, while curve (c) was obtained at a higher sensitivity.) Usually, Fresnel diffraction by a wire is associated only with a bright band along the axis of the geometric shadow. We see that coherent laser light reveals much more detail in the diffraction pattern.

The second experiment was conducted with a ball 2.4 mm in diameter. It was accurately attached by plasticine to a plane-parallel glass plate and placed into the divergent beam of light (fig. 1). A screen with a round orifice 0.1 mm in diameter was placed in front of the camera.

The results of this experiment are shown in figures 6 and 7, which clearly demonstrate two kinds of diffraction patterns as in the experiment with a cylinder. However, we can see a distinct bright point in the center of the geometric shadow, known as the Arago-Poisson spot. The history of this spot is very instructive and interesting.

The problem of light diffraction was announced by the French Academy of Sciences as a subject of competition for 1818. Most of the organizers of this competition adhered to the corpuscular theory of light, and they expected that new competitive papers would mark the final victory of their preferred theory. However, Augustin-Jean Fresnel (1788–1827) submitted a paper in which he explained all the known optical phenomena on the basis of the wave theory of light. Reading his memoir, the famous French scientist Siméon-Denis Poisson (1781–1840), a member of committee, saw that Fresnel's theory produced an "absurd" conclusion: There must be a bright spot in the center of the shadow cast by a small, round, opaque disk. This theoretical prediction moved François Arago (1786–1853) to carry out the corresponding experiment in

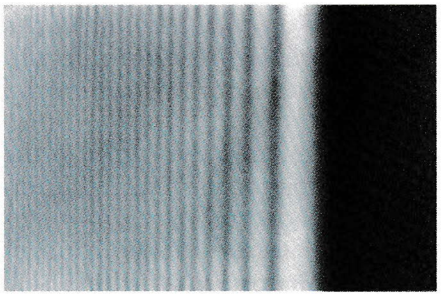


Figure 4. *Diffraction by an edge of a cylinder.*

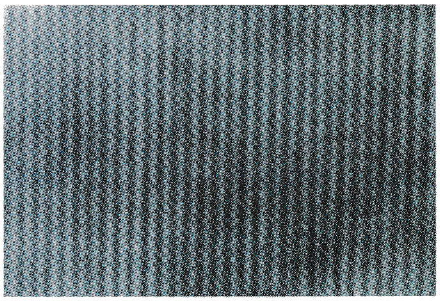


Figure 5. *Two-slit interference bands (Young's experiment).*

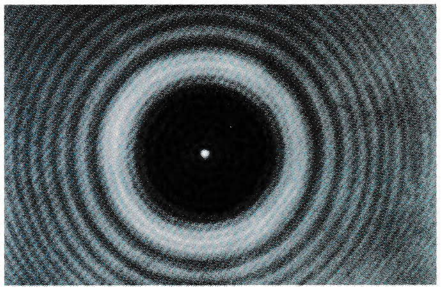


Figure 6. *Diffraction by a ball.*

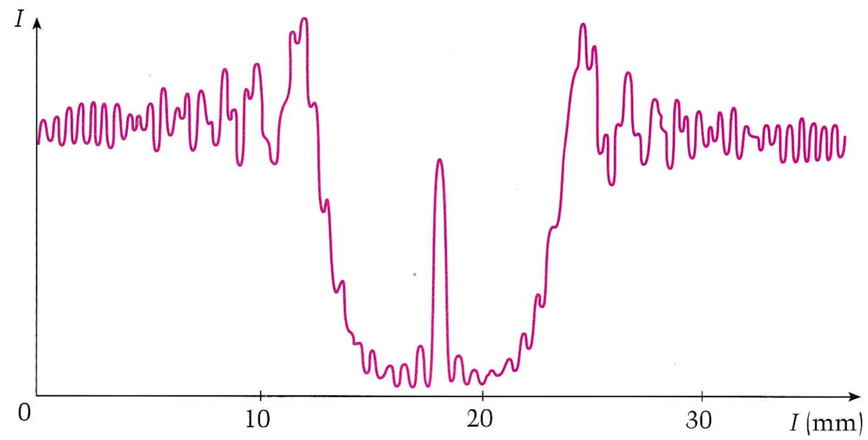


Figure 7. *Distribution of illumination for diffraction by a ball.*

which he observed a bright spot in the very center of a dark round shadow. This experiment was crucial to the acceptance of the wave theory of light.

Experiments with powerful and coherent light generated by a laser make it possible to observe rings both inside and outside the geometric shadow. The nature of these rings can also be explained along the same lines that are used to interpret the Young's double slit interference, considering that the interference pattern in the geometrical shadow is formed by an infinite number of diametrically opposite sources of light from the wave front that interacts with the ball's surface. Thus, observations with powerful and coherent laser-generated light yield fundamental knowledge of Fresnel diffraction caused by simple-shaped bodies.

Quantum on light interference:

P. V. Bliokh, "Make yourself useful, Diana", March/April 1992, pp. 34–39.

A. Eisenkraft and L. D. Kirkpatrick, "Rising Star", March/April 1995, pp. 37–38.

A. Eisenkraft and L. D. Kirkpatrick, "Color Creation", November/December 1997, pp. 32–33.

V. A. Fabrikant, "Vavilov's Paradox", July/August 1992, pp. 49–50.

The century of the cycloid

by S. G. Gindikin

THE SEVENTEENTH CENTURY was the golden age of the infinitesimal calculus, and the curve called a *cycloid* played a unique role in this era. Oddly enough, the cycloid has also come down to us in literature.

In the third part of *Gulliver's Travels*, Jonathan Swift describes Gulliver finding himself on the floating island of Laputa. The King of Laputa, a patron of the sciences, treats him to an exquisite dinner, where a mutton shoulder in the shape of an equilateral triangle is served. This geometric entrée is followed by a pudding in the shape of a cycloid. It was probably difficult to imagine a more intricate shape than the cycloid at the beginning of the seventeenth century.

About 30 years after the publication of *Gulliver's Travels*, the cycloid appeared in the novel *Tristram Shandy*, by Laurence Sterne. In it Toby Shandy, an eccentric, good-natured man, wants to create an engineering marvel. He decides to build a bridge in the shape of a cycloid, as described in the journal *Acta Eruditorum* in 1695: ". . . a lead weight is an eternal ballance, and keeps watch as well as a couple of centinels, inasmuch as the construction of them was a curve-line approximating to a cycloid,—if not a cycloid itself." Let's look at what Toby Shandy's contemporaries in the middle of the seventeenth century knew about the cycloid.

The *Histoire de la Roulette, Appellée Autrement Trochoïde ou Cycloïde* (*The History of the Roulette, Also Called the Trochoid or Cycloid*) (1658), by Amos Dettonville, is a good source of information. Amos Dettonville was actually a pen name used by Blaise Pascal (1623–1662), the French mathematician, physicist, and philosopher. Pascal wrote that the roulette is a curve so common that no figure besides the straight line and circle is encountered more often. He said that it appears before our eyes so often that it is surprising the ancients didn't notice it.

The roulette is simply the path traveled by a nail on the rim of a rolling wheel, from the moment the nail leaves the ground until it meets the ground again. It's assumed that the wheel is an ideal circle and the ground is an ideal plane. This is also the definition of a cycloid: The trajectory of a point on the circumference of a circle rolling along a line without slipping. Ascending from its lowest position, the point describes a symmetrical convex arc. Each revolution of the circle produces a new arc, and there is a cusp (a place at which the point changes direction) where the arcs touch each other (fig. 1).

The term *cycloid* (from the word *circle*) was introduced by Galileo (1564–1624), who was probably the first to notice this curve. *Roulette* is the French term (from the verb *rouler*, which means "to roll"), and *trochoid* is the corresponding Greek term. The difference in terminology reflects the disagreement between the French and Italians about who discovered the cycloid. Pascal argued that Father Marin Mersenne (1588–1648) discovered the curve in 1615, before Galileo.

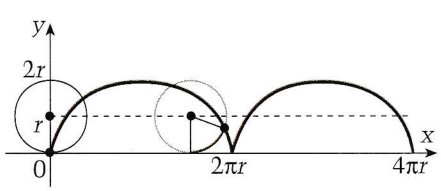


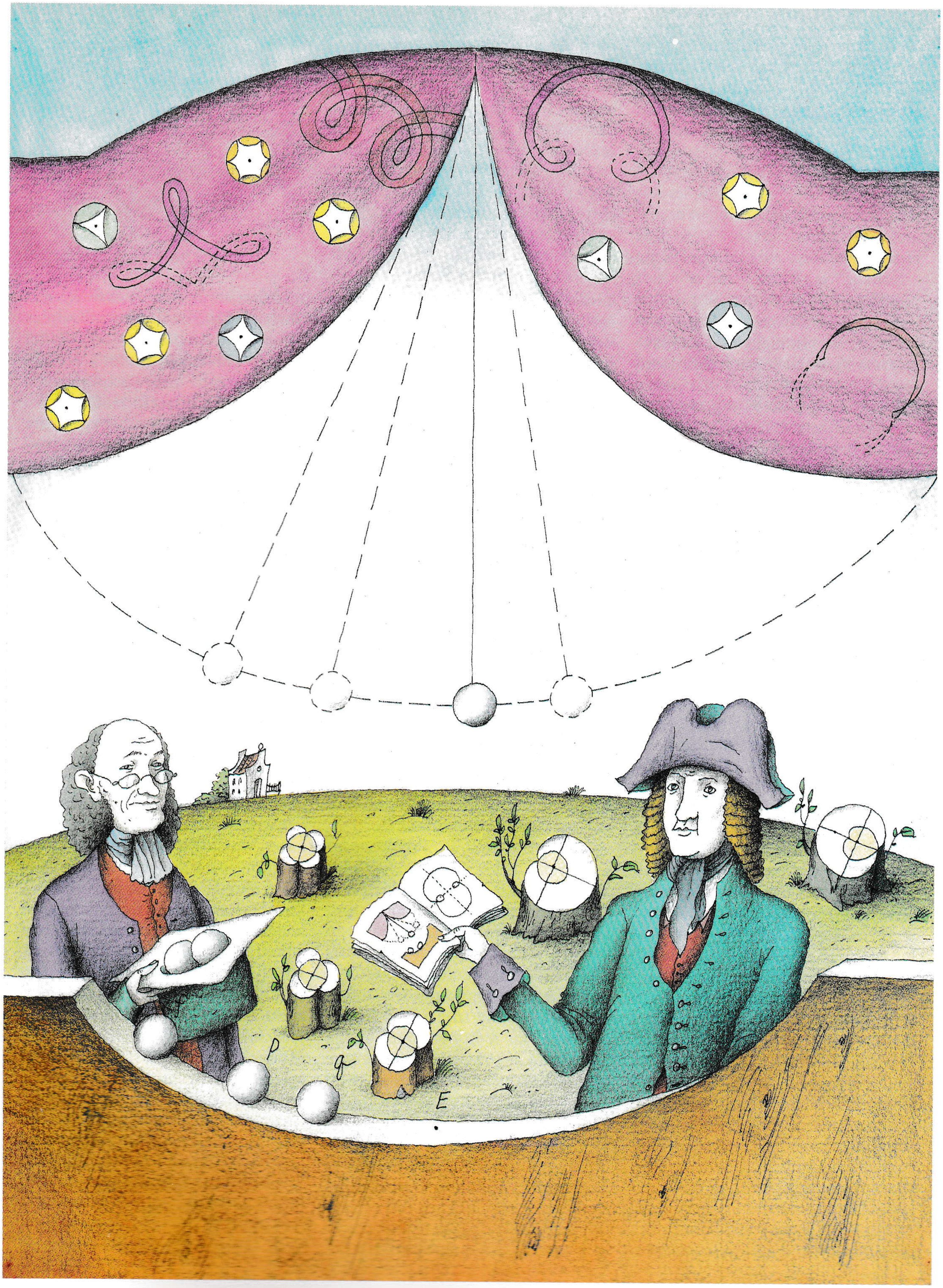
Figure 1

Epicycloids in the Ptolemaic system

So why didn't the "ancients" ever notice the cycloid? Their source of new curves was limited to problems about geometric loci, and trajectories of curvilinear motion were a bit beyond the scope of most ancient mathematicians. It was the new mechanics, which began with Galileo's work during the first years of the seventeenth century, that turned mathematicians' attention to curves of mechanical origin. Galileo introduced the first such curve—the cycloid.

Although the ancients did not know the cycloid, they knew and successfully used its close relative, the epicycloid. An epicycloid (figs. 2 and 3) can be obtained as a trajectory of a point of a circle that rolls without slipping on the outside of a second stationary circle (if the circle rolls along the inside of the second circle, a *hypocycloid* is obtained—see fig. 4). We encounter epicycloids in the geocentric (Earth-centered) model of the Solar System, elabo-

Art by V. Ivanyuk


p
q
E

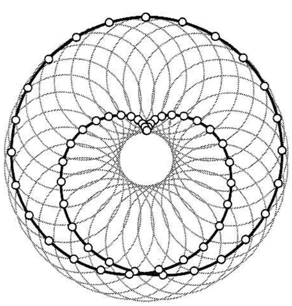


Figure 2

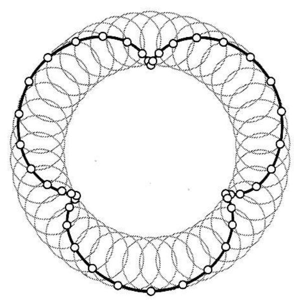


Figure 3

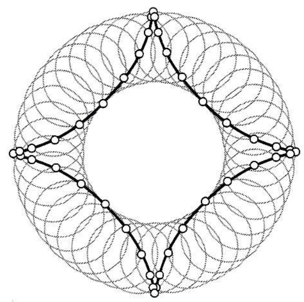


Figure 4

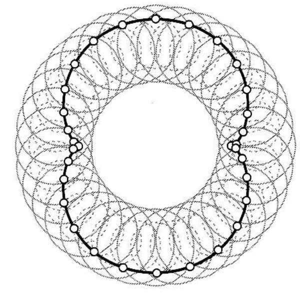


Figure 5

rated on by Ptolemy (A.D. 85?–165?).

Plato (427–347 B.C.) and Aristotle (384–322 B.C.) thought that all planets, as well as the Sun and the Moon, revolved about Earth. This theory not only contradicted the available numerical data, but also could not explain many qualitative effects. For example, Mars, which usually moves counterclockwise in our sky, sometimes moves clockwise (the so-called *retrograde motion*).

This fact can easily be explained in the context of the heliocentric (Sun-centered) system. However, the ancient scientists tried to reconcile data from astronomical observations with the geocentric system. Appollonius and then Ptolemy retained the axiom that uniform circular motion dominated the Universe. However, they argued that the motion of planets was complex: Each planet moves on a circle whose center, in turn, revolves on a large circle about the Earth. It turns out that it's possible to choose the speeds of these two rotations so that the planets' trajectories contain loops where retrograde motion occurs. These trajectories are not exact epicycloids, but are very close. If we roll a circle along the outside of another, stationary circle, a point on its boundary will describe an epicycloid, and a point on its interior will describe another curve, called a *curtailed epicycloid*, whose cusps are smoothed (fig.5). If a stick is attached to the rolling circle, the endpoint of the stick will describe a *prolate epicycloid*, whose cusps have become loops along which the stick's endpoint moves retrograde (fig. 6).

This ingenious theory not only accounted for qualitative effects such as retrograde motion, but also made it possible to account for astronomical observations made over hundreds of years. When Nicolaus Copernicus (1473–1543) suggested his heliocentric system, he had difficulty competing with the existing versions of the Ptolemaic system in accounting for astronomical observations.

It's interesting that Copernicus (and later, Galileo) did not reject the principle of uniform circular motion and retained circular orbits (which were later replaced with elliptical orbits by Johannes Kepler (1571–1630)). In particular, he wanted to construct rectilinear motion from circular motions (it was believed then that comets moved in straight lines). He was able to devise a method: Let a circumference of radius 1/2 roll on the interior side of another circumference of radius 1. Then its points describe certain degenerate hypocycloids that are identical to the diameters of the stationary circle—each point moves back and forth on its diameter. I encourage the reader to construct this beautiful, intriguing picture.

The cycloid is seen

We do not know exactly when Galileo became interested in the cycloid. Was it in his youth at the beginning of the century, when he discovered the laws of motion, or 30 years later, when he was writing his *Dialogues Concerning Two New Sciences* while struggling with increasing blindness? This book was intended to record his remarkable though yet unpublished discoveries in mechanics. At that time, Galileo was serving the sentence of the Inquisition at his villa at Arcetri, near Florence. Two of Galileo's disciples assisted him in his last years of life: young Vincenzo Viviani (1622–1647) and the more experienced Evangelista Torricelli (1608–1647). They helped their teacher complete his projects, and they were compelled to develop Galileo's ideas themselves, because Galileo hadn't the strength to do it all himself.

Galileo understood very well which problems concerning the cycloid needed to be solved first. These were the problems of constructing the tangent and of finding the area under the arch of the cycloid and the various related curvilinear figures. In modern terminology, the first problem pertains to differential calculus and the second to integral calculus.

Tangent to the cycloid

In Italy, Viviani was the first to construct the tangent to the cycloid. Later, Torricelli devised an elegant method for constructing the tangent, based on combining motions. First, he considered a parabola that is described by a body thrown horizontally (or at an angle to the horizon). The motion of this body is a combination of uniform rectilinear motion and uniformly accelerated free fall. The velocities of these motions are known, so adding them by the parallelogram rule, we obtain the velocity of the combined motion directed along the tangent to the parabola. It is very instructive to repeat this calculation, starting with the fact that the tangent to the parabola $y = x^2$ at point (x, x^2) passes through $(x/2, 0)$. It's surprising that Galileo did not find this construc-

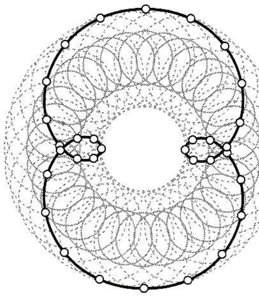


Figure 6

tion (or did he just not publish it?).

Motion along a cycloid is a combination of uniform horizontal rectilinear motion and uniform rotation. The speeds of these motions are equal (because the circle rolls without slipping). Combining the horizontal velocity with the rotational velocity vector, which is tangent to the generating circle and has the same length, we see that the tangent to the cycloid passes through the upper point of the generating circle in the corresponding position (try to reproduce this construction in fig. 1).

A similar construction was found (perhaps a little earlier) by a Frenchman, Gilles Personne de Roberval (1602–1675). Much credit should go to Roberval for developing a general technique for drawing tangents by treating a curve as a trajectory of a combined motion. His method was applicable to the cycloid and other shapes, and it successfully competed with other methods available at the time.

However, the future lay in more direct methods being developed by Pierre Fermat (1601–1665) and René Descartes (1596–1650). These methods did not require an individual approach to every curve (that is, there was no need to adjust the component motions). It was interesting to see whether these methods could be applied to the construction of tangents to curves of mechanical origin—for example, to the cycloid.

It turned out that the tangent to the cycloid could easily be constructed by Fermat's method, where Descartes's method, which worked well for polynomial curves, failed. However, Descartes was not a man to give up easily. He devised an elegant mechanical theory. Instead of having a circle rolling on a line, he suggested considering a disk. In this case, at each moment in time the motion is very close to the rotation about a certain (instantaneous) center of rotation. The velocity of each point is perpendicular to the radius drawn from the center of rotation to this point. In our case, the instantaneous center of rotation is the lowest point of the rolling disc at which it touches the directional line. The result of this line of thinking is a rule for drawing the tangent to a cycloid.

Area and the cycloid's companion

Let's turn our attention to areas. For the cycloid, French mathematicians considered problems concerning areas before those of tangents (probably because they felt more confident with areas). The first problem was to calculate the area under the cycloid's arc. Viviani and Torricelli argued that Galileo knew this area equals three times the area of the generating circle, and Roberval proved this theorem in 1634. Consider in brief his elegant reasoning.

At each moment in time, Roberval projects the observed point of the cycloid onto the vertical diameter of the rolling circle and watches the change of this projection. The curve described by this projection was called the cycloid's companion (fig. 7). Later it turned out that this symmetrical curve is nothing more than a shifted sine curve! It was in this way, and not as a plot of the sine function, that the sine curve first appeared in mathematics. The area under this curve can easily be calculated. If we cut it along the horizontal and vertical middle lines, the four components obtained compose a rectangle with sides $2\pi r$ and r (see fig. 7, where $r = 1$), Therefore, the area under the sine curve is $2\pi r^2$.

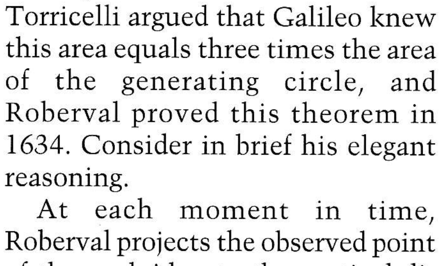


Figure 7

What about the sectors between the cycloid and its "companion"? These remaining parts were called d *Roberval's petals*, and they illustrate a method for calculating areas that was very popular at the time—Cavalieri's principle. This principle states that if two figures give equal segments when cut by an arbitrary horizontal line, their areas are also equal. Cavalieri considered figures to be composed of "indivisibles"—parallel layers of zero thickness—and stated that the area does not change when these indivisibles are shifted. Roberval's petals have the same intersection with horizontal as the half of a circle cut by its vertical diameter. Therefore, by Cavalieri's principle, their total area is the same as the area of the circle—that is, πr^2. Thus, the total area under the cycloid arc is $3\pi r^2$. An ingenious solution, isn't it?

Mersenne's problems and Pascal's competition

Thus we come to the end of the first stage of the cycloid's "life." The first problems concerning this curve had been solved elegantly, and the cycloid had become firmly established in mathematics. However, many problems that naturally occurred in the context of developing the infinitesimal calculus remained unsolved. The Franciscan monk Father Marin Mersenne (whom we mentioned earlier) played a major role in discussing new problems. Although Mersenne had several achievements in mathematics and physics to his name (for example, he measured the speed of sound rather accurately), his role in the organization of science was much more important.

At the time there were no scientific journals, and Mersenne served as a communication link between scientists throughout Europe. Scientists from different countries reported their results to Mersenne, and he communicated the results to other interested parties (his prestige guaranteed their full attention). Mersenne kept track of scientific problems and communicated them to scientists who could better solve them. Starting in 1639 he organized weekly Thursday meetings, which were attended by Etienne and Blaise Pascal (father and son), Girard Desargues, Claude Mydorge, Roberval, and others. These "Thursdays" were precursors to the creation of the French Academy of Sciences.

We can see how Mersenne nurtured the growth of young talent from his relationship with Christiaan Huygens (1629–1695). Mersenne started corresponding with Huygens in 1646. First he posed training problems, then unsolved ones, and finally he presented to Huygens the problem of the effective length of the compound pendulum. It took Huygens several decades to solve this problem.

As for the cycloid, Mersenne reasoned that its study should not be restricted to calculating the area under the entire arc, but that segments cut by different horizontal lines should be considered as well: The center of gravity and volume of the bodies of revolution generated by these segments should be found, and so on. Mersenne probably understood that with these problems, elementary manipulations with Cavalieri's indivisible layers could not give the result: The integral of the sine must be calculated in its general form. For this reason, Mersenne wanted Pascal, the most ingenious of his colleagues, to work on these problems.

However, Pascal turned to them only in spring 1658, when he was living in a convent in Port-Royal and it seemed that he would never do mathematics again. According to Pascal's sister Gilberte Perier and his niece Marguerite, he turned to mathematics when he couldn't get to sleep because of a terrible toothache. Meditating on the cycloid distracted him from the toothache and soon took up all his attention. By morning he knew the solution to the problem, and the toothache was gone. Pascal did not want to write down his result, but his friend the Duke de Roanne persuaded him to do it, saying that his return to mathematics was inspired from the heavens. The Duke was in turn inspired to donate 60 *pistoles* to organize a competition for solving Mersenne's problems.

In June 1658 a letter containing six problems concerning the cycloid was sent to Europe's most prominent mathematicians. The deadline was tight—the first of October. The letter was signed Amos Dettonville, an anagram of Pascal's pen name Louis de Montalte, under which his *Lettres a un Provincial* were published. On October 24 the results were declared. Mersenne died before his problems were solved, and his position as chairman of the jury was filled by Pierre Carcavi (1603–1684). John Wallis (1616–1703) solved all the problems, but there were objections to his solutions. Huygens solved four problems, and the prize was awarded to Dettonville, who used the prize money to publish the solutions.

The competition played an important role in scientific life. It was of major importance for the future of the infinitesimal calculus to pass from considering particular cases to general problems. The method suggested by Pascal did not use any specific features of the cycloid and could be extended to more general cases. However, Pascal did not do it himself, and only Gottfried Wilhelm Leibniz (1646–1716) appreciated this aspect of the work.

Leibniz, who developed with Sir Isaac Newton (1643–1727) the general methods of differential and integral calculus, was surprised that Pascal did not expand on his method of analyzing the cycloid to more general cases himself. It's hard to guess why. Often scientists don't immediately see solutions that appear quite natural later on. On the other hand, we can also suppose that he wasn't interested in mathematics anymore. During the last years of his life he was thinking intensively about the purpose of life and the place of humans on Earth.

Pascal's competition opened a new stage in the study of the cycloid. The participants did not restrict themselves to the problems suggested by Pascal. For example, Christopher Wren (1632–1723), a talented English mathematician and famous architect (the designer of St. Paul's Cathedral in London), didn't do very well in the competition overall. However, he did calculate the length of the cycloid's arc. He proved that this length is $8r$, and this result impressed mathematicians greatly.

The clock with the cycloidal pendulum

Another participant in the competition, the 28-year-old Huygens, digressed from his chief project—constructing the pendulum clock—to solve the cycloid problems. The first model of his clock appeared in 1658, and afterward, Huygens was dedicated to improving it.

The idea underlying the pendulum clock is that the period of pendulum oscillations is a reproducible unit of time that remains unchanged when the oscillations damp. This is the isochronous property of the pendulum discovered by Galileo (the period of oscillation is independent of the range of oscillation).

In the seventeenth century, the most important unsolved scientific puzzle was the problem of measuring longitude at sea. From ancient times, it was known how to measure latitude by looking at the positions of the stars. But the stars rotate from east to west, and can give no clue as to a ship's longitude. Leading naval powers offered huge amounts of prize money to anyone who could give them a reliable method of determining longitude at sea. If, for example, an accurate chronometer (clock) could be constructed that worked well on a rolling ship, then such a chronometer could be set to the time at the ship's port of departure. Local time could be deter-

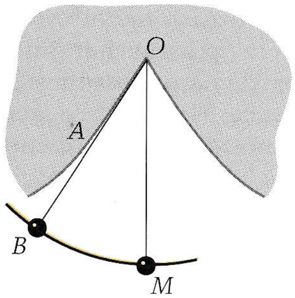


Figure 8

mined by observing the Sun or the stars, and the difference between these two would give the ship's longitude. (Today we would say that we were determinig the time zone in which the ship would be located, but there was no such global system in the seventeenth century.) Huygens intended to use his knowledge of the cycloid to construct such a chronometer.

Soon after Huygens's invention became known, it emerged that Galileo had also hit upon the idea of a pendulum clock, but only a year before his death, when he had no strength left to implement the idea. His son Vincenco, who was to finish the project after his father's death, could not solve this problem either.

In creating the first pendulum clock, Huygens discovered that Galileo's assertion that pendulum oscillations are isochronous was not completely correct. It holds only for small oscillations. Then how could he guarantee the isochronism of oscillations? Huygens knew that the length of the pendulum must be decreased as the pendulum's swing carries it away from the vertical line. But what was the exact relation between the position of the pendulum and the amount by which its length should be decreased? For his first clock, Huygens made cams that constrained the pendulum as it swung (fig. 8). He found a shape that worked by trial and error. However, he did not know what the exact shape should be. In despair, he removed the cams from his 1658 model and replaced them with an amplitude limiter. However, a year later, he reintroduced the cams. This time he knew exactly what their shape should be in order to guarantee the isochronism of oscillations. (Recall that in the meantime he had taken part in Pascal's competition.)

Galileo understood very well that the problem of the simple oscillation of a pendulum could be transformed. Imagine a heavy point particle, like a marble, rolling down a slide or chute, with the shape of the chute coinciding with the pendulum's trajectory. Then the physical constraints imposed on the point particle in the chute are identical to those imposed on the pendulum. The isochronism condition is equivalent to the condition that the particle arrives at the lowest point of the chute at the same moment in time, independent of the starting point. Such an effect is not surprising: If the particle starts from a higher point, it travels farther but picks up more speed. However, Huygens found out that, contrary to Galileo's opinion, the circular chute is not isochronous. Huygens sought the shape of the chute such that the descent time is independent of the starting point. He called such a curve a *tautochrone* (it is also called an *isochrone*). It turns out that a tautochrone curve is just an inverted cycloid.

However, the problem of the isochronous pendulum was not yet solved. It was necessary to find the shape of the cams that guaranteed that the pendulum's weight would move along a cycloid. Earlier, Huygens had studied the *developments* of various curves. Suppose a string is wrapped tightly along a curve, and one end is pulled away from the curve, so that the string "peels off" the curve. The path traced by the pulled endpoint of the string is called the curve's *development*. Studying the developments of various curves helped Huygens understand that if the pendulum's weight is to move along a cycloid, the cams themselves should be cycloidal in shape. When the pendulum's string wraps onto the cam up to its endpoint, this endpoint is at the lowest point (the turning point) of the cam. This result is known as Wren's theorem, and is usually expressed by saying that the perimeter of the cam is twice as long as the length of the pendulum.

Huygens was sure that the properties of the cycloid that he had discovered were of fundamental importance. He wrote that he had to consolidate and amplify the studies of Galileo on falling bodies and he felt that the discovery of these properties of the cycloid were the greatest achievements of this area of study.

In 1661 the test of the nautical version of the pendulum clock began, but it did not lead to the construction of a reliable sea chronometer. However, the studies in mathematics and mechanics that Huygens conducted in connection with the pendulum clock were so important that his book *Horologium Oscillatorium* (*The Pendulum Clock*) (1673) is one of the major texts in classical mechanics, along with the books of Galileo and Newton. In *Horologium Oscillatorium*, Huygens stated that the cycloid's properties had been studied better than those of any other curve.

The last mystery of the cycloid

The cycloid surprised mathematicians once more at the end of the century of the infinitesimal calculus. Newton was the first to understand the necessity of moving from individual problems to the construction of a general method. During the plague years in Lincolnshire, he essentially invented the calculus (his "method of fluxions"). However, he did not publish it at once, but used it to obtain many varied results. Many of these results appeared in the pages of his *Philosophiae Naturalis Principia Mathematica* (*Mathematical Principles of Natural Philosophy*) (1687), but the method was not presented in detail there.

In the early 1670s, Leibniz, as a distraction from his numerous state and scientific occupations, began to develop the formalism of the differential and integral calculus (under the strong influence of Pascal). He found out about Newton's work and started corresponding with him in 1676. In contrast to Newton, Leibniz actively called attention to his method using both his personal contacts and the journal *Acta Eruditorum* (as you remember, this was the journal read by Toby Shandy), which had been published since 1682.

In 1696 the notice "A New Problem that All Mathematicians Are Invited to Solve" appeared in *Acta Eruditorium*. It was presented as follows: "Two points A and B are given in a vertical plane. Find the trajectory AMB such that a body M descends from A to B along this trajectory in the shortest possible time moving under gravity." The problem gained popularity. Leibniz wrote that it was very elegant and quite new. However, in his *Dialogues Concerning Two New Sciences*, Galileo discussed the fact that a heavy particle descends along a segment of line slower than along a broken line with the same endpoints. He was quite sure that a quarter of a circle was the curve of the quickest descent. The curve of quickest descent was named the *brachistochrone*. But was it a segment of a circle?

The timeliness of the quickest descent problem manifested itself in the fact that it was solved rather quickly by a number of mathematicians: both brothers Bernoulli—Jacob (1654–1705) and Johann (1667–1748), Leibniz, Guillaume de l'Hôpital (1661–1704), and Newton (who presented an anonymous solution, but his genius betrays itself in the work). The solution by Johann Bernoulli turned out to be the most interesting. His idea was quite unexpected. He suggested replacing the mechanical problem with an optical one.

It turns out that if we assume that the speed of light at a point M is equal to $\sqrt{2gy}$, where y is the difference between the height of points A and M (and g is the acceleration due to gravity), then a ray of light passing from A to B travels along the brachistochrone. (Fermat had already established the principle that light traveling between two points seeks a path that requires the minimum time to traverse.) Then, using the laws of reflection and properties of the tangent to the cycloid (by then well understood), one can show that the brachistochrone is just an inverted cyloid with its cusp at point A!

One historian of mechanics wrote that Johann Bernoulli found his elegant solution without any general method, using only his geometric intuition and what was known in optics at that time. Thus, his artistic nature manifested itself in science. His brother Jacob was quite another kind of scientist. He had much less creative imagination, but strong critical ability. His solution was much more unwieldy, but he did not miss a chance to substantially develop a general method for solving problems of this kind. Thus, each brother possessed one aspect of scientific talent, which is combined in great scientists such as Newton.

The heroic history of the cycloid came to an end with the end of the seventeenth century. The cycloid mysteriously occurred in various problems, and nobody doubted that it plays a very important role in nature. However, some time later it became clear that the cycloid is not connected to the fundamental laws of nature as, for example, are conic sections. Problems that gave rise to the cycloid played an important role in the development of mechanics and calculus, but after these sciences were sufficiently developed, it turned out that these problems were not of the utmost importance. Thus, an instructive historical illusion occurred. However, in learning the history of the cycloid, we learn a lot about the history of science.

Divisibility rules

by Mark Saul and Titu Andreescu

THIS COLUMN WILL CONTINUE the discussion of divisibility tests started in the article "Divide and Conquer!" on page 18. We begin with a generalization of the divisibility tests for 9 and 11 given in that article, then we climb some steps to arrive at more difficult results.

Problem 1. If S is the sum of the (decimal) digits of the integer N, then $N \equiv S \pmod 9$.

Solution. We can write

$$N = a_n 10^n + a_{n-1}10^{n-1} + \ldots + a_2 10^2 + a_1 10 + a_0.$$

Now $10 \equiv 1 \pmod 9$, and therefore $10^n \equiv 1 \pmod 9$, for any positive integer n. Thus we can write

$$N \equiv a_n + a_{n-1} + \ldots + a_2 + a_1 + a_0 \pmod 9,$$

which is what we wanted to prove.

Problem 2. If S' is the "zigzag" sum of the digits of N, then $N \equiv S \pmod{11}$.

Solution. The proof is analogous to that in problem 1. We use the fact that $10 \equiv -1 \pmod{11}$. The zigzag sum results from the fact that $10^n \equiv 1 \pmod{11}$ if n is even and $10^n \equiv -1 \pmod{11}$ if n is odd.

Problem 3. (a) Show that a number consisting of evenly many identical digits is a multiple of 11. (b) The number N consists of oddly many identical digits. Show that $N - 10$ cannot be a multiple of 11.

Problem 4. A three-digit number is chosen, and a new number is formed by reversing the order of its digits. Show that the (positive) difference between these two numbers is divisible by 11.

Problem 5. (a) Find the largest prime number that divides all integers of the form AAA. (b) Find the largest prime number that divides all integers of the form $BBBB$. (c) The largest prime number that divides all integers of the form $CCCCCC$ is less than 100. Find this number.

Problem 6. Show that if a three-digit number is divisible by 37, then there exists a three-digit number consisting of the same digits in some other order, which is also divisible by 37.

Problem 7. Walter thought that he had a good test for divisibility by 7 of a three-digit number. He said, "If the sum of the three digits is a multiple of 7, then the number itself is a multiple of 7."

"Not true, 914 is not a multiple of 7, but 9 + 1 + 4 = 14 is," said Dick. "In fact, your trick works only when the tens digit and the units digit are identical."

Show that Dick is correct.

Problem 8. Alina said, "Here's how I can check for divisibility by 13. I take the units digit of the number I want and cross it out. To the new number formed, I add four times the units digit I crossed out. My original number is divisible by 13 if and only if my new number is. For example, if I want to test 1937, I form the number 193, and add $4 \cdot 7 = 28$. So I get 221. But now I must test this number."

Carol said, "Not so bad. Just do the same thing again. We get 22 + 4 = 26, which we know is a multiple of 13. Your new number is always less than the one you started with, so you will eventually end up with a two-digit number."

Show that both Carol and Alina are correct.

Problem 9. Let

$$N = a_n 10^n + a_{n-1}10^{n-1} + \ldots + a_2 10^2 + a_1 10 + a_0$$

be an integer, where n is odd. Then N is a multiple of 99 if and only if the number

$$\overline{a_n a_{n-1}} + \overline{a_{n-2} a_{n-3}} + \ldots + \overline{a_1 a_0}$$

is a multiple of 99. (The notation $\overline{ab}$ refers to the two-digit number whose tens digit is a and whose units digit is b.)

Problem 10. What is the smallest 2000-digit number that is a multiple of 99?

Problem 11: The binomial coefficient

$$\binom{99}{19}$$

is the 21-digit number

$$107196674080761936xyz.$$

What is the three-digit number xyz?

Problem 12: Show that any product of 99 consecutive integers is divisible by 99!

ANSWERS, HINTS & SOLUTIONS ON PAGE 53

AT THE BLACKBOARD III

Relativity of motion

by A. I. Chernoutsan

THE MOTION OF ANY BODY is known to be relative, because its displacement, velocity, and trajectory depend on the frame of reference from which an observer describes it. To describe a particular motion, we can use different frames of reference—both stationary and moving.

In many cases, to change from one frame of reference to another, we need only use the rule for transforming displacements, velocities, and accelerations. For displacements, we have

$$\mathbf{s}_{B1} = \mathbf{s}_{B2} + \mathbf{s}_{21},$$

where $\mathbf{s}_{B1}$ is the displacement of a body relative to the first frame of reference (for example, the stationary one), $\mathbf{s}_{B2}$ is the displacement of the body relative to the second frame of reference (say, a moving one), and $\mathbf{s}_{21}$ is the displacement of the second frame of reference measured relative to the first frame of reference. Notice that the first subscript refers to the body (or frame) being measured and the second subscript refers to the frame of reference used to make the measurement.

The analogous transformations for velocity and acceleration are

$$\mathbf{v}_{B1} = \mathbf{v}_{B2} + \mathbf{v}_{21},$$

and

$$\mathbf{a}_{B1} = \mathbf{a}_{B2} + \mathbf{a}_{21}.$$

It is easy to remember the order of the subscripts if you notice that the two outer subscripts on the right-hand side of the transformations match the subscripts on the left and the two inner subscripts on the right are the same.

As an example of using these transformations, the velocity of a boat relative to the land is equal to the velocity of the boat relative to the water plus the velocity of the water relative to the land. Symbolically, we have

$$\mathbf{v}_{BL} = \mathbf{v}_{BW} + \mathbf{v}_{WL}.$$

It should be noted that these formulas are valid for translational motion of one frame of reference relative to another (the coordinate axes of the moving frame of reference are parallel to those of the stationary frame of reference). In addition, the velocities of all moving objects should be far less than the velocity of light c = 300,000 km/s, otherwise quite different velocity transformations would come into play. These are described in Einstein's relativity theory.

Art by Sergey Ivanov

Now let's think about why it is necessary to change the frame of reference. Wouldn't it be simpler to use the same frame of reference in every case? There are many reasons not to do this.

First, in many situations we must use an alternative frame of reference because a problem cannot be solved otherwise. Consider the flight of an airplane in windy weather. The chosen course of the airplane's axis is relative to the compass needle, but the velocity of the plane is measured relative to the air. The meaning of such velocity measurements is clear in the frame of reference fixed to the moving air, where the data obtained determine the direction and value of the airplane's velocity in this frame of reference. However, we cannot use this frame of reference only, because we need to know the plane's location relative to the landmarks in order to land safely at the destination airport. So, let's write the velocity addition law

$$\mathbf{v}_{PL} = \mathbf{v}_{PA} + \mathbf{v}_{AL},$$

where P, L, and A refer to the plane, land, and air, respectively, and draw the corresponding velocities in figure 1. Usually, the wind velocity and direction are provided by a meteorological service, and other data are known beforehand: the course to the destination airport (direction of vector $\mathbf{v}_{PL}$), the velocity of the plane relative to the wind, and the airplane's velocity relative to the land (the air-traffic controllers demand that the planes land on time!). These data are quite enough to calculate all the necessary elements of a velocity triangle (say, angle θ—the correction factor to the course in windy weather) using two known sides and angle of this triangle.

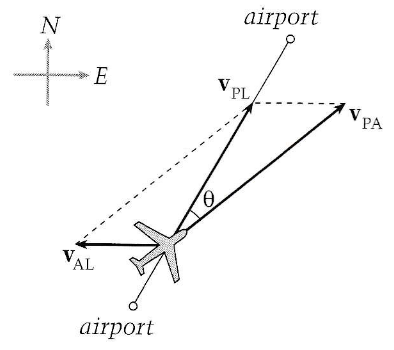


Figure 1

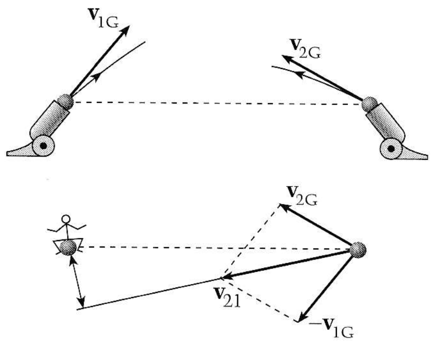


Figure 2

Second, although in many cases it is not necessary to change to another frame of reference to solve a problem, it can drastically simplify the problem and make the solution more obvious. Consider, for example, the flight of two cannonballs after the simultaneous shots of two cannons (fig. 2). How can we know the distance between the balls? The simplest way of looking at it is to use the theoretical approach of Baron von Munchausen, who liked to travel straddling a flying cannonball. The relative acceleration of the balls is zero, because both cannonballs have the acceleration due to gravity (air resistance is neglected here). Thus, from the baron's viewpoint (he sits on ball number 1) the second ball moves uniformly along a straight line with velocity

$$\mathbf{v}_{21} = \mathbf{v}_{2G} + \mathbf{v}_{1G} = \mathbf{v}_{2G} - \mathbf{v}_{1G}.$$

By determining the direction of this velocity from the plot, we can easily find the closest distance of approach of the second cannonball to the baron (we need only verify that neither of the balls hits the ground before this occurs).

Of course, we could make all of the calculations without using the baron's frame of reference, but they would be cumbersome. The "principle of Baron von Munchausen," which says that the relative motion of two freely flying bodies is uniform, helps to solve many problems. Explain, for example, why fireworks produce bright balls of fragments that grow as they fall.

It is interesting to recall that one of the dramatic episodes of human endeavor dealt with none other than the choice of the most "correct" frame of reference. The execution of Giordano Bruno (1548–1600) and the renunciation of Galileo mark the dangerous road to scientific knowledge. It was very difficult for mankind to agree that Earth is not the center of the Universe, but rather just one of the planets revolving about the Sun.

What did mankind gain by replacing the geocentric (Earth-related) frame of reference with the heliocentric (Sun-related) one? Now we understand that one of the advantages of this transition is the drastic simplification of planetary motion. This helped Johannes Kepler (1571–1630) discover in later times three famous laws of planetary motion. In turn, these laws prompted Newton to discover the law of universal gravitation.

So, what is the major advantage of the heliocentric over the geocentric frame of reference? In physics, not all frames of reference are equivalent from the dynamic viewpoint. There are so-called inertial frames of reference, where the laws of mechanics assume the simplest and sometimes even self-evident forms. By contrast to the geocentric frame of reference, the heliocentric frame of reference can be considered inertial. However, in kinematics one frame of reference is as good as another—we can use any of them, even accelerating and rotating frames of reference. However, rotating frames of reference have some unusual features, which should be considered very attentively—but this is a subject for a special discussion. ◘

Quantum on frames of reference:

G. Myakishev, "The 'most inertial' reference frame," March/April 1995, p. 48.

B. Belonuchkin, "The fruit of Kepler's struggle," January/February 1992, pp. 19–22.

A. Leonovich, "Are you relatively sure?" September/October 1996, pp. 32–33.

Bulletin Board

Website genesis

NASA's Genesis mission announces the launch of its new website, which is located at genesismission.jpl.nasa.org. The website includes interactive amusement for science enthusiasts in addition to continually updated reports of mission preparation and progress, interviews with Genesis team members, and science background on the mission.

Site highlights include a create-your-own periodic table that illuminates how the periodic table of elements came to be; an interview with Chester N. Sasaki, Genesis Project Leader; and a nine-minute video introducing the mission.

Scheduled for launch in January 2001, the Genesis mission will park a relatively low-cost spacecraft at a gravitationally stable point just beyond the orbit of the Moon, where it will collect charged atomic particles, components of solar wind. The collected materials, which will total only a few millionths of a gram, will include isotopes of oxygen, carbon, and nitrogen. Previous and current space missions studied the most abundant elements of solar wind, hydrogen, and helium.

The Genesis mission will measure the number of particles in the solar wind, their energy, and their direction of travel. Upon the spacecraft's return to Earth in 2003, scientists will analyze the data for a "weather report" on the solar wind and for clues to the formation of the Solar System.

Solar wind blows at about a million miles per hour and carries occasional energetic disturbances that can cause storms of activity in Earth's magnetosphere and knock out electrical systems on satellites. Earth-orbiting satellites are not able to capture solar wind particles easily because Earth's electrical field deflects most of the solar wind.

Mi, a name, I call myself

NASA's two new Mars probes need names, and it's turning to you. The Deep Space Mission sent off the probes last year by hitching them to another spacecraft that is also traveling to Mars. However, instead of landing gently, the miniature probes will intentionally crash into Mars' surface, enabling them to sample the sediment and test it for the presence of ice. The forebody of each probe will get buried, while the aftbody will remain on the surface and transmit the information back to Earth.

NASA is looking for pairs of names that evoke a spirit of exploration and risk for the sake of knowledge. If named after people, the people must be no longer living; names may be taken from historical, mythical, or fictional characters. NASA is not interested in names of superheroes or names with acronyms. NASA also suggests naming the probes after places or things that are related to each other, or a person and a place or thing related to that person. Send your entry with an essay of 100 words or fewer describing why the chosen names best represent the mission.

You may submit your entry electronically at the contest website or by regular mail by April 30, 1999, to Deep Space 2 Naming Contest, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Mail Stop 301-235, Pasadena, CA 91109-8099.Winners will be announced at the end of 1999, posted on the Deep Space 2 website, and contacted via postal mail. The winner and 25 finalists will each receive a Deep Space 2 poster signed by the project team. For more information, visit The Space Place at http://spaceplace.jpl.nasa.gov/ds2cntst.htm.

Not for squares

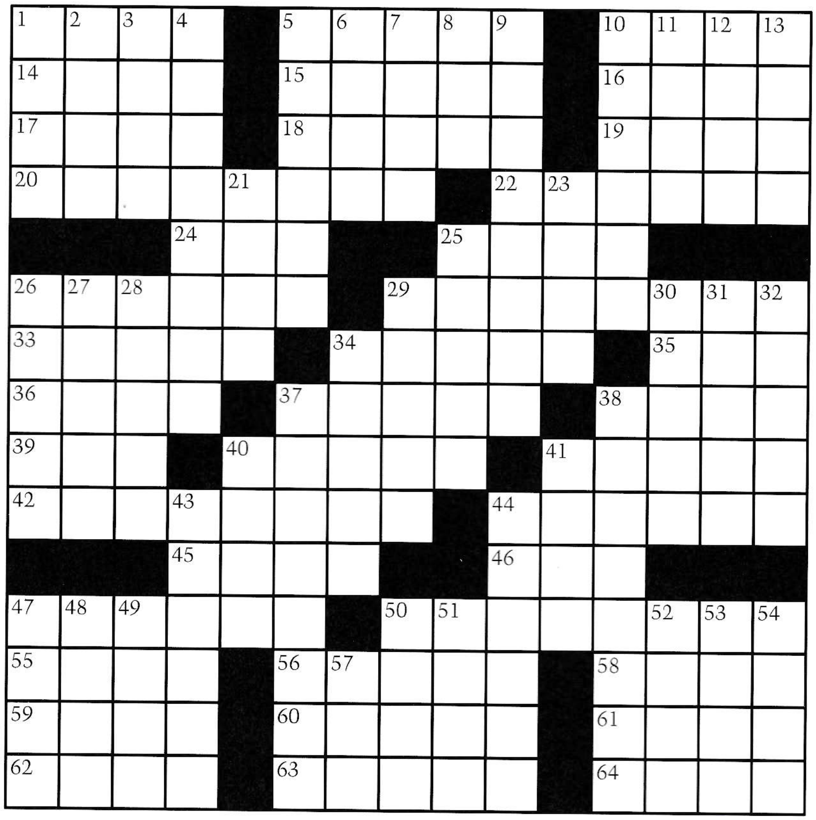
This month's CyberTeaser (B258 in this issue) proved a cinch to solve for those who could think "outside the box." Consider four rectangles at the upper left corner of the given rectangle. The ratio of the areas of rectangles 1 and 2 is equal to that of rectangles 3 and 4. Therefore, the area of rectangle 3 is six. Using similar reasoning, we can find the areas of the rectangles above the diagonal, which leads us to the area of the desired rectangle being 24. Following are the 10 speediest solvers:

Theo Koupelis (Wausau, Wisconsin)
Anastasia Nikitina (Pasadena, California)
Jim Paris (Doylestown, Pennsylvania)
Nick Bennett (Doylestown, Pennsylvania)
Elisabeth Roselle (Fairfield, Ohio)
Christopher Franck (Redondo Beach, California)
Bruno Konder (Rio de Janeiro, Brazil)
Sergio Moya (Culiacan, Mexico)
Leonid Borovskiy (Brooklyn, New York)
May Lim (Quezon City, Philippines)

Congratulations! Each of the winners will receive a *Quantum* button and a copy of the March/April issue. Everyone who submitted a correct answer in the time allotted was entered in a drawing for a copy of *Quantum Quandaries,* our collection of the first 100 *Quantum* brainteasers.

Criss×cross science

by David R. Martin



Across

1 Hawaiian goose
5 Watery soup
10 Hydrous silica
14 And others: abbr.
15 Rumanian historian Nicolae ___
16 "___ la, la, la, ..."
17 Raise to the third power
18 Yet (2 wds.)
19 Banner
20 ___'s voltage law
22 Microwave sources
24 Element 50
25 Drill
26 Startles
29 Type of diode
33 Like a rabbit
34 Filmmaker ___ Ivens
35 61
36 Give forth
37 Of charged particles
38 Pass quickly
39 A poor grade
40 Oriental
41 Sharp-edged metal
42 Nondispersive waves
44 Linger
45 Certain fly's larva
46 Intuitive letters
47 DC current generator
50 Sawtooth ___
55 Geophysicist Harry ___ (1859–1944)
56 Like 3 and 7
58 "in ___ of" (instead of)
59 Forearm bone
60 Microscopy dye
61 California wind
62 Dale
63 One cubic meter
64 Nuisance

Down

1 "___ of the woods"
2 Decorative case
3 Sodium bromide
4 Permanent electric dipole
5 Bovid ruminants
6 House top
7 *Carmina Burana* composer
8 Nitrilotriacetic acid
9 Simple ___ motion
10 ___ lithography
11 Wan
12 Winglike
13 Trails
21 Hurried
23 Circle segments
25 Engraver's tool
26 Initial numbers
27 Supporting role
28 Uranus' moon
29 ___ Salk
30 "___ ease" (2 wds.)
31 Rust, e.g.
32 Explosive ingredient
34 Elbow or knee, e.g.
37 U-235, U-238, and U-239
38 Computer circuit element
40 Elemental particle
41 ___-Einstein statistics
43 Nigerian city
44 Chemist Phoebus ___ (1869–1940)
47 Medicinal substance
48 Shout
49 Sides of a nonagon
50 Judicious
51 Moslem ruler
52 Seine tributary
53 Old cars
54 New wine
57 Curl

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE JANUARY/FEBRUARY PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M256

The desired number has the form $x = 2^a 3^b 5^c$. Then, we have $2x = 2^{a+1}3^b5^c = p^2$, and thus, a is an odd number, and b and c are even. Similarly, a is divisible by 3 and 5, $b + 1$ is divisible by 3, b by 5, c by 3, and $c + 1$ by 5. A little experimentation shows that the minimal set of a, b, and c satisfying these conditions is as follows: $a = 15$, $b = 20$, and $c = 24$.

M257

Make the substitution $y = x + 2$. Then, the equations take the form

$$y^3 - 2y - 19 = 0$$

and

$$y^3 - 2y + 19 = 0.$$

If y_0 is a root of the first equation, then $-y_0$ is a root of the second equation. In addition, we can see from the graphs that each of the equations has a single real root. The sum of these roots is zero. Therefore, the sum of the roots of the original equations is

$$(y_0 - 2) + (-y_0 - 2) = -4.$$

M258

Since a radical cannot have a negative value, we see that $x \geq 0$. From the second-level radical, we can see that $2 - \sqrt{2-x} \geq 0$, which leads to $x \leq 2$. Therefore, $0 \leq x \leq 2$, and we can make the variable substitution $x = 2\cos\phi$, $0 \leq \phi \leq \pi/2$. Then,

$$2 + x = 2(1+\cos\phi) = 4\cos^2\frac{\phi}{2}$$

and

$$\sqrt{4\cos^2\frac{\phi}{2}} = 2\cos\frac{\phi}{2}.$$

Further manipulations yield

$$2 - 2\cos\frac{\phi}{2} = 2\left(1-\cos\frac{\phi}{2}\right) = 4\sin^2\frac{\phi}{4}.$$

Thus, our equation takes the form

$$\sqrt{2+2\sin\frac{\phi}{4}} = 2\cos\phi$$

(all roots are taken with the plus sign, because $0 \leq \phi \leq \pi/2$). We can rewrite it as

$$\sqrt{2+2\cos\left(\frac{\pi}{2}-\frac{\phi}{4}\right)} = 2\cos\phi.$$

Finally, we obtain the equation

$$\cos\left(\frac{\pi}{4}-\frac{\phi}{8}\right) = \cos\phi.$$

Thus, we have

$$\frac{\pi}{4}-\frac{\phi}{8} = \pm\phi + 2\pi k.$$

Taking into account the constraints imposed on ϕ, we find that $\phi = 2\pi/9$. Answer: $x = \cos(2\pi/9)$.

M259

Let's prove an auxiliary proposition: If $ABCD$ is a rectangle and M is an arbitrary point in space, then

$$MA^2 + MC^2 = MB^2 + MD^2. \quad (1)$$

First we will prove relation (1). Figure 1 shows the special case where point M is in the plane of rectangle $ABCD$, and inside the rectangle. We draw $MP \perp AD$, and $MQ \perp AC$. Note that $AP = BQ$ and $PD = QC$. From using the Pythagorean theorem in various right triangles, we have

$$\begin{aligned} MA^2 + MC^2 &= AP^2 + PM^2 + MQ^2 + QC^2 \\ &= AP^2 + PD^2 + MP^2 + MQ^2 \\ MB^2 + MD^2 &= BQ^2 + MQ^2 + MP^2 + PD^2 \\ &= AP^2 + PD^2 + MP^2 + MQ^2, \end{aligned}$$

so equation (1) holds in this case. The reader can check that the same argument works when M is outside the rectangle (but in the plane of $ABCD$). If M is not in this plane, we can make an analogous argument by considering M', the projection of point M onto the plane of $ABCD$. We locate points P and Q by drawing perpendiculars $M'P$ and $M'Q$ as before, and a similar argument leads to the conclusion.

Now we turn to the problem under consideration. Construct a rectangular parallelepiped from the given tetrahedron in the usual way (the mutually perpendicular edges of the tetrahedron are also the edges of the parallelepiped coming from the same vertex: see figure 2).

Consider the three faces of this parallelepiped containing the three faces of the tetrahedron (fig. 1). We know

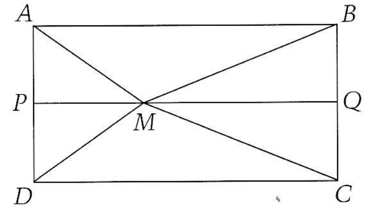


Figure 1

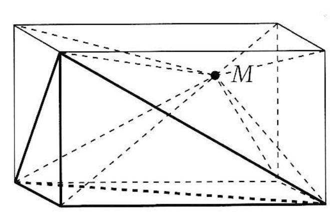


Figure 2

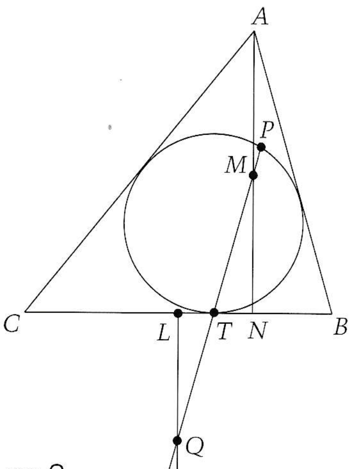


Figure 3

the distance from point M (given in the problem) to three vertices of each face. Thus, we can determine the distance from M to the fourth vertex of each of these faces. Squares of these distances are as follows: $5 + 6 - 9 = 2$, $6 + 7 - 9 = 4$, and $7 + 5 - 9 = 3$.

Therefore, we can find the distance from point M to the remaining vertex of the parallelepiped (that lies opposite the vertex of the tetrahedron with right plane angles). It turns out that this distance is zero: $2 + 3 - 5 = 0$, or $3 + 4 - 7 = 0$, or $2 + 4 - 6 = 0$. This means that point M coincides with this vertex of the parallelepiped. Therefore, the diagonal of the parallelepiped is 3, and the radius of the sphere circumscribed around the parallelepiped, and thus around the given tetrahedron, is 1.5.

M260

Let L be the midpoint of BC. Denote by Q the point of intersection of the perpendicular bisector to BC with line PT (fig. 3). Let us prove that points B, C, P, and Q lie on the same circle. First let us show that this fact proves the assertion of the problem. Note that tangents to a circle from the endpoints of a chord make equal angles with that chord (fig. 4). Now

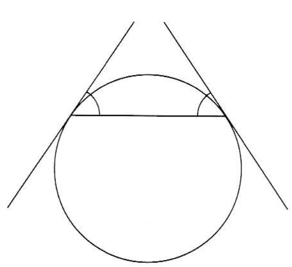


Figure 4

suppose two circles have a common point P, and we draw overlapping chords through this point. If the tangents to the two circles at the other ends of the chords are parallel, then the circles are tangent at the common endpoint of the chords. Indeed, if the two circles had different tangent lines at point P, these lines would have to make the same angle with line PAB (fig. 5).

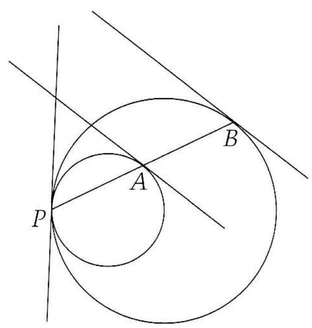


Figure 5

Now if B, C, P, and Q lie on the same circle, then the circle that passes through points B, C, and P and the circle inscribed in the given triangle have a common point P. Points P, T, and Q lie on the same line; T and Q belong to different circles; and the tangents to the corresponding circles at points T and Q are parallel. Therefore, the tangents to these circles at point P coincide—that is, these circles are tangent to each other.

Thus, we must prove that points B, C, P, and Q lie on the same circle. For this purpose, it is sufficient to prove that the following relation holds:

$$CT \cdot TB = QT \cdot TP. \qquad (2)$$

Introduce the notation $BC = a$, $CA = b$, $AB = c$, and let $2s$ be the perimeter of triangle ABC, let K be its area, let r be the radius of the inscribed circle, let $AN = h$ be the altitude drawn to side BC, and let ϕ be the angle MTN. Assume that angle B of the given triangle is acute and $b > c$ (for other cases, the reasoning and calculations are practically the same). First let us recall two facts about the circle inscribed in a triangle (fig. 6).

(i) $rs = K$. This well-known formula results from calculating K as the sum of the areas of triangles ABI, BCI, CIA.

(ii) If $s = (1/2)(a + b + c)$, then

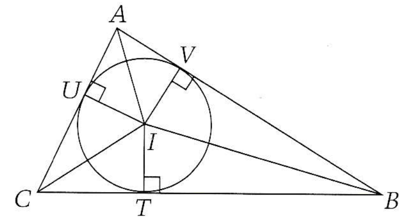


Figure 6

$$CT = CU = s - c$$
$$AU = AV = s - a$$
$$BV = BT = s - b.$$

These formulas can be proved by letting $CT = CU = x$, $AU = AV = y$, $BV = BT = z$, so that

$$x + y = ay + z = bz + z = c,$$

and solving for x, y, and z in terms of a, b, and c. We have:

$$CT = s - c = \frac{a+b-c}{2},$$
$$TB = s - b = \frac{a+c-b}{2}. \qquad (3)$$

We then find that

$$LT = LB - TB = \frac{b-c}{2} \qquad (4)$$

and

$$NB = c\cos B = \frac{2ac\cos B}{2a}$$
$$= \frac{a^2 + c^2 - b^2}{2a} = \frac{a}{2} + \frac{c^2 - b^2}{2a}.$$

The last relation, with (3), implies

$$TN = TB - NB = \frac{(b-c)(s-a)}{a}.$$

Thus,

$$\tan\phi = \frac{MN}{TN} = \frac{ha}{2(b-c)(s-a)}. \qquad (5)$$

Now we have: $TP = 2r\sin\phi$ and $TQ = LT/\cos\phi$. These relations are clear from figure 7, in which T' is the

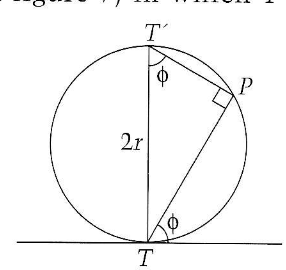


Figure 7

point diametrically opposite to T on the inscribed circle. Combined with (5), these two relations imply that

$$TP \cdot QT = r(b-c)\tan\phi$$
$$= \frac{rha}{2(s-a)} = \frac{rK}{s-a}.$$

Now, relation (2), combined with (3), takes the form

$$\frac{rK}{s-a} = (s-c)(s-b).$$

This relation follows from two formulas for the area of a triangle: $K = rs$ (the inradius times the semiperimeter) and

$$K = \sqrt{s(s-a)(s-b)(s-c)}$$

(Hero's formula). A discussion of these two formulas can be found in any book on advanced geometry.

Physics

P256

The frequency of the reflected signal doesn't coincide with that of the radar beam due to the Doppler effect. The maximum frequency shift occurs when the velocity of the reflected radio wave is directed exactly to (or from) the radar.

Let's assume that the radar emits short radio pulses with a repetition frequency f_0, while the frequency of the received pulses is f_1. The interval between the arrivals of the n-th and $(n+1)$-th pulses to the antenna is

$$\frac{1}{f_1} = \frac{1}{f_0} + T_{n+1} - T_n,$$

where T_n is the time needed for the n-th pulse to propagate from the radar to the reflecting object and back, while T_{n+1} is the corresponding period of the $(n+1)$-th pulse.

At the moment of sending the n-th pulse, the distance between the radar and the reflecting object is L_n, while at the moment of sending the next $(n+1)$-th pulse, this distance becomes L_{n+1}. We can see that

$$T_n = \frac{2L_n}{v+c},\ T_{n+1} = \frac{2L_{n+1}}{v+c},$$

where v is the velocity of the wind in a sandstorm, and c is the velocity of the radio waves (or light). The difference $L_n - L_{n+1}$ equals the displacement of the reflecting object during an interval $1/f_0$:

$$L_n - L_{n+1} = \frac{v}{f_0},$$

from which we get

$$f_1 = f_0\frac{(c+v)}{(c-v)} \cong f_0\left(1+\frac{2v}{c}\right),$$

and

$$v \cong \frac{c}{2}\frac{\Delta f}{f_0} \cong 100 \text{ m/s}.$$

P257

The initial air pressure in the tube results from the hydrostatic pressure of the mercury column of height $(H-h) = 76$ cm and atmospheric pressure. By the problem conditions, the atmospheric pressure is 10^5 Pa $\cong$ 76 cm mercury. Thus, at the beginning, the air pressure in the tube is about double that of the atmospheric pressure.

We assume that the displacement of mercury occurs slowly, so the system is always at equilibrium. Near the final state, when almost all of the mercury is squeezed out, the air pressure in the tube will be equal to the atmospheric pressure, or half the initial value. The volume of this air HS (S is the cross-sectional area of the tube) will be twice the initial volume hS. According to the ideal gas law, the air temperature in the final state must equal that in the initial state!

We can see that the mercury will not run out without the tube being heated. Therefore, our paradoxical result means that the problem cannot be solved by considering only the initial and final states: We should follow the entire process of the mercury displacement.

First, let's find how the air temperature in the tube must change in order to perform a gradual displacement of the mercury, in which the system is always at equilibrium.

If at some moment the height of the air column is z, then the air pressure in the tube $P(z)$ is

$$P(z) = P_0 + \rho g(H-z), \qquad (6)$$

where ρ is the density of mercury, and P_0 is the atmospheric pressure, which by the problem conditions is about the pressure exerted by a mercury column of height $H/2$:

$$P_0 = \rho g H/2. \qquad (7)$$

Plugging (7) into (6) yields

$$P(z) = \rho g\left(\frac{3}{2}H - z\right). \qquad (8)$$

Since we assume that the air in the tube is at thermodynamic equilibrium for any z, then the air's pressure $P(z)$, volume Sz, and temperature $T(z)$ are described by the ideal gas law:

$$\frac{P(z)Sz}{T(z)} = \frac{\frac{2P_0SH}{2}}{T_0}, \qquad (9)$$

where T_0 is the initial temperature, $2P_0$ the initial pressure, and $SH/2$ the initial volume of air in the tube. Inserting (6) and (8) into (9), we get the dependence of temperature on the height of the air column in the tube:

$$T(z) = T_0\frac{(3H-2z)z}{H^2}. \qquad (10)$$

The function $T(z)$ is plotted in figure 8. The process of displacing the mercury corresponds to the part of the parabola between the points

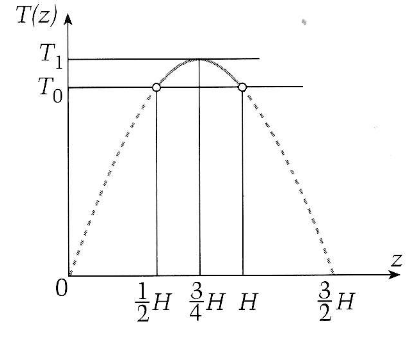


Figure 8

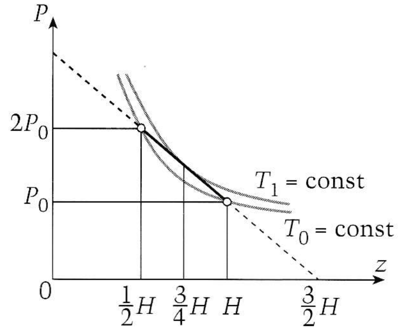


Figure 9

$z = H/2$ and $z = H$ (solid line). We can see that to carry out the slow (quasi-equilibrium) process of displacing the mercury, the temperature must rise to $T_1 = (9/8)T_0$ (at this point half of the mercury will have run out) and then fall to the starting value T_0.

Thus, the complete displacement of mercury from the tube is performed by heating the air to $T_1 \cong 326$ K. Then this temperature should be maintained (by a thermal contact with a thermal reservoir at a temperature T_1). In the following, the mercury will be displaced by the expanding air, but during this stage the process will not be quasi-equilibrium.

The *P-V* plots help clarify our reasoning (fig. 9). The equilibrium displacement is described by the linear dependence of air pressure on z described by formula (8) and shown by the solid line between the points $z = H/2$ and $z = H$. The same figure shows the isotherms corresponding to T_0 and T_1. The initial and final states lie on the same isotherm T_0 = const. Thus, at temperatures higher than T_1, no isotherm crosses the line $P(z)$. In other words, the equilibrium state of the air in the tube cannot be achieved at any value of the air column height z. Thus, if the air temperature in the tube is maintained just above the value T_1, all of the mercury will be squeezed out from the tube by the expanding air.

P258

At any arbitrary time the energy of the system is

$$E = \frac{LI^2}{2} + \frac{CV^2}{2} + \frac{mv^2}{2} - \frac{qVx}{d},$$

where L is the inductance of the coil, I is the electric current in the circuit, V is the potential drop across the capacitor, v is the velocity of the particle, and x is the particle's position. The equation describing the motion of the point-charge in the electric field is

$$ma = \frac{qV}{d}.$$

If V varies harmonically as $V = V_0 \cos \omega t$, then

$$v = \frac{qV_0}{md\omega} \sin \omega t,$$
$$x = -\frac{qV_0}{md\omega^2} \cos \omega t,$$
$$I = -CV_0\omega \sin \omega t.$$

Conservation of energy yields:

$$E = V_0^2\left(\frac{L}{2}(C\omega)^2 + \frac{m}{2}\left(\frac{q}{m\omega d}\right)^2\right)\sin^2 \omega t$$
$$+\ V_0^2\left(\frac{C}{2} + \frac{q^2}{m(\omega d)^2}\right)\cos^2 \omega t = \text{const}.$$

This equation can be satisfied only when the coefficients of $\cos^2 \omega t$ and $\sin^2 \omega t$ are equal:

$$\omega^2 = \frac{\omega_0^2}{2} + \left(\left(\frac{\omega_0^2}{2}\right)^2 + \frac{\left(\frac{q\omega_0}{d}\right)^2}{mC}\right)^{1/2}, \qquad (1)$$

where $\omega_0^2 = 1/LC$.

Let's estimate in what cases we really can neglect the forces due to the image charges, as we suggested. The order of this force is $q^2x/\varepsilon_0 d^3$, and it must be far less than the usual electrostatic force qV/d, from which we obtain

$$\frac{q^2}{\varepsilon_0 d^3 m\omega^2} << 1.$$

Since the area of each plate $S >> d^2$, and $C = \varepsilon_0 S/d$, equation (1) is valid when the point-charge doesn't change the natural frequency of the tuned circuit too much.

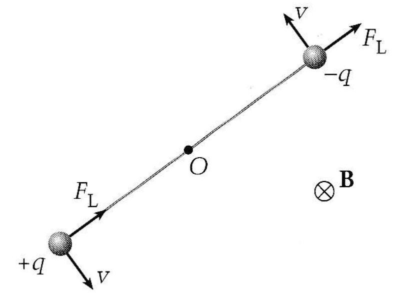


Figure 10

P259

When the magnetic field is turned on, each particle will be affected by the Lorentz force (fig. 10) directed along the rod and equal to

$$F_L = qvB = q\omega\frac{l}{2}B.$$

The resulting force acting on the dipole is

$$F = 2F_L = q\omega lB.$$

The value of this force is constant, but its direction continuously varies: Vector **F** rotates with the rod with the angular velocity ω. Therefore, the dipole's center (point O) will also revolve with the same angular velocity ω along a circle of radius r, which can be determined from Newton's second law:

$$F = 2m\omega^2 r,$$
$$r = \frac{F}{2m\omega^2} = \frac{qlB}{2m\omega}.$$

P260

The problem is solved with the laws of so-called "paraxial optics"—geometrical optics with small angles of incidence and refraction. For such

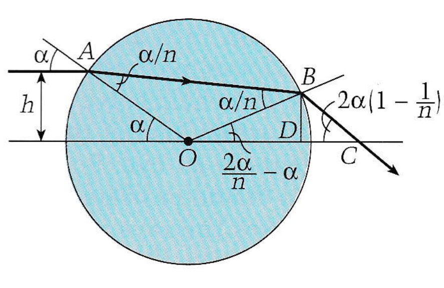


Figure 11

angles, $\sin\alpha \cong \tan\alpha \cong \alpha$, so Snell's law of refraction is reduced to the form of $n_1\alpha_1 = n_2\alpha_2$.

Let's consider an arbitrary ray incident on the ball at a distance h from the axis (fig. 11). Because $h << R$ (the beam is narrow), the angle of incidence of this ray is $\alpha \cong h/R << 1$. By plotting the further passage of this ray, we can determine all the angles in triangles AOB, OBD, and BDC:

$$\angle OAB = \angle OBA = \frac{\alpha}{n},$$

$$\angle BOD = \frac{2\alpha}{n} - \alpha,$$

$$\angle BCD = \alpha - \left(\frac{2\alpha}{n} - \alpha\right) = 2\alpha\left(1 - \frac{1}{n}\right).$$

Then,

$$BD = R\left(\frac{2\alpha}{n} - \alpha\right),$$

$$DC = \frac{BD}{2\alpha(1-1/n)} = \frac{R\alpha(2/n-1)}{2\alpha(1-1/n)} = R\frac{2-n}{2(n-1)}.$$

By the conditions of the problem, $DC = R$, from which we get $n = 4/3$.

It is crucial that the final formula does not contain the angle of incidence α, which means that all rays of the beam will be focused to a single point. This is one of the laws of paraxial optics: In a refractive system, narrow beams of parallel rays are either converged to a single point (a focus) or diverge as if they were radiated from a single point (virtual focus).

The second law of paraxial beam optics says that the rays diverging from a single point at small angles are focused by a refractive system into a point (in this case the focus may also be virtual). These laws help solve many problems in which one of the optical elements is an eye: Because of the small diameter of its pupil, the eye focuses the rays incident at small angles. For this very reason we can see a point as a point, and not as an extended source.

Brainteasers

B256

Answer:

$$2 \div (2-3\div 3)(-4 \div ((4-5)\div 5)) = 40$$

or

$$(2 \div ((2-3)\div 3) - 4) \div ((4-5)\div 5) = 50.$$

B257

See figure 12.

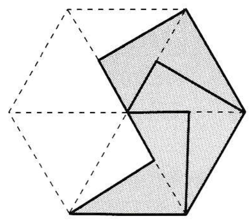


Figure 12

B258

Consider the four rectangles at the upper left corner (fig. 13). The area of one of them is not known. Let's find it. We can see that the ra-

1 2	3 ? = 6
2 1	4 3

Figure 13

tio of the areas of rectangles 1 and 2 is the same as the ratio of the areas of rectangles 3 and 4. Therefore, the area of rectangle 3 is 6. Reasoning in the same way, we can successively find the areas of the rectangles above the diagonal that runs from the upper left to the lower right. As a result, we find that the area of the desired rectangle is 24.

B259

Answer: 7 cm, because the only point D that satisfies the conditions is the vertex of the regular pyramid with triangle ABC at its base. Indeed, $6 \le DB \le 8$, and DB is an integer. Therefore, DB is either 6 cm, 7 cm, or 8 cm; the first and last variants imply that D lies on AB. (The reader can check—for example, using the law of cosines—that this does not yield an integer for DC.) Similarly, DC is either 6 cm, 7 cm, or 8 cm; and the first and last variants imply that D lies on AC. Therefore, the only remaining possibility is $AD = BD = CD = 7$ cm.

B260

The balance will not be disturbed, but the reading of the spring scale will change. By the way, will this reading increase or decrease? (Of course, it does diminish because the force of gravity tends toward zero in the center of the Earth.)

Digital world

Question 1. All you have to do to compute the perimeter is to count and add the number of outside boxes facing northward, southward, eastward, and westward. But the number of northward facing boxes is just the width of the darkened area, as is the number of southward-facing boxes. Similarly, the number of eastward-facing boxes is just the height of the darkened area, as is the number of westward-facing boxes. Thus the whole perimeter is twice the sum of the height and width of the darkened region. Because of the symmetry of the circle, this is just four times the width of the darkened region.

Question 2. If X is a sphere, the area surplus is 3/2. This can be seen by looking at X' along the voxel coordinate directions, positive and negative (fig. 14). Each of these views reveals a separate set of voxel faces, which collectively comprise

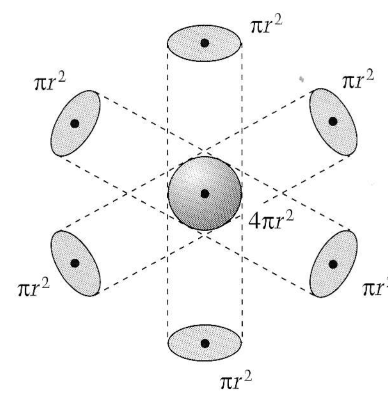


Figure 14

all the visible voxel faces—that is, the surface of X'. But these six views are circles, each with area πr^2 (where r is the radius of the sphere). The ratio of the sum of these areas ($6\pi r^2$) to the area of the sphere ($4\pi r^2$) is 3/2. You can apply the "six-views" trick to compute the area surplus of any convex object from its three projections (each viewed from two directions).

Question 3. The minimum area surplus is 1, obtained by a cube aligned with the voxel coordinate axes. The maximum area surplus is $\sqrt{3}$, obtained either by a cube voxel-aligned along its body diagonals or by an octahedron with its opposite vertices so aligned. To see the maximality of $\sqrt{3}$, consider any almost-planar "facet" of X with area A and unit normal **n** in voxel coordinates. The area of the corresponding "facet" of X' is the sum of the projected areas along the three voxel-axis viewing direction from outside the object X'. This sum is A times the sum of the direction cosines n_i in the voxel coordinate directions. The sum is maximized when the direction cosines are all equal (and thus equal to $1/\sqrt{3}$). Therefore, the maximum area surplus of the facet is $3/\sqrt{3}$, or $\sqrt{3}$. Note that this proof does not depend on the convexity of the object that is approximated by cubes, because the argument applies to each local part of the object's boundary.

Question 4. By extension of the argument in question 3, in N dimensions the minimum area surplus is 1, and the maximum is $\sqrt{N}$.

Gradus

Problem 3. (a) If a number has evenly many identical digits, its zigzag sum is 0, so it is a multiple of 11. (b) If N has oddly many digits, each equal to d, then its remainder upon division by 11 is just d. In particular, no such number can have remainder 10, since $d < 10$.

Problem 4. If the original number is $100a + 10b + c$, we can assume without loss of generality that $a \geq c$. Then we are investigating the difference $100(a - c) + (c - a) = 99(a - c)$, which is surely a multiple of 11 (and also of 9).

Problem 5. (a) $AAA = 100A + 10A + A = 111A = 3 \cdot 37A$, so the largest such prime is 37. (b) As before, $BBBB = 1111B = 11 \cdot 101B$, so the largest prime is 101. (c) $CCCCCC = 111111C = 111 \cdot 1001C = 3 \cdot 37 \cdot 7 \cdot 11 \cdot 13C$, so the answer is 37.

Problem 6. If the number is ABC, there are six candidates. In fact, the number BCA must be divisible by 37. Indeed, we have $100A + 10B + C = 37k$, for some integer k. Then $100B = 370k - 1000A - 10C$, and

$$BCA = 370k - 1000A - 10C + 10C + A = 370k - 999A.$$

Since $999 = 9 \cdot 111 = 27 \cdot 37$, this number is a multiple of 37.

Problem 7. Walter's numbers are of the form

$$100a + 10b + b = 100a + 11b \equiv 2a + 4b \pmod 7.$$

The sum of the digits is $a + 2b$. The reader can check that $a + 2b \equiv 0 \pmod 7$ *if and only if* $2a + 4b \equiv 0 \pmod 7$, thus the truth of Dick's statement. Note that the statement in italics is true because 7 is prime: For other moduli the statement may not be true.

Problem 8. Consider the number $N = 10a + b$, where b is its units digit (but a is not its tens digit). Then we are forming the number $N' = a + 4b$. Then $10N' = 10a + 40b$, and $10N' - N = 39b$, which is certainly a multiple of 13. Hence $N \equiv 10N' \pmod{13}$, and $N \equiv 0 \pmod{13}$ if and only if $N' \equiv 0 \pmod{13}$. Notice that N and N' may not be congruent modulo 13.

To see that we eventually get a two-digit number, notice that $N - N' = 9a - 3b$, and if $a > 9$ (that is, if the original number contains more than two digits), then $N - N'$ is certainly positive, so $N' < N$.

Problem 9. We use the notation of the first two problems, where S is the sum of the digits of the number N, and S' is the zigzag sum.

Suppose first that N is a multiple of 99. Then N is a multiple of 9, so S is also a multiple of 9, and $11S$ is a multiple of 99. Similarly, S' is a multiple of 11, so $9S'$ is a multiple of 99. Then $11S - 9S'$ is a multiple of 99. But this last difference is just

$$2(10a_n + a_{n-1} + 10a_{n-2} + a_{n-3} + \ldots + 10a_1 + a_0),$$

which is

$$2\left(\overline{a_n a_{n-1}} + \overline{a_{n-2} a_{n-3}} + \ldots + \overline{a_1 a_0}\right).$$

Hence the given sum must be a multiple of 99.

Now suppose that

$$\overline{a_n a_{n-1}} + \overline{a_{n-2} a_{n-3}} + \ldots + \overline{a_1 a_0}$$

is a multiple of 99. Then twice this number, or

$$2(10a_n + a_{n-1} + 10a_{n-2} + a_{n-3} + \ldots + 10a_1 + a_0) = 11S - 9S'$$

is also a multiple of 99. Then we can write $11S - 9S' = 99K$, for some integer K, and $11S = 99K + 9S'$ is a multiple of 9. It follows that S itself is a multiple of 9, and therefore so is N. Similarly, writing $9S' = 11S - 99K$, we find that S' is a multiple of 11, and therefore so is N. Hence N is a multiple of 99.

Problem 10. The smallest 2000-digit number of all is the number

$$N = \underbrace{1000\ldots000}_{1999 \text{ zeroes}}.$$

Certainly there is a multiple of 99 in the next 100 integers, so we can assume that the number we seek is of the form

$$\underbrace{1000\ldots0ab}_{1997 \text{ zeroes}},$$

for some digits a and b. The criterion of problem 9 tells us that the sum $10 + 00 + 00 + \ldots + ab$ must be a multiple of 99, so the smallest possible values of a and b are 8 and 9. Answer:

$$\underbrace{1000\ldots089}_{1997 \text{ zeroes}}.$$

Problem 11. We have

$$\binom{99}{19} = \frac{99 \cdot 98 \cdot 97 \cdot \cdots \cdot 81}{19 \cdot 18 \cdot 27 \cdot \cdots 3 \cdot 2 \cdot 1},$$

and a quick count of the factors of the numerator and denominator will show that this number is a multiple of 2, of 3^3, and of 11, but not of 5. Thus z is a nonzero even digit. So for this integer,

$$S = 1+0+7+1+9+6+6+7 \\ +4+0+8+0+7+6+1 \\ +9+3+6+x+y+z$$

is a multiple of 9, and

$$S' = 1-0+7-1+9-6+6-7 \\ +4-0+8-0+7-6+1 \\ -9+3-6+x-y+z$$

is a multiple of 11. It follows that $x+y+z$ is a multiple of 9 and $x-y+z$ is a multiple of 11, and these two sums must have the same parity. (We can see this, for example, by noting that their sum is $2x+2z$, which must be even.) Now $x+y+z$ cannot be 0, because z is a nonzero digit. If $x+y+z=9$, then $x-y+z=11$ (they must have the same parity), but the second number cannot be larger than the first. Also, $x+y+z$ cannot be 27, for x, y, and z are at most 9, and z is even. Hence $x+y+z=18$, and $x-y+z$ must equal 0.

It follows that $y=9$ and $x+z=9$, and the only possible candidates for $\overline{xyz}$ are 792, 594, 396, and 198. Finally, we note that

$$\binom{99}{19}$$

is divisible by 27, as is the number 107196674080761936 (we can verify this easily by dividing by 9, then by 3). Therefore their difference, which is $\overline{xyz}$, is also divisible by 27. The only candidate which satisfies this criterion is 594.

Problem 12. Without loss of generality, we can assume that the 99 integers are all positive, and we can write them as $k+1, k+2, \ldots k+99$ for some nonnegative integer k.

Now we look at the quotient

$$\frac{(k+1)(k+2)\cdots(k+99)}{99!} = \binom{k+99}{99},$$

and since any binomial coefficient is an integer, the conclusion follows.

Sky

by David Arns

See the wonders of the sky—wondrous sky!
Astronomic marvels everywhere do meet the eye!
They are waiting, waiting, waiting
for us just to take a glance,
And behold them in their splendor,
Filled with awe that they engender,
as we gaze in dreamlike trance;
And we stare, stare, stare
through the icy winter air,
And are dazzled at the glory
of the heav'nly inventory
Of the sky, sky, sky, sky, sky, sky, sky—
At the heav'nly inventory of the sky.

See the mighty galaxies—galaxies!
Twinkling at us shyly through the branches of the trees.
How they shimmer, shimmer, shimmer
(or so it appears to us),
But their distance is enormous,
So astronomers inform us:
Some, a billion light-years plus.
But they shine, shine, shine,
with a radiance benign,
That belies the brilliant, blinding,
awful glare that we'd be finding,
Were we near, near, near, near, near, near, near—
The glare that we'd be finding were we near.

Think about exploding stars—dying stars,
Throwing through the heavenlies their luminescent scars.
See them glimmer, glimmer, glimmer
with a wispiness of light
From their tendrils filamental
Made from gases elemental
in the darkness of the night,
Reaching out, out, out,
on their interstellar route
Leaving light-years far behind them
where the gravity confined them
In the stars, stars, stars, stars, stars, stars, stars—
In the fusion-heated centers of the stars.

Or think about our neighborhood—our neighborhood!
We have asteroids and planets to explore (and yes, we should).
They are spinning, spinning, spinning
as they orbit round the Sun
And their paths, which are elliptic,
All are close to the ecliptic
As they make their annual run
They go round, round, round,
Yet in vacuum, make no sound,
But continue their rotation,
their precession and nutation
Through the years, years, years, years, years, years, years—
They continue their rotation through the years.

Dutch treat

by Dr. Mu

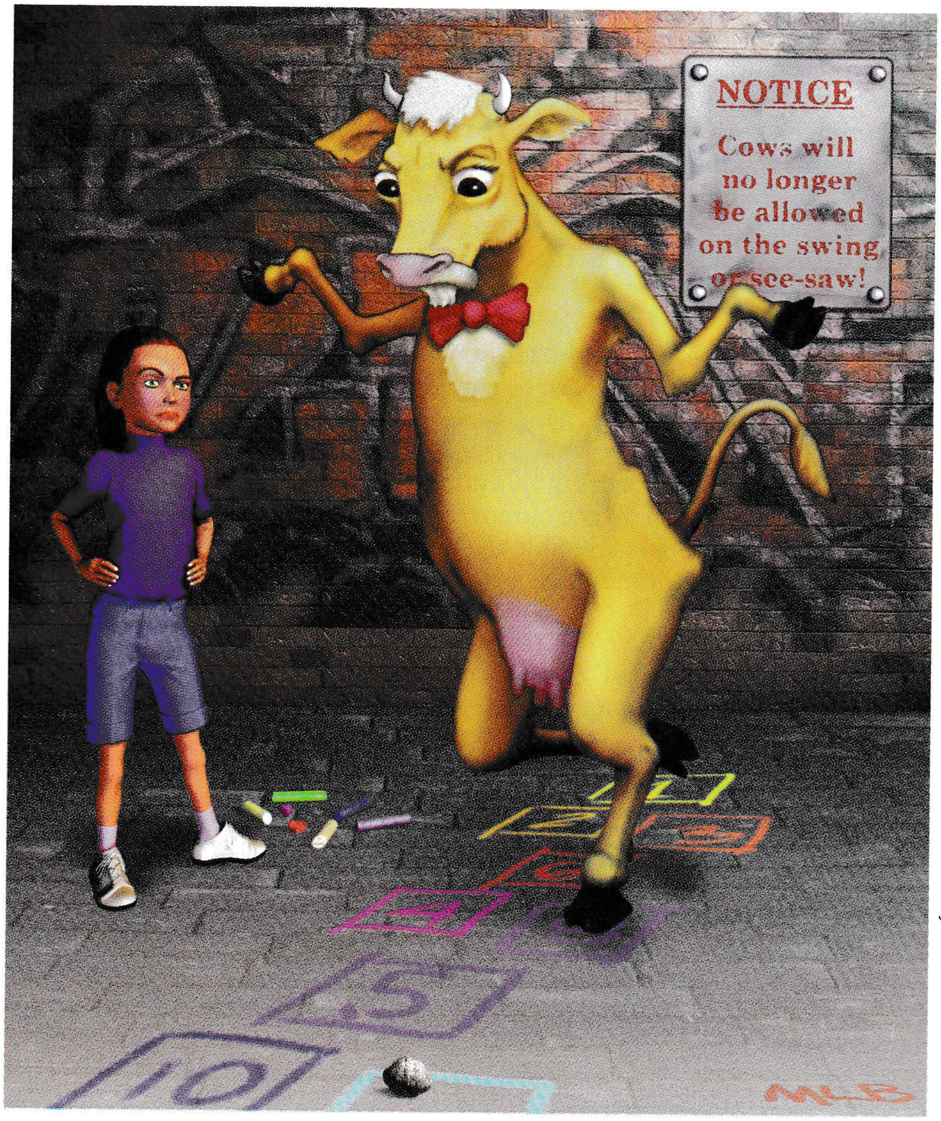
WELCOME BACK TO COWCULATIONS, THE column devoted to problems best solved with a computer algorithm. The herd survived the blizzard of '99. We stayed warm and cozy in the barn while the snow drifts piled up outside, some as high as 10 feet. The milk truck was delayed one day when the Interstate highway was closed by whiteout conditions that caused a 50-car pileup. Farmer Paul worked overtime keeping the driveway leading up to the barn snow free, food in the trough, and the gutters clean. Our milk production stayed on course, and we passed the time surfing the Internet and catching up on our e-mail.

One evening I noticed a new message in my inbox from Dr. Tom, a friend from the Netherlands. "Buck up, Dr. Mu," he wrote, "spring is just around the corner and soon you'll be back outside playing hopscotch again. But while you're waiting for the green grass to return to the pasture, I have a programming problem for you."

Dr. Tom is a professor of Mathematics and Computer Science at the Einhoven University of Technology, in the Netherlands. He maintains the website of the International Science Olympiads which includes links to all Science Olympiad sites: Mathematics, Physics, Chemistry, Informatics (Computer Science), Biology, and Astronomy. It's the first place to go to find all links to International and National Science & Mathematics Olympiads. I met Dr. Tom at the 1995 International Olympiad in Informatics, which was held in Einhoven.

The message continued, "Consider the sequence 1, 2, 3, 6, 4, 8, 5, 10, 7, 14, 9, 18, . . . and write a *nice* and *efficient* and *documented* program to generate it. I'm sending this out as a *Dutch Treat* in place of a Christmas & New Year's card this year. Happy New Year, Dr. Tom."

My barnmates and I went after the challenge like black flies on a fresh cow pie. Such a simple and tasty



Art by Mark Brenneman

treat, surely it can be downed in just a few bytes. After an hour or so we compared programs. Some were simple, others not. Some were fast, some were not. One solution, however, had everything. It was simple, fast, and ingenious. I wondered to myself, can the readers of this column come up with a similar elegant solution? This suggested the next problem, which, you guessed it, is the next Challenge Outta Wisconsin.

COW 15

Write a program that generates the terms in the sequence 1, 2, 3, 6, 4, 8, 5, 10, 7, 14, 9, 18, ... Test it by finding the 100,000th term. Speed and elegance count.

Throw your rock then hippity hop.
Cover the squares and don't stop.
Watch the pattern under your feet.
Discover the rule in this Dutch Treat.
Write the code and make it fast.
Find the number that comes up last.
—Dr. Mu

COW 13

In Cow13 you were asked to write a program that would transform any number into its corresponding binary representation and produce a frequency count of consecutive bit strings for any specified length. You were instructed to test the program on the one millionth prime number raised to the 100th power, examining bit strings of length 4. The millionth prime to the 10th power was given, so finding the test number was not the problem—just raise the given number to the 10th power.

Solution

If you have *Mathematica*, the millionth prime number raised to the 100th power can be computed directly.

```
n = Prime[10^6]^100;
```

Since there are over 700 digits in this number, let's save some space and shorten up the output showing the beginning and ending digits and hide the middle 664 digits. This is done in *Mathematica* as follows:

```
n // Short
```

```
985238294443356323041458107 <<664>>
  765401675758180369471876400l
```

The development of a fast algorithm for converting any integer into its base b expansion was given with the problem in the November/December 1998 column. Here is the resulting *Mathematica* code:

```
baseExpansion[x_, b_] := Module[{q = x,
    ans = {}},
  While[q ≠ 0, AppendTo[ans, Mod[q, b]];
    q = Floor[q/b]];
Reverse[ans]]
```

Now we can find the binary representation of our number.

```
n2 = baseExpansion[n, 2];
```

We have over 2300 digits in this base 2 expansion, so again we show a shortened part of the number.

```
FromDigits[n2] // Short
```

```
101011100011101010111110010 <<2335>>
  011011011001100001111100001
```

To look at all bit strings of length 4, we need only partition n_2 into lists of length 4 offset by just one bit.

```
bitStrings = Partition[n2, 4, 1];
bitStrings // Short
```

```
{{1, 0, 1, 0}, {0, 1, 0, 1},  <<2382>>,
  {0, 0, 0, 0}, {0, 0, 0, 1}}
```

We find the unique expansions by taking the **Union** of the bitStrings. It is not a big surprise that we find that 16 possible bit strings occur somewhere in n_2.

```
uniqueBitStrings = Union[bitStrings]
```

```
{{0, 0, 0, 0}, {0, 0, 0, 1}, {0, 0, 1, 0},
 {0, 0, 1, 1}, {0, 1, 0, 0}, {0, 1, 0, 1},
 {0, 1, 1, 0}, {0, 1, 1, 1}, {1, 0, 0, 0},
 {1, 0, 0, 1}, {1, 0, 1, 0}, {1, 0, 1, 1},
 {1, 1, 0, 0}, {1, 1, 0, 1}, {1, 1, 1, 0},
 {1, 1, 1, 1}}
```

Now we can do the frequency count on the occurrence of each four bit string in n_2 and make a table of the results.

```
Table[{uniqueBitStrings[[i]], Count[bitStrings,
  uniqueBitStrings[[i]]]}, {i, 1, 16}] //
MatrixForm
```

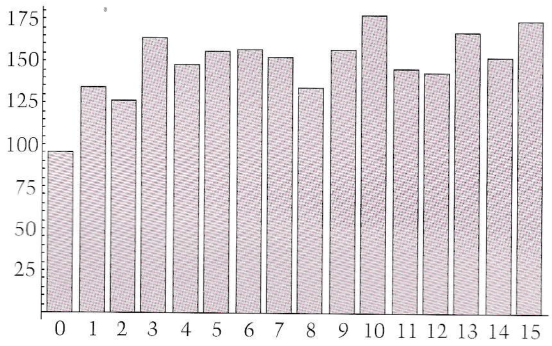
$$\begin{pmatrix} \{0,0,0,0\} & 96 \\ \{0,0,0,1\} & 134 \\ \{0,0,1,0\} & 126 \\ \{0,0,1,1\} & 164 \\ \{0,1,0,0\} & 148 \\ \{0,1,0,1\} & 156 \\ \{0,1,1,0\} & 157 \\ \{0,1,1,1\} & 153 \\ \{1,0,0,0\} & 134 \\ \{1,0,0,1\} & 157 \\ \{1,0,1,0\} & 178 \\ \{1,0,1,1\} & 146 \\ \{1,1,0,0\} & 143 \\ \{1,1,0,1\} & 167 \\ \{1,1,1,0\} & 153 \\ \{1,1,1,1\} & 174 \end{pmatrix}$$

We assigned each of the bit patterns its corresponding decimal number

$\{0, 0, 0, 0\} \to 0$, $\{0, 0, 0, 1\} \to 1$,
$\{a, b, c, d\} \to a2^3 + b2^2 + c2^1 + d, \ldots \{1, 1, 1, 1\} \to 15$,

and graph the results.

```
Needs["Graphics`Graphics`"]
BarChart[Table[{Count[bitStrings,
   uniqueBitStrings[[i]]], i - 1}, {i, 1, 16}]]
```



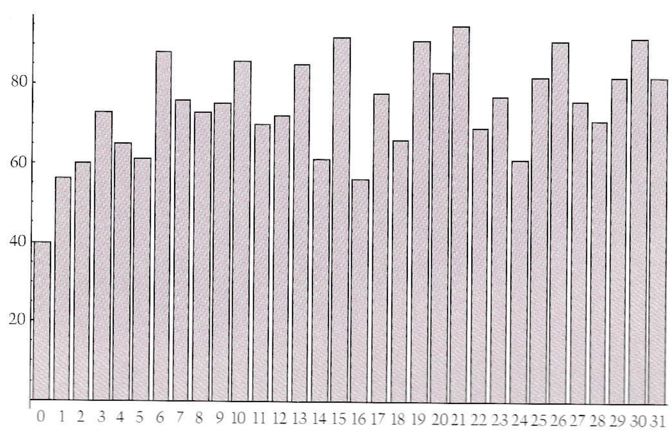
A new function, called bitStringDistribution, can be built in *Mathematica* to carry out the steps just presented.

```
bitStringDistribution[number_,
  bitStringLength_] := Module[{bitStrings,
  uniqueBitStrings},
bitStrings = partition[baseExpansion[
    number, 2], bitStringLength, 1];
  uniqueBitStrings = Union[bitStrings];
BarChart[Table[Count[bitStrings,
  uniqueBitStrings[[i]]], i - 1},
  {i, 1, 2^bitStringLength}]]]
```

Let's try it out by examining the bit strings of length 5 in the binary expansion of Prime$[10^6]^{100}$.

```
bitStringDistribution[Prime[10^6]^100, 5]]
```

It was a surprise to me that the distribution was not more uniform. Why should 11111 occur twice as often as 00000 in the binary expansion of Prime$[10^6]^{100}$?



And Finally . . .

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

If you are interested in learning more about any International Science or Mathematics Olympiad, hopscotch over to Dr. Tom's website at http://olympiads.win.tue.nl/. As always, the USA Computing Olympiad web site is http://www.usaco.org.

Index of Advertisers