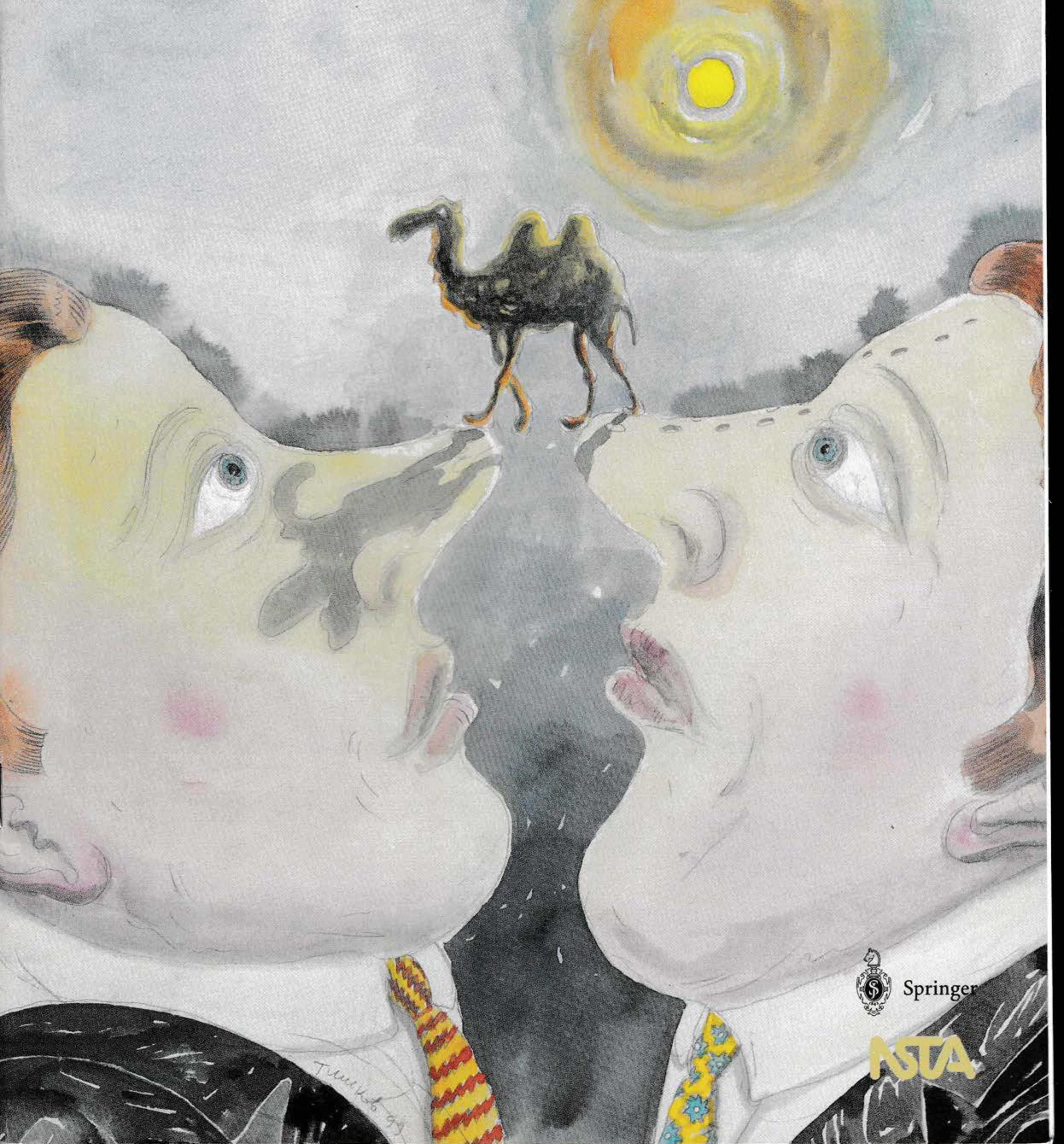
QUANTUM

SEPTEMBER/OCTOBER 1999

US $7.50/CAN $10.50



Springer

NSTA

Oil on canvas, 54 1/8 × 67 7/8, Samuel H. Kress Collection. © 1999 Board of Trustees, National Gallery of Art, Washington, D.C.

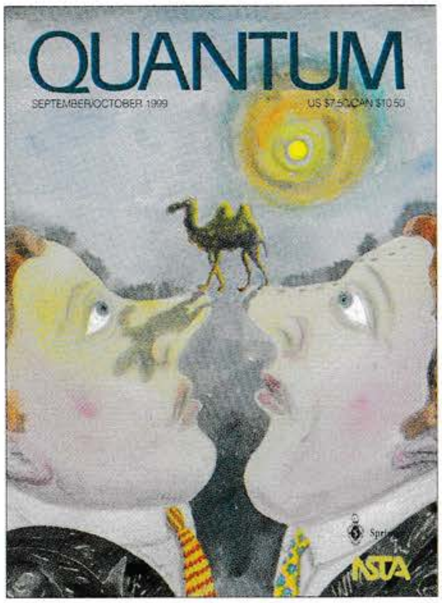
Laocoön (c. 1610–1614) by El Greco

LAOCOÖN, A MYTHICAL PRIEST OF TROY, IS depicted above after incurring the wrath of the goddess Athena. His crime? Hurling a spear at the Trojan horse to prove it was hollow. Unfortunately for Laocoön, the horse had been dedicated to Athena and she took offense at his desecration of the object. Consequently, vipers were sent to dispatch Laocoön and his sons.

Because Laocoön's warnings were dismissed, Troy fell and its denizens disappeared into the collective melting pot of time. However, one only has to look skyward (along with the reclining Laocoön) to find Trojans in our modern world. Learn how the ancients' spheres of influence continue to interact by reading "When Trojans and Greeks Collide" on page 16.

QUANTUM

SEPTEMBER/OCTOBER 1999 VOLUME 10, NUMBER 1



Cover art by Leonid Tishkov

One hump or two? Unfortunately, the answer may not always be on the tip of your nose. If the camel in question is some distance away, the laws of physics and optics dictate how close you would have to be to distinguish between the two varieties.

Fortunately, calculating the distance is a bit easier than, say, trying to lead a camel through the eye of a needle. To find out just how close you'd have to be to count the protrusions, turn to "Physical Optics and Two Camels" on page 44.

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

FEATURES

DEPARTMENTS

NSTA *Quantum* (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 10 (6 issues) will be published in 1999–2000. *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: $49. 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

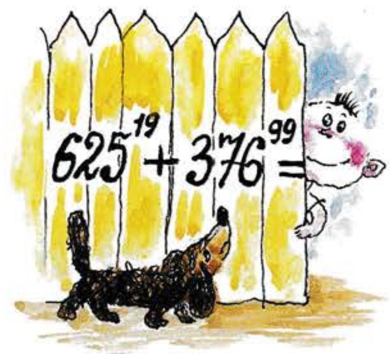
A publication of the National Science Teachers Association (NSTA)
& Quantum Bureau of the Russian Academy of Sciences
in conjunction with
the American Association of Physics Teachers (AAPT)
& the National Council of Teachers of Mathematics (NCTM)

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!

B271

Positive outlook. Do six different positive numbers exist such that their sum equals their product?



B272

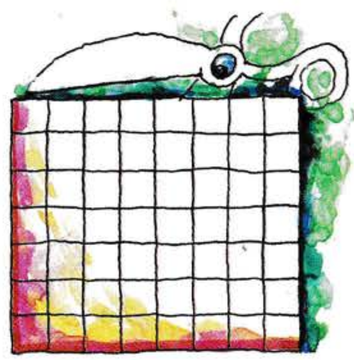
Ultimate trio. Find the last three digits of the sum $625^{19} + 376^{99}$.

B273

Radical pattern. Find $\sqrt{12345678987654321}$.

B274

Choice cuts. A 7×8 rectangle is cut from a sheet of graph paper. Cut this rectangle into polygons consisting of no more than 5 squares each in such a way that the total length of the cuts is minimized (the cuts must follow the lines on the paper).



Art by Pavel Chernusky

B275

Dry the hay. To dry freshly mown hay, one must stir and turn it over frequently. Why?

ANSWERS, HINTS & SOLUTIONS ON PAGE 54

Electric multipoles

How a little order can weaken your potential

by A. Dozorov

THE ELECTRIC POTENTIAL GENERATED BY A point charge at some point in space is inversely proportional to the distance between the point and the charge. At first glance, it seems that any arbitrary set of electric charges located in some region would create a potential that is also inversely proportional to the distance to this cluster of charges. However, in general this is not true. If we arrange the charges in a certain order, we can obtain a potential that is inversely proportional to any integer power of the distance.

To prove this curious feature of the electric potential, we need only one mathematical fact: if the absolute value of $|x|$ is less than one, the following formula is valid:

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots \qquad (1)$$

This is the well-known formula for the sum of an infinite descending geometric progression.

Now we consider various systems of electrical charges. The potential ϕ generated by a point charge q at some distance r is inversely proportional to the distance $\phi = kq/\varepsilon r$, where the proportionality coefficient k depends on the accepted system of units. In SI (International System of Units), $k = 1/4\pi\varepsilon_0$. However, theoreticians prefer to use CGSE (centimeter-gram-second electrostatic system), because in this system $k = 1$. To further simplify our problem, we assume all charges to be located in vacuum ($\varepsilon = 1$).

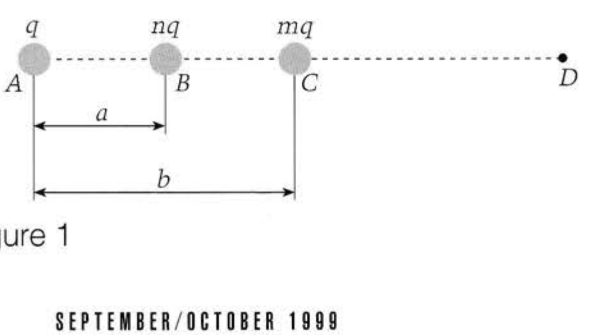


Figure 1

For example, consider three point charges located on the same line at some distance from each other (fig. 1). Let $AB = a$, $AC = b$, and the charges be q, nq, and mq, where n and m are integers. Now we calculate the electric potential at a point D located on the extension of segment AC at a distance r from point A that is sufficiently far from all three charges. The latter condition can be written as $r >> a + b$. According to the superposition principle, the potential ϕ at point D equals the algebraic sum of the potentials generated by each charge:

$$\phi = \frac{q}{r} + \frac{nq}{r-a} + \frac{mq}{r-b} = \frac{q}{r}\left(1 + \frac{n}{1-\frac{a}{r}} + \frac{m}{1-\frac{b}{r}}\right).$$

Since we consider the case when $r >> a + b$, we can apply formula (1). Thus, we have

$$\begin{aligned}\phi &= \frac{q}{r}\left[1 + n\left(1 + \frac{a}{r} + \frac{a^2}{r^2} + \frac{a^3}{r^3} + \ldots\right) + m\left(1 + \frac{b}{r} + \frac{b^2}{r^2} + \frac{b^3}{r^3} + \ldots\right)\right] \\ &= q\frac{1+n+m}{r} + q\frac{na+mb}{r^2} + q\frac{na^2+mb^2}{r^3} + q\frac{na^3+mb^3}{r^4} + \ldots \end{aligned} \qquad (2)$$

This equation shows that at large distances the absolute value of each subsequent term is much smaller than that of the previous one, provided the numerator of the previous item is not zero. For example, if the sum of the charges is not zero ($q + nq + mq \neq 0$), the major

Art by Sergey Ivanov

role in equation (2) is played by the first term, so the total potential of the system of charges will be inversely proportional to the first power of distance. By contrast, if the sum of the charges is zero (the system is neutral as a whole), the key role will be played by the second term: the potential will be inversely proportional to the square of the distance between the charges and the test point. However, we can place the charges in such a way that the first and second terms are both zero. Equation (2) shows that the two requisite conditions are

$$q + nq + mq = 0 \qquad (3a)$$

and

$$q(na + mb) = 0. \qquad (3b)$$

Condition (3a) specifies that the total charge of the system must be zero, so the charges cannot all have the same sign. Eliminating the arbitrary charge q from equations (3a) and (3b), we get two equations with four parameters n, m, a, and b. This means that there is an infinite number of variants of arrangements and values of the charges that satisfy the set of equations (3). We can arbitrarily choose any two parameters. Let $b = 2a$ (that is, $AB = BC$). In this case equations (3a) and (3b) yield $n = -2$ and $m = 1$.

Therefore, charge q must be placed at point A, charge $-2q$ at point B, and charge q at point C (figure 1). The distances between the neighbors must be equal: $AB = BC = a$.

Such a system of charges generates a potential described (at large distances) by the following formula (see formula (2)):

$$\phi = \frac{2qa^2}{r^3} + \frac{6qa^3}{r^4} + \ldots .$$

Here the major role is given to the first term, so the potential is inversely proportional to the third degree of the distance.

When the following requirement is met in addition to (3a) and (3b):

$$na^2 + mb^2 = 0, \tag{3c}$$

formula (2) yields a potential proportional to r^{-4}. However, the system of equations (3a–3c) has no solutions. Indeed, let's multiply equation (3b) by a and compare the result with (3c). In this way we obtain $a = b$, and (3b) reduces to the equation $n + m = 0$, while (3a) leads to the contradiction $1 = 0$ (provided $q \neq 0$, but the case $q = 0$ is of no physical interest). We obtain a similar result for all other terms in formula (2).

Thus, depending on their values and relative locations, the three charges shown in figure 1 can generate only potentials that at large distances are proportional to $1/r$, $1/r^2$, or $1/r^3$.

A field generated by an arbitrary system of electric charges can be considered in a similar way. In addition, there exists an elegant method of constructing a system of charges that at large distances generates a field with the potential

$$\phi_n = \frac{C_n}{r^{n+1}}, \tag{4}$$

where n is an integer and C_n is a constant determined by the values of the charges and their arrangement. The charge system that generates a field with potential (4) is called an n-order electric multipole or 2^n-pole. The simplest case $n = 0$ (0th order multipole) corresponds to a single point charge. In the general case the n-th order multipole is formed by 2^n charges. It turns out that if we have an n-th order multipole (2^n-pole), it is easy to construct the $(n + 1)$-th order multipole. To this end we supply the initial n-th multipole (2^n-pole) with the same n-th order multipole (2^n-pole) but shifted symmetrically to some distance and composed of the opposite charges. In this way we obtain a system composed of two opposite 2^n-poles. The resulting system forms a multipole of the $(n + 1)$-th order with $2(n + 1)$ charges (2^{n+1}-pole).

Let's consider some examples. A point charge $-q$ is a multipole of the 0th order. As a first step, we use it to construct a multipole of the first order (2-pole), called a *dipole*. To this end, we shift the charge $-q$ a distance l to the right and change its sign. As a result, we obtain the dipole shown in figure 2. It is characterized by the so-called dipole moment $\mathbf{d} = q\mathbf{l}$ (vector $\mathbf{l}$ is directed from the negative to the positive charge, its value being equal to the distance between the charges).

Let's calculate the potential ϕ_1 generated by the dipole at a distant point D lying on the extension of the line segment connecting the charges ($AD = r >> l$):

$$\phi_1 = \frac{-q}{r} + \frac{q}{r-l} = \frac{q}{r}\left(-1 + \frac{1}{1-\frac{l}{r}}\right).$$

Using formula (1) we get the major term of potential generated by a dipole:

$$\phi_1 \cong \frac{ql}{r^2} = \frac{d}{r^2}. \tag{5}$$

If the observation point does not lie on the extension of the line connecting the charges, two coordinates determine its position. The first coordinate is the distance to one of the charges (for a distant observation point where $r >> l$, it doesn't matter which of the two charges is used to this end). The second coordinate may be the angle between the dipole moment and the radius-vector drawn from a charge (again for $r >> l$ it doesn't matter which charge is chosen) to the observation point. So, point C (fig. 2) is characterized by the distance $\rho = BC$ and angle θ. In this case calculating the potential is more difficult. The result is

$$\phi_1 = \frac{d\cos\theta}{\rho^2}. \tag{5a}$$

To make the next step and construct a second-order multipole (2^2-pole or quadrupole), we must add to our dipole a similar dipole with opposite charges and shift it some distance (note, we do not rotate the original dipole: both its charges are shifted identically). Let's shift the dipole a distance l along the extension of the line connecting its charges (figure 3a). We could make a smaller shift, but the calculations are lengthier.

When the charges are situated on the same line, the multipole is called *axial*. Figure 3b shows such an axial quadrupole originated from the charge system given in figure 3a. In the general case (figure 3c) a quadrupole is composed of unlike charges of equal value located at the

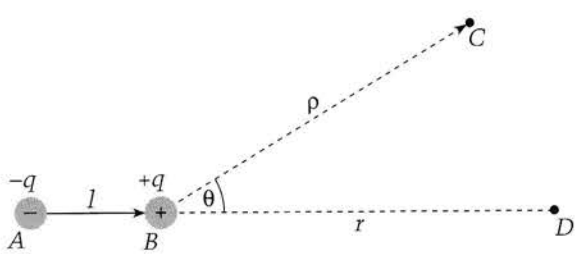


Figure 2

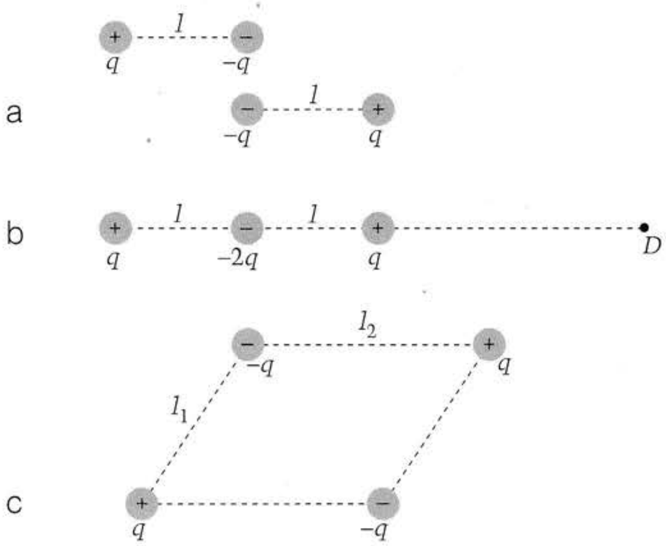


Figure 3

vertices of a parallelogram. According to formula (4), the potential of a quadrupole ϕ_2 must be proportional to $1/r^3$. Therefore, in series (1) we must take into account only the numbers up to x^2.

Now let's calculate the coefficient C_2 for the axial quadrupole shown in figure 3b. For a distant point D on the axis of the quadrupole we have

$$\phi_2 = \frac{q}{r} + \frac{(-2q)}{r-l} + \frac{q}{r-2l}$$
$$\cong \frac{q}{r}\left[1 - 2\left(1 + \frac{l}{r} + \frac{l^2}{r^2}\right) + \left(1 + \frac{2l}{r} + \frac{4l^2}{r^2}\right)\right] = \frac{2ql^2}{r^3}.$$

The next step is to form a third-order multipole usually called an *octupole*, because in general it has 8 equal-magnitude charges. To this end we supply the axial quadrupole (figure 3b) with the symmetrical axial quadrupole as shown in figure 4a. As a result, we get an

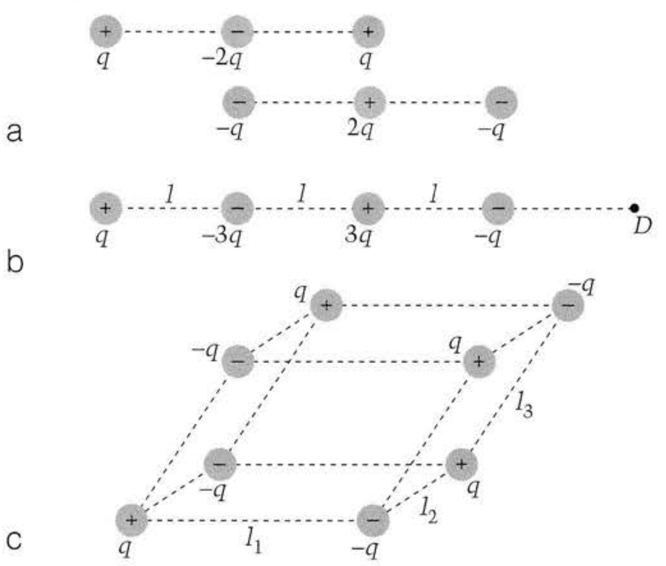


Figure 4

axial octupole (figure 4b). In the general case a quadrupole (figure 3c) produces an octupole (figure 4c) with the charges located at the vertices of a parallelepiped. According to (4), the potential ϕ_3 of the electrostatic field generated by an octupole must be proportional to $1/r^4$. Let's find the potential generated by the axial octupole at point D:

$$\phi_3 = \frac{q}{r} + \frac{(-3q)}{r-l} + \frac{3q}{r-2l} + \frac{(-q)}{r-3l}$$
$$\cong \frac{q}{r}\left[1 - 3\left(1 + \frac{l}{r} + \frac{l^2}{r^2} + \frac{l^3}{r^3}\right) + 3\left(1 + \frac{2l}{r} + \frac{4l^2}{r^2} + \frac{8l^3}{r^3}\right) - \left(1 + \frac{3l}{r} + \frac{9l^2}{r^2} + \frac{27l^3}{r^3}\right)\right]$$
$$= -\frac{6ql^3}{r^4}.$$

The fields and potentials of the higher order multipoles are calculated in a similar way.

The field of any system of electric charges can be represented at large distances as a sum of the fields generated by multipoles of different orders. The higher the order of the multipole approximating the analyzed system, the more "neutral" this system is and the more rapidly its field decreases with distance.

We have considered the electric potentials generated by axial multipoles along their axes. For an arbitrary point, the potential calculations are made in a similar way, but they are rather cumbersome.

If the potential of a multipole is known ($\phi_n = C_n/r^{n+1}$), it determines the values of the electrical field E and the force $F = qE$ affecting a probe charge q by the multipole. Since $n \geq 0$ for any multipole,

$$E \cong \frac{1}{r^{n+2}}. \qquad (6)$$

Circumventing the mathematical details, let's consider the lines of force of a dipole and an axial quadrupole shown in figures 5a and 5b, respectively. For the dipole (figure 5a) the electric field is obtained at each point by computing the vector sum of the fields $\mathbf{E}_+$ and $\mathbf{E}_-$ generated, respectively, by the positive and negative charges: $\mathbf{E} = \mathbf{E}_+ + \mathbf{E}_-$. Figure 5a shows these vectors for one point in space. A similar plot is shown in figure 5b for an axial quadrupole. At each point the total field $\mathbf{E}$ is the sum of three vectors: two intensity vectors $\mathbf{E}_+$ produced by the positive charges, and one vector $\mathbf{E}_-$ corresponding to a negative charge with a magnitude equal to the sum of the positive charges. Three-dimensional plots of the lines of force for the electric dipole and quadrupole are obtained by rotating figures 5a and 5b about the corresponding axes of symmetry.

In addition to electrical systems, we can consider magnetic multipoles—systems composed of magnets or

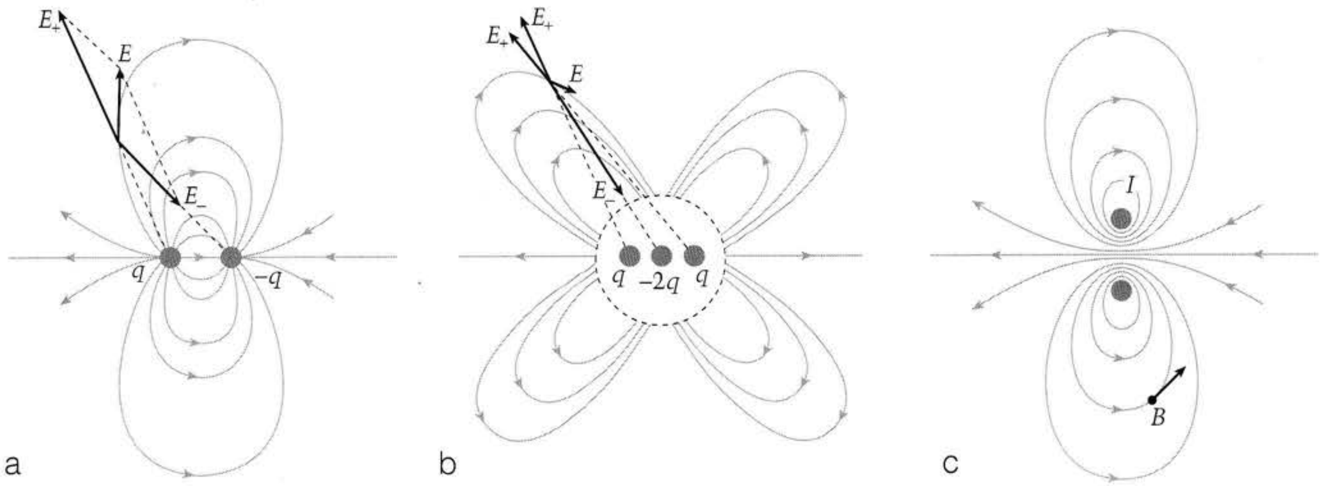


Figure 5

closed currents. However, a magnetic system has an important distinction: single magnetic charges (monopoles) have not been found in nature, so the elementary unit of the magnetic system is a magnetic dipole. Figure 5c shows the lines of forces produced by a magnetic dipole generated by a circular current I whose plane is perpendicular to the plane of the page. At large distances the field plots shown in figures 5a and 5c are identical.

The similarity between magnetic and electric dipoles can be illustrated by another example. The needle of a magnetic compass is a magnetic dipole oriented along the lines of a magnetic field. In a similar way, an electric dipole turns in an electric field: it also assumes the direction along the lines of forces, thereby playing the role of an "electrical compass" (figure 6).

Examples of electric multipoles are atoms and molecules. If during the formation of a molecule, the electrons are redistributed between the atoms in such a way that the centers of "gravity" of the positive and negative charges do not coincide, the molecule acquires an intrinsic dipole moment and is called *polar*. For example, molecules of hydrochloric acid and water are polar (figure 7). If the unlike charges in such a molecule are spaced by a distance about the radius of a hydrogen atom ($l = 0.5 \cdot 10^{-8}$ cm), the dipole moment of the molecule is on the order of $d = el \cong 2.4 \cdot 10^{-18}$ cm CGSE unit charge. (In the CGSE system, the charge of an electron is $e = 4.8 \cdot 10^{-10}$ CGSE). The order of magnitude of this value corresponds to the experimental data shown in figure 7.

If a molecule is composed of similar atoms (O_2, H_2, or Cl_2), the electrons cannot "recognize" their "native" atom and thus they are located homogeneously between both atoms, which yields a dipole moment of zero. Such molecules are called *nonpolar*. Figure 7 shows a linear molecule of carbon dioxide, CO_2, which is nonpolar. The appearance of this molecule suggests that the redistribution of electrons between the atoms should produce an axial quadrupole (compare it with figure 3b), while the electric field generated by a carbon dioxide molecule should look like that shown in figure 5b (but in the opposite direction).

More complicated distributions of atoms and electrons in molecules yield multipoles of higher orders.

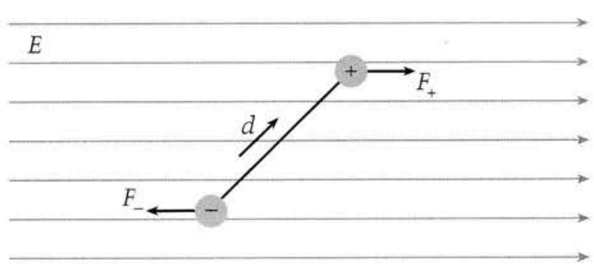


Figure 6

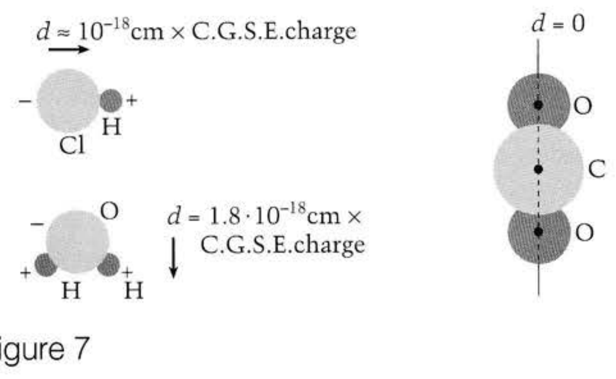


Figure 7

Quantum on electro- and magnetostatics:

A. Eisenkraft, L. D. Kirkpatrick, Physics Contest—series of installments on electrostatics: July/August 1992, p. 24; January/February 1993, p. 44; November/December 1993, p. 46; May/June 1994, p. 40.

A. Stasenko, "Love and hate in the molecular world," November/December 1994, pp. 10–13.

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

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

A. Leonovich, "Do you have potential?" November/December 1998, pp. 28–29.

HOW DO YOU FIGURE?

Challenges

Math

M271

Systems engineering. Solve the following system of equations:

$$\begin{cases} x+[y]+\{z\}=3.9 \\ y+[z]+\{x\}=3.5 \\ z+[x]+\{y\}=2, \end{cases}$$

where $[x]$ denotes the integer part of x (that is, the maximum integer value not exceeding x) and $\{x\} = x - [x]$ denotes the fractional part.

M272

Looking for x. Solve the equation

$$x+\sqrt{x^2-9}=\frac{2(x+3)}{(x-3)^2}.$$

M273

Locus pocus. An isosceles triangle ABC ($AB = BC$) is given in a plane. Find the locus of points M in the plane such that $ABCM$ is a convex quadrilateral and $\angle MAC + \angle CMB = 90°$.

M274

Cubic edge. The distances from all vertices of a cube and from the centers of its faces to a certain plane (14 quantities in all) take two different values, and the lower value is 1. What can the length of the cube's edge be?

M275

What's your angle? In a triangle ABC, angle B is obtuse and its measure is α. The bisectors of angles A and C intersect the opposite sides at points P and M, respectively. Points K and L are taken on side AC such that $\angle ABK = \angle CBL = 2\alpha - 180°$. Find the angle between lines KP and LM.

Physics

P271

Pinocchio's cap. Old Geppetto made a cap out of thin tin for his beloved creation Pinocchio . The cap had a conical shape with height H = 20 cm and vertex angle α = 60°. Will the cap be in stable equilibrium sitting on Pinocchio's head if his head is a sphere of diameter D =15 cm? (S. Krotov)

P272

De nihilo per astra. According to one cosmological hypothesis, stars are formed from interstellar gas and dust due to compression produced by gravitation. Evaluate the time needed to make a star from a gigantic spherical cloud with density $\rho = 2 \cdot 10^{-20}$ g/cm^3. (We assume that under compression the particles do not pass one another. The gravitational constant $G = 6.67 \cdot 10^{-11}$ N·m^2/kg^2.) (V. Skorovarov)

P273

Loop on a soap film. A loop of thread of length l floats on a soap film (figure 1). The part of the film located inside the loop is carefully pierced. What geometrical shape will the loop assume? What is the equilibrium tension in the thread if the coefficient of surface tension for the soap solution is σ? (A. Buzdin, S. Krotov)

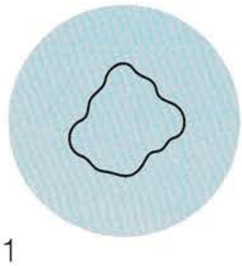


Figure 1

P274

Electron in a magnetic field. An electron flies into a homogeneous magnetic field. At point A its velocity is $\mathbf{v}$, which makes an angle α with the direction of the magnetic field (figure 2). For what values of the magnetic field will the electron arrive at point C? The electron's charge is e, its mass m, and the distance AC is L.

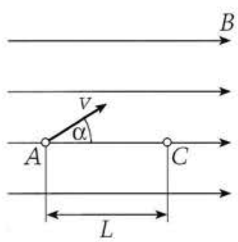


Figure 2

P275

A lake in a desert. Travelers in the desert sometimes observe what appears to be a sea or lake. At what distance from the observer does such a mirage appear? Assume that the speed of light near the ground in the desert varies according to the formula $c(z) = c_0(1 - az)$, where c_0 is the speed of light at the ground and z is the altitude above the ground. (B. Klyachin)

ANSWERS, HINTS & SOLUTIONS ON PAGE 51

Thoroughly modern Diophantus

The arithmetic of elliptic curves

by Y. Solovyov

MODERN MATHEMATICS has inherited from antiquity several great writings. One of them is the *Arithmetica* by Diophantus of Alexandria. Written in the third century A.D., it disappeared for more than 1000 years and was believed to be lost. It was not seen again until 1464, when the German scientist Regiomontanus (1436–1476) found 6 of the 13 volumes of *Arithmetica*. The first Latin translation of this book was printed in 1575. When the edition prepared by Claude-Gaspar Bachet de Méziriac appeared in 1621, it became the reference book for many mathematicians, such as Pierre de Fermat (1601–1665) and René Descartes (1596–1650).

The book did not seem at all obsolete, despite the thousand years of oblivion. In fact, it left the best works on algebra of the sixteenth century far behind. For example, unlike European algebraists of that time, Diophantus operated freely with negative and rational numbers, used letter notation for equations, and most important, was able to find integer and rational solutions of linear, quadratic, and cubic equations and of systems of equations with integer coefficients in two or more variables. The solution of such equations (now called *diophantine*) has ever since remained an important subject of mathematical investigations.

Now we are ready to consider the solutions of several diophantine equations (I've tried to choose the most beautiful of them). To solve these equations, one must not only read Diophantus' great book carefully, but also get in touch with the latest events in modern mathematics.

Diophantus' method of secants

Let's illustrate this method using a particular case of a problem solved by Diophantus in his *Arithmetica*. Consider the following equation:

$$x^2 - y^2 = 1. \qquad (1)$$

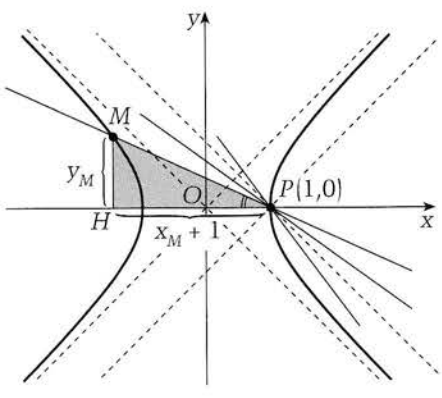


Figure 1

Suppose that we must find all its rational solutions—that is, all the ordered pairs

$$(x,y) = \left(\frac{a}{b}, \frac{c}{d}\right), a, b, c, d \in \mathbf{Z},$$

that satisfy equation (1).

We can consider equation (1) (or any other equation in the variables x and y) as a curve on the coordinate plane. In the present case, it is a hyperbola (fig. 1). The solution (1, 0), corresponding to the point P in which the hyperbola meets the x-axis, strikes our eyes at once. Let's draw a secant line through this point, with slope k. It's equation will be

$$y = k(x - 1). \qquad (2)$$

Now let us find the second point where this line intersects the curve with equation (1). We just substitute the right side of equation (2) for y in equation (1) and solve the resulting quadratic equation for x. We find that

$$x = \frac{-k^2 \pm 1}{1 - k^2}.$$

We already know one of the roots, namely $x_1 = 1$ (it corresponds to the point (1, 0)), and the second root

Art by V. Ivanyuk

$$x_2 = \frac{k^2+1}{k^2-1}$$

gives us the desired second point of intersection:

$$(x_2, y_2) = \left(\frac{k^2+1}{k^2-1}, \frac{2k}{k^2-1}\right). \quad (3)$$

For all rational k ($k \neq \pm 1$), this formula determines a point on the curve and thus a rational solution of the given equation. (When $k = \pm 1$, the secant intersects the hyperbola only at the point P (fig. 1).) And conversely, for any rational solution (for any rational point M on the curve), secant PM is determined by equation (2) with *rational* k (because the legs of the right triangle PMH are rational in this case).

Thus, *when k takes all possible rational values* ($k \neq \pm 1$), *equation (3) gives all possible rational solutions of the equation (1).*

Diophantus did not introduce a coordinate system, nor did he consider the curve corresponding to the given equation. In fact, the coordinate approach in geometry first appeared in the seventeenth century in the work of Descartes. Diophantus introduced the substitution from equation (2) in a purely algebraic way and then obtained equation (3) (in different notation, of course). Moreover, he realized that the method illustrated was applicable not only to the polynomial $x^2 - y^2 - 1$ but also to a general second degree polynomial in two variables:

$$p(x, y) = ax^2 + bxy + cy^2 + dx + ey + f,$$

where $a, b, \ldots, f$ are integers or rational numbers, if one has found at least one rational root of the equation.

Not every curve defined by a second-degree polynomial contains rational points. For example, there are no such points on the circle $x^2 + y^2 = 3$ or on the ellipse $x^2 + 82y^2 = 3$. Still, there are rational points $(a/c, b/c)$ on the circle $x^2 + y^2 = 1$. A triple of integers (a, b, c) defined by such a point is called a *Pythagorean triple*, because it satisfies the relation $a^2 + b^2 = c^2$, which appears in the statement of the Pythagorean theorem. We can find all Pythagorean triples using the method of secants (as well as in other ways).

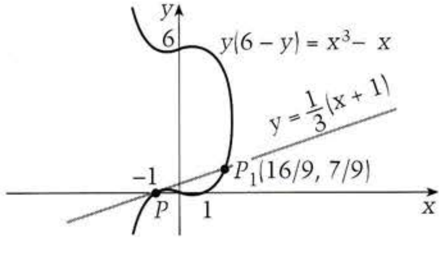


Figure 2

The problem of the existence of a rational point on a second-degree curve turned out to be very difficult. The first nontrivial advances were made by Indian mathematicians Brahmagupta (598–c. 665) and Bhaskara II (1114–c. 1185), and the complete answer was not found until 1768 by the famous French mathematician Joseph-Louis Lagrange (1736–1813).

Diophantus did not confine himself to second-degree equations. He successfully coped with cubic equations and formulated a general approach to such equations, as we will see in the following section.

Tangent to a curve

In one problem from *Arithmetica*, one must find a rational solution of the equation

$$y(6 - y) = x^3 - x. \quad (4)$$

Diophantus' solution is brief and brilliant. Let's try, he wrote, the substitution $x = 2y - 1$. We obtain

$$6y - 6y^2 = 8y^3 - 12y^2 + 4y.$$

If the 6 were a 4, then the linear terms would have nicely vanished! But the 4 appeared from the 2 in the substitution $x = 2y - 1$. So, let's replace 2 with 3—that is, let $x = 3y - 1$. Then the linear terms disappear and we obtain

$$y^2(9y - 7) = 0, \quad (5)$$

and thus $y = 7/9$ and $x = 16/9$. We've found the rational solution (16/9, 7/9) of cubic equation (4).

At first glance, there is nothing special in this solution. We simply guessed that we should use the substitution $x = 3y - 1$, which helped us find the solution. What profound idea is hidden here? To answer this question, let's look at the coordinate plane once again and draw the graph of equation (4) (see fig. 2). (Later in the article we explain how to draw such graphs.) The gray line in figure 2 is the line $x - 3y + 1 = 0$. It is tangent to our curve at the point $P(-1, 0)$ (in fact, equation (5) has, in addition to the root $y = 7/9$, two "equal roots" $y^2 = 0$).

We could continue this procedure by drawing another tangent to curve (4) through the rational point (16/9, 7/9). The reader can confirm that this line meets the curve in a third rational point, and so on. But Diophantus didn't take this step, and more than 1500 years passed until mathematicians could use Diophantus' ideas in their full generality.

Curves of degree 3

Continuing the geometric approach we've used thus far, let's concentrate not on the solution of the third-degree equations, but on the following equivalent question: *What rational points are there on the plane curve determined by the third-degree equation*

$$f(x, y) = ax^3 + bx^2y + \ldots + hx + iy + j = 0$$

with integer coefficients?

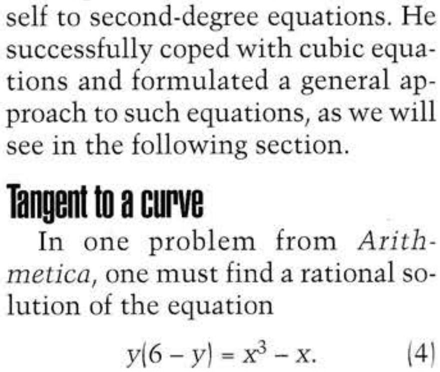


Figure 3

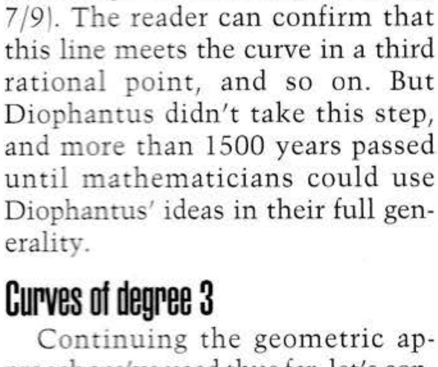


Figure 4

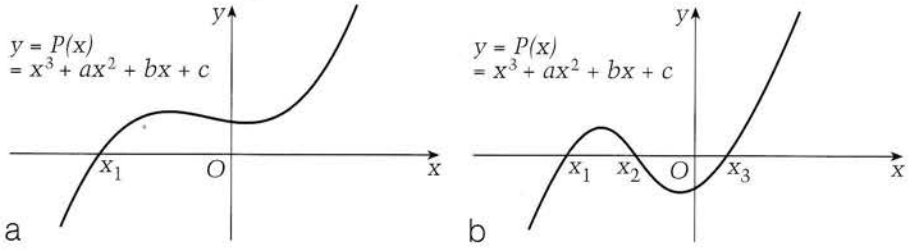


Figure 5

We can divide all curves of this sort into two huge classes. The first class is composed of all the curves with cusps (the reader is invited to check that the origin is such a point on the curve $y^2 = x^3$) or self-intersections (fig. 3) as well as all curves for which there exists a decomposition (factorization)

$$f(x, y) = f_1(x, y) \cdot f_2(x, y),$$

where $f_1(x, y)$ and $f_2(x, y)$ are polynomials of smaller degrees (fig. 4). Such curves are called *degenerate*. The second consists of all non-degenerate curves determined by third-degree polynomials with integer coefficients. Such curves are called *elliptic*.[1] It is this, the most general class, that attracts our attention. All the elliptic curves that we will consider will be given in the *canonical form*

$$y^2 = x^3 + ax^2 + bx + c \qquad (6)$$

with integer coefficients a, b, and c, such that the polynomial

$$P(x) = x^3 + ax^2 + bx + c$$

has no multiple roots.

The assumption that the equation of each of our curves can be written in this canonical form does not decrease the generality of our reasoning: Each nondegenerate curve $f(x, y) = 0$ can be rewritten in the form of equation (6) using a suitable substitution. If the coefficients of $f(x, y)$ are integers, then the problem of finding all rational points on the curve $f(x, y) = 0$ can be reduced to the similar problem for a curve written in the form of equation (6) with integers a, b, and c.

Graphs of elliptic curves

First, let's find out what curve (6) looks like. The easiest way to draw it is as follows. Take the graph of the function

$$y = \sqrt{x^3 + ax^2 + bx + c}$$

and reflect it with respect to the x-axis. To draw the graph of this function, we start with the graph of $y = x^3 + ax^2 + bx + c$. It is well known that every third-degree polynomial (without multiple roots) has either one or three real roots. Therefore, the graph of $y = x^3 + ax^2 + bx + c$ looks as it is shown in figures 5a and 5b. Now we easily obtain the graph of the function

$$y = \sqrt{x^3 + ax^2 + bx + c}$$

(see fig. 6a) and thus the shape of the elliptic curve $y^2 = x^3 + ax^2 + bx + c$ (fig. 6b). Figure 6 illustrates the case corresponding to figure 5a. We suggest that the reader draw the curve corresponding to figure 5b. It will consist of two parts (see fig. 9).

Note that the graphs of the functions $y = \sqrt{P(x)}$ and $y = -\sqrt{P(x)}$ join smoothly, without forming corners, at the points x_1, x_2, and x_3. This happens because the tangent to the graph $y = \sqrt{P(x)}$ is vertical at these points.

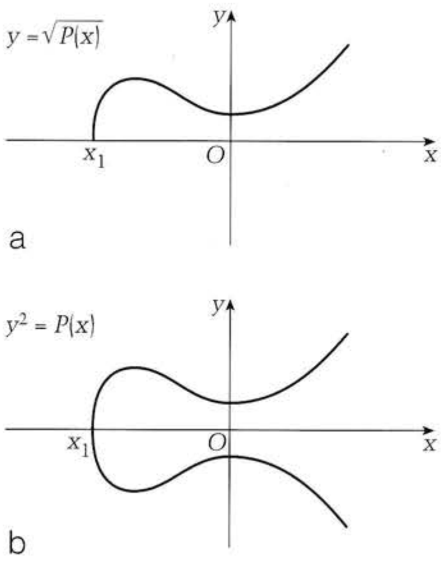


Figure 6

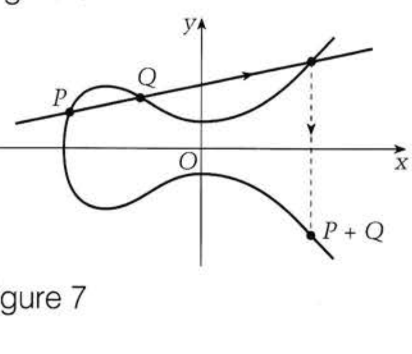


Figure 7

Addition of points on an elliptic curve

When applied to elliptic curve C, the method of secants gives an unexpected result. It turns out that we can "add" the points of C. That is, we can define an operation, which we will call "addition," on the points of C based on the graphical representation of the curve (fig. 7). Take two points P and Q on C and draw a line through them. This line will meet curve C at a third point. Reflect this point with respect to the x-axis. The result is called the *sum* of P and Q. In figure 7 it is denoted by $P + Q$. (However, not every line through two points of C meets C at a third point—for example, a vertical line does not.)

Let's study the properties of our new operation and compare it with the operation of addition of numbers. The latter operation is *commutative*—that is, $a + b = b + a$—and *associative*—that is, $(a + b) + c = a + (b + c)$. Furthermore, this operation has an *identity element*—a number O such that $a + O = a$ for all a. Finally, for each number a there is an *inverse* number—a number $(-a)$ such that $a + (-a) = O$.

And what happens on an elliptic curve? First of all, *the addition of points is commutative*. In fact, to find $Q + P$, we start with the same line as for $P + Q$. Therefore, $Q + P = P + Q$.

[1]Not that these curves are not themselves ellipses. The connection between ellipses and these curves, which is preserved in their name, might make a good subject for another article.

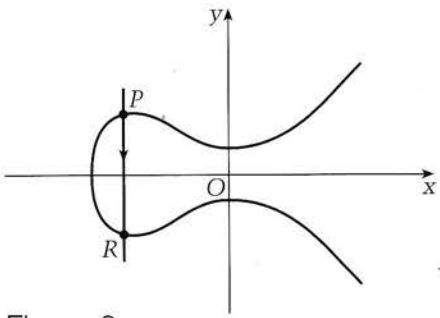


Figure 8

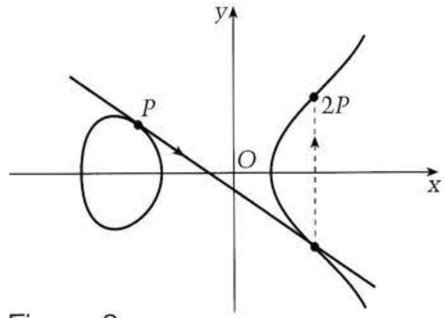


Figure 9

The associativity of the addition of points of an elliptic curve holds too, but it is not easy to prove this. We can see a geometric interpretation of this fact with the help of a drawing.

Now let us take up the question of the existence of an identity element: a point E of the curve such that $P + E = P$, for any other point P of the curve. How can we find such a point? Look at figure 8. Take an arbitrary point P on the curve. We want to find something such that if we draw a line through P and the "something," take the intersection of this line with the curve, and reflect it with respect to the x-axis, then we will come back to P. Let R denote the point symmetric to P with respect to the x-axis. It follows that the line through P and the "something" must pass through P and R, which is to say it must be vertical. Therefore, if a point E exists such that $P + E = P$, it can't lie on the plane, because it must belong to the curve and to the vertical line at the same time.

Since there is no such point E in the plane (and because we need it badly), we will simply attach it to the plane and call it the *point at infinity*. What properties must it have? Every vertical line tends to infinity in two directions: up and down. Let's require that all these infinities correspond to one and the same point E. In other words, we will regard E as *the point where all vertical lines meet*. Point E, the identity with respect to addition, is correctly determined by this requirement. By our definition of E, a vertical line through P also passes through E. Therefore, R, the second point where this line meets the elliptic curve, satisfies the relation $P + R = E$; thus, it is the inverse of P. On the other hand, R is the point symmetric to P with respect to the x-axis. Therefore, every point P of the curve has an inverse $-P = R$. Thus, we've checked that *the addition of points of an elliptic curve satisfies all the properties of the addition of numbers.*

How does one compute $P + P$? When points were different, we drew a secant. And now that they coincide, it's clear that we must draw a tangent (fig. 9).

And what about $3P$? It is very simple: we add $2P$ to P. Similarly, $4P = 3P + P$, $5P = 4P + P$, and so on.

Searching for rational points

Now that we are armed with the operation of addition of points, we will look for the rational points. Let $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ be two rational points on the elliptic curve $y^2 = x^3 + ax^2 + bx + c$, where a, b, and c are integers, and let the line through P and Q meet the curve at a third point $R = (x_3, y_3)$. Then R is a rational point, also.

It is not at all difficult to prove this statement. Indeed, if the line is determined by the equation

$$y = kx + d, \tag{7}$$

then k and d must be rational, because we can express them in terms of the coordinates (x_1, y_1) and (x_2, y_2) of the points P and Q by the formulas

$$k = \frac{y_1 - y_2}{x_1 - x_2},$$

$$d = y_1 - kx_1 = \frac{x_1 y_2 - x_2 y_1}{x_1 - x_2}.$$

Substituting (7) into the equation of the elliptic curve, we obtain the following third-degree equation in x with rational coefficients:

$$(kx + d)^2 = x^3 + ax^2 + bx + c,$$

which can be rewritten as

$$x^3 + (a - k^2)x^2 + (b - 2kd)x + c - d^2 = 0.$$

The relationship between the roots and coefficients of a polynomial equation implies that

$$x_1 + x_2 + x_3 = k^2 - a.$$

Since both x_1 and x_2 are rational, x_3 is rational, too, and thus so is $y_3 = kx_3 + d$.

Using this reasoning, we can easily derive the formula for the coordinates of the point $P + Q$. By definition, $P + Q$ is the symmetric image of R with respect to the x-axis. Therefore, the coordinates (u, v) of the point $P + Q$ are given by the formulas

$$u = k^2 - a - x_1 - x_2$$

and

$$v = -ku - d = -[k(u - x_1) + y_1].$$

Substituting k and d for the expressions we've found above, we finally have

$$\begin{aligned} u &= \frac{(y_1 - y_2)^2}{(x_1 - x_2)^2} - (a - x_1 - x_2), \\ v &= \frac{y_1 - y_2}{x_1 - x_2}(x_1 - u) - y_1. \end{aligned} \tag{8}$$

Clearly, when $x_1 = x_2$, these formulas are meaningless. In this case we must replace the equation of a secant (7) with the equation of a tangent and repeat our reasoning. At last, we obtain

$$\begin{aligned} u &= -2x_1 + a - \left(\frac{3x_1^2 + 2ax_1 + b}{2y_1}\right)^2, \\ v &= y_1 + \frac{3x_1^2 + 2ax_1 + b}{2y_1}(u - x_1). \end{aligned} \tag{9}$$

Thus, if we know at least one rational point P of an elliptic curve, we can use the above formulas to compute $2P$, $3P$, and so on. Suppose, for example, that the curve is given by the equation $y^2 = x^3 - 2$ and $P = (3, 5)$. Then we can find the new rational point

$$2P = \left(\frac{129}{100}, -\frac{383}{1000}\right).$$

Now we can compute $3P$, $4P$, and so on. Note that the numbers we have to deal with grow rapidly. If u_n denotes the first coordinate of the point nP, then

$$u_1 = 3,$$
$$u_2 = \frac{129}{100},$$
$$u_3 = \frac{164323}{29241},$$
$$u_4 = \frac{2340922881}{58675600},$$
$$u_5 = \frac{307326105747363}{160280942564521},$$

Further coordinates grow even faster. For instance, u_{11} has 71 digits in its numerator.

At the present time, no general procedure is known that would allow us to find all rational solutions of the equation $y^2 = x^3 + ax^2 + bx + c$. In the example we've considered, we just guessed the first solution (3, 5) of the equation $y^2 = x^3 - 2$. In the general case no universal method exists for finding such a first solution. Finding an effective procedure of getting an initial rational solution of an elliptic equation is one of the greatest problems of number theory. However, if we know one solution, we can find others by means of formulas (8) and (9).

Order of points on an elliptic curve

Consider the sequence of points nP, "multiples" of the point P. We must distinguish two essentially different cases. First of all, we might obtain identity at some finite stage n. In other words, there might exist a number n such that $nP = E$. If $mP \neq E$ for all $m < n$, then we say that point P has *finite order n*. For instance, point $P = (0, 2)$ on the curve $y^2 = x^3 + 4$ has order 3, point $P = (2, 3)$ on the curve $y^2 = x^3 + 1$ has order 6, and point $P = (3, 8)$ on the curve $y^2 = x^3 - 43x + 166$ has order 7. The question then arises: how many points with finite order are there, and what are their orders?

In 1976 an outstanding result in this field was attained by the American mathematician B. Mazur,[2] who showed that *if P is a rational point of the n^{th} order, then $n \leq 10$ or $n = 12$; on the other hand, there are at the most 16 rational points of finite order on an elliptic curve.*

The second case we must consider is that the points P, $2P$, $3P$, $4P$, and so on, are all different. In 1901 the famous French mathematician Henri Poincaré (1854–1912) formulated the hypothesis that *for every elliptic curve there is a finite number of points $P_1, \ldots, P_r$ such that any rational point P of the curve is expressible in their terms—that is, we can represent it in the form*

$$P = n_1P_1 + \ldots + n_rP_r + Q,$$

where $n_1, \ldots, n_r$ *are integers defined in a unique way by* P, *and* Q *is a point of finite order.* And it is impossible to express one of the points $P_1, \ldots, P_r$ as a combination of the others. The number r is known as the *rank* of the curve.

In 1922 the young English mathematician L. Mordell proved Poincaré's conjecture, but his reasoning suggests no constructive method for calculating the rank. It isn't even known nowadays whether elliptic curves of arbitrarily large rank exist. Still, it was proved that the rank of a curve can be estimated from the coefficients a, b, and c of the equation $y^2 = x^3 + ax^2 + bx + c$, and therefore any curve of a large rank must have large coefficients. For example, one of the curves with rank $r \geq 8$ has

$$a = -3^2 \cdot 1487 \cdot 1873,$$

$$b = 2^5 \cdot 3^2 \cdot 5 \cdot 151 \cdot 14551 \cdot 33353,$$

$$c = 2^8 \cdot 3^4 \cdot 5^2 \cdot 7 \cdot 151^2 \cdot 193 \cdot 273 \cdot 156307.$$

Curves of arbitrary degrees

Here we've confined ourselves to curves (and thus to diophantine equations) of degrees 2 and 3. What happens for the degrees $n \geq 4$? In this case it is natural to point out the class of non-degenerate curves of the n^{th} degree (one typical representative of such curves is $x^n + y^n = 1$). When $n > 3$ the picture changes drastically. As early as 1931 Mordell conjectured that *the number of rational points on such curves is always finite.* For more than half a century Mordell's hypothesis remained in the center of mathematical studies all over the world. The Russian mathematicians I. R. Shafarevich, Y. I. Manin, S. Y. Arakelov, A. N. Parshin, and Y. G. Zarhin contributed to its solution. But the honor of giving the final solution of this problem fell in 1983 to the young German mathematician Gerd Faltings. At the 1986 International Mathematical Congress in Berkeley, California, he was awarded the highest mathematical prize, the Fields Medal, for this achievement.

[2]See his article "Questioning Answers" in the January/February 1997 issue of *Quantum*.—*Ed.*

When Trojans and Greeks collide

The challenge of multi-body systems

by I. Vorobyov

IS IT POSSIBLE TO DESCRIBE the trajectories of bodies that interact according to the law of universal gravitation? For example, how do the planets move? Assume that we know the precise coordinates and velocities of n planets at some moment of time. In addition, we assume that the planets are affected only by the forces of gravitational attraction of the other $(n - 1)$ planets. We are to find the subsequent positions of the planets (that is, their trajectories) using the initial conditions.

The laws that describe planetary motion are well known: these are Newton's second law and the law of universal gravitation. No extra laws are necessary to describe (and predict) the trajectories of moving planets. Is it an easy problem?

For a system composed of two bodies, this problem was solved for the first time in history by Sir Isaac himself. In this very simple case, knowing the initial conditions (positions and velocities of the planets), one can determine the future state of the system at any moment and with any precision.

However, when scientists considered systems composed of more than two bodies, they met with huge mathematical difficulties. One cannot obtain the general solution even for the case of three bodies ($n = 3$). Even to the present this three-body problem has not been strictly solved with known mathematical methods, notwithstanding 250 years of efforts by mathematicians around the world. For this reason, "the n-body problem" (this is how specialists refer to this problem in the general case) continues to attract the attention of mathematicians.

Another stimulus is the immense importance of the n-body problem for celestial mechanics and astronautics. No wonder that the total number of papers devoted to this venerable problem of mechanics amounts to 2000, and every year another 15–20 works are added.

At present there are a number of methods for obtaining approximate solutions to the n-body problem for a limited period of time, and for plotting the trajectories of any specific system of bodies with sufficient accuracy, provided the initial conditions are specified. The effectiveness and reliability of these methods has been proved many times—for example, in calculations of the trajectories of spaceships and satellites. Using such methods, John Adams and Urbain Leverrier discovered the planet Neptune "at the end of a writing pen."

O.K., but why are the Trojans mentioned in the title of this article, the reader may be wondering. The point is that astronomers are devoted connoisseurs of mythology, and there is a group of asteroids with this name that relates to the theme of this article very closely. However, we will talk about the Trojans somewhat later, and now continue the story of the n-body problem.

The problem of Lagrange

The exact solution of the n-body problem doesn't exist. However, 200 years ago, the outstanding French mathematician Joseph Louis Lagrange found the exact solutions for a system of three bodies characterized by some "specific" initial conditions. His solutions (now called "Lagrangean") are the only exact formulas for this problem known up to now.

Consider three bodies not located on the same line that rotate with the same angular velocity along concentric circles lying in the same plane. These bodies can be treated as the points of an imaginary solid body that rotates around the motionless axis (this axis passes through the common

Art by Vera Khlebnikova

center of the concentric circles, the "orbits" of the bodies).

For what locations of the masses is such motion possible? What must the angular velocity of the imaginary solid body be in order to maintain its integrity and not fly to bits? This was how the problem was posed by Lagrange himself. Let's try to solve it, following the steps of this great master of science.

Triangle of masses

Let three bodies with masses m_1, m_2, and m_3 revolve along concentric circles with a common center at point O (fig. 1). The radii of the orbits r_1, r_2, and r_3 and the distances between the bodies are assumed to be much greater than their sizes. The acceleration of any body results from its attraction to the other two bodies, and it is determined by the vector sum of the respective forces of gravitational attraction. The equations of motion of three bodies may be written in vector form as

$$\begin{cases} \mathbf{F}_{12} + \mathbf{F}_{13} = m_1\mathbf{a}_1, \\ \mathbf{F}_{21} + \mathbf{F}_{23} = m_2\mathbf{a}_2, \\ \mathbf{F}_{31} + \mathbf{F}_{32} = m_3\mathbf{a}_3. \end{cases}$$

Here

$$\left|\mathbf{F}_{ij}\right| = G\frac{m_i m_j}{r_{ij}^2}$$

is the attraction—that is, the force with which body m_j acts on body m_i. Since all the bodies revolve with the same angular velocity (by the statement of the problem), we have $\mathbf{a}_1 = -\omega^2 \mathbf{r}_1$, $\mathbf{a}_2 = -\omega^2 \mathbf{r}_2$, and $\mathbf{a}_3 = -\omega^2 \mathbf{r}_3$. Here r_i is the radius-vector of point m_i drawn from the center of revolution O. We call the bodies "points" because the distances between them are far greater than their sizes. Our further reasoning is purely "mathematical," and at first glance does not seem related to the problem. However, this is the way that leads to the solution of the first part of the problem and results in determining the configuration of the bodies.

Let's add the equations of system (1):

$$\begin{aligned}\mathbf{F}_{12} + \mathbf{F}_{13} + \mathbf{F}_{21} + \mathbf{F}_{23} + \mathbf{F}_{31} + \mathbf{F}_{32} &\\ = m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + m_3\mathbf{a}_3 &\\ = -\omega^2\left(m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + m_3\mathbf{r}_3\right).&\end{aligned}$$

According to Newton's third law, the sum of all the internal forces in a closed system is zero. More formally, the top line can be rewritten as

$$\sum_{i,j=1}^{3}\mathbf{F}_{ij} + \sum_{i,j=1}^{3}\mathbf{F}_{ji} = 0,$$

because $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$. Thus, we get

$$m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + m_3\mathbf{r}_3 = 0, \qquad (2)$$

because $\omega \neq 0$. Equation (2) describes the position of the center of revolution: it is located at a point O such that the radius vectors of the bodies m_1, m_2, and m_3 drawn from this point must satisfy equation (2).

Now let's introduce vectors $\mathbf{R}_2$ and $\mathbf{R}_3$ as shown in figure 2. We have

$$\begin{aligned}\mathbf{r}_1 &= \mathbf{r}_2 + \mathbf{R}_3,\\ \mathbf{r}_1 &= \mathbf{r}_3 + \mathbf{R}_2.\end{aligned} \qquad (3)$$

The next step is to express $\mathbf{r}_1$ in terms of $\mathbf{R}_2$ and $\mathbf{R}_3$. To do this, we multiply the first equation of (3) by m_2, the second by m_3, and add the resulting equations:

$$\begin{aligned}m_2\mathbf{r}_1 + m_3\mathbf{r}_1 &\\ = m_2\mathbf{r}_2 + m_2\mathbf{R}_3 + m_3\mathbf{r}_3 + m_3\mathbf{R}_2.&\end{aligned}$$

Taking into account that $m_2\mathbf{r}_2 + m_3\mathbf{r}_3 = -m_1\mathbf{r}_1$ (which follows from equation (2)), we get

$$\left(m_1 + m_2 + m_3\right)\mathbf{r}_1 = m_2\mathbf{R}_3 + m_3\mathbf{R}_2,$$

from which we obtain

$$\mathbf{r}_1 = \frac{m_2\mathbf{R}_3 + m_3\mathbf{R}_2}{m_1 + m_2 + m_3}.$$

Thus, vector $\mathbf{r}_1$ is the sum of two vectors

$$\rho_2 = \frac{m_3}{m_1 + m_2 + m_3}\mathbf{R}_2$$

and

$$\rho_3 = \frac{m_2}{m_1 + m_2 + m_3}\mathbf{R}_3,$$

which are parallel to the sides R_2 and R_3 of the triangle of masses.

Now we draw the vectors for the forces that act on the body m_1 on the plot of the triangle of masses (fig. 3). Here $\mathbf{F}_{12}$ is the force acting on it from body m_2, and $\mathbf{F}_{13}$ from body m_3. These forces are directed along the sides R_2 and R_3 of the mass triangle. Their resultant force is directed along vector $\mathbf{r}_1$ to point O (that is, the vector of acceleration of body m_1 is directed to the center of revolution). This means that the parallelogram of forces and the parallelogram constructed on the vectors ρ_2 and ρ_3 are similar! Therefore,

$$\left|\mathbf{F}_{12}\right| : \left|\rho_3\right| = \left|\mathbf{F}_{13}\right| : \left|\rho_2\right|,$$

and

$$\begin{aligned}G\frac{m_1 m_2}{R_3^2} &: \frac{m_2}{m_1 + m_2 + m_3}R_3\\ = G\frac{m_1 m_3}{R_2^2} &: \frac{m_3}{m_1 + m_2 + m_3}R_2.\end{aligned}$$

Finally, cancellation yields $R_2 = R_3$.

In a similar way we can find the expressions for $\mathbf{r}_2$ and $\mathbf{r}_3$ via the sides of the mass triangle (to this end, vectors $\mathbf{R}_3$ and $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_1$ should be introduced with corresponding directions). By considering the forces affecting the body m_2 (or m_3) we obtain the equalities $R_1 = R_3$ and $R_1 = R_2$.

Thus, we have arrived at a wonderful result: $R_1 = R_2 = R_3$, so the configuration of all the bodies is an equilateral triangle! The first problem is solved, and we invite the reader to solve the second one (see exercise 1). Now we are ready to visit the Trojans.

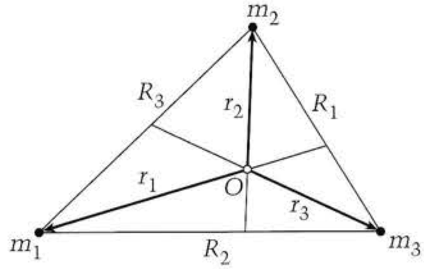


Figure 1

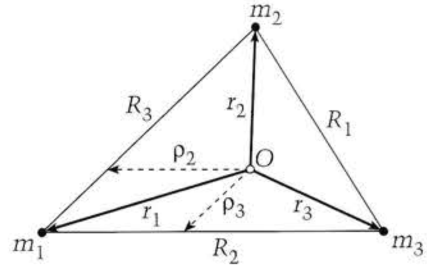


Figure 2

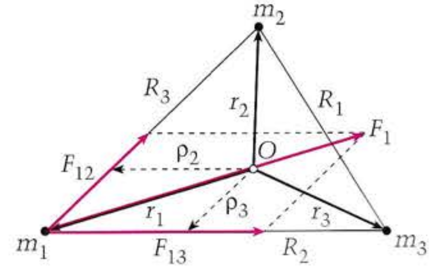


Figure 3

Trojans in a gravitational trap

On February 2, 1906, the German astronomer Maximilian Wolf discovered the asteroid Achilles. According to preliminary data, it moved with a speed of 13 km/s in a nearly circular orbit. However, Jupiter has a similar quasi-circular trajectory—and with the very same speed! Thus, two celestial bodies, large and small, shared the same orbit around the Sun.

Indeed, if a planet of mass m revolves around the Sun along an orbit of radius r, it is attracted to the Sun with the force

$$F = G\frac{mM}{r^2}$$

(M is the Sun's mass), which gives the planet a centripetal acceleration $a = v^2/r$, where v is the speed of the planet. That is,

$$G\frac{mM}{r^2} = \frac{mv^2}{r}.$$

Consequently, the orbital radius

$$r = G\frac{M}{v^2}$$

doesn't depend on the planet's mass and is determined entirely by its speed.

Perhaps the orbital planes of Jupiter and Achilles are inclined to each other? No, researchers quickly found that the asteroid really moved along Jupiter's orbit and ahead of the venerable planet by 55.5°. Of course, the scientists immediately recalled the work of Lagrange. The Sun, Jupiter, and Achilles formed a nearly equilateral triangle, revolving around one of its vertices, the Sun.

Later new asteroids were found at the vertices of two equilateral triangles with a common Sun-Jupiter base. They were named after the heroes of the ancient Trojan War. Five of them (this family is called the Trojans) are the rear-guard, which lags behind Jupiter, while the advance-guard is formed of 10 warriors (called the Greeks, fig. 4). These asteroids are rather big, and the larg-

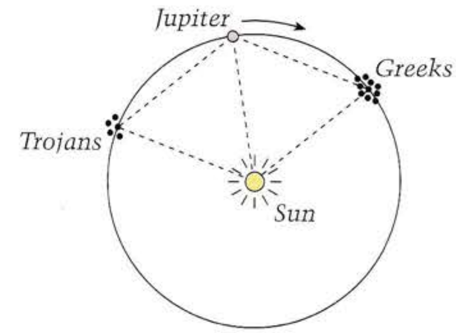


Figure 4

est of them, Patroclus, is 216 km across; the size of eight others is smaller than 100 km.

In 1959 the Polish astronomer Pan Kordylevsky found vast clouds of space dust in the vertices of the equilateral triangles drawn on the Earth-Moon axis. The role of the Trojans was played by myriad dust specks trapped by the combined attraction of Earth and the Moon.

Thus, the mathematical solutions published by Lagrange in 1772 and considered by him to be purely theoretical results were related to later astronomical discoveries.

Stability of motion

Did we consider all possible solutions to the three-body problem? No, there is a linear arrangement of three bodies that can rotate as a single body. However, such a configuration is unstable. If a body is slightly shifted from the line, the balance of forces will be disturbed and deviation will increase.

Is the triangle configuration of gravitating bodies stable? Not always. Gachot, in 1843, was the first to formulate the conditions of stability for the Lagrangean triangle. He showed that the configuration is stable provided the masses of the two large bodies that form the base of the triangle are such that m_1/m_2 is a sufficiently small ratio and

$$\frac{(m_1 + m_2)^2}{m_1 m_2} > 27.$$

If the mass of a space particle is much smaller than m_1 and m_2, and it enters the region of the vertex of any equilateral triangle constructed on an m_1–m_2 base with a relatively small speed, it will be "trapped." The particle can circumscribe quite an intricate trajectory about the vertex of the Lagrangean triangle, but the whole configuration will rotate as a single body.

The mass of Jupiter is 1000 times smaller than that of the Sun. The total mass of the Trojans is negligible in comparison with the Sun or Jupiter, so the requirements of stability are met with a high stability margin. The inequality is true also for the Earth-Moon system: their masses differ by 81 times.

How could the conditions of stability be deduced? Perhaps there is a simple deduction, but in general, the problems of stability are rather complicated and should be considered in a special article.

Exercises

1. Prove that the angular velocity of rotation of a Lagrangean triangle is

$$\omega^2 = G\frac{m_1 + m_2 + m_3}{R^3},$$

where R is the distance between the bodies.

2. We considered the triangle of masses as a rigid body. However, if the values of the velocity vectors are proportional to the distances to the center of revolution and the vectors are directed at the same angle to the line segments connecting the bodies with the center, then the configuration will be congruent to the initial arrangement at any time. Thus, the equilateral triangle will expand or contract during its revolution. Try to prove this on your own.

Quantum on planetary motion:

Y. Osipov, "Catch as Catch Can," January/February 1992, pp. 38–43.

A. Byalko, "A Flight to the Sun," November/December 1996, pp. 16–20.

V. Surdin, "Swinging from Star to Star," March/April 1997, pp. 4–8.

V. Mozhaev, "In the Planetary Net," January/February 1998, pp. 4–8

A Chebyshev polyplayground

Recurrence relations applied to a famous set of formulas

by N. Vasilyev and A. Zelevinsky

ONE POPULAR IMAGE OF THE MATHEMATICIAN is a person who is constantly performing tedious calculations, writing down and transforming cumbersome formulas. Although some elegant and important fields of mathematics exist that do without formulas, this opinion does have certain foundations. The ability to look at formulas from an original point of view, transform them, find new formulas, and discover various relations among them plays an important role in the work of a mathematician. In this article, we consider a series of formulas related to the so-called Chebyshev polynomials and some powerful mathematical ideas that underlie them.

Two remarkable sequences of polynomials

The polynomials in question occur in many problems of mathematical analysis, computational mathematics, and algebra. These polynomials were first considered by the prominent Russian mathematician Pafnuty Chebyshev in 1854 in connection with the following question.

n	T_n	U_n
0	1	1
1	x	$2x$
2	$2x^2 - 1$	$4x^2 - 1$
3	$4x^3 - 3x$	$8x^3 - 4x$
4	$8x^4 - 8x^2 + 1$	$16x^4 - 12x^2 + 1$
5	$16x^5 - 20x^3 + 5x$	$32x^5 - 32x^3 + 6x$
6	$32x^6 - 48x^4 + 18x^2 - 1$	$64x^6 - 80x^4 + 24x^2 - 1$
7	$64x^7 - 112x^5 + 56x^3 - 7x$	$\vdots$
	$\vdots$	

Table 1. *Chebyshev polynomials of the first and second kind. If we multiply each polynomial by 2x, then subtract the one above it in the table, we obtain the next polynomial.*

Consider various polynomials of degree n with a leading coefficient of 1. Which of them deviates least from zero in the interval $[-1, 1]$? We will answer this question in two ways, depending on how we interpret it.[1] Suppose the polynomial is $F_n(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0$. Then we can interpret the phrase "deviate least from 0" to mean that the quantity

$$c_n = \max_{[-1,1]}\left|F_n(x)\right|$$

takes on the minimum value.

It turns out that this polynomial is

$$\tilde{T}_n(x) = \frac{1}{2^{n-1}}T_n(x).$$

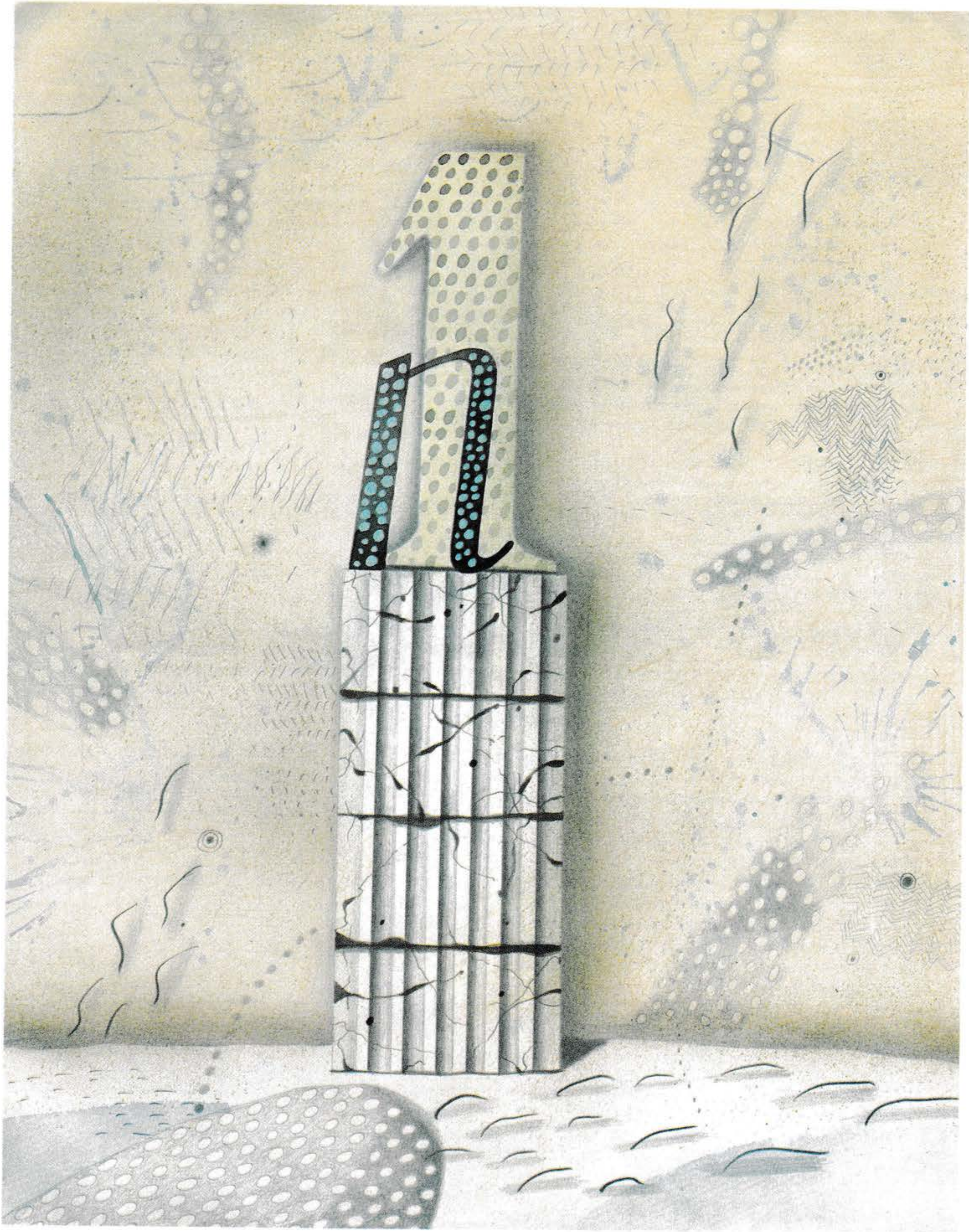
That is, it is the polynomial shown in the left column of table 1 divided by its leading coefficient. For example, among quadratic trinomials, the one that deviates least from zero, c_2, is $1/2$; and for any other trinomial $x^2 + px + q$, this deviation is greater. The cubic polynomial with the least deviation from zero is $x^3 - (3/4)x$. Its deviation from zero is $c_3 = 1/4$. In general, the polynomial $\tilde{T}_n(x)$ deviates from zero by $c_n = 1/(2^{n-1})$, which is less than for any other polynomial $F_n(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0$ of degree n.

We can measure the deviation of a polynomial from zero in a different way, by requiring that the area between the graph of the polynomial and the x-axis, and between the lines $x = 0$ and $x = 1$, be minimal. If you know calculus, you know that this is the quantity

$$I_n = \int_{-1}^{1}\left|F_n(x)\right|dx.$$

[1]However, we will not give proofs that these answers are correct. An elementary proof can be found in reference 1 at the end of the article.

Art by Vasily Vlasov



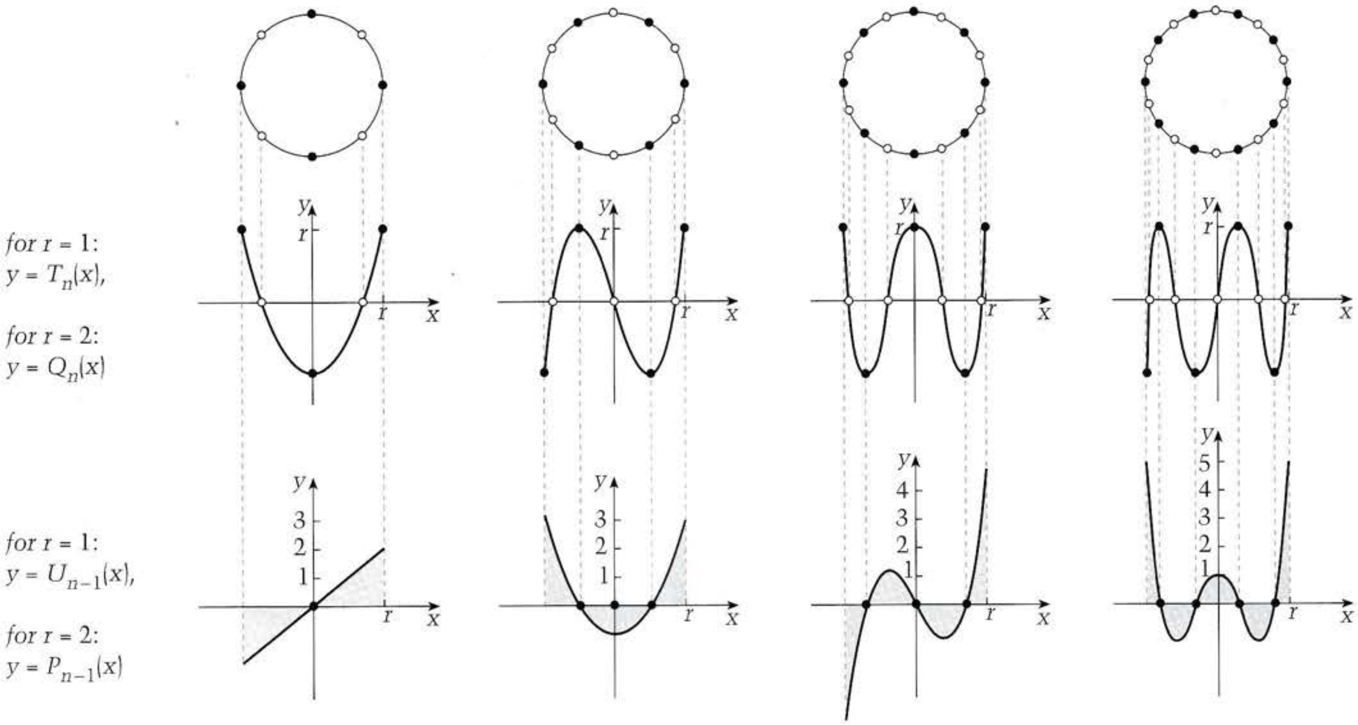


Fig. 1. *Take a transparent sheet of paper ($0 \le x \le 2\pi r$, $-r \le y \le r$) with the graph of the function $y = r \cos nx$ drawn on it. Roll up this sheet into a cylinder (with diameter and height 2h); then, look at it from its side in such a way that the graphs on the front and back parts of the cylinder coincide—you will see the graph of the nth Chebyshev polynomial of the first kind. For n = 2, 3, 4, 5, these graphs are shown in the upper row of the figure (for r = 1, we obtain the graphs of $y = T_n(x)$ and for r = 2, the graphs of $y = Q_n(x)$—see exercise 3). Under the graph of the nth polynomial, its derivative divided by n is graphed. This is the (n – 1)th Chebyshev polynomial of the second kind—all figures painted blue have equal areas.*

For this interpretation of "least deviation from zero," the polynomial

$$\tilde{U}_n(x) = \frac{1}{2^n} U_n(x),$$

where $U_n(x)$ is shown in the right-hand column of table 1, has the minimum deviation. For $U_n(x)$, the quantity l_n (the blue area in fig. 1) equals 2. Thus, it is $1/2^{n-1}$ for $\tilde{U}_n$, and is greater for any other polynomial $F_n(x) = x^n + \ldots$ of degree n (this is the content of the *Korkin–Zolotarev* theorem).

The facts mentioned are related to the following properties of the Chebyshev polynomials:

(1) The values of T_n are equal in absolute value for all the turning points and for the endpoints of the interval [–1, 1]. The area of each of the $n + 1$ regions bounded by the graph of the polynomial

$$U_n(x) = \frac{1}{n+1} T'_{n+1}(x),$$

the x-axis, and lines $x = \pm 1$ is the same (see fig. 1).

As it turns out, the only polynomials that possess this property are those obtained from the equations $y = U_n(x)$ and $y = T_n(x)$ by a linear change of the variables x and y.

Property **(1)** follows from the following basic relations:

(2) $T_n(\cos\phi) = \cos n\phi$, $\sin\phi \cdot U_{n-1}(\cos\phi) = \sin n\phi$.

In addition to the trigonometric relations **(2)**, which determine the values of polynomials T_n and U_n for $|x| \le 1$, there exist quite different identities for $|x| > 1$:

(3)
$$T_n(x) = \frac{\left(x + \sqrt{x^2 - 1}\right)^n + \left(x - \sqrt{x^2 - 1}\right)^n}{2},$$
$$U_n(x) = \frac{\left(x + \sqrt{x^2 - 1}\right)^{n+1} - \left(x - \sqrt{x^2 - 1}\right)^{n+1}}{\sqrt{x^2 - 1}}.$$

The roots of T_n and U_n can be found from the following identities:

(4)
$$T_n(x) = 2^{n-1}\left(x - \cos\frac{\pi}{2n}\right)\left(x - \cos\frac{3\pi}{2n}\right)\ldots\left(x - \cos\frac{(2n-1)\pi}{2n}\right),$$
$$U_n(x) = 2^n\left(x - \cos\frac{\pi}{2n+1}\right)\left(x - \cos\frac{2\pi}{2n+1}\right)\ldots\left(x - \cos\frac{n\pi}{n+1}\right)$$

Thus, the roots and turning points of polynomial T_n are the projections of the vertices of a regular $4n$-gon inscribed in a circle whose diameter has endpoints (–1, 0) and (1, 0) onto this diameter (see fig. 1).

In the following, we will prove, among other relations, identities **(2)**–**(4)** and use them to demonstrate some important methods of algebraic transformations.

Any of the above formulas can be taken for the definition of the Chebyshev polynomials. In this article, we will take as our definition the simple recurrence relation presented in the caption to table 1. We will derive all our other formulas from this relation.

It is more convenient for our purpose to deal with polynomials obtained from T_n and U_n by scaling (see fig. 1): $P_n(x) = U_n(x/2)$, $Q_n(x) = 2T_n(x/2)$. For these polynomials, the interval $[-2, 2]$ plays the same role that interval $[-1, 1]$ plays for T_n and U_n. The convenience of the new polynomials is that their coefficients are integers and the leading coefficients are 1. Naturally, using the inverse transformations $P_n(2x) = U_n(x)$ and $Q_n(2x) = 2T_n(x)$, we can return to the original polynomials T_n and U_n whenever we like. As a rule, in this article, we will prove various propositions for P_n. We invite the reader to prove similar propositions for Q_n. We recommend carrying out all calculations in detail, using small values of $n = 2, 3, 4$ first (until everything becomes clear).

Recurrence relations and induction

Set $P_0(x) = 1$, $P_1(x) = x$, and

$$P_{n+1}(x) = x \cdot P_n(x) - P_{n-1}(x). \qquad (1)$$

Write down the first several elements of this sequence:

$$P_2(x) = x^2 - 1,$$

$$P_3(x) = x(x^2 - 1) - x = x^3 - 2x,$$

$$P_4(x) = x(x^3 - 2x) - (x^2 - 1) = x^4 - 3x^2 + 1,$$

$$P_5(x) = x(x^4 - 3x^2 + 1) - (x^3 - 2x) = x^5 - 4x^3 + 3x,$$

and so on (you can check your results up to P_{12} using table 2).

The polynomials $P_n(x)$ occur in various situations. For example, consider the fractions

$$R_1(x) = x,\ R_2(x) = x - \frac{1}{x},\ R_3(x) = x - \cfrac{1}{x - \cfrac{1}{x}},$$

$$R_4(x) = x - \cfrac{1}{x - \cfrac{1}{x - \cfrac{1}{x}}},\ R_5(x) = x - \cfrac{1}{x - \cfrac{1}{x - \cfrac{1}{x - \cfrac{1}{x}}}} \ldots$$

These *continued fractions* provide a useful tool for solving various problems involving the approximation of numbers and functions (Chebyshev studied continued fractions as well).

Performing various manipulations, we obtain

$$R_2(x) = \frac{x^2 - 1}{x},\ R_3(x) = \frac{x^3 - 2x}{x^2 - 1},$$

$$R_4(x) = \frac{x^4 - 3x^2 + 1}{x^3 - 2x},\ R_5(x) = \frac{x^5 - 4x^3 + 3x}{x^4 - 3x^2 + 1}, \ldots$$

$n \backslash k$	0	1	2	3	4	5	6
0	1						
1	1	1					
2	1	2	1				
3	1	3	3	1			
4	1	4	6	4	1		
5	1	5	10	10	5	1	
6	1	6	15	20	15	6	1
7	1	7	21	35	35	21	7
8	1	8	28	56	70	56	28
9	1	9	36	84	126	126	84
10	1	10	45	120	210	252	210
11	1	11	55	165	330	462	462
12	1	12	66	220	495	792	924
	...	...	...	...	...	...	...

Table 2. *Pascal's triangle. The numbers that appear in the nth red-marked diagonal taken with alternate signs are coefficients of polynomials $P_n(x)$. Exercises 6a and 6b concern the sums of binomial coefficients.*

(we invite the reader to check this). We notice that the numerators and denominators of these fractions are just polynomials $P_n(x)$.

Here is another example. Consider the function $\sin n\phi$ and try to represent it in terms of $\sin\phi$ and a polynomial in $\cos\phi$:

$$\sin 2\phi = \sin\phi \cdot 2\cos\phi,$$
$$\sin 3\phi = \sin\phi \cdot (4\cos^2\phi - 1),$$
$$\sin 4\phi = \sin\phi \cdot (8\cos^3\phi - 4\cos\phi).$$

It turns out that $\sin n\phi = \sin\phi \cdot P_{n-1}(2\cos\phi)$ for all $n \ge 1$. In other words, for $\sin\phi \ne 0$,

$$P_n(2\cos\phi) = \frac{\sin(n+1)\phi}{\sin\phi}. \qquad (2)$$

These relations can be readily obtained by induction and formula (1). Indeed,

$$R_{n+1}(x) = x - \frac{1}{R_n(x)}.$$

Therefore, if we assume that, for a certain n,

$$R_n(x) = \frac{P_n(x)}{P_{n-1}(x)},$$

a similar relation for $n + 1$ can be easily obtained from (1):

$$R_{n+1}(x) = x - \frac{P_{n-1}(x)}{P_n(x)} = \frac{xP_n(x) - P_{n-1}(x)}{P_n(x)} = \frac{P_{n+1}(x)}{P_n(x)}.$$

Similar manipulations can be carried out for sines. If we assume that for $k = n - 1$ and $k = n$

$$\sin(k+1)\phi = \sin\phi \cdot P_k(2\cos\phi),$$

it follows from (1) that

$$\begin{aligned}\sin\phi\cdot P_{n+1}(2\cos\phi) &= 2\cos\phi\cdot\sin\phi\cdot P_n(2\cos\phi)\\ &\quad -\sin\phi\cdot P_{n-1}(2\cos\phi)\\ &= 2\cos\phi\cdot\sin(n+1)\phi-\sin n\phi\\ &= \sin(n+2)\phi.\end{aligned}$$

Here we used the identity $2\cos\alpha\cdot\sin\beta=\sin(\alpha+\beta)+\sin(\beta-\alpha)$.

Notice that we carried out the inductive step from $n-1$ and n to $n+1$. In most proofs, the inductive step is made from n to $n+1$. Because of our unusual "double step," we must in this case verify the first two identities (for $n = 0$ and $n = 1$) separately.

Exercises

1. Prove by induction using the recurrence relation (1) that the following identity holds for $|x| > 2$:

$$P_n(x)=\frac{\left(x+\sqrt{x^2-4}\right)^{n+1}-\left(x-\sqrt{x^2-4}\right)^{n+1}}{2^{n+1}\sqrt{x^2-4}}. \tag{3}$$

This identity will be also considered below.

2. Prove that (a) $P_n(2) = n + 1$ and (b) (a) $P_n(-2) = (-1)^n\cdot(n+1)$. (Prove it in three different ways: using (1), passing to the limit in equation (2) as $\phi\to 0$ and $\phi\to\pi$, and passing to the limit in (3) as $x\to\pm 2$.)

3. Consider a sequence of polynomials $Q_0(x)$, $Q_1(x)$, $Q_2(x)$, . . . satisfying equation (1) and the initial conditions $Q_0(x) = 2$ and $Q_1(x) = x$. Write down the first six polynomials $Q_n(x)$. Prove the following identities:

(a) $$\frac{Q_n(x)}{Q_{n-1}(x)}=x-\cfrac{1}{x-\cfrac{1}{\ddots-\cfrac{1}{x-\cfrac{2}{x}}}}$$ (there are $n-1$ minus signs);

(b) $2\cos n\phi = Q_n(2\cos\phi)$; (2′)

(c) for $|x| > 2$,

$$Q_n(x)=\frac{\left(x+\sqrt{x^2-4}\right)^{n}+\left(x-\sqrt{x^2-4}\right)^{n}}{2^{n}}. \tag{3′}$$

4. Prove that any sequence of polynomials $R_0(x)$, $R_1(x)$, . . . satisfying equation (1) can be expressed in terms of the sequence $P_n(x)$ as follows:

$$R_n(x) = R_1(x)\cdot P_{n-1}(x) - R_0(x)\cdot P_{n-2}(x).$$

In particular, $Q_n(x) = xP_{n-1}(x) - 2P_{n-2}(x) = P_n(x) - P_{n-2}(x)$. Derive all the identities in exercise 3 from this formula.

Roots of polynomials and products

Many interesting formulas that involve symmetric expressions depending on n numbers (or variables) can be easily explained if these numbers are considered as the roots of a certain polynomial of degree n. For n numbers

$$\gamma_k=2\cos\frac{k\pi}{n+1}$$

(where $k = 1, 2, ..., n$), $P_n(x)$ serves as such a polynomial. Indeed, substituting

$$\frac{\pi}{n+1},\frac{2\pi}{n+1},\ldots,\frac{n\pi}{n+1}$$

for ϕ in (2), we see that

$$\gamma_k=2\cos\frac{k\pi}{n+1}$$

are the roots of $P_n(x)$. Now we recall the *factor theorem* of elementary algebra:[2] *if γ is a root of a polynomial $P(x)$, then $P(x)$ is divisible by $x-\gamma$.* The polynomial $P_n(x)$ must be divisible by each of the binomials $x-\gamma_k$, and hence, by their product. Since our polynomial is of degree n and its leading coefficient is 1, it is just the product

$$\prod_{1\le k\le n}(x-\gamma_k).$$

Thus,

$$P_n(x)=\prod_{1\le k\le n}\left(x-2\cos\frac{k\pi}{n+1}\right). \tag{4}$$

Exercise 5. (a) Prove the identity

$$Q_n(x)=\prod_{1\le k\le n}\left(x-2\cos\frac{(2k-1)\pi}{2n}\right). \tag{4′}$$

(b) Verify identities (4) and (4′) for $n = 2, 3, 4$, and 5.

Consider an interesting identity that can be derived by comparing formulas (4) and (2).

For $m > 0$, calculate $P_{2m}(0)$ in two different ways and equate the expressions obtained. On the one hand, we obtain from (2):

$$P_{2m}(0)=P_{2m}\left(2\cos\frac{\pi}{2}\right)=\frac{\sin\frac{(2m+1)\pi}{2}}{\sin\frac{\pi}{2}}=\sin\left(\frac{\pi}{2}+m\pi\right)=(-1)^m.$$

On the other hand, we have from (4):

$$P_{2m}(0)=\prod_{1\le k\le 2m}\left(-2\cos\frac{k\pi}{2m+1}\right).$$

Substituting

$$\left(-\cos\left(\pi-\frac{k\pi}{2m+1}\right)\right)$$

for each

$$\cos\frac{k\pi}{2m+1}$$

for $m + 1 \le k \le 2m$, we obtain

[2]For many problems involving the factor theorem, see Gradus ad Parnassum in *Quantum*: March/April 1998, May/June 1998, and July/August 1998.—*Ed.*

$$P_{2m}(0) = (-1)^m \cdot \left[2^m \cdot \prod_{1 \le k \le m} \cos\frac{k\pi}{2m+1} \right]^2.$$

The expression in brackets is positive, because it involves cosines of acute angles only, and therefore, it equals 1. Thus we obtain

$$\prod_{1 \le k \le m} \cos\frac{k\pi}{2m+1} = \frac{1}{2^m}. \qquad (5)$$

This relation has an elegant verbal formulation: *for $m > 0$, the geometric mean of the cosines of the acute angles that are multiples of $\pi/(2m+1)$ is 1/2.*

Exercises

6. (a) Find $P_n(1)$, $P_n(-1)$, $Q_n(1)$, and $Q_n(-1)$.

Prove the following identities that are similar to (5):

(b) $\displaystyle\prod_{1 \le k \le m} \sin\frac{k\pi}{2m} = \frac{\sqrt{m}}{2^{m-1}}$ $(m \ge 1)$;

(c) $\displaystyle\prod_{1 \le k \le m} \tan\frac{k\pi}{2m+1} = \sqrt{2m+1}$ $(m \ge 1)$;

(d) $\displaystyle\prod_{1 \le k \le m} \cos\frac{(2k-1)\pi}{4m} = \frac{\sqrt{2}}{2^m}$ $(m \ge 1)$.

7. Find all values of m and n for which (a) P_n is divisible by P_m; (b) Q_n is divisible by Q_m.

Generating functions, power series, and coefficients

In this section, we consider a very fruitful method that is widely used in calculus, combinatorics, and probability theory: the method of generating functions. In some cases, this method allows us to find the separate elements of a sequence and construct it piece by piece, just as a building is constructed of bricks.

Consider a sequence $a_0, a_1, a_2, \ldots$ The following expression is called the *generating function* of this sequence:

$$f(z) = a_0 + a_1 z + a_2 z^2 + \cdots = \sum_{n \ge 0} a_n z^n.$$

Expressions of this kind are called *formal power series*. Such series may be added, subtracted, and multiplied like ordinary polynomials. One series may be divided by another, if the constant term of the divisor is not zero. Such a series may also be differentiated or integrated term by term.[3] All these operations can be used to obtain new sequences from those already analyzed. It is often possible to find a simple expression for the generating function from the recurrent relation that defines a sequence. Conversely, sometimes it proves possible to find a general formula for an element of the sequence or a relation connecting several elements of the sequences given its generating function.

For a finite sequence $a_0, a_1, a_2, \ldots, a_n$, the polynomial $f(z) = a_0 + a_1 z + \ldots + a_2 z^n$ serves as the generating function. For example, the polynomial $f_n(z) = (1+z)^n$ is the generating function for the sequence of *binomial coefficients* $C_n^0, C_n^1, \ldots, C_n^n$; these coefficients appear in the nth row of Pascal's triangle (see table 2):

$$\sum_{0 \le k \le n} C_n^k z^k = (1+z)^n. \qquad (6)$$

We differentiate this identity k times and set $z = 0$ to obtain

$$C_n^k = \frac{n(n-1)\ldots(n-k+1)}{1 \cdot 2 \cdot \ldots \cdot k}.$$

Let's consider the identity $(1+z)(1+z)^n = (1+z)^{n+1}$ and write it as

$$(1+z)\left(\sum_{0 \le k \le n} C_n^k z^k \right) = \sum_{0 \le k \le n+1} C_{n+1}^k z^k.$$

By removing the parentheses and collecting the terms z^m, we obtain the important relation $C_n^m + C_n^{m-1} = C_{n+1}^m$.

Among infinite sequences, the geometric progression b_0, $b_n = q b_{n-1}$ has an especially simple generating function. In the sum

$$f(z) = \sum_{n \ge 0} b_n z^n = b + \sum_{n \ge 1} b_n z^n,$$

replace each b_n with $q b_{n-1}$ to obtain

$$f(z) = b + qz \sum_{n \ge 1} b_{n-1} z^{n-1} = b + qz\, f(z).$$

We have $f(z)(1 - qz) = b$, from which we obtain

$$f(z) = \sum_{n \ge 0} b_n z = \frac{b}{1 - qz}. \qquad (7)$$

This is the well-known formula for the sum of the infinite geometric progression (for $|qz| < 1$). The same method can be used to obtain the generating functions for the sequence of polynomials $P_n(x)$.

Define

$$\Phi(z) = \sum_{n > 0} P_n(x) z^n = 1 + xz + \sum_{n \ge 2} P_n(x) z^n.$$

(Here, x plays the role of a parameter, and we will write $P_n, P_{n-1}, \ldots$ instead of $P_n(x), P_{n-1}(x), \ldots$ for the sake of brevity.) By relation (1), for $n \ge 2$, we replace each P_n for

[3]See "Generating Functions," by S. M. Voronin and A. G. Kulagin in the May/June 1999 issue of *Quantum*.—*Ed.*

$xP_{n-1} - P_{n-2}$. Then,

$$\begin{aligned}\Phi(z) &= 1 + xz + \sum_{n\geq 2} xP_{n-1}z^n - \sum_{n\geq 2} P_{n-2}z^n \\ &= 1 + xz + xz\cdot\sum_{n\geq 2} P_{n-1}z^{n-1} - z^2\cdot\sum_{n\geq 2} P_{n-2}z^{n-2} \\ &= 1 + xz + xz\left(\Phi(z) - 1\right) - z^2\Phi(z).\end{aligned}$$

Therefore, $\Phi(z)\cdot(z^2 - xz + 1) = 1$ and

$$\Phi(z) = \frac{1}{z^2 - xz + 1}. \qquad (8)$$

This simple formula hides the entire intricate sequence of polynomials P_n. We can extract the individual P_n from this sequence in two different ways.

(1) For $|x| > 2$, the quadratic equation $z^2 - xz + 1$ has two roots:

$$u = \frac{x + \sqrt{x^2 - 4}}{2},\ v = \frac{x - \sqrt{x^2 - 4}}{2}. \qquad (9)$$

From the factorization $z^2 - xz + 1 = (z - u)(z - v)$, remembering that $uv = 1$, we have

$$\begin{aligned}\Phi(z) &= \frac{1}{(u - z)(v - z)} = \left(\frac{1}{v - z} - \frac{1}{u - z}\right)\frac{1}{u - v} \\ &= \left(\frac{u}{1 - zu} - \frac{v}{1 - zv}\right)\frac{1}{u - v} = \sum_{n\geq 0}\frac{u^{n+1} - v^{n+1}}{u - v}z^n.\end{aligned}$$

That is, $P_n(x) = (u^{n+1} - v^{n+1})/(u - v)$. This is relation (3).

(2) The coefficients of any polynomial $P_n(x)$ can be found from (8) in the following way:

$$\begin{aligned}\Phi(z) &= \frac{1}{1 - \left(xz - z^2\right)} = \sum_{k\geq 0}\left(xz - z^2\right)^k = \sum_{k\geq 0}\left(\sum_{0\leq j\leq k}(-1)^j C_k^j x^{k-j}z^{k+j}\right) \\ &= \sum_{n\geq 0} z^n\left(\sum_j (-1)^j C_{n-j}^j\cdot x^{n-2j}\right)\end{aligned}$$

(here, we used the formulas for the sum of the infinite geometric progression (7), binomial coefficients (6), and found the coefficient of z^n, which is the required $P_n(x)$). Therefore,

$$P_n(x) = \sum_j (-1)^j C_{n-j}^j x^{n-2j}. \qquad (10)$$

For example,

$$P_6(x) = C_6^0 x^6 - C_5^1 x^4 + C_4^2 x^2 - C_3^3 = x^6 - 5x^4 + 6x^2 - 1.$$

Certainly, formulas (3) and (10) can be proved by induction without using generating functions. However, the way they have appeared almost by themselves from the short formula (8) is wonderful.

All our manipulations with infinite series need to be vindicated. The justification can be performed in two different ways. First, we notice that all series considered are convergent for small absolute values of z (for example, series (7) is convergent for $|z| < 1/|q|$). Therefore, they are well-defined functions of z. The other way is to check that formally defined operations on infinite series (addition, multiplication, and so on) possess all usual properties (when these operations are performed, every coefficient of the resulting series is expressed in terms of a finite number of coefficients of the operands, with z considered merely as a symbol).

Exercises

8. Consider a sequence of *Fibonacci numbers*

$$u_0 = 0,\ u_1 = 1,\ u_{n+1} = u_n + u_{n-1}.$$

(a) Prove that the generating function of this sequence is

$$\frac{z}{1 - z - z^2}.$$

(b) Binet's formula gives the value of the nth Fibonacci number explicitly in terms of n:

$$u_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right].$$

Derive this formula from the generating function of the Fibonacci sequence.

(c) Prove the identity

$$u_n = \sum_j C_{n-j-1}^j.$$

9. (a) Find the generating function for the sequence of polynomials $Q_n(x)$ defined in exercise 3 and prove that, for $|x| > 2$,

$$Q_n(x) = u^n + v^n,$$

where u and v are defined by (9) (this formula was also proved in exercise 3c).

(b) (for those who are acquainted with complex numbers) Verify that, for $|x| < 2$, (3) and (3′) turn into (2) and (2′), respectively. (*Hint:* if $x = 2\cos\phi$, $u = \cos\phi + i\sin\phi$ and $v = \cos\phi - i\sin\phi$.)

References

1. Yaglom, A. M. and Yaglom, I. M., *Challenging Mathematical Problems with Elementary Solutions.* Translated by James McCawley. San Francisco: Holden-Day, 1957.

2. Polya, G. and Szegö, G., *Problems and Theorems in Analysis.* Berlin: Springer-Verlag, 1978.

3. Vilenkin, N.Ya., *Combinatorics.* Translated by A. Shenitzer and S. Shenitzer. New York: Academic Press, 1971.

8. S. M. Voronin and A. G. Kulagin, "Generating Functions." *Quantum,* May/June 1999, pp. 8–13.

New CONTEST FORMAT NAMES

NEW CONTEST, NEW FORMAT, NEW NAMES

The American Mathematics Competitions is pleased to announce a new contest as well as new names and a modified format for two of the current exams. The AJHSME is now the American Mathematics Contest ⇨8 (AMC ⇨8) and the AHSME is now the American Mathematics Contest ⇨12 (AMC ⇨12). The new contest is the American Mathematics Contest ⇨10 (AMC ⇨10), for students in grades 10 and below. This new contest will give more young students a chance to successfully participate in a significant mathematical problem solving experience.

Why should my school sign up? Because the AMC ⇨10 and AMC ⇨12 provide an excellent opportunity to challenge your students' mathematical abilities. It is but a means for furthering mathematical interest and development.

The AMC ⇨10 and AMC ⇨12 will each be 75 minutes long and will consist of 25 questions each. Each correct answer is worth 6 points and a blank is worth 2 points. The AMC ⇨10 and AMC ⇨12 will have several questions in common and will be given at the same time, on the Tuesday before the third Monday in February (the current AHSME date). The students should choose between AMC ⇨10 and AMC ⇨12. Students in 10th grade and under may take either the AMC ⇨10 or AMC ⇨12, but 11th and 12th grade students may not take the AMC ⇨10. The school team score will be determined from the AMC ⇨12. To qualify for the AIME a student must score at least 100 points on the AMC ⇨12 or be in the top 1% of the AMC ⇨10 participants.

The registration fee for one or both contests is $30.00. One bundle of ten AMC ⇨12 is $12.00 and one bundle of ten AMC ⇨10 is $10.00. The first bundle of the AMC ⇨10 will be free for the year 2000 only.

<u>2000 AMC exam dates:</u>

AMC⇨**8** - TUESDAY, November 14, 2000
AMC⇨**10** - TUESDAY, February 15, 2000
AMC⇨**12** - TUESDAY, February 15, 2000
AIME - TUESDAY, March 28, 2000
USAMO - TUESDAY, May 2, 2000

www: **http://www.unl.edu/amc**

Titu Andreescu, Director
American Mathematics Competitions
University of Nebraska, Lincoln
P.O. Box 81606
Lincoln, Nebraska 68501-1606
phone: 800-527-3690
fax: 402-472-6087
e-mail: titu@amc.unl.edu

A Program of the
Mathematical Association of America

Do you really k

THE QUEST FOR THE TRUE nature of time has occupied humanity for thousands of years. Today there is no shortage of speculation about time. Looking through the pages of popular science books, we encounter the notions of "four-dimensional space-time," "the arrow of time," "waves of time," "irreversibility of time," "time-traveling machines," "curvature of space and time," and the like.

Admittedly, I approached the problem of time with considerable trepidation, so different and limitless it seems. However, some hints from scientific treatises encouraged me. For example, Newton differentiated between "granted from the heavens" absolute time and "relative time," which is "self-evident" and "usual," and can be measured by a clock. Richard Feynman proposed that instead of racking our brains searching for a definition of time we should simply learn how to measure it properly.

To this end, people invented clocks, and what marvelous devices they are! And how many types have been constructed! Water, solar, sand, mechanical, quartz-stabilized, atomic . . . To make these clocks, scientists found suitable periodic processes, chose time standards, and got the hang of measuring very small fragments of time. So, are we equal to the problem of time, after all?

Let's not jump to conclusions: time still retains many of its mysteries. Solving them would mean great progress in science. At present, being in the framework of our "usual time," let's consider if time is really so "commonplace," even in school physics problems. As Naum Korzhavin said:

Time? It is a given essence
not to be discussed.
Should you think of the evanescence
of your life that quickly passed?

Problems and questions

1. How should a sundial be constructed to provide correct readings at any time of year?

2. Will ink drops that fall from a dropper attached to a uniformly moving cart (fig. 1) really mark equal

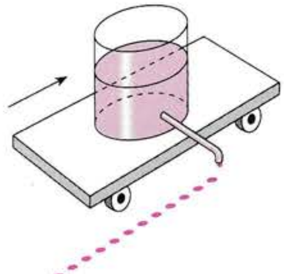


Figure 1

distances on the table (a well-known high school physics experiment)?

3. Two boats traveled on a river in the same direction with different speeds. Simultaneously they met a raft that was floating downstream. After half an hour the boats turned back and traveled with the same (relative to the water) speeds. Which boat met the raft first?

4. Figure 2 shows the dependence of velocity on time for two bodies moving along the x-axis. What is the physical meaning of the intersection point of the figure? Is it possible to determine from the figure when the bodies will meet?

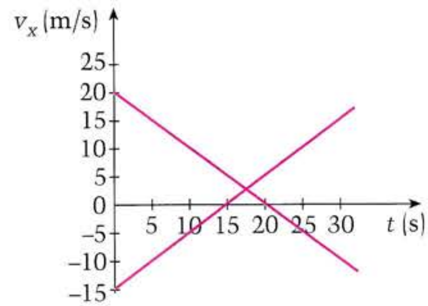


Figure 2

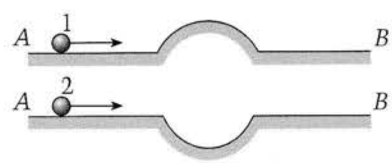
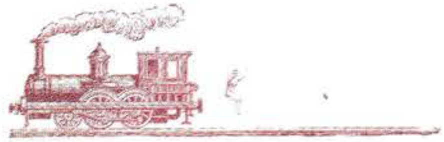


Figure 3

5. Two balls rolling on the surfaces shown in figure 3 pass points A simultaneously with the same velocity. Will they arrive at points B simultaneously? Neglect friction.

6. A body falls from the window of a train. Will the time to free fall from the same height be equal in the following cases: (a) the train is at rest, (b) the train moves with a constant velocity, (c) the train is uniformly accelerated.



7. An object is hurled at an angle to the horizon. What will take a longer time, the rise or fall of the object, provided air resistance is not neglected?

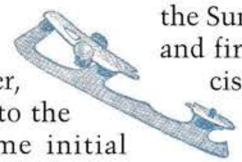
8. A fly sits on the bottom of a closed tube. The tube falls freely, always in the vertical position.

y know time?

How will the time of fall change if the fly flies from the bottom to the top of the tube during the fall?

9. A long rope passes over a stationary pulley. Two gymnasts of the same mass hang onto the ends of the rope at the same height above the ground. The first gymnast begins to climb with a constant speed relative to the rope, and the second gymnast starts to descend at one-half this speed. Who will reach the pulley first?

10. Does the weight of an hourglass depend on whether the sand flows in it?



11. Molly and Amy skated across a frozen river, rushing from one shore to the other after gaining some initial speed. Molly stopped, and the ice cracked under her. Why did this occur? The thickness of ice was identical everywhere.

12. Find the period of a mathematical pendulum in a spacecraft after the engines are turned off.

13. A bob oscillates on a vertical rubber cord. By how many times will the period of oscillation of the bob change if it is suspended on the same cord folded in two?

14. Why do radars emit electromagnetic waves in short wave packets and not continuously?

15. Why is lightning seen for a very short time, but its thunder lasts for a much more longer period?

Microexperiment

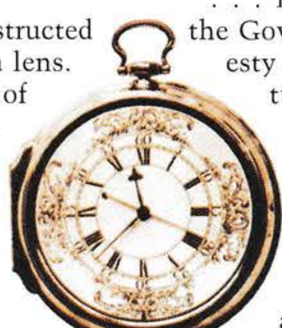
Pass a rope through a hook on the ceiling, attach a small bob to its end, and start it swinging with a small amplitude. Then gradually pull the other end of the rope, thereby lifting the bob. How will the period of oscillation of the bob change?

It is interesting that . . .

. . . in the second century B.C. an ancient Greek astronomer Hipparchus managed to calculate the duration of the Earth's year with marvelous accuracy: his figure was only 6 minutes too long.

. . . during the first millenium A.D., the Chinese reformed their calendar 70 times, and their system of chronology was modified 13 times.

. . . a sun-clock was constructed that employed a gun and a lens. The lens focused the rays of the Sun on the gun's primer and fired the gun at a precise moment of time. In this way the time was announced all through the neighborhood.



. . . an ancient Greek water clock (*clepsydra*) measured time according to the water level in a vessel, which had a small, leaky orifice. To make the water leak out uniformly, the shape of the vessel must satisfy a fourth-degree equation.

. . . In a watchmaker's shop in some Alpine country there was a banner on the wall: "This clock reads the precise time." The watchmaker corrected his clock daily according to the chime of a bell in the cloistral observatory. As it turned out, the inhabitants of this cloister determined time not from observations of the heavens, but from the clock of the rural horologist.

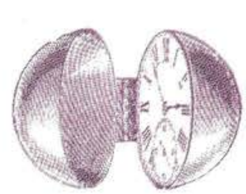
. . . in 1232 Emperor Friedrich II got a present from an Egyptian Sultan. It was a clock "with wheels and weights." In addition to time, it showed the motions of the Sun, Moon, planets, and stars.

. . . it was not till 1659, when Christiaan Huygens solved an important problem of constructing clocks. The clock was tuned by changing the pendulum's length. However, there was no shortage of attempts to dispute his priority; the Italians insisted that the pendulum clock was invented by Galileo himself.

. . . in 1714 the Government of Her Majesty of Great Britain instituted a reward for the construction of a marine clock suitable for the precise measurement of longitude. The size of the reward depended on the accuracy of the device.

. . . in the seventeenth century variational principles played an important role in the development of optics and mechanics. The most well-known of these are Pierre de Fermat's principle, which says that light always chooses the path corresponding to the minimum traveling time, and Johann Bernoulli's problem on the brachistochrone, the curve of the most rapid descent.

. . . the famous Russian self-educated inventor Ivan Kulibin spent two years constructing a unique egg-shaped clock. It consisted of 427 parts discernible only with a magni-



CONTINUED ON PAGE 34

Breaking up is hard to do

by Arthur Eisenkraft and Larry D. Kirkpatrick

"THE FISSION OF THE URAnium nucleus can be considered a very interesting paragraph (but only a paragraph) in the story of physics." So stated George Gamow, a noted physicist of this century, in 1961. The technological products of the discovery of fission, notably the atomic bomb and nuclear power, have greatly elevated its importance in our culture.

The entire development of the bomb cannot be understood without a comprehensive knowledge of the events of World War II. The history of the discovery of fission includes many aspects of the politics prior to the declaration of war. We recommend that readers turn to Richard Rhodes's outstanding book, *The Making of the Atomic Bomb*, for a stirring account of the history of physics in this era. We also recommend trying to locate *Moments of Discovery, The Discovery of Fission.* This out-of-print audiotape history published in 1984 by the American Institute of Physics (AIP) includes recordings of many of the principal players including Einstein, J. J. Thomson, Rutherford, Bohr, Hahn, Frisch, Compton, Szilard, and Fermi. It is quite a thrill to hear Rutherford state that "a nucleus is a very small thing."

Our chore is somewhat modest in comparison to both the science history and the political history surrounding nuclear fission. We wish to explore the details of fission, including when it occurs and how we can explain the enormous amounts of energy that are available.

O crucified Jove, do you turn your just eyes away from us or is there here prepared a purpose secret and beyond our comprehension?
—Dante

Let's begin with a recipe. The ingredients are 6 protons, 6 neutrons, and 6 electrons; the product is a carbon-12 atom.

6 protons	$6(1.007276\ \text{u}) = 6.043656\ \text{u}$
6 protons	$6(1.008665\ \text{u}) = 6.051990\ \text{u}$
6 protons	$6(0.000549\ \text{u}) = \underline{0.003294\ \text{u}}$
	total mass $= 12.098940\ \text{u}$

where $\text{u} = 1.66 \cdot 10^{-27}$ kg is the atomic mass unit.

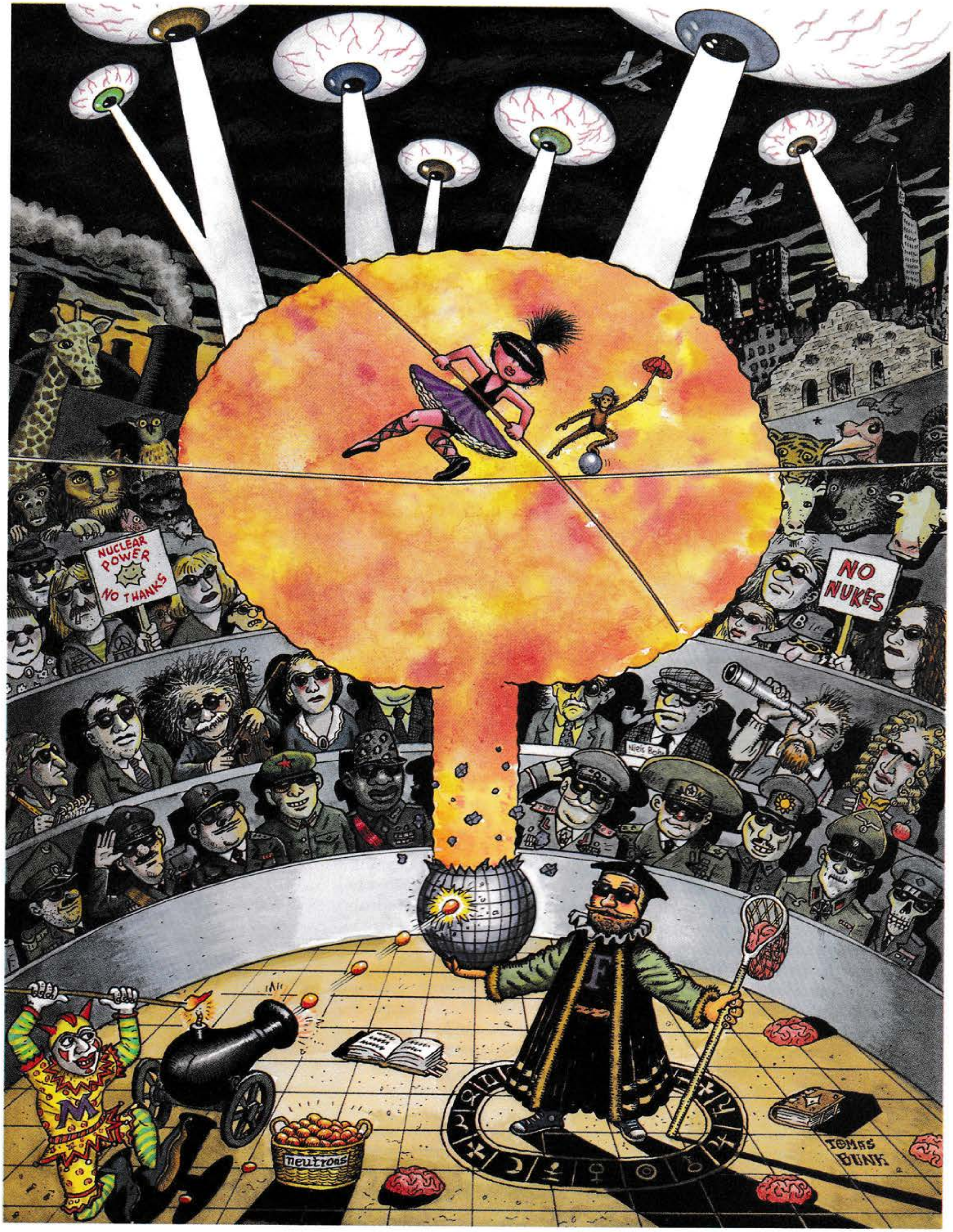
Surprise! The total mass of a neutral carbon-12 atom is exactly 12 u (by definition). So, where did the mass go? The missing mass, the mass defect, is actually released as energy when the nucleus is formed. From a different perspective, the carbon nucleus has a binding energy that holds the nucleus together. Removing a proton or neutron or separating all of the protons and neutrons requires an expenditure of energy. We have here a direct appreciation of Einstein's startling discovery in 1905 that mass and energy are one and the same and that the conversion factor for changing mass units into energy units is the square of the speed of light. This most celebrated equation of all of science is $E = mc^2$.

This equation shows us that a mere 1 g (10^{-3} kg) of mass yields $9 \cdot 10^{13}$ J of energy. To put this in perspective, if sold as electricity, this energy has a value of more than 2 million dollars. Applying Einstein's equation, we find that 1 u yields 931.5 MeV, where $1\ \text{MeV} = 1.6 \cdot 10^{-13}$ J.

Returning to the carbon nucleus, the mass defect is 0.098940 u. This has a corresponding binding energy of 92 MeV. If we repeat the analysis for carbon-11, we find that the mass of the atom is 11.011433 u and the mass of the constituent parts is 11.090275 u. The mass defect of carbon-11 is 0.078842 u or 73 MeV. The removal of this neutron must have required an expenditure of 19 MeV. Students of introductory physics are probably more familiar with the energy required to ionize a hydrogen atom, 13.6 eV. Thus, 13.6 eV of energy must be given to the electron to free it. Most of chemistry deals with the exchange of electrons and effectively deals with energies on the order of a few eV per atom. In contrast, changes in nuclear structure have corresponding energies of millions of eV.

It is informative to compare the average binding energy per nucleon of the two isotopes of carbon. Carbon-12 has an average binding energy of 92/12 MeV or 7.7 MeV per nucleon. Carbon-11 has an average binding energy of 73/11 MeV or 6.6 MeV per

Art by Tomas Bunk


NUCLEAR POWER NO THANKS
NO NUKES
Niels Bohr
neutrons
TOMAS BUNK

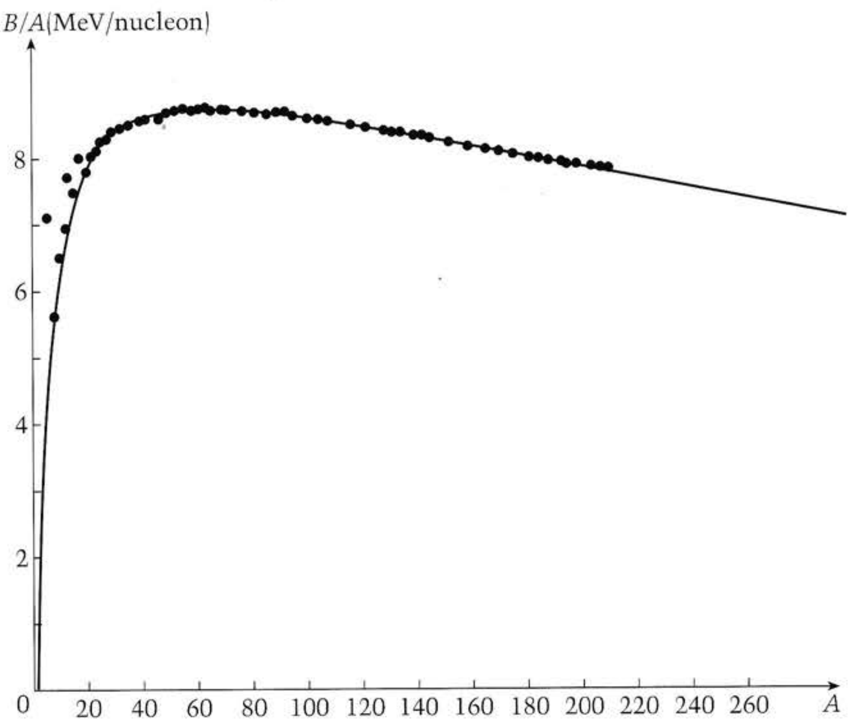


Figure 1. *Average binding energy per nucleon vs. mass number, for naturally occurring isotopes. The dots are based on measured binding energies; the smooth curve is based on the liquid-drop model. The curve has a maximum at A = 56 for iron-56.*

nucleon. A similar calculation can be done for isotopes of all elements, and the curve that is generated is shown in figure 1. The average binding energy for most nucleons is approximately 8 MeV/nucleon. The curve has the interesting feature of having a maximum at iron-56. Nuclei with less mass and nuclei with more mass have less binding energy per nucleon. Thus, light nuclei combining to form a heavier nucleus will release energy in a process called fusion. A heavy nucleus will release energy when it splits into two lighter nuclei in a process called fission.

The binding energy curve came about after the discovery of fission. When Hahn and Strassman first recognized the fission of uranium, it startled them and the scientific world. The discovery of the neutron by Chadwick in 1932 had provided a new research tool for nuclear physics. The neutron could enter the nucleus without having to overcome the Coulomb repulsive force that a proton would experience. Enrico Fermi, soon after the discovery of the neutron, began bombarding elements with neutrons and produced many new isotopes. Since many of the isotopes Fermi produced emitted beta particles, when he bombarded uranium he thought that he had discovered a transuranic element.

Otto Hahn and his colleagues Lise Meitner and Fritz Strassman had been chemically analyzing radioactive elements for some time. Unfortunately, Meitner was forced by the Hitler regime to leave Germany in July 1938. Hahn and Strassman found that one of the elements emerging from the uranium nucleus was barium. Their first hypothesis was that it was radium, which would have required two alpha particles to leave the uranium nucleus. Even this seemed unlikely—a low energy neutron knocking two alphas from the nucleus was beyond expectation. Barium was even more unlikely, but barium it was. Meitner and her nephew Otto R. Frisch were quick to deduce that the addition of the neutron caused instability and the uranium nucleus broke into two parts. If barium was one piece, the other must be krypton. This element had indeed also been detected. Frisch mentioned fission to Niels Bohr, who was on his way to America, and en route Bohr and a colleague, Leon Rosenfeld, mapped out the liquid drop model of the atom that could predict this surprising behavior.

The liquid drop model of the nucleus treats the nucleus as a droplet of nuclear material. The nucleons on the surface are held to the drop by a surface tension. The model is quantitative and leads to an equation that can accurately predict the binding energy curve. Our explanation of the relevant equation follows that of Ohanian in his text *Modern Physics*.

To begin, we must recognize that the nuclear radius is proportional to $A^{1/3}$, where A is the atomic mass. This relationship is the result of many scattering experiments. That being the case, the volume of a nucleus ($4/3\pi R^3$) is proportional to A and the density of nuclear matter (the ratio of mass to volume) must be constant for all nuclei.

The most important term in the derivation of a binding energy equation is associated with the number of nucleons, since each nucleon attracts every other one through the short-ranged strong force. The binding energy must contain a term proportional to A. Since the nucleons on the surface do not have as many neighbors as nucleons within the interior, the second term must account for this decrease in the binding energy. Since R^2 is proportional to $A^{2/3}$, the correction term is proportional to $A^{2/3}$ and is negative. The Coulomb repulsion force between all of the protons tends to drive the nucleus apart. This term is proportional to $Z^2/A^{1/3}$, where Z is the number of protons in the nucleus.

Finally, there is a quantum-mechanical correction that takes into account the exclusion principle. Just as electrons cannot all be in the same quantum state, but fill successive shells, the nucleons must also fill shells. This leads to a term that is related to the numbers of protons and neutrons and the total number present. The constants of proportionality are found through numerous experimental data. Weizsacker's semi-empirical formula for the binding energy is

$$B = \left[15.753A - 17.804A^{2/3} - 0.7103\frac{Z^2}{A^{1/3}} - 94.77\frac{\left(\frac{1}{2}A - Z\right)^2}{A}\right]\text{MeV}.$$

The smooth curve in figure 1 is based on the liquid-drop model and can be seen to fit the data exceedingly well.

Assuming that a nucleus splits into two equal parts ($A/2$, $Z/2$ for each product), we can calculate the difference in binding energies:

$$B_{A,Z} = \left[15.753A - 17.804A^{2/3} - 0.7103\frac{Z^2}{A^{1/3}}\right]\text{MeV},$$

$$B_{A/2,Z/2} = \left[15.753\frac{A}{2} - 17.804\left(\frac{A}{2}\right)^{2/3} - 0.7103\frac{(Z/2)^2}{(A/2)^{1/3}}\right]\text{MeV},$$

$$2B_{A/2,Z/2} - B_{A,Z} = \left[-4.6A^{2/3} + \frac{0.26Z^2}{A^{1/3}}\right]\text{MeV}.$$

We have ignored the quantum mechanical term, which is small in comparison. If the nucleus were to undergo fission, the electrostatic force must be greater than the surface tension.

$$-4.6A^{2/3} + \frac{0.26Z^2}{A^{1/3}} > 0,$$

$$\frac{Z^2}{A} > 18.$$

In the case of ^{238}U, we have $Z = 92$, $A = 238$, and $Z^2/A = (92)^2/238 = 35.6$, which is certainly larger than 18. Fission does not proceed directly but requires an elongation of the nuclear drop. This is more likely to occur when the additional neutron is added to the uranium nucleus.

Enrico Fermi and Emilio Segre did not discover the fissioning of uranium, although fission did indeed occur during their 1934 experiments. Segre is quoted as saying, "The whole story of our failure is a mystery to me. I keep thinking of a passage from Dante: 'O crucified Jove, do you turn your just eyes away from us or is there here prepared a purpose secret and beyond our comprehension?'" (from AIP's *Moments of Discovery* audiotape). The discovery of fission in 1939 led immediately to the development of the atomic bomb effort, which included Fermi, who was then living in the United States. How might world history have been altered if the discovery of fission had occurred before the emigration of physicists to the United States and well before the start of World War II? What does this suggest about the role of chance in history?

The contest problem this month includes analysis of a number of features of the fission process.

A. It is not hard to follow the reasoning Frisch and Meitner used to calculate the energy released in fission. Consider a typical fission reaction:

$$^{235}_{92}\text{U} + {}^{1}_{0}\text{n} \rightarrow {}^{140}_{54}\text{Xe} + {}^{94}_{38}\text{Sr} + 2{}^{1}_{0}\text{n}.$$

The Xe rapidly decays into $^{140}_{58}$Ce and the Sr into $^{94}_{40}$Zr, with the emission of electrons of negligible mass. We now know the following masses:

$^{235}_{92}$U	235.004 u
$^{1}_{0}$n	1.009 u
$^{140}_{58}$Ce	139.905 u
$^{94}_{40}$Zr	93.906 u

Calculate the energy released in the fission reaction.

B. The discovery and the exploitation of fission did not require knowledge of $E = mc^2$. In fact, at the time the masses of the radioactive daughter nuclei were not known well enough to make a good calculation. Frisch and Meitner calculated the energy release by a second method (which was the only method Joliot used).

Calculate the radii of the Ce and Zr nuclei above using the approximate equation, $R = KA^{1/3}$, where $K = 1.0 \cdot 10^{-15}$ m. Assume that at the moment the uranium breaks into these fragments, the distance between the centers of the two fragments is equal to the sum of their radii. Calculate the electrostatic repulsion between them. Using the electrostatic potential, calculate the work done to separate these two fission products. Compare this total energy with that found in part A.

C. Surprisingly, the uranium rarely breaks into two equal products. Use the semi-empirical binding energy equation to show that the energy released is greatest for the symmetric rare fission.

Elevator physics

Sid Govindan and Japeck Tang, two students of Art Hovey, physics teacher at Amity Regional High School in Woodbridge, Connecticut, submitted correct solutions to the contest problem in the March/April 1999 issue. They reported that they particularly enjoyed the problem in part C.

In part A our readers were asked to show that a ball dropped in an elevator accelerating in the upward direction will return to its original height relative to the elevator floor if its collision with the floor is completely elastic. Let's begin by writing equations for the positions of the ball and elevator after the collision with the floor using the same notation we used in developing the problem:

$$y'_b = y_h + (v_h + v_r)t - \tfrac{1}{2}gt^2$$
$$y'_f = y_h + v_h t + \tfrac{1}{2}at^2,$$

where y'_b and y'_f are the positions of the ball and floor, respectively, y_h is the position of the floor at the time

$t = 0$ of the collision, v_h is the velocity of the floor at $t = 0$, v_r is the relative velocity of the ball and the floor at $t = 0$, and a is the upward acceleration of the elevator.

The height of the ball h' above the elevator floor is given by the difference of the coordinates of the ball and floor:

$$h' = y'_b - y'_f = v_r t - \tfrac{1}{2} g' t^2,$$

where $g' = g + a$ is the effective acceleration due to gravity in the elevator. Earlier we calculated $v_r = g' t_d$, where t_d is the time for the ball to fall to the floor. Substituting this expression and using our results that the time t_u for the ball to rise to its highest position above the floor equals t_d, we have

$$h' = \tfrac{1}{2} g' t_d^2.$$

Lastly, we use

$$t_d = \sqrt{\frac{2h}{g'}}$$

to show that $h' = h$. Once again, we obtain the same result that we would get on the ground: the ball returns to its original height.

Part B asks about dropping a ball in a train with a constant horizontal acceleration a. The effective gravity $\mathbf{g'}$ is given by the vector difference of $\mathbf{g}$ and $\mathbf{a}$ and makes an angle θ with the vertical such that $\tan\theta = a/g$. The ball dropped in the train falls along the direction of $\mathbf{g'}$ just like a ball dropped while standing on the ground falls along the direction of $\mathbf{g}$.

We use the idea of an effective gravity to solve a very interesting problem in a simple way. The door on a car is slightly ajar. If the car accelerates uniformly from rest, how far will the car travel before the door slams shut? We model the car door as a rectangle with a uniform mass distribution and a length L from front to back and treat the door as being acted on by a gravitational force in the backward direction. We can ignore the real gravitational force, because the hinges do not allow motion up and down.

The torque acting on the door is given by

$$\tau = \tfrac{1}{2} Lma \sin\theta,$$

where θ is the angle between the door and the side of the car. If we assume that the door is thin, we can think of the door as being constructed from a column of thin rods. The moment of inertia of the door about its hinges is then

$$I = \tfrac{1}{3} mL^2.$$

Newton's second law for rotational motion yields

$$\tau = I\alpha = \tfrac{1}{2} Lma \sin\theta = \tfrac{1}{3} mL^2 \alpha.$$

Therefore,

$$\alpha = \frac{3a}{2L} \sin\theta.$$

This is just the equation for a simple harmonic oscillator if we make the approximation that $\sin\theta \cong \theta$. The period for this motion is

$$T = 2\pi \sqrt{\frac{2L}{3a}}.$$

The time for the door to close is just one-fourth of this period.

Therefore, the distance traveled by the car before the door slams shut is

$$d = \tfrac{1}{2} at^2 = \tfrac{1}{8} aT^2 = \tfrac{1}{12} \pi^2 L.$$

CONTINUED FROM PAGE 29

fying glass. The clock struck on every quarter hour, and every hour it showed a performance of a miniature theatre with music.

. . . bubble chambers, which are used to detect elementary particles, made it possible even in the 1950s to determine the mean lifetime of a particle with an accuracy of about 10^{-11} s.

. . . one of the wonderful predictions of the theory of relativity, the dilation of time in a gravitational field, was not experimentally proved until 1960. To measure this effect, a fantastic accuracy of $3 \cdot 10^{-12}$ percent was achieved, which placed this experiment into the record book of modern physics.

. . . the shortest interval of time that manifests itself in experiments is less than $3 \cdot 10^{-27}$ s. This is the time needed for light to travel an electron's diameter, known to be less than 10^{-18} m.

. . . until recently the most precise atomic clock was made in the United States: it could measure 3 million years with an accuracy of 1 s. However, Germany is ready to beat the record and to measure 1 billion years with the same 1-second accuracy.

—A. Leonovich

Quantum on clocks, time, eternity:

A. I. Chernoutsan, "Time Machines and the Theory of Relativity," September/October 1992, pp. 50–51.

I. Lalayants and A. Milovanova, "Physics Frights Frauds," January/February 1993, pp. 11–16.

V. M. Babovic, "Confession of a Clock Lover," September/October 1996, pp. 44–48.

A. A. Mikhailov, "The Long Road to Longitude," March/April 1997, pp. 42–47.

V. I. Kuznetsov, "A Clock Wound for All Time," May/June 1997, pp. 26–30.

AT THE BLACKBOARD I

Alexandrian astronomy today

by Case Rijsdijk

THE APPROACH OF THE LUnar eclipse of April 1996 seemed an opportune time to develop an exercise to calculate the size of, and distance to, the Moon as part of the South African Astronomical Observatory's Science Education Initiative.

This project led me to rediscover some aspects of early Greek astronomy as practiced by astronomers who worked in the Great Library of Alexandria from about 300 B.C. to about A.D. 150. Aristarchus used a lunar eclipse to estimate the size of the Moon; Hipparchus later improved on this estimation. The same geometry was used again much later by Copernicus in "De Revolutionibus."

If the size of Earth is calculated using Eratosthenes' method (which is explained later in the article) and an eclipse of the Moon is observed and photographed, then the distance to the Moon and the size of the Moon can be found using some simple equipment, straightforward geometry, basic trigonometry, and a little ingenuity.

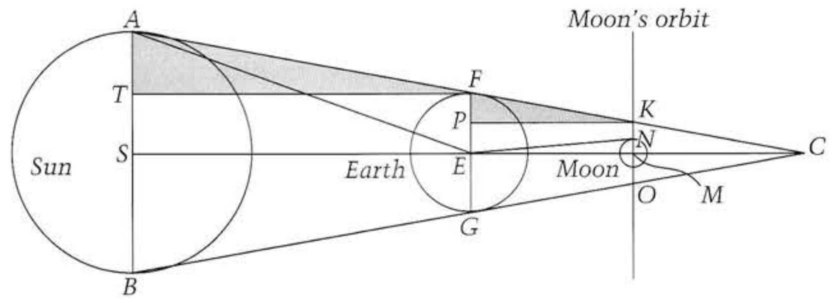
Lunar eclipse geometry

Triangle FGC (fig. 1) represents Earth's shadow, and line KMO represents the Moon's orbit. Assuming that the Sun is n times farther from Earth than the Moon is, we have

$$SE = n \cdot EM,$$

and substituting variables as indicated below the figure yields

$$n = \frac{SE}{EM} = \frac{D}{d}. \qquad (1)$$



Earth–Sun distance ES = D
Earth–Moon distance EM = d
Sun's radius AS = R
Earth's radius EF = r
Moon's radius MN = a
Radius of Earth's shadow MK = s

Figure 1

As can be seen during an eclipse of the Sun, the Sun and Moon appear to be the same size, because they subtend the same angle (approximately) in the sky, so $\angle AES = \angle MEN$. Therefore, triangles ASE and MNE are similar. (They don't look similar, because the diagram is not drawn to scale.) Then, using equation (1), we have

$$AS = n \cdot NM.$$

Substituting variables yields

$$n = \frac{AS}{MN} = \frac{R}{a}. \qquad (2)$$

Triangles ATF and FPK are also similar, and we obtain

$$\frac{TF}{PK} = \frac{AT}{FP},$$
$$\frac{SE}{EM} = \frac{AS - TS}{FE - PE}.$$

Therefore,

$$\frac{SE}{EM} = \frac{AS - FE}{FE - MK}. \qquad (3)$$

Using the assigned symbols and substituting equations (1) and (2) into equation (3) gives us

$$\frac{D}{d} = \frac{nd}{d} = \frac{R - r}{r - s},$$

or

$$n = \frac{na - r}{r - s}. \qquad (4)$$

We can rearrange equation (4) to obtain

$$n(a + s) = r(n + 1).$$

Further manipulation yields

$$a\left(1 + \frac{s}{a}\right) = r\left(1 + \frac{1}{n}\right). \qquad (5)$$

So to find the distance to the Moon d, we must find values for r, n, and a in terms of d and the ratio s/a. This ratio can be found during an eclipse of the Moon. The radius of the Moon a can be found in terms of d by measuring the angular diameter of the Moon. Using Eratosthenes' method, Earth's radius r can be found and n, well, that can be overcome!

Earth's shadow and the Moon compared

The Moon is photographed during an eclipse to get an image that looks similar to figure 2. This occurs when the Moon is entering the um-

Figure 2

bra. It is best to take a series of photographs with a large telephoto lens or telescope so that there is a selection to choose from. The photograph is enlarged so that accurate measurements can be made from it. If there isn't a convenient eclipse to photograph, it is also possible to copy a picture from a magazine or book of a previous eclipse.

The photograph is needed to find the ratio of the radius of the Moon to that of Earth's shadow, or how many times the diameter of Earth's shadow is larger than the diameter of the Moon. There are several ways to do this, but the best way is to use the geometry of the circle and a scale drawing. By making as large a photocopy of the photograph as possible, points C and D are marked in such a way that they are as far apart as possible along an imaginary radius extended.

Points A and B are the points of intersection of the circles. The perpendicular bisector of the chord of a

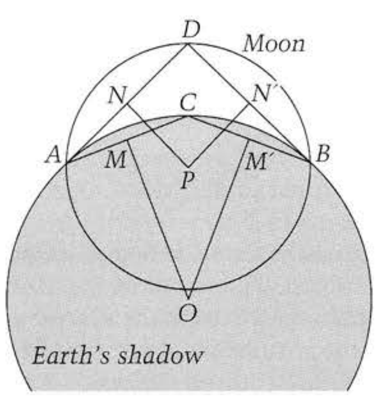


Figure 3

circle passes through the center of that circle. Referring to figure 3, we can see that N is the midpoint of AD and N' is the midpoint of DB. The perpendiculars from these points meet at P. Segment DP would then be the radius of the Moon. Similarly, OC is the radius of Earth's shadow at the distance of the Moon's orbit, and we obtain the ratio

$$\frac{OC}{DP} = \frac{s}{a}. \qquad (6)$$

Angular size of the Moon

We can measure the angular size of the Moon directly or indirectly. To measure it directly, a slider is made to move smoothly along a meterstick. A small ball bearing (about 6 mm in di-

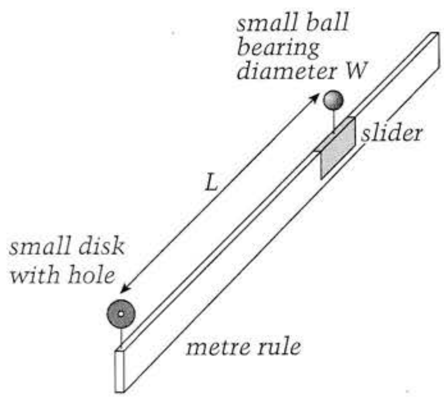


Figure 4

ameter) is mounted on the slider. A disk with a small hole in it is attached to one end as shown in figure 4. The meterstick is rested on something firm, and the Moon is sighted through the small hole in the fixed disk. The slider is now moved so that the ball bearing exactly covers the Moon. The angle subtended by the ball bearing is now the same as that subtended by the Moon:

$$\tan \alpha = \frac{W}{L}.$$

Therefore,

$$\alpha = \tan^{-1} \frac{W}{L}. \qquad (7)$$

Great care must be exercised when taking these measurements, as this result is crucial to obtaining an accurate result.

An indirect measurement of the angular size of the Moon is possible with the help of a pinhole camera. Since the Sun and the Moon appear to be the same size (subtend the same angle), as shown during a solar eclipse, it is possible to construct a good size pinhole camera and get a reasonable image of the Sun that can then be marked. The ratio of image size to image distance will yield the required angle α.

Emulating Eratosthenes

Eratosthenes' experiment can easily be repeated using some shadow sticks. First another school is found on a N–S line at least 500 km away. Once contact has been established, students at each location measure the length of the shadow of a vertical stick at the same time. Care must be taken to ensure that the stick is vertical by using a simple plumb line made from a "bulldog" clip, some string, and a lead sinker (fig. 5). For ex-

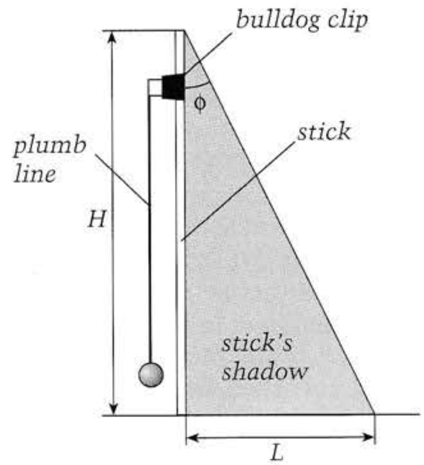


Figure 5

ample, two places W and T have been chosen (fig. 6), and at each place the length L of the shadow and height H of the stick is measured at, say, 12:00

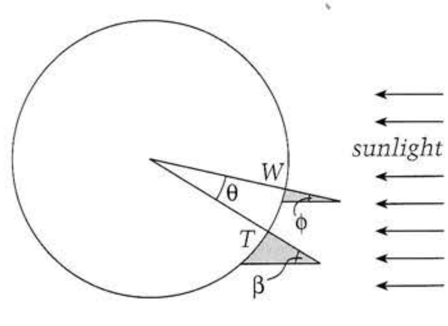


Figure 6

noon. The angle ϕ can then be found:

$$\tan\phi = \frac{L}{H},$$

or

$$\phi = \tan^{-1}\frac{L}{H}.$$

Similarly, the angle β at T can be calculated. We can then use geometry to show that

$$\theta = \beta - \phi.$$

The distance $WT = S$ is found from an atlas, and then by simple proportions we obtain

$$\frac{S}{C} = \frac{\theta}{360} = \frac{S}{2\pi r},$$

where C is Earth's circumference and r its radius. This is rearranged to get

$$r = \frac{S \cdot 360}{2\pi\theta}. \qquad (8)$$

There are two main sources of error in this experiment. The first is that, due to diffraction, there is some difficulty in seeing a clear shadow and in determining exactly where the shadow ends; it usually covers about 2 cm or so. The second source of error is the stick not being vertical. With some care this problem can be minimized. The first source of error is best confronted by making the stick as long as possible and securing a T piece to its top. Accuracy will also be improved if many different groups of students do this part of the experiment and average the results.

Earth-to-Moon distance

To find the distance from Earth to the Moon, we begin by using equation (5) and substituting the following values that have been found:

- from equation (7), the angle α subtended by the Moon. This was the angle subtended by the diameter; for the radius half this value is required. That is,

$$\tan\frac{\alpha}{2} = \frac{a}{d}.$$

Therefore,

$$a = d\tan\frac{\alpha}{2}. \qquad (9)$$

- the ratio s/a that was obtained from the eclipse and using equation (6).
- the value for r obtained using the shadow sticks and equation (8).

This leaves n. Aristarchus realized that the Sun was farther away from Earth than the Moon was, but he did not know how many times farther. He tried to work it out using geometry, but the value he got, 20, was too small. However, if it is assumed that n is very large, then the ratio $1/n$ is very small and can be neglected. This gives a minimum value for the radius of the Moon of

$$a\left(1 + \frac{s}{a}\right) = r,$$

from which we obtain

$$a = \frac{r}{1 + \frac{s}{a}}.$$

The value for a can now be found, since all the other values are known. Once a has been found, d can be calculated using equation (9), from which we get

$$d = \frac{a}{\tan\frac{\alpha}{2}}.$$

Alexandrian values

Hipparchus used the following values:

$$s/a = 8/3,$$
$$\alpha = 31',$$

so $\alpha/2 = 15'5$, and thus

$$a \cong d/220.$$

He used Eratosthenes' value for r of 6,500 km. Substituting into equation (5) gives the following result:

$$\frac{d}{220}\left(1 + \frac{8}{3}\right) = (6500 \text{ km})\left(1 + \frac{1}{n}\right).$$

Therefore,

$$d = (390{,}000 \text{ km})\left(1 + \frac{1}{n}\right).$$

Then if, as assumed, n is large:

$$d = 390{,}000 \text{ km}.$$

Using the fact that $\alpha = 31'$, the diameter of the Moon is

$$(390{,}000 \text{ km}) \tan\alpha = 3{,}516 \text{ km}.$$

This figure doesn't compare too badly with presently accepted figures:

diameter: 3,476 km
mean distance: 384,404 km (range of 356,400 to 466,700 km).

data collected	school 1	school 2
ratio of Earth shadow/Moon, (s/a)	2.67	2.67
angular diameter of Moon	26′	31′
Earth's radius	6,383 km	6,384 km
calculated values		
distance to Moon from Earth	457,731 km	386,161 km
diameter of Moon	3,461 km	3,476 km

Table 1

Student astronomers

Several schools in South Africa took part in the project, and the two results in table 1 are fairly typical.

While not accurate, these results do give figures that are not too far from the true values. It is fairly obvious that the critical result in this experiment is the angular size of the Moon, and the statistical mean from a large number of readings appears to be the best way to obtain an accurate result.

Case Rijsdijk is responsible for science and astronomy education at the South African Astronomical Observatory in Observatory, South Africa. He can be reached at case@saao.ac.za.

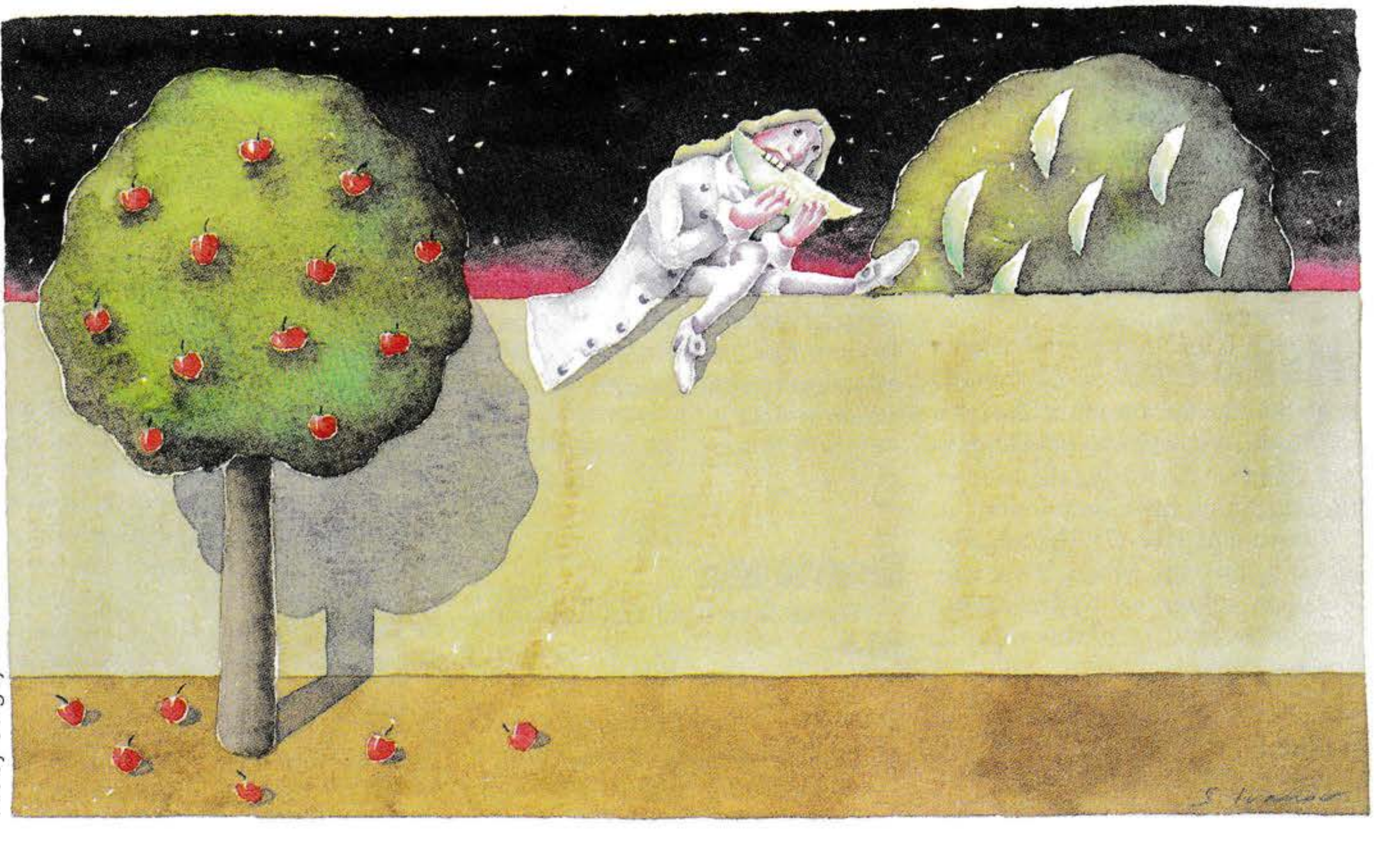
The great law

by V. Kuznetsov

A GREAT CALAMITY OCcurred in Great Britain in 1665 and 1666: the plague ran its deadly course through the population. Citizens of large towns left their homes to save themselves in the small villages. This is why the young philosopher and mathematician Isaac Newton left London and returned to his native town of Woolsthorpe. In this period he was interested in the problem of why the Moon revolves about the Earth. What force keeps it in orbit? For if there were no force, he reasoned, the Moon would have left Earth long ago.

We can imagine the course of Newton's reasoning. He thought of an apple as an astronomical body. An apple always falls downward. Why? It seems as if Earth attracts it. If one body attracts another body, then the first body should have an attractive force proportional to its mass. If this is true, the attractive force generated by Earth and applied to the Moon must be formidable. Thus, both the apple and the Moon are attracted by Earth. Perhaps this is how Newton conceived of the law of universal gravitation.

Let's make a calculation analogous to that made by Newton. The Moon's orbit is almost circular. Its period of revolution T_M around Earth is 27.3 days. The distance between Earth and the Moon is about 60 times greater than Earth's radius R_E. If an object moves along a circular trajectory with a speed v, it has a centripetal acceleration v^2/R. The radius of the lunar orbit is $R_M = 384{,}400$ km and the Moon's orbital velocity



Art by Sergey Ivanov

is $v = 2\pi R_M/T_M = 3680$ km/h, so the centripetal acceleration is $a = v^2/R_M = 0.0027$ m/s^2. We know another value of centripetal acceleration: it is the acceleration of free fall at Earth's surface $g = 9.81$ m/s^2, which is larger by far than the Moon's centripetal acceleration.

Newton was not the only person who thought about the attractive ability of Earth. His colleague and opponent Robert Hooke (1635–1703) tried to measure how the force of gravity changes with altitude. To this end he used a spring scale (one of his inventions) and carried it to the top of a hill. However, on the hilltop the load stretched the spring the same amount as at sea level. Still Newton felt that the Earth-generated attraction must decrease with distance.

The empirical laws of planetary motion found by Johannes Kepler (1571–1630) helped Newton find the simple law of how the gravitational attraction decreases with distance: If the centers of two balls with homogeneously distributed masses are a distance r apart, the attractive force F between them is directed along the line connecting their centers, and the value of this force is directly proportional to the product of the balls' masses m_1 and m_2, and inversely proportional to the square of the distance r:

$$F = G\frac{m_1 m_2}{r^2}. \qquad (1)$$

This law contains the gravitational constant G that must be determined by experiment.

Now we return to the attractive force acting on the Moon. The Moon is situated at a distance from Earth that is 60 times Earth's radius. Therefore, the force of attraction and the acceleration due to Earth-generated gravity at the distance of the Moon's orbit are 60^2 times smaller than at Earth's surface:

$$\frac{a}{g} = \frac{1}{3600},$$

or

$$a = 0.0027 \text{ m/s}^2.$$

This value agrees with the previously calculated lunar centripetal acceleration.

In this way Newton could test the law of universal gravitation. However, his calculation did not demonstrate an exact agreement between the astronomical and theoretical values for the lunar centripetal acceleration: living in Woolsthorpe, Newton did not know the precise value of Earth's radius. This value was not measured with reasonable accuracy until years later.

Notwithstanding the rough agreement of both values, Newton was not satisfied. There was the attraction of the apple to Earth to be understood. Can one consider that Earth attracts the apple with the same force as if all of Earth's matter were concentrated in one point, at the center of the planet? In reality, the nearer parts of Earth attract the apple much more strongly than the more distant parts.

What to do? The creative genius of Newton couldn't rest, so he switched his focus to optics. He ground lenses, constructed an excellent telescope, and devotedly studied optical spectra.

Still, his thoughts kept gravitating back to the problem of the apple. He spent years inventing integral calculus and with this powerful tool proved a wonderful theorem: a spherical shell with homogeneously distributed mass attracts a body in the same way as if the entire mass of the shell were concentrated at its center (figure 1a). Earth can be considered to be a set of concentric spherical shells (figure 1b). The force generated by each shell doesn't depend on the other shells, and the gravitational force is not diminished (screened) by intervening shells. If we accept these two conditions (the second condition is the most wonderful feature of gravitation), then it is clear that Earth attracts an apple just as Newton supposed.

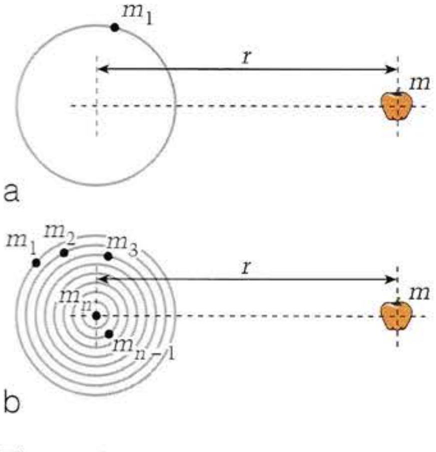


Figure 1

So, there is a common cause of an apple's fall and the attraction of the orbiting Moon by Earth: the force of gravity generated by our planet. What forces control the motions of the planets themselves? Since they orbit the Sun, the gravitational force must be generated by the Sun. Taking a small step further, we can generalize: any two bodies attract each other with the force of gravity described by equation (1).

Johannes Kepler discovered the laws of planetary motion at the beginning of the seventeenth century. Kepler's first law says that the planets move along ellipses, the Sun being at one of the foci. According to his second law, the planets move faster near the Sun and slower at larger distances from it. Kepler found a mathematical law to describe this feature of planetary motion: the line connecting the Sun and a planet sweeps equal areas per unit time. Kepler's third law compares the orbits of different planets: the squares of the periods of revolution of the planets are proportional to the cubes of their mean distances from the Sun.

Kepler found these laws on the basis of many years of astronomical records. By contrast, Newton showed that these laws could be deduced from his laws of motion and universal gravitation.

Once upon a time, three celebrities of the Royal Society met in a London tavern. They were the famous architect Sir Christopher Wren (1632–1723) and the naturalists Robert Hooke and Edmond Halley (1656–1742). Halley said that he managed to obtain Kepler's third law by combining the law of universal gravitation with Newton's laws of motion—but only in the case of circular orbits. Indeed, the circular motion is caused by the centripetal force Mv^2/R that pulls a planet toward the center of the circle. This force is generated by the Sun, so

$$G\frac{M_S M}{R^2} = \frac{Mv^2}{R}.$$

For circular motion, $v = 2\pi R/T$. Plugging this into the formula, we immediately obtain Kepler's third law:

$$G\frac{M_S}{R} = \frac{4\pi^2 R^2}{T^2},$$

or

$$\frac{R^3}{T^2} = \text{const.}$$

When Halley finished his story, Sir Christopher raised his glass and announced a prize to anyone who could prove that the law of universal gravitation also agrees with the elliptical orbits of the planets.

After a while, Halley asked Newton about this problem. In November 1684 Halley received Newton's manuscript with the solution of the problem. Several years were necessary to edit and publish the manuscript. At last, in September 1687, Newton's treatise was published. It was the famous "Philosophiae Naturalis Principia Mathematica."

The laws applied to the motion of the planets yielded some important predictions. For example, Sir Isaac "weighed" the Sun by expressing its mass in units of Earth's mass. Indeed,

$$G\frac{M_S M_E}{R_E^2} = M_E\frac{v^2}{R_E} = M_E\frac{4\pi^2 R_E^2}{R_E T_E^2},$$

$$M_S = \frac{4\pi^2 R_E^3}{G T_E^2},$$

$$G\frac{M_E M_M}{R_M^2} = M_M\frac{v^2}{R_M} = M_M\frac{4\pi^2 R_M^2}{R_M T_M^2},$$

$$M_E = \frac{4\pi^2 R_M^3}{G T_M^2},$$

where R_E and R_M are the orbital radii of Earth and the Moon, respectively. We get an interesting result. The Moon is the satellite of Earth, and Earth is a satellite of the Sun. In every equation the mass of a satellite is canceled out. In the first equation where the law of universal gravitation is applied to the Sun and Earth, it is the mass of Earth that is canceled out, while in the second equation the canceled term is the mass of the Moon. Therefore, the masses of Earth and the Sun are expressed in terms of values that are readily measured by astronomers: the orbital radii and the periods of revolutions. Only one value remained unknown—the gravitational constant G. However, if we take the mass ratio of the Sun and Earth, the constant G is also canceled out, so we get the formula

$$\frac{M_S}{M_E} = \frac{R_E^3 T_M^2}{R_M^3 T_E^2} = \left(\frac{R_E}{R_M}\right)^3 \left(\frac{T_M}{T_E}\right)^2.$$

To calculate the Sun's mass in common units, we need to know the mass of Earth. Newton knew only its volume, because the mean density of Earth's matter wasn't known in his time. Only one thing was firmly established: the continents were denser than the oceans. But by what factor was the mean density of Earth larger than that of the oceans? Newton concluded that the average density of Earth was somewhere between 5 and 6 g/cm^3. Eighty years later Sir Henry Cavendish (1731–1810) "weighed" Earth and obtained the constant G using a torsion balance. It turned out that Earth's density is 5.5 g/cm^3.

How can one determine the mass of the Moon? Unfortunately, it cancels from all the equations. Still Sir Isaac found a way to estimate this value as well.

He turned his mind to the phenomenon that had puzzled humankind for thousands of years. What causes the tides in the oceans? The Romans attributed the tides to the position of the Moon in the sky. "It was a full Moon and a great tide," wrote Julius Caesar himself. Still, nobody guessed what a close connection the Moon had with the tides. Newton was the first to find the correct answer.

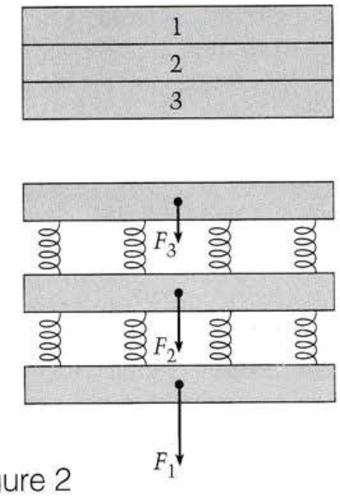


Figure 2

Let's consider the motion of three plates of the same mass that lie freely one upon another (figure 2). At some time, three forces start to act on these plates: $F_1 > F_2 > F_3$. All the plates will be set in motion. However, plate 3 will lag behind the central plate 2, and plate 1 will move ahead of it. If the outer plates are connected to the central plate with some springs, the stretching of the springs will counterbalance the net forces $F_1 - F_2$ and $F_2 - F_3$.

Earth and its hydrosphere can be imagined as three bodies falling toward the Moon: the hard core and two layers of water: one facing the Moon and the other located on the opposite side of Earth. The Moon attracts these imaginary "bodies" differently. Let's first consider the effect of lunar attraction upon a unit of terrestrial mass (figure 3). At an

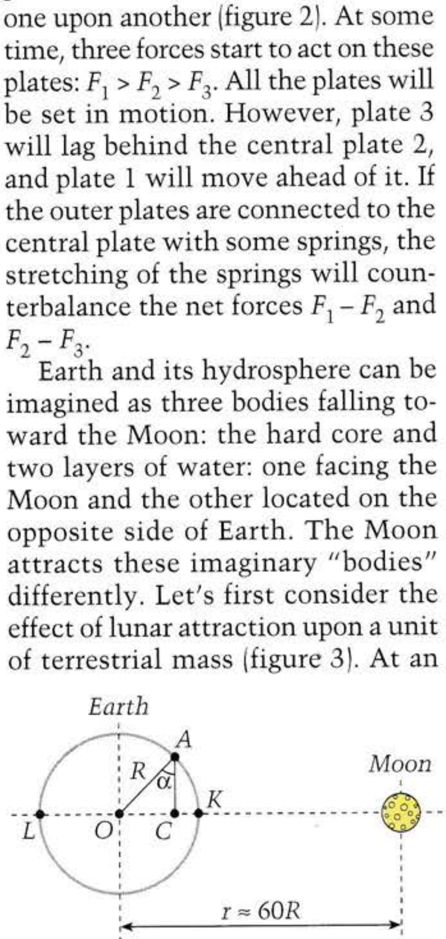


Figure 3

arbitrary point A this force is

$$F_A = G\frac{M_M}{R^2\cos^2\alpha + (r - R\sin\alpha)^2} \cong G\frac{M_M}{r^2}\left(1 + 2\frac{R}{r}\sin\alpha\right).$$

At point K, $\sin\alpha = 1$, and the attraction force is

$$F_K = G\frac{M_M}{r^2}\left(1 + 2\frac{R}{r}\right).$$

Similarly, at point L, $\sin\alpha = -1$, so

$$F_L = G\frac{M_M}{r^2}\left(1 - 2\frac{R}{r}\right).$$

Thus, the various parts of Earth are attracted to the Moon by different forces.

The Earth's solid mass and its watery covering are connected by "springs" of gravity. In the center of Earth, a unit mass is affected by the force $F_c = GM_M/r^2$ of lunar gravitation. This is the very force that determines the motion of Earth's core. At the point L the ocean water "lags behind" in the process of Earth "falling" toward the Moon, because here a unit mass is affected by a force smaller than F_c: $F_L - F_c = -2GM_MR/r^3$. By contrast, at point K this force is larger than F_c, so in the vicinity of this point, water tries to run faster than Earth's core. However, the "springs" of terrestrial gravitation don't allow it to do so. The lunar gravitation slightly "stretches" these springs with the tidal force $2GM_MR/r^3$. Therefore, tidal "humps" appear on opposite sides of Earth. They try to maintain the same position relative to the Moon. If the Moon didn't move relative to Earth, and Earth did not rotate about its axis, the water shell would maintain its shape elongated toward the Moon. However, due to Earth's rotation, the tidal wave moves relative to the continents at 1800 km/h and lags relative to the Moon's motion.

While solar attraction is much stronger than lunar attraction, it is more homogeneous due to the large distance to the Sun relative to Earth's size. Thus, the solar tidal force $F_S = 2GM_SR/r_{SE}^3$ is smaller than the lunar tidal force. Indeed, the Sun's mass is 27 million times greater than the Moon's, but the distance between the Sun and Earth is 389 times the distance between Earth and the Moon. Therefore,

$$F_S = 2G\frac{27\cdot 10^6 M_M}{(389r)^3}R \cong 0.45F_M.$$

Thus, the solar tides are weaker than lunar tides. Twice a month the Sun, Earth, and Moon are situated on the same line. In this case the solar and lunar tidal forces interfere constructively and generate a large tide. When the Earth-Sun line is perpendicular to the Earth-Moon direction, the solar and lunar tidal forces act destructively and produce a small tide. Near small isolated islands lost in an ocean where the tides are not distorted by the continental shores, large tides raise the water level 1.30 m, and small tides lift it only 0.65 m. The corresponding calculations make it possible to evaluate the ratio of solar and lunar masses using the parameters of the tides.

Thus, the tides helped Newton calculate the mass of the Moon. His calculations were not particularly precise. The problem is complicated by the friction of huge water masses with the oceanic bed, as well as by other processes that cannot easily be analyzed. The precise value of the lunar mass was found only with the help of artificial lunar satellites. The trajectories and periods of revolution of the lunar orbiters yielded data to determine a mass of the Moon that was only roughly estimated by Sir Isaac. Moreover, the lunar orbiters showed that the Moon's mass is distributed unevenly throughout its volume, so the lunar satellites do not strictly follow Kepler's laws.

Before Newton's time astronomers thought comets paid only a single visit to Earth. However, Newton showed that comets could move along closed, elliptical orbits. The specific feature of these orbits is a pronounced elongation. This is why the comets fly away to great distances from the Sun. Accordingly, they have a long period of revolution. Edmond Halley calculated the moment of return of a famous comet, whose appearance could be traced in the ancient chronicles. The prediction was a striking success: the comet returned periodically at the calculated times. It can be seen every 76 years. Only one astronomer, Johann Gottfried Galle (1812–1910), who lived almost 100 years, saw this comet two times. The recurrence of comets is a strong argument in favor of the law of universal gravitation.

Quantum on Kepler's laws, gravitation, and space flight:

A. Eisenkraft and L. D. Kirkpatrick, "When Days Are Months," May 1990, p. 34.

S. Filonovich, "The Modest Experimentalist, Henry Cavendish," January/February 1991, pp. 41–44.

B. E. Belonuchkin, "The Fruits of Kepler's Struggle," January/February 1992, pp. 19–22.

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

Y. Osipov, "Catch as Catch Can," January/February 1992, pp. 38–43.

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

Y. Danilov, "The Act of Divine Providence," May/June 1993, pp. 41–45.

Y. Smorodinsky, "Late Light from Mercury," November/December 1993, pp. 41–43.

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

A. Byalko, "A Flight to the Sun," November/December 1996, pp. 16–20.

A. Eisenkraft and L. D. Kirkpatrick, "Mars or Bust," March/April 1997, pp. 34–36.

V. Surdin, "Swinging from Star to Star," March/April 1997, pp. 4–8.

V. Mozhaev, "In the Planetary Net," January/February 1998, pp. 4–8.

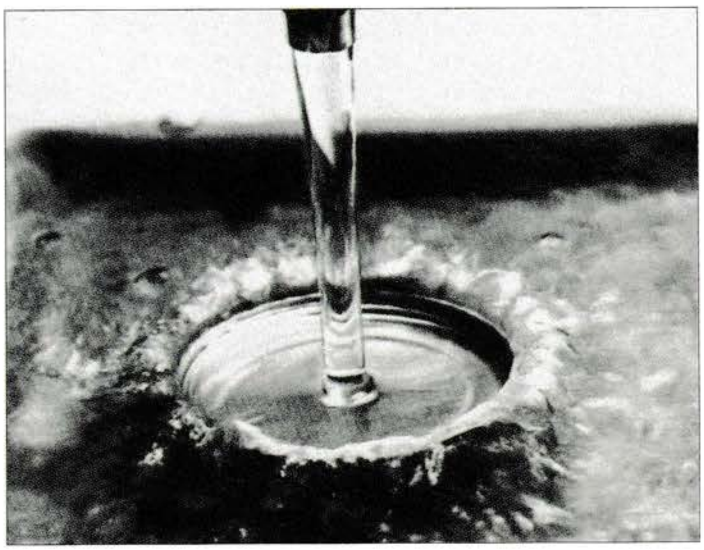
V. E. Belonuchkin, "Turning the Tides," May/June 1998, pp. 10–14.

An unsinkable disk

by A. Luzin

FOR THE EXPERIMENTS WE NEED A THIN disk made of a material denser than water. We can take a sheet of soft steel with a thickness of 0.2–0.3 mm, an aluminum sheet of 0.2–1.9 mm, or a sheet of Plexiglas of 0.3–6.0 mm. The most suitable disk will have a diameter of 90–150 mm. In the center of the disk a depression should be made with a diameter of 12–15 mm and a depth of 1.5–3.0 mm.

If we place such a disk on the surface of water, it floats because of the surface tension of the liquid. Splash some water on it, and the disk will surely sink, because it is more dense than water. Now hold the disk in your hand and place it under a jet of water. Note that water pushes down on the disk rather strongly (this is the hydrodynamic pressure of the water jet). There is nothing supernatural in the results of our experiments—physics can explain them.



Place the disk on the water's surface and guide a strong vertical jet into the depression at its center. You will see a paradoxical phenomenon—the jet pushes down on the disk but cannot sink it! A spectacular experiment of this kind can be made with a colored Plexiglas disk, through which the small objects on the bottom of a vessel can be seen clearly. This experiment can be easily demonstrated provided the water jet is sufficiently even, smooth, and free of visible vortices.

Why doesn't the disk sink? Watch it for a few minutes and note that there is a circular hump of water on the disk's surface that is pushed by the diverging thin layer of running water far from the center. As a result, a region of decreased pressure appears between the disk's center and the hump, and the difference of the forces applied to the upper and lower sides of the disk can be quite enough to keep it afloat, even with a rather strong water jet. This is an interesting manifestation of the relationships given by Bernoulli's law, which says that a thin, fast jet of liquid flowing radially on a disk pushes away a high water hump in which the velocity of the flow is small.

In hydraulics, the observed water "hump" has a special name, the *hydraulic jump*. It is a sharp, steplike rise of the water level in an open waterway where the character of motion changes from rapid and turbulent into tranquil, laminar, and steady flow. The hydraulic jump can be clearly seen on the downstream wall of a dam, but in that case its form is not circular but linear. A circular hydraulic jump can be seen even without a disk by letting the vertical water jet fall onto the bottom of a sink, even if the sink bottom is not precisely flat.

The mathematical description of a hydraulic jump is based on Bernoulli's law, which in the case of a fluid flowing in a horizontal channel with vertical walls and an open water surface can be written as

$$h_1 + \frac{v_1^2}{2g} = \frac{v_2^2}{2g} + h_2,$$

or

$$E_1 = E_2,$$

where h is the depth of the flow, v is its velocity at some cross-section of the channel, and $E = h + v^2/2g$ is the value known in hydraulics as the *specific cross-sec-*

tional energy. If the liquid is non-ideal but has a rather small internal friction, then the above equation is replaced by the inequality $E_1 > E_2$. In the simplest case of a rectilinear hydraulic jump, when the channel width b is constant, the inequality can be rewritten as $E(h_1) > E(h_2)$, where

$$E(h) = h + \frac{Q^2}{2b^2gh^2},$$

and $Q = vbh$ is the flow rate, which in the case of steady flow is the same at any cross-section of the channel.

The specific cross-sectional energy E can be considered to depend only on h, and this dependence is nonmonotonic. At some critical depth of flow h_c, the specific cross-sectional energy is minimal, while as $h \to \infty$ it grows due to the first term, and as $h \to 0$ it also grows due to the second term. The critical flow depth can be obtained from the equation $E(h_c) = 0$, which says that the first derivative of E with respect to h must be zero:

$$h_c = \sqrt{\frac{Q^2}{b^2g}}.$$

We can see that two types of steady flow are possible, corresponding to the cases when (a) the velocity of the flow is rather high, but its depth is relatively small ($h < h_c$), the latter growing slowly along the channel, while the specific cross-sectional energy slowly decreases, and (b) the velocity of the flow is rather small, but the depth is sufficiently large ($h > h_c$) and slowly decreases along the channel together with the specific cross-sectional energy. Experience shows that a "composite" steady flow is possible that consists of the flows of first and second types and a rather narrow region of hydraulic jump between them.

It is not a simple problem of how such a complex flow is formed and how the hydraulic jump in it is generated. In this article we use only approximate and qualitative conceptions. Clearly a hydraulic jump forms in a channel with a flat horizontal bottom, provided the kinetic energy of the fluid at some cross-section is sufficiently high ($h < h_c$). The energy gradually decreases along the flow at larger distances from this place due to viscosity, and the flow depth gradually increases and reaches the critical value $h = h_c$. The flow depth cannot grow any more, because this would mean an increase in E, which contradicts the inequality. Fluid is decelerated in the place where $h \cong h_c$, and it accumulates there until the flow becomes stable. This process leads to a drastic increase in the flow depth.

A more detailed theoretical consideration of the described experiments can be made using the so-called "wave theory" of the hydraulic jump, which calculates the height of the step, the abrupt change of pressure, and the generated "lifting" force that supports the disk on the surface of water. You can read about it in problem 4.58 in the wonderful book by Jearl Walker called *The Flying Circus of Physics* (New York: John Wiley & Sons, 1977).

Quantum on fluid mechanics:

L. Guryashkin and A. Stasenko, "The history of a fall," March/April 1995, pp. 10–15.

S. Kuzmin, "Spinning in a jet stream," September/October 1994, pp. 49–52.

L. Leonovich, "Fluids and gases on the move," January/February 1996, pp. 28–29.

A. Mitrofanov, "Against the current," May/June 1996, pp. 22–29.

J. Raskin, "Foiled by the Coanda effect," January/February 1994, pp. 5–11.

H. Schreiber, "A viscous river runs through it," November/December 1995, pp. 43–46.

A. Stasenko, "Whirlwinds over the runway," July/August 1997, pp. 36–39.

I. Vorobyov, "Canopies and bottom-flowing streams," July/August 1995, pp. 45–47.

V. E. Belonuchkin, "Turning the tides," May/June 1998, pp. 10–14.

Physical optics and two camels

by A. Stasenko

WHAT IS THE FARTHEST distance that you could be from a camel and be able to distinguish with unaided eyes whether it has one or two humps? What could be difficult in answering a question of such vital importance to used camel salespersons? Well, it turns out that the laws of physical optics impose a principal limitation on the very possibility of accomplishing this task.

Let's start from the very beginning. A broad parallel beam of light with intensity I_0—the energy flowing through unit area per unit time, $J/(s \cdot m^2)$—shines from the left on a nontransparent screen that has an infinitely long slit of width d (fig. 1).

Geometrical optics says that the light behind the screen is also a parallel beam of width d. This width doesn't vary with distance, so if we place a white screen perpendicular to the beam at any distance from the slit, we will see a light band of the same width d and illumination I_0 (right side of fig. 1).

However, light has a spatial character, parameterized by its wavelength λ. At this point an experienced *Quantum* reader may conclude that in the case under consideration, the dimensionless ratio λ/d must play an important role. And our wise reader would be absolutely correct.

Let's divide the slit into two luminous bands of width $d/2$. Then we concentrate the energy of these bands into two infinite luminous threads separated by a distance $d/2$ (figure 2, left). Now the problem is reduced to *two-beam interference*, and it becomes similar to the famous two-slit experiment of Thomas Young. Now let's determine what pattern will appear on a white screen placed at a distance x to the right of the slit.

First, we calculate the path difference $\Delta = r_2 - r_1$ from threads A and B to the point with coordinate y on the screen. Since the slit is narrow ($AB << x$), the two triangles in figure 2 yield

$$r_1^2 = r^2 + \left(\frac{d}{4}\right)^2 - 2r\frac{d}{4}\sin\theta,$$

$$r_2^2 = r^2 + \left(\frac{d}{4}\right)^2 + 2r\frac{d}{4}\sin\theta.$$

By subtracting the first equation from the second, we get

$$r_2^2 - r_1^2 = rd\sin\theta.$$

The left side can be rewritten as $(r_2 - r_1)(r_2 + r_1)$, where the first factor is the path difference and the second factor $(r_2 + r_1) \cong 2r$. Now we have

$$\Delta = \frac{d}{2}\sin\theta.$$

The same result can be obtained immediately by replacing the "curved" triangle ABC by the rectangular triangle ABC'.

Recall the basic concept of interference: if the path difference from two sources to the same point is equal to an integer number of wavelengths λ, these waves augment each other (constructive interference), but if the path difference is an odd number of half-wavelengths, the waves cancel each other (destructive interference). Thus, instead of a steplike luminosity function shown on the screen in figure 1 by a lover of geometrical optics, a far more complicated interference pattern ap-

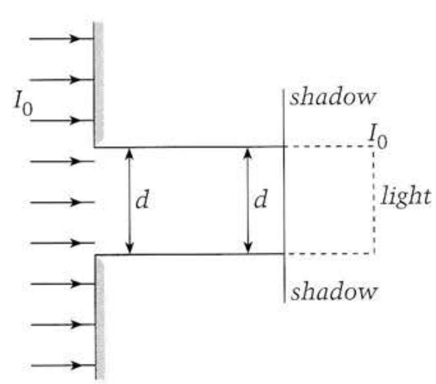


Figure 1

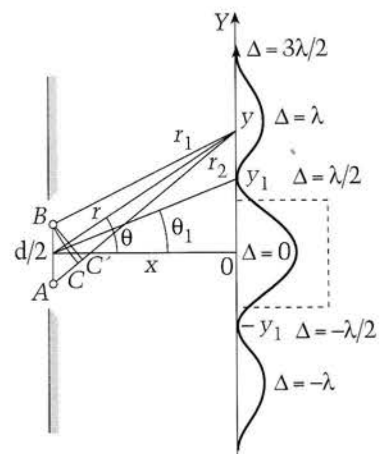
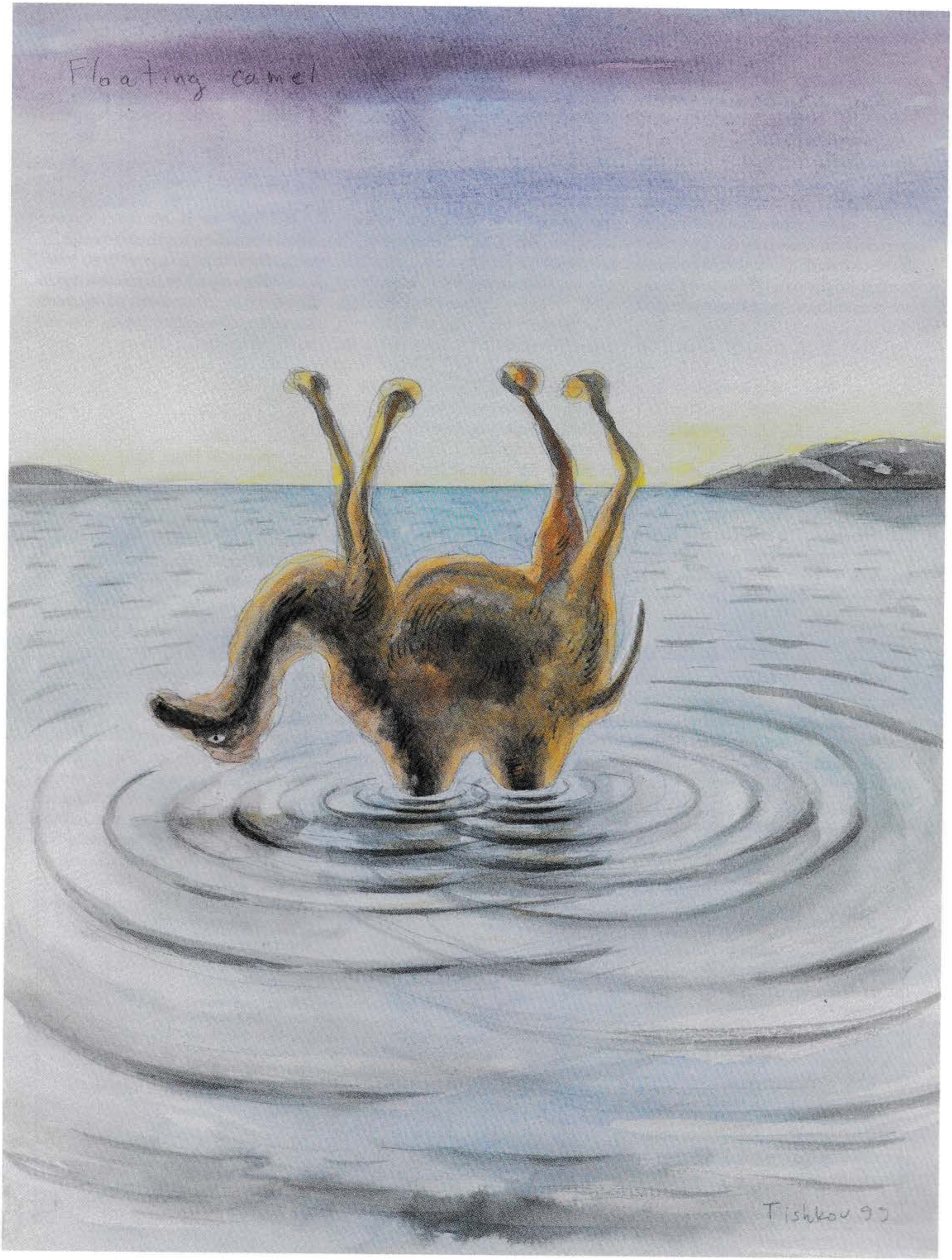


Figure 2

Art by Leonid Tishkov


Floating camel
Tishkov 93

pears that consists of alternating bright and dark bands (figure 2). The brightest illumination will be in the middle of the screen on the line located just opposite the slit ($y = 0$). The brightness of other bands will decrease with distance from the middle line, because they are situated farther from the luminous slit.

The most interesting thing for our problem concerns the locations of the two dark bands that border the central bright band. It follows from our reasoning that at these places

$$\frac{d}{2}\sin\theta_1 = \pm\frac{\lambda}{2},$$

or

$$\sin\theta_1 = \frac{\pm y_1}{\sqrt{y_1^2 + x^2}} = \pm\frac{\lambda}{d}.$$

This formula is a triumph for our experienced reader: indeed, the interference pattern generated by a slit depends on that very important parameter, the dimensionless ratio of the light's wavelength to the width of the slit that the light passes through!

Now let's return to our camels. In this case the pupil of an eye plays the role of the "slit," although in reality it is a round orifice, not an infinite slit. The retina of the eye plays the role of the screen on which the interference pattern was observed (figure 2). It turns out that a similar interference pattern appears on the retina! Of course, in this case it is not composed of parallel bands, but of concentric bright and dark rings surrounding the central bright spot. The radius of the first dark ring corresponds to an angle (a little bit larger than in the previous case) given by

$$\sin\theta_1 = 1.22\frac{\lambda}{d},$$

where d is the diameter of the pupil.

Each "point" of the remote object (a camel), which sends a nearly parallel beam of reflected solar light to the eye, is projected onto the retina as a bright spot surrounded by a set

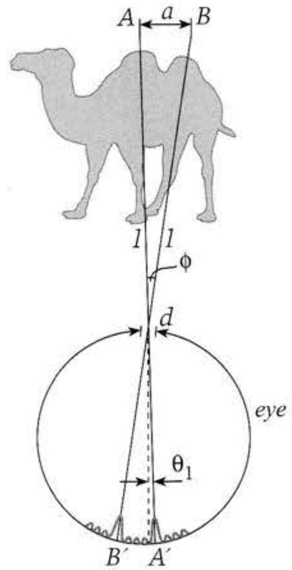


Figure 3

of rings. (Doesn't it look like the wave pattern on a pond's surface produced by a pebble?) We must distinguish (physicists say "resolve") two points of the object.

The interference pattern produced by these points is qualitatively shown in figure 3. This figure explains the condition of resolution: if the maximum illumination from the second point (B) coincides with the minimum illumination produced by the first point (A) or is farther from it, these points may be considered as separate (that is, they are resolved). Otherwise, these points blur into a single spot.

Note that the clearest image is produced on the retina when the pupil's diameter d = 3 mm. At this diameter the angular resolving power of the eye as determined by the laws of physical optics is on the order of

$$\phi = \frac{1.22\cdot 5\cdot 10^{-7}\text{ m}}{3\cdot 10^{-3}\text{ m}} \cong 50''.$$

Here we have used the characteristic wavelength $\lambda = 5\cdot 10^{-7}$ m = 500 nm. When people want to clearly scrutinize an object, they turn their eyes in such a way that the image forms on the so-called *macula lutea* (yellow spot) of the retina, which contains about 15,000 cones (sensitive cells) that occupy an area with an angular size of about 1.5°. This is the area of greatest concentration of cones in the eye. In the *macula lutea* every cone subtends an angle of about

$$\phi_1 = \frac{1.5°\cdot 3600''/°}{\sqrt{15000}} \cong 50''.$$

It looks as if Mother Nature knows interference theory perfectly well, doesn't it?

In this respect it is interesting to read from Jonathan Swift's *Gulliver's Travels*:

> *Nature hath adapted the eyes of the Lilliputians to all objects proper for their view: they see with great exactness, but at no great distance. And to show the sharpness of their sight towards objects that are near, I have been much pleased with observing a cook pulling a lark, which was not so large as a common fly; and a young girl threading an invisible needle with invisible silk.*

At the same time, Gulliver ascertained that he was a dozen times larger than a Lilliputian, and that all objects in Lilliput were smaller than ours by the same factor. Thus, a Lilliputian had eyes with pupils 1/12 the diameter of ours. Therefore, the interference angle for Lilliputians is 12 times that of ours, which means that the image of a point light source on his retina is also bigger than ours by the same factor. Why should Nature produce retina cells smaller than that in Gulliver? It would be a waste of cells! Will the Lilliputians' vision really be better than ours even at short distances?

So, to resolve two points, the (approximate) requirement

$$\phi \ge \theta_1 \text{ or } \frac{a}{l} \ge 1.22\frac{\lambda}{d}$$

must be met, from which we get

$$l \le \frac{ad}{1.22\lambda}. \qquad (*)$$

Now we are ready to make a numerical estimation. Let the distance between the camel's humps be $a \cong 0.5$ m, the pupil's diameter be $d \cong 1$ mm (remember that there is bright sunlight in the desert!), and the mean wavelength of the sunlight be $\lambda \cong 0.5$ µm. This yields

$$l \leq \frac{0.5 \cdot 10^{-3}}{1.22 \cdot 0.5 \cdot 10^{-6}} \text{ m} \cong 800 \text{ m}.$$

Are the humps of a camel the most interesting objects of observation? Perhaps not. It is no coincidence that telescope makers try to make a larger pupil (objective): indeed, the smallest angular distance (ϕ) between two stars must be no less than θ_1. It is also clear why an eagle must have a large pupil to discern a mouse on the ground from high in the sky.

What would happen if we put together a telescope and a microscope? At first glance, each component of this system magnifies the image by a thousand times—so the entire "supertelescope" will magnify it by a million times! Would we see the pebbles on the Martian soil?

Alas, the image will be helplessly spoiled by interference even at the objective of the telescope, so the details of objects smaller than that specified by *Rayleigh's condition* (*) will be lost forever. In addition, there is interference at the objective of a microscope, which prevents the resolution of two points spaced at a distance less than the wavelength of the illuminating light (this is why microbes cannot be observed in optical microscopes).

So what have we learned? Were optical devices invented simply to observe interference patterns on their "pupils"? Certainly not. Those who invented the first microscopes and telescopes knew nothing of light interference, because the advent of the wave theory of light was far over the horizon. It seemed that the plots of the light rays drawn according to the laws of geometrical optics opened an unlimited vista to increased magnification by proper choices of objectives and eyepieces, as well as by their separations. However, as usually occurs in physics, the new theory revealed the limits of an older, simplified theory. The camels have served to remind us of this old truth.

Quantum on light interference:

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

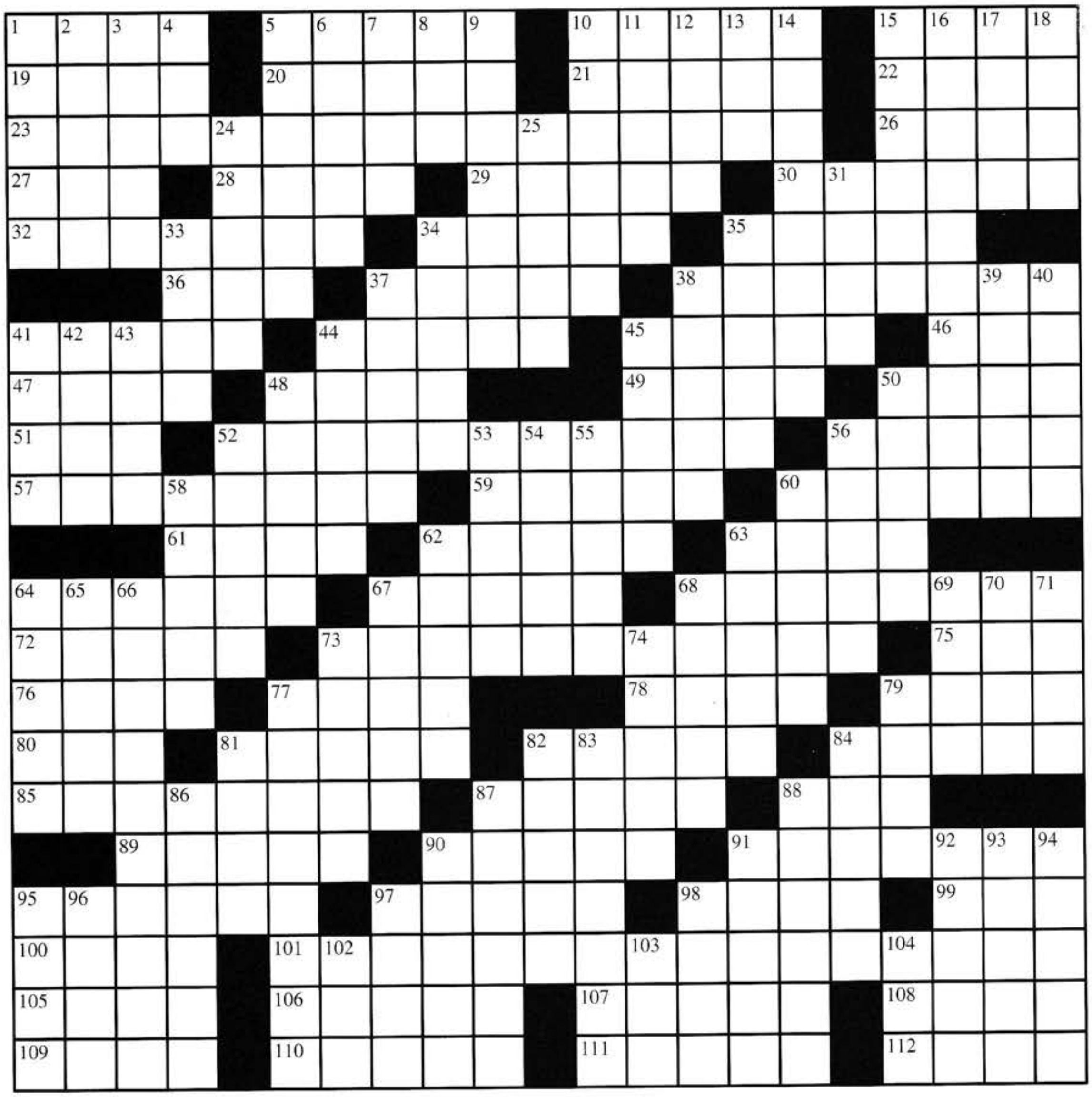
A. Eisenkraft and L. D. Kirkpatrick, "Rising Star," March/April 1995, pp. 37–38.

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.

Criss × cross science

by David R. Martin



Across

1 Steals
5 Muscle pain
10 Head skin
15 Mincingly cute
19 Oriental nurse
20 Perch
21 Prank
22 Jewish philosopher ___ Haam
23 Space-time mathematician
26 Ascent
27 Geologic time period
28 Superman's girl
29 Scottish island
30 Expert
32 Superconductivity theorist
34 Filleted
35 Sandstone
36 Physics assoc.
37 A race of gods (Scand. myth.)
38 Compounds with element 7
41 ___ metal (used in flints)
44 Lane-___ equation (for stellar structure)
45 Indian's home
46 Moist
47 Width times length
48 "To ___." (perfect)
49 Imitates
50 Protected
51 Type of gate: abbr.
52 Black hole theorist
56 Cardinal number
57 Long wavelength radiation
59 Fossil resin
60 Stun with noise
61 Cavity: comb. form
62 French city
63 Gloom
64 National park in Maine

67 Sick person
68 One dyne/cm^2
72 Linearly independent vector set
73 Heisenberg's ___ principle
75 Silver iodide: abbr.
76 ___ Struve (1897–1963)
77 Cleveland's lake
78 Simple math operator
79 64,459 (in base 16)
80 Mal de ___ (seasickness)
81 Former
82 Glances at
84 Diving birds
85 ___ motion (of tiny particles)
87 1972 Chem. Nobelist Stanford ___
88 Element 5: comb. form
89 Horse sound
90 Puerto ___ (Chilean city)
91 Universe origin
95 "Athlete of the Half-Century"
97 ___ Lama
98 City near Madras
99 Twelve grams of carbon-12
100 Church song
101 X-ray star discoverer
105 To shelter
106 Happen
107 Weapon supplier
108 44,525 (in base 16)
109 1/760 atmosphere
110 Hammer's partners
111 Like a shade tree
112 Apportion

Down

1 Harlot of Jericho
2 Psi follower
3 More naked
4 Sinusoidal motion: abbr.
5 Slender, open boats
6 Kidney secretion
7 Appendages
8 1002
9 Carbohydrate with 5 carbon atoms
10 Point maker
11 Cried like a crow
12 Orbital point
13 Albanian money
14 Humans and monkeys
15 Mathematician Alfred ___
16 Type of star
17 No difficulty
18 German river
24 ___ null
25 Blood peptide
31 Land measurement
33 56,010 (in base 16)
34 Make wet
35 Window cleaners need
37 Fix
38 8.686 decibels
39 978,926 (in base 16)
40 Dutch painter (1626–1679)
41 Long skirt
42 Element in steel
43 Vassal
44 Former hypothetical medium of space
45 Stories
48 Major blood vessel
50 Military cap
52 Singer ___ Joplin
53 Ethiopian emperor ___ Selassie (1892–1975)
54 Type of wheat
55 The late Siskel's partner
56 Football's Bradshaw
58 Type of galaxy
60 Conduits
62 Kinder
63 78A's partner
64 Nuclear weapon
65 Provide food
66 Star gazer
67 Negative particle
68 Kinematic relativity's Edward ___ (1896–1950)
69 German chemist Lambert ___ (1818–1899)
70 Silver cyanide
71 BBQ favorite
73 Bathsheba's husband
74 Separated
77 Fleabane and horseweed oil
79 Nongrass herb
81 Type of transistor
82 Submarine finder
83 Having high water coincidence
84 15D's forte
86 Black hole theorist ___ Israel
87 Teeth
88 Like some stars
90 Spot: comb. form
91 Momentary
92 Form of ultrasonic tomography
93 Musical group of nine
94 Soar
95 "All's well ___ ends well."
96 Wood: comb. form
97 701
98 Latin consonant sound
102 Resort in SW Peru
103 Mine output
104 Engine shaft

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE JULY/AUGUST PUZZLE

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

Bulletin Board

GIFTs for the taking

Applications for the 2000 Growth Initiatives For Teachers (GIFT) Program, a $12,000 grant for secondary science and math teachers, are now available. Through the GIFT Grant Program, the GTE Foundation promotes excellence in math and science education. The GIFT program is open to public and private school math and science teachers, grades 7 to 12, in 35 eligible states and the District of Columbia.

Each winning team shares a $12,000 GIFT grant—$7,000 to implement the school enrichment project and $2,500 to each of the participating teachers to help them pursue professional development activities.

To apply for a GIFT grant, team members jointly submit a proposal and budget for a school enrichment project that must be based in the team's classrooms or school and directly involve both math and science students for the 2000–2001 school year. Also, as part of the application, team members submit proposals and budgets for individual professional development activities that directly support implementation of the school enrichment project and address their own needs for professional growth, especially regarding technology in education.

Gift Fellows participate in a yearlong program designed to inspire their teaching efforts through opportunities for professional growth. The year begins with the GIFT seminar, held in June, for which the GIFT Fellows travel to Boston and Washington, D.C. During the week-long seminar, teachers attend presentations by noted scientists and educators, tour GTE labs, and interact with GTE employees and government officials. All expenses are underwritten by GTE.

To be eligible for the GIFT grant, applicants must:

- hold a bachelor's degree and state certification (where applicable),
- have completed at least one year of full-time paid teaching in the same school district by July 1, 1999,
- teach grades 7 to 12 (grade 6 only if in a middle or junior high school) at a regionally or state-accredited non-profit school (public or private),
- carry a full teaching load, with more than half of the teaching schedule in math and/or science, both in the current year and the year in which the grant will be applied,
- teach in Alabama, Arizona, Arkansas, California, Colorado, Connecticut, District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kentucky, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Nebraska, New Hampshire, New Mexico, North Carolina, Ohio, Oklahoma, Pennsylvania, South Carolina, Tennessee, Texas, Virginia, Washington, West Virginia, or Wisconsin.

Teachers who would like to receive an application by mail are encouraged to call (800) 315-5010, or send email to gift@gte.com. Applications are also available online at www.gte.com/aboutegte/community/gte_foundation/opportunities/gift.html. Deadline for submissions is January 4, 2000.

What a difference a word makes

Due to a clerical error on our part, a crucial word was left out of the initial posting of this month's CyberTeaser. The problem is correctly worded in this issue (Brainteaser B271); on the web, the word "different" made a belated appearance. Luckily, some of our ambitious readers assumed the more difficult wording and we received answers that satisfied both conditions (denoted with an asterisk below). This month's winners are

Bruno Konder (Rio de Janeiro, Brazil)*
Jorge G. Moya (Culiacan, Mexico)*
Jerold Lewandowski (Troy, New York)*
Adam Cabrera (Billings, Montana)*
Theo Koupelis (Wausau, Wisconsin)*
May Lim (Quezon City, Philippines)
Patrick Maxfield (Folsom, California)*
Vladimir Novakovski (Springfield, Virginia)
Clarissa Lee (Petaling Jaya, Selangor, Malaysia)
Manny Dekermenjiian (Sunnyvale, California)

Congratulations to our winners, who will receive a *Quantum* button and a copy of this issue.

Everyone who submitted a correct answer before it was posted at our Web site was eligible to win a copy of our brainteaser collection *Quantum Quandaries*. Visit http://www.nsta.org/quantum to find out who won the book, and while you're there, try your hand at the new CyberTeaser!

ANSWERS, HINTS & SOLUTIONS

Math

M271

Add all three equations. Since $[x] + \{x\} = x$, we obtain

$$2(x + y + z) = 9.4,$$

from which we get $x + y + z = 4.7$. Now we add the first two equations to obtain $x + y + z + [y] + \{x\} = 7.4$. Therefore, $[y] + \{x\} = 7.4 - 4.7 = 2.7$. Thus, $[y] = 2$ and $\{x\} = 0.7$. Now, adding the first equation to the third one and the second equation to the third one, we find that $[x] = 1$, $\{z\} = 0.2$, $[z] = 0$, and $\{y\} = 0.8$. Answer: $x = 1.7$, $y = 2.8$, and $z = 0.2$.

M272

Denote the left-hand side of the given equation by y. Then, $y - x = \sqrt{x^2 - 9}$. Therefore, $y - x \geq 0$. Square the last equation and represent x in terms of y:

$$x = \frac{y^2 + 9}{2y}$$

(we can easily check that $y \neq 0$). The inequality $y - x \geq 0$ implies

$$y - \frac{y^2 + 9}{2y} = \frac{y^2 - 9}{2y} \geq 0 \qquad (1)$$

(a result that will be useful later). We find that

$$x + 3 = \frac{(y + 3)^2}{2y}$$

and

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

We then substitute these expressions in the original equation to obtain

$$y = \frac{4y(y + 3)^2}{(y - 3)^4}.$$

Now, our equation breaks into two equations: $(y - 3)^2 = 2(y + 3)$ and $(y - 3)^2 = -2(y + 3)$ or $y^2 - 8y + 3 = 0$ and $y^2 - 4y + 15 = 0$. The second equation has no real roots, and the first one has two roots: $4 \pm \sqrt{13}$. Only the greater of these roots satisfies condition (1). Now we can return to the original unknown and find that $x = 8 - \sqrt{13}$.

M273

Let O be the center of the circumcircle of triangle AMC (fig. 1). The conditions of the problem imply that $\angle MAC$ is acute. The choice of point O implies that $\angle MOC = 2\angle MAC$, and point O lies on the same side of AM as points B and C. It is clear from the equality $MO = CO$ that

$$\begin{aligned}\angle OMC &= 90^\circ - (1/2)\angle MOC \\ &= 90^\circ - \angle MAC = \angle BMC.\end{aligned}$$

Therefore, the line MB contains point O. If O does not coincide with B, then BM is the perpendicular bisector of AC (and segment BM must intersect segment AC). If O coincides with B, then point M lies on the corresponding arc of the circle. The desired locus is shown in figure 2.

M274

Consider three diagonals of the cube's faces incident to the same vertex. At least one of them is not parallel to the given plane: let it be diagonal AC, and let O be its center (fig. 3). Then, the given plane must intersect AC at the center of one of the segments AO or OC. Otherwise, the distances from A, O, and C to this plane would differ from each other (since the minimum of the given distances is 1, the plane cannot contain any of these points). In general, any diagonal of any other face of the cube either intersects the given plane in the manner described or is parallel to this plane.

Now, it is not difficult to conclude that there are only two possibilities: (1) the given plane is parallel to two faces of the cube and divides the perpendicular edges in the proportion 1:3, or (2) the intersection of the given plane with the cube is the regular hexagon whose vertices are the midpoints of certain of its edges. In the first case, the cube's edge is 4; in the second case, it is $2\sqrt{3}$.

M275

Notice that

$$\angle CBK = \alpha - (2\alpha - 180^\circ) = 180^\circ - \alpha,$$

which is to say that it is equal to the exterior angle of triangle ABC at vertex B. Thus, BC is the bisector of the exterior angle of triangle ABK at vertex B. By assumption, AP is the bisector of angle BAK. Therefore,

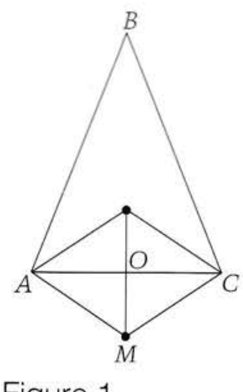


Figure 1

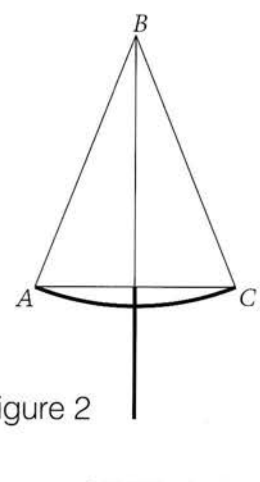


Figure 2

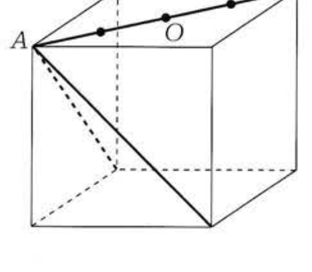


Figure 3

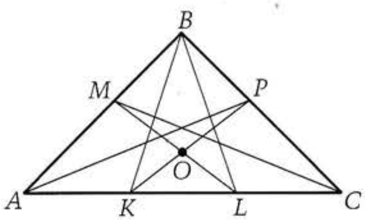


Figure 4

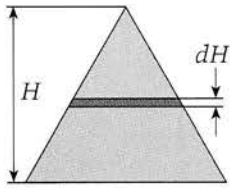


Figure 5

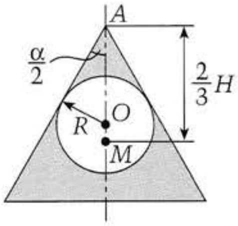


Figure 6

point P is equidistant from lines AB, BK, and AC. Consequently, KP is the bisector of angle BKC. Similarly, LM is the bisector of angle BLA.

Consider the situation illustrated in figure 4. In this case, $\alpha \geq 2(2\alpha - 180°)$, which means that $\alpha \leq 120°$. Therefore, $\angle KBL = 360° - 3\alpha$. Suppose, for some numbers β and γ, that $\angle BKL = 2\gamma$, $\angle BLK = 2\beta$. Then, from triangle BKL, $2\beta + 2\gamma + (360° - 3\alpha) = 180°$. Therefore, $\beta + \gamma = 3\alpha/2 - 90°$. From triangle KOL we find that $\angle KOL = 180° - (\beta + \gamma) = 270° - 3\alpha/2$. Now, the angle between two lines is defined as the smallest of the angles formed at their intersection. Thus it cannot be obtuse. In our case, $\angle KOL$ is not acute. Therefore, if $\alpha \leq 120°$, the angle between lines KP and ML is $3\alpha/2 - 90°$, and if $\alpha > 120°$, this angle is $270° - 3\alpha/2$.

Physics

P271

As a first step, we must determine the position of the cap's center of gravity CG. To this end we "cut" the cone into a set of thin rings of the same width dH (fig. 5). The mass of a ring grows linearly from the vertex to the base of the cap. The center of gravity of any ring lies on the cone's axis.

Now we "flatten" the cap in such a way as to transform every ring into an equilateral trapezoid and the entire cone into an isosceles triangle. The center of gravity of any composite part of the cap will remain at the same place (on the axis), so the CG of the system will not change. It is known that the CG of a triangular plate is located at the point of intersection of its medians. Therefore, the CG of the cap is located on its axis at the distance $(2/3)H$ from the vertex.

The state of equilibrium will be stable if a small displacement of the cap from this position would raise the CG (and increase the potential energy of the system). In this case an unconstrained system will return to the initial equilibrium state.

In order to have stable equilibrium for Pinocchio's cap on his head, the cap's CG (point M in figure 6) must be lower than the CG of Pinocchio's head (point O). Hence, the inequality $AM > AO$ must be satisfied:

$$\frac{2}{3}H > \frac{R}{\sin\frac{\alpha}{2}},$$

or $H > 3R = 22.5$ cm. We know from the problem statement that $H = 20$ cm. Therefore, the cap will not be in stable equilibrium sitting on Pinocchio's head.

P272

Let's consider a particle located at a distance R from the spherical cloud's center. It is known that the resulting force affecting the particle can be determined by accounting only for that part of the spherical cloud located inside a sphere of radius R. (see A. Stasenko's "The New Earth" in the July/August 1999 *Quantum*).

From the problem statement, we know that the compressing particles do not pass each other, so the total mass that attracts the probe particle remains constant. We may propose that this mass is concentrated in the center of the cloud. Now the problem simplifies to calculating the time for a particle to fall to the central mass.

We will consider the trajectory of the particle as a part of a very elongated ellipse with semimajor axis $R/2$. We compare the motion along this trajectory with the revolution along the circular orbit of radius R. According to Kepler's third law,

$$\frac{T_c^2}{T_e^2} = \frac{R^3}{\left(\frac{R}{2}\right)^3},$$

where T_c and T_e are the periods of revolution of the circular and elliptic orbits, respectively. The period T_c can be easily found with the help of Newton's second law and the law of universal gravitation:

$$\frac{mv^2}{R} = F_g = G\frac{mM}{R^2} = G\frac{m\rho\frac{4}{3}\pi R^3}{R^2},$$

from which we obtain

$$T_c = \frac{2\pi R}{v} = \sqrt{\frac{3\pi}{G\rho}},$$

and

$$T_e = \frac{T_c}{2^{3/2}} = \sqrt{\frac{3\pi}{8G\rho}}.$$

We see that the period T_e doesn't depend on R. Therefore, the duration of the fall of a particle to the gravitating center (the time of formation), which is equal to half the period of revolution along the elliptic orbit, doesn't depend on the size of the cloud and is given by the formula

$$\tau = \frac{T_e}{2} = \sqrt{\frac{3\pi}{32G\rho}} \cong 1.5 \cdot 10^{13} \text{ s}$$
$$\cong 10^6 \text{ years}.$$

P273

The remaining film will minimize its surface after being pierced. Therefore, the hole will form the geometric figure of maximum area for the constant perimeter l. It is known from geometry that a circle fulfills these requirements (figure 7). In our case, the radius R is

$$R = \frac{l}{2\pi}.$$

To find the tension in the thread, let's consider a small element with length $\Delta l = R\Delta\phi$ (figure 8). It is affected by two tensions $\mathbf{T}_1$ and $\mathbf{T}_2$ ($T_1 = T_2 = T$) that act tangentially on the element from neighboring parts of the thread and the force $2\mathbf{F}_{st}$ due to the soap film. ($\mathbf{F}_{st}$ is the surface tension acting on the element due to one surface of the film.) From the equilibrium condition for the element along the radial direction, we get

$$2T\frac{\Delta\phi}{2} - 2F_{st} = 0.$$

Here we considered the small size of the element, and used the approximation $\sin\phi = \phi$. To find the force F_{st}, we allow the film to contract (virtually) even more, so the element Δl will be displaced by ΔR. The energy of the film will decrease by $\Delta E = 2\sigma\Delta l\Delta R$ due to the work performed by the surface tension of the film $\Delta W = 2F_{st}\Delta R$, from which we get

$$\Delta E = \Delta W,$$
$$2\sigma\Delta l\Delta R = 2F_{st}\Delta R,$$
$$F_{st} = \sigma\Delta l.$$

Thus, the coefficient of surface tension, which in this case characterizes one of the surfaces of the soap film, may be thought of as the force affecting a unit length of the boundary in the direction normal to the boundary:

$$\phi = \frac{F_{st}}{\Delta l}.$$

Since $T\Delta\sigma = 2F_{st}$, the tension of the thread will be

$$T = 2\frac{F_{st}}{\Delta\phi} = 2\sigma R = \sigma\frac{l}{\pi}.$$

We have considered this problem in detail because it demonstrates the origin of surface tension. It turns out that a useful notion for calculations is that the surface energy is equivalent to the stretching of the surface of a liquid. The stretching forces can be considered as being applied to a linear boundary of the film. If we subdivide (virtually) a film into two parts, they affect each other through forces normal to their common boundary. At first glance, this looks very similar to the stretching of a rubber film. However, there is a key distinction: when liquid film is stretched, the forces of surface tension remain constant, while the analogous forces in a rubber film increase. Thus, there is an analogy between liquid and rubber film, but it is not ideal, so we must use it carefully.

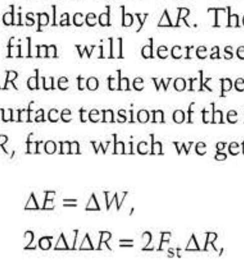


Figure 7

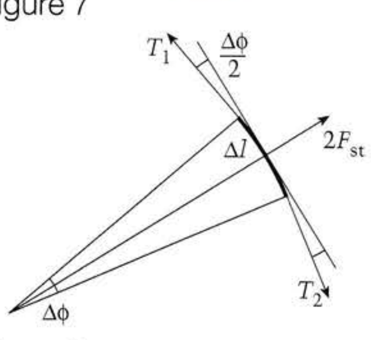


Figure 8

P274

An electron moving in a magnetic field is affected by the Lorentz force. Since this force is always perpendicular to the magnetic field, its projection onto the direction of the field is always zero. Therefore, the electron moves with constant velocity $v\cos\alpha$ in the direction of the magnetic field. The projection of its trajectory onto the plane perpendicular to the vector $\mathbf{B}$ yields a circular motion with speed $v\sin\alpha$, which is described by the equation

$$\frac{mv^2\sin^2\alpha}{R} = eBv\sin\alpha,$$

from which the radius R and period of revolution T can be found:

$$R = \frac{mv\sin\alpha}{eB},$$
$$T = \frac{2\pi R}{v\sin\alpha} = \frac{2\pi m}{eB}.$$

Thus, the electron moves along a helix: in the direction of the magnetic field, its motion is a uniform translation, while in the plane normal to the field, the electron's trajectory is a circle.

Assume that the electron arrives at point C having performed n complete revolutions. Denoting the corresponding magnetic field by B_n, we obtain the time for the electron to travel between points A and C:

$$\tau = \frac{L}{v\cos\alpha}.$$

On the other hand, the time necessary to perform n complete revolutions is

$$\tau = nT,$$

or

$$\frac{L}{v\cos\alpha} = n\frac{2\pi m}{eB_n},$$

where $n = 1, 2, 3, \ldots$ As a result, we have a set of quantized values for B_n, which ensure that the electron hits the target point C:

$$B_n = \frac{2\pi mv\cos\alpha}{eL}n.$$

P275

The dependence of the speed of light in the air on altitude results in the bending (refraction) of the light. This phenomenon underlies the observation of "lakes" in deserts.

Consider an observer of height h standing on the ground (figure 9). Let's trace the rays emitted from point A. The beam that forms an angle α_0 with the vertical curves in such a way that its trajectory contacts the ground tangentially and then heads to the sky. Rays emitted at the angles $\alpha > \alpha_0$ also go to the sky. By contrast, rays emitted at the angles $\alpha < \alpha_0$ hit the ground. Using the principle of reversibility of rays, we realize that the observer sees sand up to the angle α_0, but at larger distances (and larger angles) the observer sees blue sky. Paradoxically, the "sky rays" arrived at the observer from beneath, so it looks like a blue lake is spreading in the distance (subconsciously, a human being is aware that light always propagates along a straight line).

Now let's calculate the distance H. According to the generalized law of refraction for stratified media,

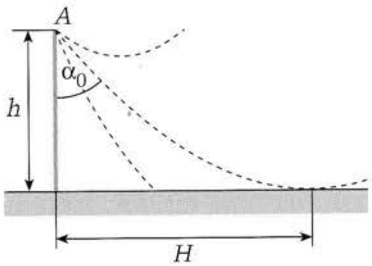


Figure 9

$$\frac{\sin\alpha_0}{\sin\beta_0}=\frac{c(h)}{c_0},$$

where $c(h)$ is the speed of light at altitude h, β_0 is the angle between the tangent to the beam and the vertical. Since $\beta_0 \cong 90°$ at the ground, we get

$$\sin\alpha_0=\frac{c(h)}{c_0}=1-ah.$$

We must also take into account that the speed of light changes very slowly with altitude, so the curvature of the rays is very small. Therefore, we can assume that the beam travels to point H along a straight line. Figure 9 shows that $H = h \tan \alpha_0$. Using the law of refraction, we get:

$$H=\frac{h(1-ah)}{\sqrt{ah(2-ah)}}.$$

Brainteasers

B271

Yes. For example, consider the numbers 1, 2, 3, 4, 5, and x. Then, x can be found from the equation

$$1+2+3+4+5+x=1\cdot 2\cdot 3\cdot 4\cdot 5\cdot x.$$

B272

When 625 is raised to any power, the result ends in 625. When 376 is raised to any power, the result ends in 376. To verify these facts, it is sufficient to square each of these numbers. Therefore, the last three digits of the sum are 001.

B273

It is easy to verify that $111111111^2 = 12345678987654321$.

B274

Each square of the graph paper inside the rectangle either belongs to a cutting line or not. We need the total length of all segments that do not belong to any cutting lines to be the maximum. Consider all possible figures composed of no more than 5 squares, and calculate the ratio of the number of segments inside a figure to the number of squares in it. The maximum value of this ratio is 1, and it is reached for two figures: a square of size 2×2 and the same square with an additional grid square (see figure 10). Therefore, if we cut the given rectangle into such figures, we obtain the desired result. This is easy to achieve by cutting the rectangle as shown in figure 11.

B275

The surface area of the hay, from which the water evaporates, is increased by stirring it up. That's why it dries more quickly and uniformly.

Kaleidoscope

1. The rod that casts the shadow must be directed toward the North Pole (in the Northern Hemisphere, of course).

2. No, because the time intervals between drops will increase due to the decreasing water level in the dropper.

3. The boats met the raft simultaneously.

4. The point of intersection indicates the time when the bodies have equal velocities. The figure cannot be used to determine when the bodies meet.

5. The paths traveled by the balls are identical. Since the mean velocity of the second ball is higher, it will arrive at point B first.

6. The motion of the train doesn't affect the vertical motion of the object. Therefore, in all three cases the time will be the same.

7. Due to air resistance, the vertical component of acceleration of the body at any altitude will be

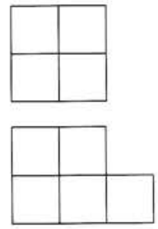


Figure 10

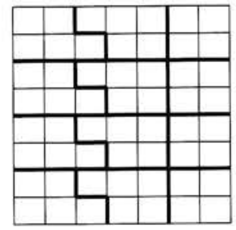


Figure 11

larger during the rise than during the fall. Thus, the time of rise will be smaller than the time of fall.

8. When the fly is moving upward, the bottom of the tube sinks relative to the center of mass of the system, which falls with the acceleration due to gravity. Thus, the bottom of the tube will hit the ground earlier in comparison with the case when the fly is motionless.

9. The force of gravity and the tension of the rope are the same for both gymnasts. Therefore, they will arrive to the pulley simultaneously.

10. No, it doesn't, because the decrease in the weight of the sand is counterbalanced by the force of the falling sand hitting the bottom.

11. A person at rest affects the ice for a longer time, thus producing a larger deformation.

12. The period of oscillation will be infinitely large. In other words, the oscillation stops in the state of weightlessness.

13. The period of oscillation will be half as large, because the spring constant of the cord will increase by a factor of 4.

14. To measure the period between the transmission and reception of the wave.

15. As the extent of a lightning is rather great, sound generated from the distant parts of it will arrive later, so the thunder is protracted in time. By contrast, the speed of light is so large that the corresponding delays are negligible, therefore we observe lightning as a single flash of light.

Microexperiment

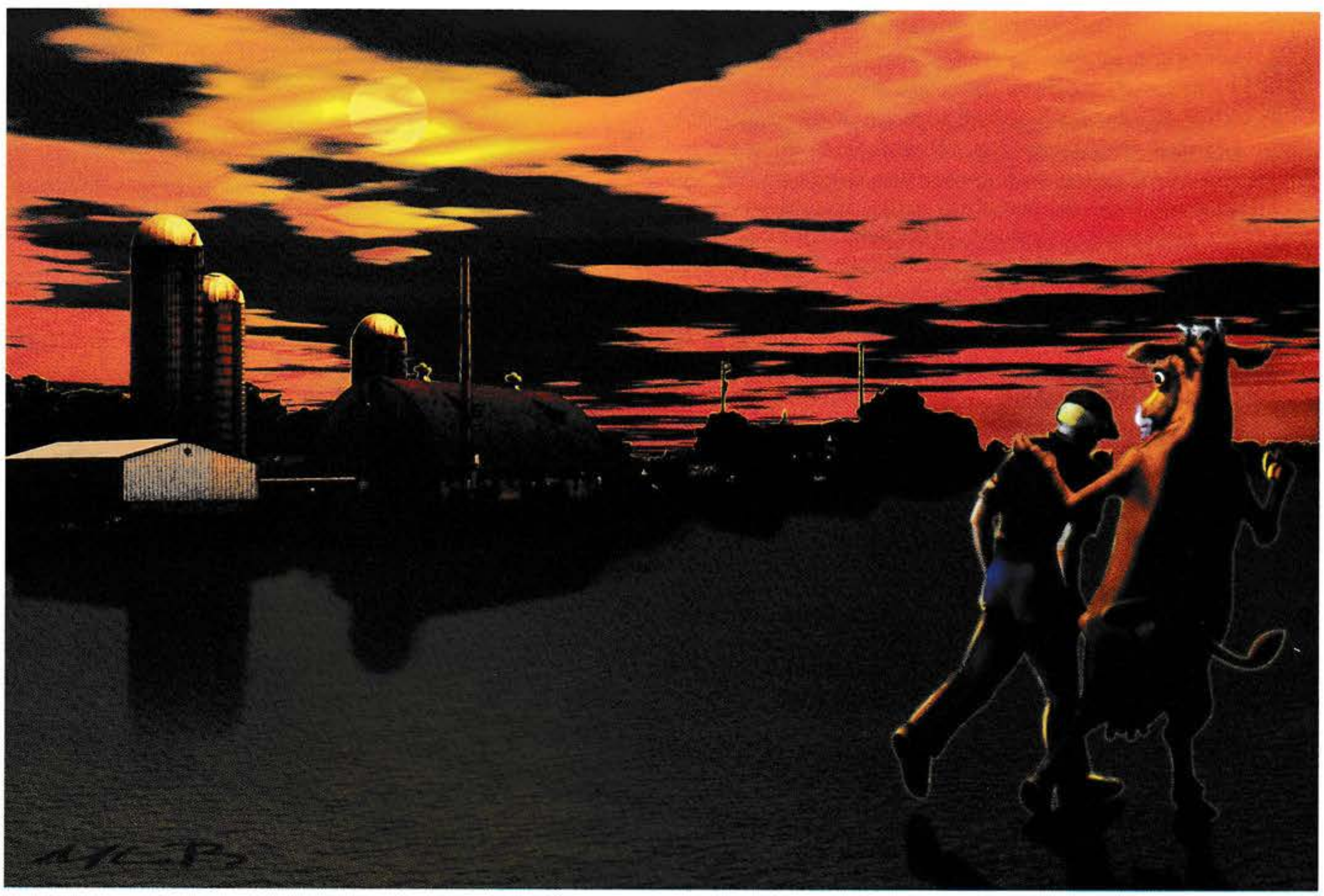
The period of oscillation decreases, because the length of the pendulum decreases.

Out to Pasture

by Dr. Mu

WELCOME BACK, FOR THE LAST TIME, TO Cowculations, the column devoted to problems best solved with a computer algorithm. Why the last time? Well, in October a windstorm blew through the farm and took out the sick calf barn and one silo. The estimated replacement costs came to over $30,000, which was enough to trigger a life altering decision to end our single family dairy farm in Wisconsin. As his father told farmer Paul when passing the farm on to him, the hardest decision he would have to make is when to quit. The Nielsen farm has been in operation for three generations, starting with Paul's grandfather who bought it in 1936 and began milking with 24 Guernseys. By the late 60's, Paul had transformed the herd into registered Holsteins that were sold all over the world. Today, the rest of the world can breed its own registered cows so that market is gone. And there isn't enough money in raising cows for milk alone, so Paul, with Mother Nature's help, made the tough decision.

On March 22, a truck pulled into the driveway to pick up the last of the herd. We were all being relocated to other farms throughout the state. I was leaving for the green hills of western Wisconsin along the Mississippi and would be put out to pasture. Farmer



Art by Mark Brenneman

Paul was taking a regular job at the Schmidt Farm Implement Company in town. We parted as old friends and promised to write often. I left a final tribute to Paul on the barn door.

Dear Paul,
Into this barn we came each day
To eat our ration of grain and hay.
We made a pact and kept our word:
In exchange for milk, you fed the herd.
We worked together, taking and giving,
And created a life as well as a living.
Your milking chores are over now.
Good Luck Paul, from your last COW.
—Dr. Mu

COW 18

Here is the last Challenge Outta Wisconsin. The digital product of a positive integer N is defined as the product of its nonzero digits. The digital product root of a positive integer N is obtained by repeatedly taking digital products of each digital product until a single digit is obtained. For example, the digital product root of 123456789 is 8 by the following series of products:

$x \to y$ means take the product of the nonzero digits in x and create y.

$123456789 \to 362880 \to 2304 \to 24 \to 8$.

Find the digital product root of 12345^{6789}. In Mathematica, a one-line program can be written that finds the answer in less than one second.

COW 17

In COW 17 you were asked to write a program that takes as input an integer n, $n \le 9$, and finds all ways to insert pluses or minuses between the digits 12345...n, so the resulting expression sums to zero. $1-2-34+5+6+7+8+9 = 0$ is one solution for $n = 9$. Find all of them.

Solution

The solution, submitted by Eric Rimbey, is a good example of how clean a solution can be if constructed functionally. Let's see how it works. First, notice that every possible expression to evaluate must have either a +, a −, or a blank between two consecutive digits. Thus the maximum number of expressions that will need to be examined when $n=9$ is 3^8 or 6561. So, a brute force approach will easily work. We need only to generate them all and check out their total.

Begin by defining the list chars

```
chars = {"+", "-", ""}
```

and a digit sequence from 1 to 9

```
digitsequence = ToString /@ Range[9]

{1, 2, 3, 4, 5, 6, 7, 8, 9}
```

Next, generate all the possible sequences of chars separated by commas and a comma at the end.

```
charsequences =
 Distribute[Join[Table[chars, {8}],
 {""}], List];
```

There are 6,561 such character sequences. Here are a few of them:

```
Short[charsequences, 6]

{{+, +, +, +, +, +, +, +, }, {+, +, +, +, +, +, +, -, },
 {+, +, +, +, +, +, +, }, {+, +, +, +, +, +, -, +, },
 {+, +, +, +, +, +, -, -, }, {+, +, +, +, +, +, -, , },
 <<6550>>, {, , , , , , -, -, }, {, , , , , , , -, , },
 {, , , , , , , +, }, {, , , , , , , -, }, {, , , , , , , , }}
```

Now place the characters between the digits and form the possible strings.

```
digitchar =
 Flatten[Transpose[{digitsequence,
 #} ]]& /@charsequences;
```

Here are a few of them:

```
Short[digitchar, 6]

{{1, +, 2, +, 3, +, 4, +, 5, +, 6, +, 7, +, 8, +, 9, },
 {1, +, 2, +, 3, +, 4, +, 5, +, 6, +, 7, +, 8, -, 9, },
 {1, +, 2, +, 3, +, 4, +, 5, +, 6, +, 7, +, 8, , 9, }, <<6556>>,
 {1, , 2, , 3, , 4, , 5, , 6, , 7, , 8, -, 9, },
 {1, , 2, , 3, , 4, , 5, , 6, , 7, , 8, -, 9, }}
```

Next, join the digits and characters together to get all possible expressions.

```
possible = StringJoin /@ digitchar;

Short[possible, 4]

{1+2+3+4+5+6+7+8+9, 1+2+3+4+5+6+7+8-9,
 1+2+3+4+5+6+7+89, 1+2+3+4+5+6+7-8+9,
 1+2+3+4+5+6+7-8-9, 1+2+3+4+5+6+7-89,
 1+2+3+4+5+6+78+9, 1+2+3+4+5+6+78-9,
 1+2+3+4+5+6+789, <<6544>>, 1234567+8-9,
 1234567+89, 1234567-8+9, 1234567-8-9,
 1234567-89, 12345678+9, 12345678-9,
 123456789}
```

Finally, select out the expressions that sum to zero.

```
solutions = Select[possible,
 ToExpression[#] == 0&]

{1+2-34-56+78+9, 1-2-34+5+6+7+8+9,
 1-23-4-56-7+89, 12+3+4-5-6-7+8-9,
 12+3-4+5-6+7-8-9, 12+3-45+6+7+8+9,
 12+34-56-7+8+9, 12-3+4+5+6-7-8-9,
 12-3+4+56-78+9, 12-3-4-5+6-7-8+9,
 12-3-4-5-6+7+8-9}
```

These steps can all be gathered together to create a new function called ZeroSum.

```
ZeroSum[k_Integer? (0 < # < 10&)] :=
 Module[ {chars = {"+", "-", ""},
   digitsequence, charsequences},
  digitsequence = ToString/@ Range[k];
  charsequences =
  Distribute[Join[Table[chars, {k - 1}],
  {""}], List];
 Select[StringJoin /@
  (Flatten[Transpose[{digitsequence,
  #}]]& /@ charsequences), ToExpression[#]
  == 0&]]
```

All solutions for $n = 9$ can be found in about one second.

```
ZeroSum[9] // Timing
```

```
{1.15 Second, {1+2-34-56+78+9,
 1-2-34+5+6+7+8+9, 1-23-4-56-7+89,
 12+3+4-5-6-7+8-9, 12+3-4+5-6+7-8-9,
 12+3-45+6+7+8+9, 12+34-56-7+8+9,
 12-3+4+5+6-7-8-9, 12-3+4+56-78+9,
 12-3-4-5+6-7-8+9, 12-3-4-5-6+7+8-9}}
```

The problem was also solved by Joseph Post.

COW 18 solution

The one-line solution to the last COW is:

```
DigitalProductRoot[n_] :=
 FixedPoint[Apply[Times,
 Select[IntegerDigits[#], Positive]]&, n]
```

It works as follows: select the integer digits that are positive, multiply them, and repeat this process until the answer remains fixed. What could be more natural! To get the answer to the last COW, the command DigitalProductRoot[12345^{6789}] took .77 seconds in Mathematica 4.0 to run. The output is equal to farmer Paul's lifetime ranking as a Wisconsin dairy farmer—number 1.

Thank you, Mark Brenneman, for your inspired illustrations. Goodbye, it was a kick.

Index of Advertisers

For free information on advertisers in this issue, circle the appropriate reader service number on the reader service card, then mail the postage-paid card or fax it to 1-856-786-4415, or visit us at http://www.pub-serv.com/rs/nsta