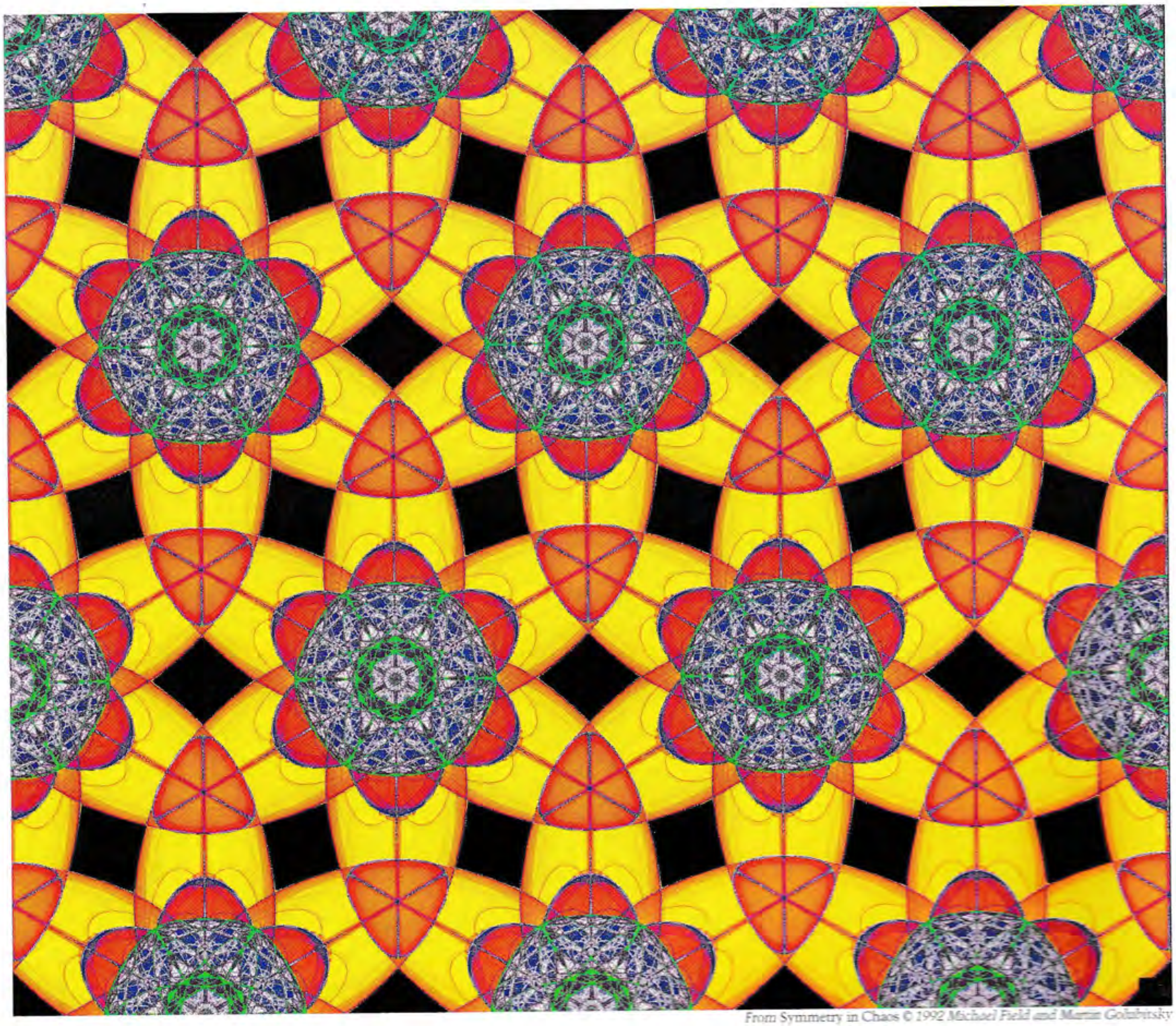
QUANTUM

MAY/JUNE 1997

US $7.50/CAN $10.50

Springer

GALLERY



Dutch Quilt (1992) by Michael Field and Martin Golubitsky

SYMMETRY HAS ALWAYS BEEN AN IMPORTANT element in art. In their book *Symmetry in Chaos: A Search for Pattern in Mathematics, Art and Nature*, mathematicians Michael Field and Martin Golubitsky provide numerous examples of such symmetry, from a rose window at Chartres to the hood ornament on a Mercedes Benz.

Nature is also full of symmetry—in fact, one is led to wonder if our art would be so symmetric if nature were not. But to a degree not found in art, nature is also chaotic, in the recently coined technical sense. A chaotic system is unpredictable, complex, and sensitive to initial conditions. It is often described by a remarkably concise set of "rules," yet minute variations in these rules can produce drastically different results.

In addition to presenting a visual feast of their own *symmetric icons*, *quilts*, and *symmetric fractals*, Field and Golubitsky provide the mathematics behind the pictures, including the QuickBasic code used to generate them. Many images from the book and the computer code can be found on the World Wide Web at http://math.uh.edu/~chaos.

Although the authors called the above design "Dutch Quilt," it is also reminiscent of the kind of pattern you might see in a kaleidoscope. Like the kaleidoscope image, "Dutch Quilt" is based on a hexagonal tiling of the plane. For a look at kaleidoscopes that goes well beyond the simple tubular toy, turn to page 4 and keep turning . . .

QUANTUM

MAY/JUNE 1997

VOLUME 7, NUMBER 5

Cover art by Yury Vashchenko

As any chef knows, it takes time to reduce a liquid stock properly. It can't be hurried, even if *you're* in a hurry. Sometimes the stuff seems "irreducible." The wait can drive you crazy! But the results are worth it.

If you patiently work through the article that begins on page 22, you'll end up with a strong sense of the connection between *irrationality* and *irreducibility* in mathematics, if not the culinary arts. Perhaps the best bet for cooks is to keep a sharp knife handy and chop off those pesky prefixes before starting any recipe.

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

From Cherokee math to tubby genes

Notes from the Hard Drive C@fé

"TECHNOLOGY, CONTENT, and access are coming together," said Linda Roberts of the U.S. Department of Education on April 14, moments before she and Senator Jeff Bingaman of New Mexico cut a "virtual ribbon" to open the 2nd Annual Hard Drive C@fé, hosted by the National Geographic Society and organized by the Eisenhower National Clearinghouse for Mathematics and Science Education (ENC). The event showcased 32 World Wide Web sites devoted to education (including those of the National Science Teachers Association and the National Council of Teachers of Mathematics), and it offered some encouragement to those who feel that technology will do just fine, and that access is on its way—it's the content we're worried about.

One reason content is troublesome on the Web is that it's hybrid medium with a high visual content. While a picture can convey information more quickly in some situations, in most others the written word is much more efficient. You can take in far more information reading for a half hour than watching the evening news on TV. What is it about moving images in a box that transfixes us? A simulated emotional bond is created, and for some it is very powerful indeed.

A generation ago, Marshall McLuhan said, "The medium is the message," and spawned new way of looking at television and other mass media. Many people have teased many messages out of that Delphic utterance. One interpretation says that a medium—for instance, television—not only affects the way content is presented, it selects the content. The most newsworthy "event" on a given day might be the release of statistics showing a decrease in violent crime, yet more air time will likely be devoted to a live shot of a reporter recounting the details of an actual crime, or, better yet, a crime in progress. Television wants to do what it does best, which is show pictures. There's nothing inherently wrong with that, as long as we remind ourselves of the limitations of pictures—their lack of context, their ambiguity, and so on.

Late last year, as the holiday gift season approached, something called Web TV was introduced. It's a television set with Internet capabilities built in. Interestingly, it was sold without a keyboard (although one could buy one as an accessory). The implication is that most Web "surfers" roam from site to site much like "couch potatoes" flip from channel to channel. Web TV seems more TV than Web, more passive than active, more entertaining than intellectually satisfying.

Meanwhile, the technology continues to flourish. Java, Shockwave, and other Web browser enhancements are already adding movement and sound (and who knows, maybe even smell) to some high-voltage sites. "Design" takes the lead, and wordsmiths struggle to stay relevant. Web design becomes a big production, with budgets to match.

Is that the message of this new medium?

Happily, many of the pioneers who headed into cyberspace have been busy creating interesting, useful, and yes, entertaining content on shoestring budgets. A highlight of the ENC Hard Drive C@fé was the presentation by the National Indian Telecommunications Institute (NITI). NITI trains Native American teachers in the basics of Web work. The teachers then go home and produce Web pages of their own. In one instance, a teacher has combined Cherokee language instruction (which can't be taught in the schools) with mathematics. It's a fascinating niche that shows how the Web can be used to provide a small, scattered audience with unique educational material at very little cost.

Some of the Web sites featured at the Hard Drive C@fé

Access Excellence
www.gene.com/ae

Challenger Center OnLine
www.challenger.org

ENC Online
www.enc.org

Lawrence Hall of Science
www.lhs.berkeley.edu

National Indian Telecommunications Institute
numa.niti.org

NPR Science Friday Kids Connection
www.npr.org/programs/sfkids

USA Today Classline
classline.usatoday.com

Several "graduates" of Genentech's Access Excellence program—teachers who are scattered across the United States—are collaborating to put data from the Human Genome Project into a form that is useful to teachers and students. As they explained their plans for activities involving the "tubby gene" to a C@fé visitor, their enthusiasm was evident—and contagious.

It was encouraging to see that some of the bigger media players like *USA Today*, *National Geographic*, and National Public Radio (NPR) are producing Web sites aimed at students, teachers, and parents. Even more heartening is the response of the scientific community to such efforts. The NPR site engages the services scientists who act as "mentors," answering questions from students. Ira Flatow, host of NPR's Science Friday, says that interest is so great among working scientists, the producers of his Web site have to turn away mentors (they have about 40).

It was striking that, despite the availability of powerful Web search engines, one still encounters sites with long lists of links to other sites. True, someone has waded through the hundreds of hits and picked the "best," but the notion that someone knows what's best for me goes against the Web grain. Also, Web pages come and go, and links to them must share their fate.

As the World Wide Web evolves, we'll continue to find sites to give us the latest baseball scores and stats, and sites that want to sell us something. But it seems likely that we'll also be able to find sites that provide thoughtfully produced, thought-provoking content, especially in the science area.

If you would like to sample the entrées and à la carte dishes served at the Hard Drive C@fé, you can visit the ENC's page at www.enc.org/hdcafe/, or e-mail me for the Web addresses of all the participating organizations.

—Tim Weber
Managing Editor, *Quantum*
tweber@nsta.org

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.

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 7 (6 issues) will be published in 1996–1997. *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.

Advertising: MPA

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

Printed on acid-free paper.

Periodicals postage paid at New York, NY, and additional mailing ofices. **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: $17.70; Personal rate (non-student): $23.70; Institutional rate: $45.00; 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. Send orders to *Quantum*, Springer-Verlag New York, Inc., P.O. Box 2485, Secaucus NJ 07096-2485; or call 1-800-SPRINGER (777-4643) (in New York, call (201) 348-4033).

All Countries Outside North America: Personal rate is US$31.70 including surface delivery charges; institutional rate is US$45.00; airmail delivery is US$13.00 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.

Рис. 2. Обычный калейдоскоп

On kaleidoscopes

A mathematician looks at them in all their dimensions

by E. B. Vinberg

THE KALEIDOSCOPE (FROM THE GREEK FOR "viewing a beautiful form") is a children's toy in which pieces of colored glass, reflected many times in three mirrors, produce amazing patterns. These mirrors are arranged as the three side faces of a regular triangular prism, so that the angles between them are equal to $\pi/3$. If the values of the angles were different, then, generally speaking, the images would overlap and no symmetric pattern would appear. Still, there are some exceptions, which we'll find later in this article.

The ordinary kaleidoscope, described above, is actually two-dimensional, since we see just a plane pattern in it. One can imagine a three-dimensional kaleidoscope as a polyhedral chamber with mirror sides. An observer placed inside it would see repeated images of all the items lying inside the chamber. The images would overlap, except for a few particular cases (we'll list them below) when this doesn't occur—instead, a symmetric, three-dimensional pattern appears.

Leaving aside the question whether it can be realized in practice, one can speak about multidimensional kaleidoscopes, as well as non-Euclidean kaleidoscopes—that is, kaleidoscopes on a sphere and in Lobachevskian space. A comprehensive description of all kaleidoscopes in Euclidean space and on a sphere of an arbitrary dimension was given in 1934 by the English mathematician H. S. M. Coxeter. The cover of the Russian translation of his book shows a kaleidoscope on an ordinary (two-dimensional) sphere (fig. 1). There is an intimate connection between spherical kaleidoscopes and regular polyhedrons, which we'll discuss in more detail below.

Kaleidoscopes on the Lobachevskian plane were used by Poincaré and Klein at the end of the last century in their research on the theory of automorphic functions of a complex variable. In 1958–60 the eminent Dutch artist M. C. Escher created several intriguing designs based on these kaleidoscopes.

Since 1965 kaleidoscopes in Lobachevskian space have become the subject of intensive research in connection with certain problems in group theory. A complete description of such kaleidoscopes in an arbitrary dimension is far from being completed. There is a surprising theorem (proved by the author of this article) that asserts that no kaleidoscope exists in n-dimensional Lobachevskian space if $n \geq 30$. Examples of such kaleidoscopes are known only for $n \leq 8$.

Art by V. Ivanyuk

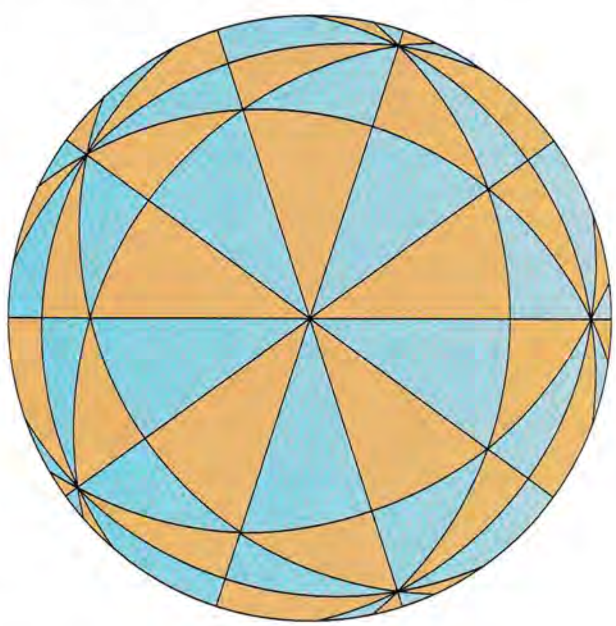


Figure 1

In addition to their applications in geometry (regular polyhedrons), the theory of functions of a complex variable, and group theory, kaleidoscopes play an important role in number theory, the theory of Lie algebras, algebraic geometry, and many other branches of mathematics. However, I should point out that the word "kaleidoscope" is not used in the mathematical literature. Mathematicians speak of a "discrete group generated by reflections" instead.

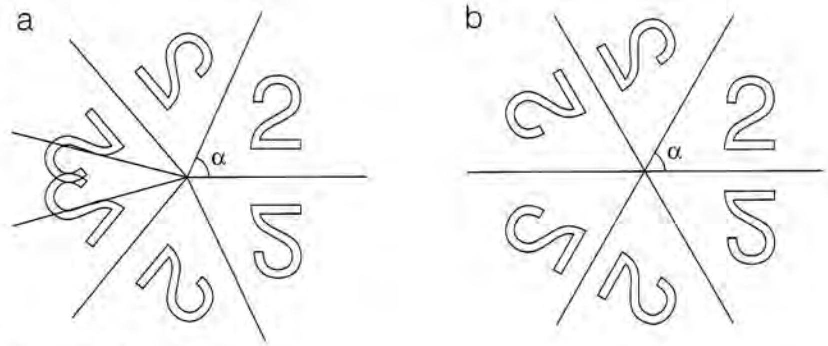


Figure 2

In this article, we won't have an opportunity to examine the applications of kaleidoscopes (except for the connection between spherical kaleidoscopes and regular polyhedrons). However, the investigation of kaleidoscopes themselves occupies a strong niche in the field of geometry.

Fundamental property

To begin our survey of kaleidoscopes, let's consider the simplest situation: two mirrors set at an angle α to each other. If $\alpha = \pi/k$ for some integer k, we will say that α is an *integral submultiple* of π. If α is not an integral submultiple of π, then (see figure 2a) images of an item placed between them overlaps, so that you see images of two different points in one point. (As a matter of fact, you see images of two different points simultaneously only if you change your point of view. However, this has no bearing on our theoretical discussion.) If, on the other hand, α is an integral submultiple of π, this overlapping does not occur (fig. 2b).

Since the images of any point do not leave the plane perpendicular to the common axis of the mirrors (this is the plane depicted in figure 2), the phenomenon discovered above is planar in nature. We can speak about reflections of plane figures with respect to straight lines, and figure 2 demonstrates that repeated mirror images of a plane figure with respect to the sides of an angle do not overlap if and only if the value α of the angle is an integral submultiple of π. More precisely, if $\alpha = \pi/k$, where $k \geq 2$, is a positive integer, then the plane splits into $2k$ congruent angular domains with a common vertex, so that an image of the original angular domain appears in each of them. In half of these domains, the picture will be inverted; in the other half, including the original one, it will be normal.

Imagine a convex polygonal domain formed by mirrors. When will the images obtained by repeated reflections with respect to the sides not overlap? It follows from our reasoning above that all the angles of the polygon must be integral submultiples of π. It's possible to prove the sufficiency of this condition. If the condition holds, the plane breaks up into polygons congruent to the original one, so that any two of them that have a common side are symmetric with respect to this side. Each polygon in this tiling contains an image of the original domain. A tiling of the plane obtained from an equilateral triangle is shown in figure 3. This is, in fact, the tiling you see in an ordinary kaleidoscope.

Similarly, repeated images of the interior of a convex polyhedron, after reflections with respect to its faces, do not overlap if and only if all dihedral angles of the polyhedron are integral submultiples of π. This theorem holds even for non-Euclidean polygons and polyhedrons.

A polygon (or polyhedron) whose (dihedral) angles are all integral submultiples of π is called a *Coxeter polygon (polyhedron)*. So the task of describing all theoretically possible kaleidoscopes is equivalent to describing all Coxeter polygons and polyhedrons.

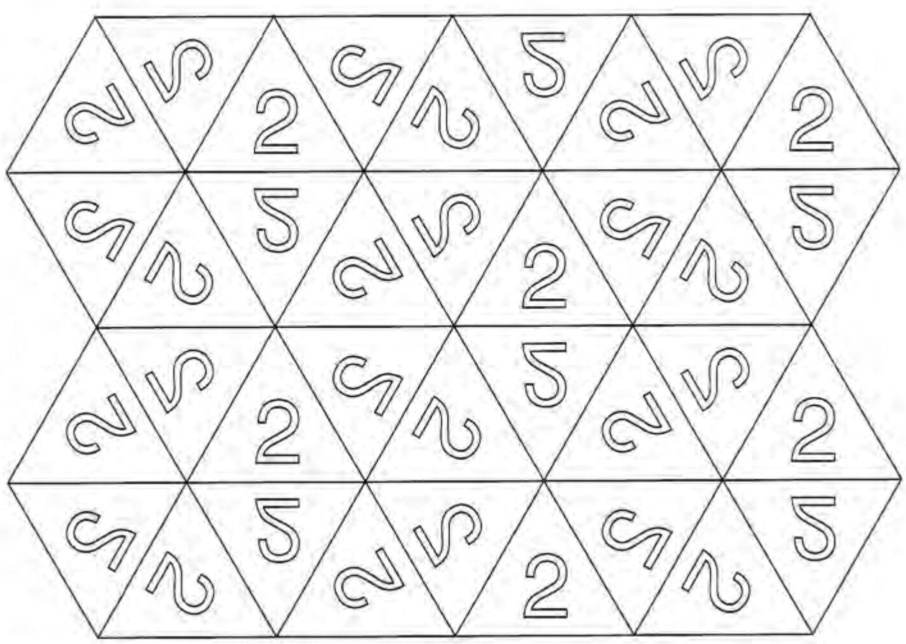


Figure 3

Two-dimensional kaleidoscopes

It's not difficult to find all Coxeter polygons on the Euclidean plane. We know that sum of the angles of a Euclidean polygon is equal to $\pi(n-2)$. Thus the average value of its angles is $\pi(1-2/n)$, which gives $\pi/2$ for $n = 4$. But it follows from the definition of a Coxeter polygon that none of its angles can exceed $\pi/2$. Therefore, the only Coxeter quadrangle is a rectangle, and there are no Coxeter polygons with more than four sides.

Further, since the sum of the angles in a triangle is π, we have the following Diophantine equation for a Coxeter triangle with angles π/k, π/l, π/m:

$$\frac{1}{k}+\frac{1}{l}+\frac{1}{m}=1. \tag{1}$$

Up to permutations of k, l, and m, this equation has three solutions:

$$(3, 3, 3), (2, 4, 4), (2, 3, 6).$$

Thus there are exactly three different Coxeter triangles: equilateral triangle, isosceles right triangle, and right triangle with acute angles equal to $\pi/3$ and $\pi/6$. Figure 4 shows the corresponding tilings of a plane. These tilings, together with that of the rectangle, constitute the four types of two-dimensional Euclidean kaleidoscopes.

In a similar way, we can find all two-dimensional spherical kaleidoscopes. It can be proved that the sum of the angles of a spherical n-gon is greater than $\pi(n-2)$. What can we say, then, about the number of angles in a Coxeter spherical polygon? Simply this: there are no Coxeter spherical polygons other than triangles. For the case of a Coxeter spherical triangle, equation (1) is replaced by the inequality

$$\frac{1}{k}+\frac{1}{l}+\frac{1}{m}>1, \tag{2}$$

which has four solutions:

$$(2, 2, m), (2, 3, 3), (2, 3, 4), (2, 3, 5).$$

The first of these solutions corresponds to the tiling of a sphere with $4m$ "birectangular" triangles, produced by the equator and $2m$ meridians equidistant from one another. The solution (2, 3, 5) corresponds to the tiling shown earlier in figure 1.

We'll return to spherical kaleidoscopes below in connection with regular polyhedrons.

As far as kaleidoscopes on the Lobachevskian plane are concerned, they are much more diverse. Everything that is impossible on the Euclidean plane or on a sphere becomes possible on the Lobachevskian plane. In fact, the sum of the angles of an n-gon on the Lobachevskian plane is less than $\pi(n-2)$. Thus on the Lobachevskian plane there is an n-gon with angles π/k_1, π/k_2, ..., π/k_n for all k_1, k_2, ..., k_n satisfying the inequality

$$\frac{1}{k_1}+\frac{1}{k_2}+\cdots+\frac{1}{k_n}<n-2,$$

This inequality holds automatically for $n > 4$, and for $n = 4$ it holds only if $(k_1, k_2, k_3, k_4) \neq (2, 2, 2, 2)$. If $n = 3$, we obtain the following inequality for a Coxeter triangle with angles π/k, π/l, π/m:

$$\frac{1}{k}+\frac{1}{l}+\frac{1}{m}<1. \tag{3}$$

Its solutions include all triplets (k, l, m) except for the solutions of equation (1) and inequality (2).

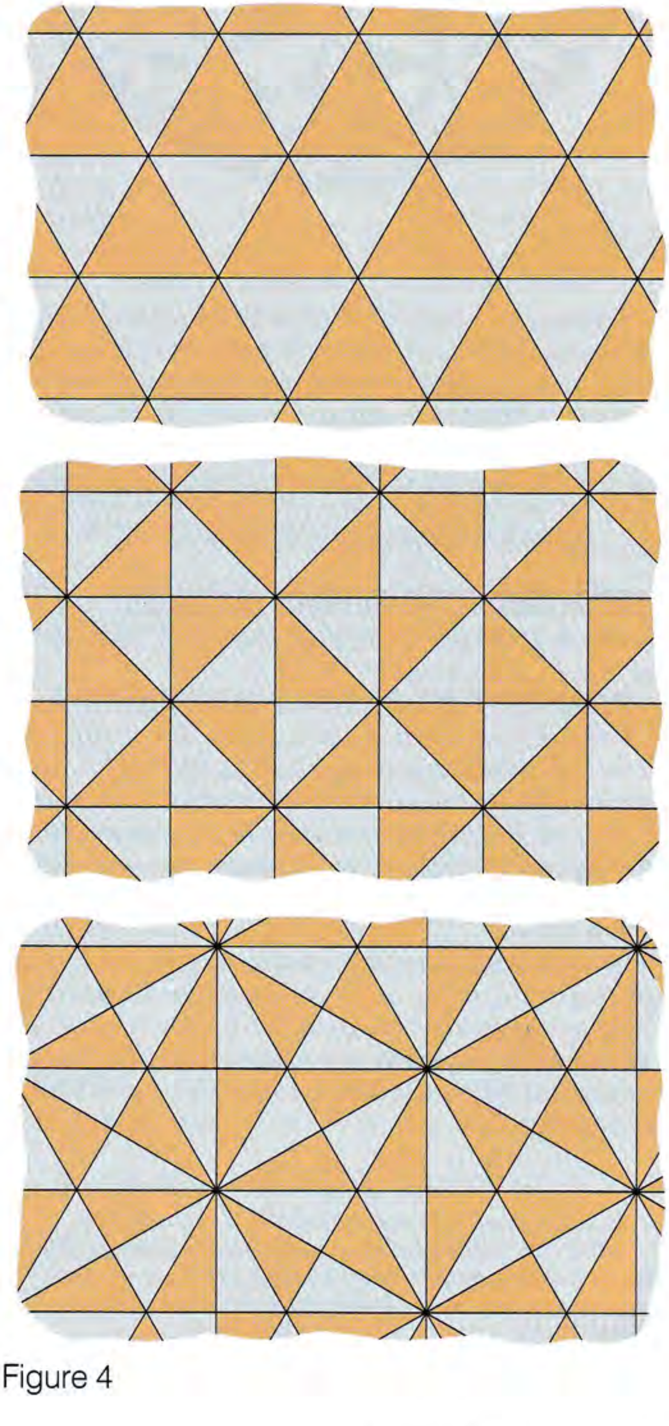


Figure 4

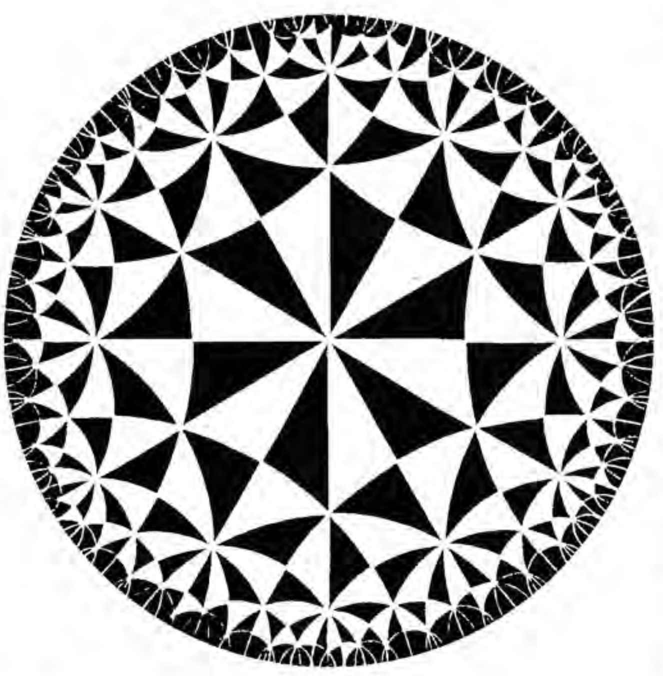


Figure 5

For example, on the Lobachevskian plane there is a triangular kaleidoscope with angles $\pi/2$, $\pi/4$, $\pi/6$. The tiling corresponding to it is shown in figure 5. Here we use the so-called Poincaré model, in which the Lobachevskian plane is represented by an open disk and the straight lines on it are represented by its diameters and arcs of circles, perpendicular to its border; the angles coincide with their Euclidean counterparts.[1]

Spherical kaleidoscopes and regular polyhedrons

Any regular polyhedron can be associated with a spherical kaleidoscope.

Let M be a regular polyhedron with center at O. Let A be the center of one of its faces, B the midpoint of an edge adjacent to this face, and C one of the two vertices that belong to the edge. We'll call the trihedral cone K with vertex at O and edges passing through points A, B, and C, respectively, the *fundamental cone* of the polyhedron M (see figure 6, where M is a cube).

If we vary the faces, their edges, and the vertices belonging to the edges, we'll obtain many different fundamental cones from any given polyhedron. They do not overlap, and their union covers all of space. We can find the number N of fundamental cones from one of the following formulas:

$$N = 2pF = 4E = 2qV, \tag{4}$$

where we use the following notation:

F is the number of faces of the polyhedron M;
E is the number of its edges;
V is the number of its vertices;
p is the number of sides of (each) face;
q is the number of edges emerging from each vertex.

Thus for a cube,

$$F = 6, E = 12, V = 8, p = 4, q = 3, N = 48.$$

The polyhedron M is symmetric with respect to every plane containing a face of a fundamental cone, and any two adjoining fundamental cones are symmetric to each other with respect to their common face. For example, if we consider a cube, then the plane OAC is its symmetry plane, passing through two opposite edges; and the cone K' symmetric to the cone K with respect to this plane is the fundamental cone whose edges pass through the points A, C, B' (see figure 6).

The edge OA of the fundamental cone K is a common edge of $2p$ different fundamental cones, and their dihedral angles at this edge are equal. Thus the dihedral angle at edge OA of cone K is equal to π/p. Similarly, the dihedral angle at edge OB is equal to $\pi/2$, and the angle at the edge OC is equal to π/q. Thus we conclude that a fundamental cone cuts from a sphere concentric to the polyhedron M a right spherical triangle with acute angles π/p and π/q. Any two of these triangles having a common side will be symmetric with respect to this side. In this way we obtain a spherical kaleidoscope.

When we pass from a regular polyhedron to the spherical kaleidoscope, a piece of information is lost. We cannot say which edge (OA or OB) of a fundamental cone passes through the center of a face, and which edge passes through a vertex of the polyhedron M. The same kaleidoscope will correspond to the regular polyhedron M' whose vertices coincide with centers of the faces of polyhedron M. Regular polyhedrons M and M' of this sort are called duals of each other. For example, a cube is the dual of an octahedron. The tetrahedron is its own dual (or, it's the dual of a regular polyhedron, which is a tetrahedron). The numbers p

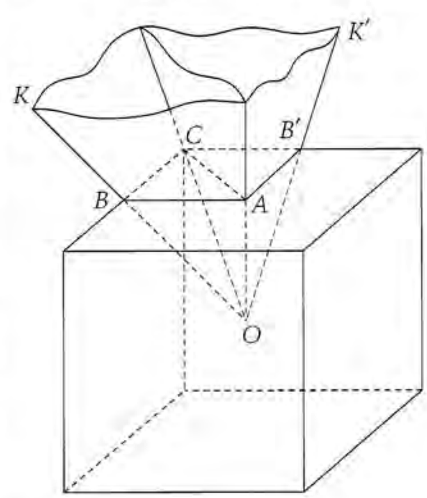


Figure 6

[1]M. C. Escher's fanciful tilings on the hyperbolic plane can be found in almost any book devoted to his artistry. See, for instance, *M. C. Escher Kaleidocycles* by Doris Schattschneider and Wallace Walker (Pomegranate Artbooks, 1977)—the image on page 19 uses squares and triangles in a tiling of Escherian fishes.

and q switch when we pass from M to M', as do the numbers F and V.

Every spherical kaleidoscope, determined by one of the solutions

$$(2, 3, 3), (2, 3, 4), (2, 3, 5)$$

of inequality (2), corresponds to a pair of dual regular polyhedrons. These pairs are tetrahedron–tetrahedron, cube–octahedron, and dodecahedron–icosahedron, respectively. There is no regular polyhedron corresponding to the solution $(2, 2, m)$, since its existence implies that $p, q \geq 3$.

It's known that the area of a spherical triangle is equal to its angular excess—that is, to the sum of its angles minus π. In particular, the area of a right spherical triangle with acute angles π/p and π/q equals $(1/p + \ldots + 1/q - 1/2)\pi$. Recalling that area of the whole sphere is 4π, we obtain another formula for calculating the number N:

$$N = \frac{4}{1/p + 1/q - 1/2}. \tag{5}$$

(compare this with equation (4) above).

A similar connection exists between regular n-dimensional polyhedrons and kaleidoscopes on an $(n-1)$-dimensional sphere. It's amazing that while there are only five regular polyhedrons in three-dimensional space, there are six of them in four-dimensional space, and there are only three of them in n-dimensional space if $n > 4$ (they are analogues of the tetrahedron, cube and octahedron).

Three-dimensional kaleidoscopes

The task of finding all Coxeter polyhedrons is complicated by the fact that the relations between dihedral angles of a polyhedron are not as simple as the relations between the angles of a polygon.

The intersection of a convex polyhedron M and a small sphere with center C at one of its vertices defines a convex spherical polygon whose angles are equal to the dihedral angles at the corresponding edges of M. Therefore, if the number of edges emerging from the vertex C is q, then the sum of the dihedral angles at these edges is greater than $\pi(q-2)$. This implies that if all the dihedral angles of the polyhedron M are not greater than $\pi/2$ (in particular, if it is a Coxeter polyhedron), then there are only three edges emerging from any of its vertices. Polyhedrons that satisfy this last condition are called *primitive*. Thus, the cube and tetrahedron are primitive polyhedrons; the octahedron is not.

However, these simple inequalities do not exhaust the set of relations between the dihedral angles of a convex polyhedron.

Consider the simplest case, when M is a triangular pyramid. Let's assign arbitrary numbers to its faces and denote the angle between the ith and jth faces by $\alpha_{ij} = \alpha_{ji}$. Using linear algebra, we can show that the angles of a Euclidean triangular pyramid comply with the following relation:

$$\begin{vmatrix} 1 & -\cos\alpha_{12} & -\cos\alpha_{13} & -\cos\alpha_{14} \\ -\cos\alpha_{12} & 1 & -\cos\alpha_{23} & -\cos\alpha_{24} \\ -\cos\alpha_{13} & -\cos\alpha_{23} & 1 & -\cos\alpha_{34} \\ -\cos\alpha_{14} & -\cos\alpha_{24} & -\cos\alpha_{34} & 1 \end{vmatrix} = 0. \tag{6}$$

(The determinant on the left is called the *Gramm determinant* of the system of unit vectors orthogonal to the pyramid's faces. It vanishes because these vectors are linearly dependent.)

Note that we can similarly prove that the angles α, β, γ of a Euclidean triangle satisfy the relation

$$\begin{vmatrix} 1 & -\cos\alpha & -\cos\beta \\ -\cos\alpha & 1 & -\cos\gamma \\ -\cos\beta & -\cos\gamma & 1 \end{vmatrix} = 0.$$

However, if the sum of any two of the angles is less than π, this equation is equivalent to the equality $\alpha + \beta + \gamma = \pi$. (Try to prove it!) As far as the equation (6) is concerned, it cannot be reduced to such a simple form.

Equation (6), together with the inequalities derived above, are necessary and sufficient for the existence of a triangular pyramid with dihedral angles α_{ij} in Euclidean space. Using this fact, we can find all Euclidean triangular pyramids whose dihedral angles are integral

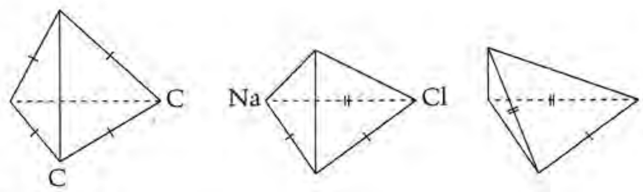


Figure 7

submultiples of α. It turns out that there are exactly three of them. They are shown in figure 7, where the following notation is adopted: dihedral angles at the clear edges are equal to $\pi/2$, and at the edges with one or two strokes they are equal to $\pi/3$ and $\pi/4$, respectively. We can see that the first of the pyramids in figure 7 is cut by its plane of symmetry into two pyramids similar to the second. The third pyramid can be obtained in the same way from the second.

Except for these three, there are only four other kaleidoscopes in Euclidean space that can be reduced in some sense to planar kaleidoscopes. They are made of right prisms whose bases form a two-dimensional kaleidoscope.

Three-dimensional Euclidean kaleidoscopes are directly related to crystallography. Some crystal lattices can be obtained if we place several atoms in a certain way in such a kaleidoscope and consider all their images, which appear as repeated reflections with respect to the sides of the kaleidoscope. Thus the lattice of a diamond appears from the first kaleidoscope shown in figure 7 if we place atoms of carbon at the two vertices marked in the figure. The lattice of table salt appears from the second if atoms of sodium and chlorine are placed at the vertices indicated.

We can also find kaleidoscopes on a three-dimensional sphere. All these kaleidoscopes are nothing but (spherical) triangular pyramids. The equal sign in equation (6) is replaced by the "greater than" sign, just as the sum of the angles of a triangle becomes greater than π when we pass from the plane to the sphere.

Andreyev's theorem

In Lobachevskian space, the equal sign in equation (6) is replaced by the "less than" sign. We can find all the Coxeter polyhedrons in Lobachevskian space, which are triangular pyramids. However, they make up only a trifling part of all Coxeter polyhedrons in this case. Just as on the Lobachevskian plane, there are Coxeter polygons with an arbitrarily large number (as a matter of fact, with an arbitrary number) of sides, in Lobachevskian space there exist Coxeter polyhedrons with arbitrarily large number of faces. Still, unlike polygons, their combinatoric structure might be very complicated. Therefore, it's difficult to give a complete description of them.

It appears that the most complete description of all possible Coxeter polyhedrons in Lobachevskian space is contained in the theorem proved by E. M. Andreyev in 1970. This is a general theorem, concerning not only Coxeter polyhedrons but all polyhedrons whose dihedral angles do not exceed $\pi/2$. Polyhedrons of this sort are called *acute-angled* (though they might have right dihedral angles). As we proved above (no corrections to the proof are necessary in the case of Lobachevskian space), every acute-angled polyhedron is simple.

Andreyev's theorem suggests the necessary and sufficient conditions for the existence of an acute-angled polyhedron of the given combinatoric structure (other than that of a triangular pyramid) in Lobachevskian space. These conditions are as follows:

1. If three faces of the polyhedron meet in a vertex, then the sum of the angles formed by them is greater than π (the necessity of this condition was proved in the previous section).
2. If three faces of the polyhedron are adjacent to each other but do not meet in a vertex, then the sum of the dihedral angles between them is less than π.
3. If four faces adjoin each other "in a circle" (like the lateral faces of a quadrangular prism), then there are dihedral angles different from $\pi/2$ between them.
4. If the polyhedron is a triangular prism, then some of angles between its bases and its lateral faces are different from $\pi/2$.

Andreyev's theorem is in a certain sense analogous to the famous theorem of A. D. Alexandrov concerning the existence of a Euclidean polyhedron with the given development. Still, there is no exact Euclidean analogue of this theorem (and there cannot be). It's one of the theorems that are peculiar to Lobachevskian geometry, like the criterion of congruence of triangles with equal angles.

Using Andreyev's theorem, we can prove, for example, that there exist "rectangular" polyhedrons (that is, polyhedrons whose dihedral angles are all right angles) with an arbitrarily large number of faces in Lobachevskian space. (The reader is invited to try to prove this.) And so we have many different rectangular kaleidoscopes in Lobachevskian space. Since we can't look at all of them in detail here, I'll confine myself to noting that since the end of the last century, kaleidoscopes in Lobachevskian space have been applied to the arithmetic of quadratic forms, and during the past 15 years they have found application in three-dimensional topology.

Just for the fun of it!

B201

Strange painting. There is a painting on the wall of Dr. Smile's waiting room. The unusual thing about this painting is the way it's hung. Dr. Smile hammered two nails (instead of one) into the wall. He says that he has wound the picture wire around these nails in such a way that the painting would fall if either the nail were pulled out. How did he do it? (A. Spivak)

B202

Missing digit. How would the number 1/1996 change (that is, will it increase or decrease, and by what factor will it be multiplied) if the first nonzero digit in its decimal notation were omitted? (D. Averiyanov)

B203

Oil and vinegar. Imagine you're preparing a picnic basket for yourself and a friend and you'd like to have a tossed salad. The problem is, you like your salad with just vinegar on it, and your friend likes it with just oil, but you don't have room for two containers. Since oil and vinegar don't mix, you fill a single bottle with both liquids. Can you take a little vinegar to prepare your salad and a spoonful of oil for your friend's salad in such a way as to leave the rest of oil and vinegar in the bottle?

B204

Triangles in a parallelogram. Two arbitrary points are taken inside a parallelogram. Line segments connecting them to all the vertices of the parallelogram are drawn (see the figure above). Prove that the sum of the areas of the two red triangles is equal to the sum of the areas of the two blue triangles. (I. Sharygin)

B205

Family planning. A family of four (father, mother, son, and daughter) went on a hike. They walked all day long and, when evening was already drawing on, came to an old bridge over a deep gully. It was very dark and they had only one lantern with them. The bridge was so narrow and shaky that it could hold no more than two persons at a time. Suppose it takes the son 1 minute to cross the bridge, the daughter 3 minutes, the father 8 minutes, and the mother 10 minutes. Can the entire family cross the bridge in 20 minutes? If so, how? (When any two persons cross the bridge, their speed is equal to that of the slower one. Also, the lantern must be used while crossing the bridge.)

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

Does elementary length exist?

Some surprising implications of the theory of relativity and quantum mechanics

by Andrey Sakharov

SCIENTISTS ALL OVER THE world expect the physics of elementary particles to provide results of great practical and even philosophical importance, perhaps defining more exactly the basic notions of time, space, and causality. There is no reason to expect such changes in the fundamental principles in other branches of physics, in which the atomic particles (electrons, photons, nuclei) can be considered the basic ones. In optics, biophysics, molecular and crystal physics, and most other areas, the basic principles of quantum mechanics, statistical physics, and the theory of relativity provide the firm and reliably proved foundation for the theoretical description and explanation of observed phenomena, and for new predictions, discoveries, and practical applications (like the transistor, laser, electroluminescence, paramagnetic resonance, the Mossbauer effect, holography, and so on). We are certain that in these branches of physics any new phenomenon can be comprehensively described on the basis of known principles, although we may sometimes need to use powerful computing resources and additional experimental data (as recently happened, for example, with the phenomena of superfluidity and superconductivity).

However, when physicists try to explain the nature of mass, electric charge, and other properties of elementary particles themselves, as well as their mutual interactions and conversions one into the other, an impression arises that in this branch of physics we lack certain fundamental principles. Research conducted with particle accelerators, as well as experiments with "natural accelerators" (cosmic rays), have produced one surprise after another. In just the last 10 years, dozens of new elementary particles with queer properties were discovered, including two "types" of neutrino (of electron and muon origin).[1] Moreover, the list of discoveries contains violations of the symmetrical nature of natural law for mirror reflection, for the conversion of particles into antiparticles, and for reversing the direction taken by physical processes. This last violation of symmetry is the most surprising, since it still lacks even a phenomenological description.

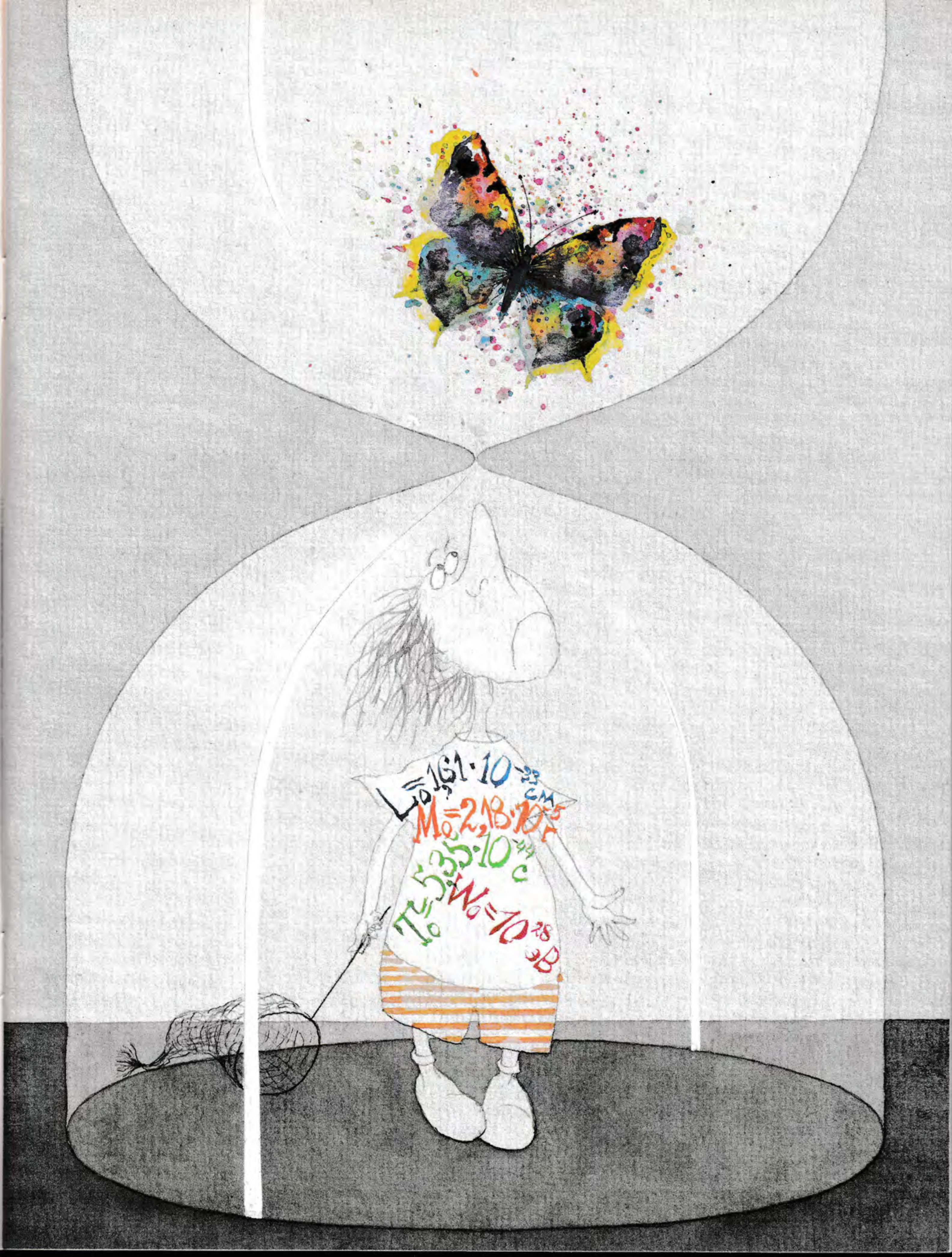
Only by a stretch of the imagination can the author consider himself an expert on the physics of elementary particles. Nevertheless he dares to discuss one of the basic problems in this field—the problem of elementary length. This has to do with the supposed existence of a limitation in principle on the application of modern science's basic ideas about space and causality (that is, the theories of quantum mechanics and relativity). It involves the need to describe "small-scale" phenomena lying beyond a certain limit by means of some new, more abstract and more fundamental physical concepts and mathematical methods.

This article will not describe new and wonderful discoveries. Its basic assertion is of a rather negative nature. Nevertheless the author feels that in addressing the problem of the

From an article written in 1968 by the great scientist and human rights activist Andrey Dmitryevich Sakharov for the journal *Physics in High School*. Published in *Quantum*'s sister magazine *Kvant* in 1991.—*Ed.*

[1]Many changes have taken place in the field of elementary particles since 1968. There are now three types of neutrino, the muon is no longer a meson, and quarks mediate the strong force. However, in the interest of the historical perspective, we have chosen to leave the article in its original form.—*Ed.*

Art by Pavel Chernusky



$L_0 = 1{,}61 \cdot 10^{-33}$ см
$M_0 = 2{,}18 \cdot 10^{-5}$ г
$T_0 = 5{,}35 \cdot 10^{-44}$ с
$W_0 = 10^{28}$ эВ

first principles of science (such unshakable concepts as length and an interval of time), any step forward and any more precise definition of nuances should be of interest not only for specialists. So, however unclear the situation is, the author has decided to talk about the tortuous drama of ideas in one corner of modern theoretical physics, as he sees it.

Even before the creation of quantum theory, when trying to describe the electron as a point particle, scientists encountered an obstacle in calculating its electrostatic energy. Let's recall that the electrostatic energy of a sphere with a uniform surface charge density equals $U = e^2/(2r)$, where e is the charge and r is the radius.[2] For an arbitrary distribution of the charge density along the radial direction, we have $U \sim e^2/r$. The point electron corresponds to the approximation $r \to 0$, which yields $U \to \infty$. According to Einstein's famous formula, the energy U is related to the rest mass as $m = U/c^2$, so the mass of the point electron must be infinite. Inserting the experimentally obtained electron mass into the formula $m = e^2/(rc^2)$ yields $r = 2.8 \cdot 10^{-13}$ cm. This value is known as the classical radius of the electron.

The situation became even more complicated with the advent of quantum mechanics. On the one hand, quantum effects result in far smaller numerical values for the electromagnetic energy of the electron for the same r, but the problem of infinite energy $W \to \infty$ at $r \to 0$ still exists, though now W is proportional to $\ln(r^{-1})$ and not to r^{-1}. On the other hand, the principal difficulties in considering the electron as a point particle arise when calculating the other basic theoretical values: the force of interaction between particles, the probabilities of scattering or decay, and so on. Still, it is very difficult to reconcile the notion of a nonpoint particle with the principles of relativity theory—indeed, signals could travel with a velocity greater than that of light along the solid body of an extensive particle.

It was proposed that the quantum theory of elementary particles is incomplete both logically and mathematically. This thought was formulated most clearly in the 1930s by the outstanding German physicist and theorist Werner Heisenberg. Here is the line of his argument. In his opinion, the difficulties of the theory of elementary particles had deep, intrinsic roots; they touch on fundamental principles, just as the seemingly unsolvable problems of the electromagnetic theory of moving bodies did before the creation of relativity theory, or the paradoxes of atomic phenomena before the era of quantum mechanics.

The difficulties of electrodynamics could not be overcome without revising and delineating such a seemingly self-evident notion as simultaneity. The new formulas of relativity are only a secondary result of such an epistemological revision of basic notions. The paradoxes of wave–particle duality gave rise to even more profound ideas—the complementarity principle and the statistical interpretation of the wave function. According to Heisenberg, the inconsistencies in considering elementary particles as point objects, the absence in the modern theory of any criteria that would determine the numerical values for the mass and charge of elementary particles—these are manifestations of the incomplete and inexact character of the very notions of space, time, and causality for "small-scale" phenomena.

Heisenberg noted that Einstein's theory of relativity differs from the ideas of Galileo and Newton on space and time by postulating the existence of an absolute unit of velocity, which in Einstein's theory is the maximum velocity for the propagation of interactions, numerically equal to the velocity of light in vacuum ($c = 3 \cdot 10^{10}$ cm/s). At velocities that are far less than c, the pre-Einsteinian concepts describe reality correctly. Similarly, the boundary between quantum and classical (that is, "nonquantum") theories is determined by another basic constant, which has the dimension "energy × time": Planck's constant $\hbar$, which is the proportionality factor for the difference in the energies of two quantum levels and the electromagnetic frequency related to the quantum transition:

$$E_1 - E_2 = \hbar\omega.$$

If ω is measured in angular units rad/s, the numerical value of the constant $\hbar$ equals $1.05 \cdot 10^{-27}$ erg · s. Planck himself measured the oscillation frequency $\nu = \omega/2\pi$ in s^{-1}, so he defined as the constant $h = 2\pi\hbar = 6.6 \cdot 10^{-27}$ erg · s. The definition and designation of $\hbar$ were introduced by Dirac.[3]

The classical notions correspond to reality when one is dealing with macroscopic processes—for instance, investigating the transmission of radio waves by an antenna, when the emitted energy E is much greater than the energy of an individual quantum $\hbar\omega$. However, the classical approximation is absolutely useless when one is considering the emission of a single photon by an excited atom.

Heisenberg noted further that difficulties in the quantum theory of elementary particles arose in analyzing problems where the transition of a large amount of energy or momentum was important—that is, during collisions of particles whose location was severely restricted in space and so had a very small de Broglie wavelength. Thus Heisenberg advanced a hypothesis that at some elementary length l_0 (in the first variant of his concept he considered this to be the classical radius of the electron r), the known laws of quantum and relativistic theories lose their power, and that to describe such a tiny world we need new notions, even more

In this system of units, Coulomb's constant $k = {}^1/_4\pi\varepsilon_0$ is set equal to 1.—*Ed.*

[3]As a rule, physicists do not use the SI system of units, preferring instead the CGS system (the first letters of the fundamental units: centimeter, gram, second). In this system, the erg is the unit of energy (1 erg = 10^{-7} J). If some value is measured in a unit that has no special name, physicists write "CGS unit."—*Ed.*

abstract than the ones used in these theories.

According to Heisenberg, it is the magnitude l_0 that also determines the characteristic scale for the mass of elementary particles. Taking as the unit of mass $\hbar/cl_0 = 70\ \text{MeV}/c^2$, we obtain the mass of particles at rest with a high degree of accuracy (the following list is based on the current set of elementary particles):

μ-meson	3/2
π-meson	2
K-meson	7
η-meson	8
proton, neutron	13.5
Λ-hyperon	16
Σ-hyperon	17
Ξ-hyperon	19
electron	1/137
photon, neutrino, graviton	0
etc.	

Digressing a bit, let's note that the availability in modern theory of two "natural" units (their dimensions are $[c]$ = length/time and $[\hbar]$ = energy × time) leads to the situation where, among the three basic units forming the basis of any system of units (for example, in SI, m, s, kg), only one unit (say, the length L) must be considered arbitrary. The unit of time can be defined as $T = L/c$; the unit of mass as $M = \hbar/Lc$; the unit of energy as $E = \hbar c/L$; and so on. In theoretical physics it is a common practice to accept $\hbar = c = 1$ and to measure all the physical parameters as powers of length. This trick greatly simplifies the formulas, from which the coefficients $\hbar$ and c have disappeared. In this system, momentum p, mass m, and energy E are expressed in the reciprocal units of length—say, cm^{-1}. The relativistic formulas for energy and momentum look like this:

$$E = \frac{m}{\sqrt{1-v^2}} = \sqrt{m^2 + p^2},$$

$$p = \frac{mv}{\sqrt{1-v^2}}.$$

The magnetic moment is expressed in units of length or in the reciprocal units of mass. For example, the magnetic moment of the electron (also known as the Bohr magneton—see below) is equal to $e/(2m)$. Other physical magnitudes can be expressed in a similar way. The application of this "one-dimensional" system of units is rather effective, provided the magnitudes for unit mass or length are characteristic of the phenomenon being investigated. Now let's return to Heisenberg's ideas.

When Heisenberg advanced his notions, the list of elementary particles included only the electron (and its antiparticle, the positron), proton, neutron, and photon. Nowadays this list is expanded to include dozens of particles. Among the added items are the μ-meson (muon) and two "sorts" of neutrinos, which together with the electron and the corresponding antiparticles form the family of weakly interacting particles, the leptons. In addition, a number of new, strongly interacting particles were discovered. These included particles with a very short lifetime; they were called resonance particles (for example, the η-meson in the list above). The strongly interacting particles (hadrons) are subdivided into two large groups: the so called baryons, which are similar in their properties to the proton and neutron (the long-lived baryons Λ, Σ, Ξ were named hyperons); and the mesons—typical examples are the π- and ρ-mesons, which are responsible for nuclear forces, as well as *K*- and η-mesons (which were also listed above).

Now we have no reason to suppose that the mass of any natural particle is necessarily of the order of $1/l_0 = 70$ MeV (taking $\hbar = c = 1$, we use 1 MeV as a unit not only of energy but of mass, momentum, and reciprocal length). For example, we have every reason to believe that there are particles (probably unstable) with much larger masses. So this "empirical" argument in favor of Heisenberg's numerical value for elementary length now seems not very convincing. The argument based on the classical estimate of electromagnetic mass is also not convincing due to the aforementioned decrease in this value in quantum theory. The last point strikes one as being particularly important.

Heisenberg supposed that there will be drastic deviations from modern theory in the laws of interaction of elementary particles at energies larger than $1/l_0 = 70$ MeV. Initially, when new particles were discovered in cosmic rays, which had great penetrating abilities, it was thought that they were electrons that possessed high energy and thus did not "obey" the laws of quantum electrodynamics. However, it soon became clear, that they were just ordinary particles whose mass was 200 times that of electrons, and this "trivial" property was responsible for their unique penetrating power. Nowadays there are no phenomena that could confidently be interpreted as an overt violation of modern theory. Let's examine this in more detail.

Modern physics knows of four kinds of interaction:

1. "Strong" interactions (the nuclear forces are a typical example);
2. Electromagnetic interactions;
3. "Weak" interactions (which are responsible for the processes of beta-decay);
4. Gravitational interactions.

A comprehensive quantitative theory and extensive experimental data exist for electromagnetic interactions, so this is a good field in which to look for possible deviations from modern theory. Until now all such attempts have produced negative results. I will describe some of them, because even a negative result is important for such an important problem—an analysis of experimental accuracy yields an estimate of the possible limit to the validity of modern views. It is also important that this question has various links with other fields of modern physics, so it is interesting in itself.

At present, among the electromagnetic properties of elementary particles, the magnetic moment is the most studied. According to a hypothesis advanced in 1925 by

Uhlenbeck and Goudsmit, the electron is similar to a tiny top—it has an angular moment equal to 1/2 (in units of $\hbar$) and also has a dipole magnetic moment $e/(2m)$. Much evidence was collected in favor of this hypothesis in the course of research on magnetic phenomena in spectroscopy. Later the outstanding English physicist Paul Dirac showed that Uhlenbeck and Goudsmit's hypothesis was compatible with the description of the electron as a point charged particle obeying the equations of quantum mechanics and relativity.

Surprisingly, in the 1930s it was found that the magnetic moment of the proton was 2.9 times larger than $e/(2m_p)$, where m_p is the proton's mass. In addition, the Russian theorists Tamm and Altshuler predicted, and the American scientist Luis Alvarez experimentally detected, the existence of magnetic moment of the neutron, which is electrically neutral and according to the formula above should not have any magnetic moment at all. Now the accepted practice is to call the magnitude $\mu_0 = e/(2m)$ the normal magnetic moment, and to treat any extra moment as "anomalous." According to modern views, the anomalous magnetic moment of the proton and neutron is caused by their inner structure, but a theory for this phenomenon is still lacking—as well as a theory for the strong interacting particles.

Until 1947 it was thought that the electron had no anomalous moment. However, a study of the interaction energy between the electron's magnetic moment and that of the proton resulted in some discrepancies. (By the way, this interaction is responsible for the electromagnetic waves (wavelength $\lambda = 21$ cm) radiated by atomic hydrogen in the cosmos, which play a very important role in radio astronomy.) The American theorist Gregory Breit proposed—and shortly thereafter his compatriots, the experimentalists Kusch and Foley—found a tiny anomalous magnetic moment in the electron. The relative value of this moment was about $1.2 \cdot 10^{-3}$. The theory of the anomalous magnetic moment had been created by the outstanding American experimentalist Julian Schwinger in 1948 as the result of the great advances in the mathematical apparatus of quantum electrodynamics made in 1940s by Schwinger and independently by Sinitiro Tomonaga, Hans Bethe, Hendrik Kramers, Richard Feynman, Freeman Dyson, and others.

According to Schwinger, the relative anomalous moment is given by the formula

$$a = \frac{\mu - \mu_0}{\mu_0} = \frac{e^2}{2\pi} = 1.16 \cdot 10^{-3}$$

and results from the interaction of an electron or μ-meson with the electromagnetic quantum fluctuations (or zero-point vibrations) of a vacuum.

In quantum theory, a vacuum is not synonymous with emptiness. For any system, this theory introduces the concept of energy levels (Bohr's hypothesis). Extrapolating this approach to the vacuum yields an interpretation of the photon as an excited state of one of the vacuum's electromagnetic oscillatory degrees of freedom. The basic state (level) of every degree of freedom corresponds to the absence of a photon with a given wavelength. Although the average value of the quantum-mechanical electric field in this system is zero at any moment of time, the field does exist, because its amplitude, which corresponds to the given degree of freedom, cannot be equal to zero and so undergoes quantum zero oscillations (quantum fluctuations), creating a "cloud of probability" near the average (equilibrium) state. The full energy of the interaction between a charged particle and the zero oscillations of the vacuum derives from interaction with zero oscillations of different wavelengths, and the change in this energy in the "external" magnetic field was interpreted by Schwinger as being caused by the anomalous magnetic moment.

The interaction energy of electrons with the vacuum's zero oscillations can be expressed by an integral taken over all possible values of the momentum p (reciprocal wavelengths) of these oscillations, where p_0 is the assumed limit to the applicability of current concepts. Thus the energy is proportional to

$$m_e \sim e^2 \int_0^{p_0} dp \frac{m}{\sqrt{p^2 + m^2}} \cong \int_m^{p_0} e^2 m \frac{dp}{p}$$
$$= me^2 \ln \frac{p_0}{m}.$$

For dimensional reasons, when a magnetic field of intensity H is present, the expression under the integral sign will change by a value proportional to e^3H/p^2. Therefore, the change in the electron energy in a magnetic field, which in accordance with Schwinger's idea we consider equal to $\mu - \mu_0$, is proportional to $me^3H \int_m^{p_0} \frac{dp}{p^3}$. Thus

$$\mu - \mu_0 \sim me^3 \left(\frac{1}{m^2} - \frac{1}{p_0^2} \right).$$

According to Schwinger, the proportionality factor in this formula is $1/4\pi$. Earlier we wrote this formula as

$$\mu - \mu_0 = \frac{e^2}{2\pi} \mu_0 = \frac{e^3}{4\pi m}$$

—that is, without the factor $(1 - m^2/p_0^2)$, which corresponds to $p_0 \to \infty$. When $p_0 \neq \infty$, we have corrections to the anomalous moment proportional to m^2/p_0^2. Denoting by a_t the theoretical value obtained by Schwinger and other theorists, which gave a more precise estimate within the framework of modern theory, we have (in order of magnitude)

$$\delta = \frac{a - a_t}{a_t} \cong \frac{m^2}{p_0^2},$$

or

$$p_0 = \frac{m}{\sqrt{\delta}}.$$

This formula shows that the most "promising" object for studying violations of quantum electrodynamics is the heaviest particle among the known ones—the μ-meson, whose anomalous moment fluctuates (this was noted by the Soviet physicist Berestetsky).

The first experiments that detected the anomalous moment of the electron were done by the molecular beam method. The credit for the development of this method, which dates back to the classical experiment of Stern and Gerlach, belongs mostly to the American physicist Isidor Rabi. However, the most precise measurements of a (with a relative accuracy $\delta = 2 \cdot 10^{-5}$ for the electron and $\delta = 4 \cdot 10^{-3}$ for the μ-meson) were done in a number of American laboratories much later and by a different approach. These experiments showed that $a = 1.162 \cdot 10^{-3} \pm 0.004 \cdot 10^{-3}$ (these data are for μ^+-mesons; similar results were obtained by Farley and Brown for μ^--mesons). With all known corrections, the theoretical value a_t equals $1.1654 \cdot 10^{-3}$—that is, it coincides with the experimental value to within the accuracy of the measurements. Therefore, the value $\delta = (a - a_t)/a_t$ defined above is certainly less than $4 \cdot 10^{-3}$. Thus quantum electrodynamics is undoubtedly correct for energies and momenta less than $p_0 = m/\sqrt{4 \cdot 10^{-3}}$ —that is, when these values are less than a few GeV.

Another method for studying the applicability of quantum electrodynamics is based on the collisions of electrons with electrons and electrons with positrons in so-called colliding beams. Why do we need colliding beams? The theory of relativity combines the vector of (kinetic) momentum $\mathbf{p}$ and the energy of the particle E into the so-called four-vector. Three-dimensional vectors have the property of preserving their scalar product during rotation of the three-dimensional coordinate axes. However, given a more general "Lorentzian transformation" of the reference system, which takes into account not only the rotation of the coordinate axes but also the shift into another inertial reference frame, a more general invariant emerges: the Einstein–Minkovski scalar product of four-vectors. For two colliding particles, the four-dimensional scalar product of the energy–momentum vectors is given by

$$I = E_1E_2 - p_{1x}p_{2x} - p_{1y}p_{2y} - p_{1z}p_{2z}.$$

Clearly all qualitative theoretical assertions, and in particular the deviations from modern theory, can depend only on an invariant value. When an electron at rest ($p_1 = 0$) collides with an electron having a momentum $\mathbf{p}_2 = \mathbf{p}$, we have

$$I_1 = m\sqrt{m^2 + p^2}.$$

On the other hand, for colliding beams of electrons having momenta $\mathbf{p}_1 = \mathbf{p}$ and $\mathbf{p}_2 = -\mathbf{p}$, the invariant is

$$I_2 = m^2 + 2p^2.$$

If $p = 10^3m$ (that is, has an energy of 500 MeV), then $I_2 = 2 \cdot 10^3 I_1$. The advantage of the colliding beam method is obvious when we compare I_1 and I_2.

Experiments with colliding beams have been conducted in Novosibirsk (Russia) under the guidance of Budker and seem quite promising. Within the accuracy of the measurements, these experiments showed no deviations from modern theory.

Thus the body of theoretical and experimental arguments forces us to admit that Heisenberg's theoretical limit $l_0 = r$ must be shifted to far greater energies. Although negative in character, this result is very important for the modern physics of elementary particles.

Long ago the American physicist Eugene Wigner noted that the very notion of measuring tiny intervals of time and space ($\Delta x \le L_0 = 10^{-33}$ cm; $\Delta t \le L_0/c = 10^{-44}$ s) encounters difficulties in principle, if one simultaneously takes quantum mechanical effects and gravitation into account. The distance and time interval between any two points in Einstein–Minkowski space (that is, between two "events") must be subjected to quantum fluctuations, or to zero quantum oscillations, just as any other physical value. In this respect the gravitational field cannot differ qualitatively from the electromagnetic or any other field. Note that L_0 can be estimated by dimensional analysis.[4] In his time Max Planck noted that, using the numerical value of the gravitational constant $G = 6.67 \cdot 10^{-8}$ CGS, as well as the constants $\hbar$ and c, one can construct a system of "natural" units for all physical magnitudes (in other words, substitute the "one-dimensional" system described above for the "zero-dimensional" system of units). For instance, the unit of length L_0 can be defined as

$$L_0 = G^{1/2}\hbar^{1/2}c^{-1/2} = 1.61 \cdot 10^{-33} \text{ cm}.$$

Correspondingly, the unit of time will be

$$T_0 = \frac{L_0}{c} = G^{1/2}\hbar^{1/2}c^{-3/2} = 5.35 \cdot 10^{-44} \text{ s},$$

that of energy

$$E_0 = \frac{\hbar}{T_0} = G^{-1/2}\hbar^{1/2}c^{3/2} = 2 \cdot 10^{16} \text{ erg} = 10^{28} \text{ eV},$$

and that of mass

$$M_0 = \frac{E_0}{c^2} = G^{-1/2}\hbar^{1/2}c^{-1/2} = 2.18 \cdot 10^{-5} \text{ g}.$$

It turns out that Wigner's work (mentioned above) leads to the extraction of these two magnitudes, L_0 and T_0, as the boundaries of our modern view of the nature of time and space. Some scientists (including the Russian theorist Kompaneyets) have pointed out that using L_0 as the effective radius of the electron does not lead to an excessively large electromagnetic mass in quantum electrodynamics, contrary to what would happen in classical electro-

[4]For a primer on dimensional analysis, see "The Power of Dimensional Thinking" by Yuly Bruk and Albert Stasenko (*Quantum*, May/June 1992).—*Ed.*

dynamics. The reason is that the aforementioned fact that electromagnetic mass in quantum electrodynamics is proportional to $\ln(r^{-1})$. Recently the Russian scientist Markov offered the hypothesis that L_0 (and thus the related magnitude $M_0 = 1/L_0$) determines also the maximum possible mass of an elementary particle. He coined the term "maximon" for such a particle. It is known that the formation of stable particles from component parts that can themselves be unstable results in a decrease in the total mass (the mass "defect" that arises in nuclear physics as a small correction to Prout's law). So, following Markov, we should not be surprised that the observable stable particles (electrons, protons, and so on) have masses that are far less than the "natural" unit of mass $M_0 = 2 \cdot 10^{-5}$ g.

More and more physicists now think that this boundary L_0 will lead to the most drastic changes in our views. Still it is very important to be sure that there are no other characteristic values between $r = 2.8 \cdot 10^{-13}$ cm and $L_0 = 1.61 \cdot 10^{-33}$ cm that could play a similar fundamental role. At presently this question can be answered only by indirect theoretical considerations. Here is one of the arguments, which is taken from an analysis of the principles of the general theory of relativity.

It is known that the motion of material bodies in a gravitational field is described in Einstein's theory as movement along the shortest line in "curved" space–time. Because of this "curvature," the shortest distance is not a "straight" line but a "curved" line in space–time described by the set of equations

$$x = f_1(t),\ y = f_2(t),\ z = f_3(t),$$

where f_1, f_2, f_3 are nonlinear functions.

In Einstein's theory, the degree of the curvature of space is found by a condition that can be qualitatively described in the following way. In the vicinity of bodies possessing mass (or energy, which is the same thing), space is affected by a curving "force" (of course, the term "force" is used here in a certain generalized sense). At the same time, space has the property of "elasticity" that "works against" the curving force. The balance of these two "forces" determines the degree of curvature. Usually the deviations of the properties of space from the properties described by Euclidean geometry are rather small—that is to say, the "elasticity" of space is very large.

What determines the "elasticity" of a vacuum? We might suppose that it is variations in the quantum fluctuations of the vacuum. Earlier we discussed these fluctuations in connection with Schwinger's theory of the anomalous magnetic moment. Here, we might say that when space–time is being "curved," the fluctuations become "cramped" and that they "violate" the boundaries, which results in an increase in the vacuum's energy. In a formal sense, this effect is infinite if fluctuations of the shortest wavelength are taken into account. The value of the gravitational constant (the reciprocal of the "coefficient of elasticity of space") will have the correct numerical value only if the fluctuations have a wavelength greater than $L_0 \sim 10^{-33}$ cm. The future will say whether this reasoning is correct.

So—what is there beyond the limit set by L_0? What modifications to quantum theory (if any) will be necessary for processes occurring at distances less than 10^{-33} cm, or characterized by energies larger than 10^{28} eV? Nobody knows. We should probably agree with those scientists who expect profound, fundamental changes in the way we think about physics. The value 10^{28} eV is so much higher than the range of energy currently studied (the largest Russian ring accelerator, at Serpukhov, has an energy of "only" $7 \cdot 10^{10}$ eV) that the final clarification of this set of problems may remain out of reach for the near future.

Challenges in physics and math

Math

M201

Mind your minuses. Find the unique real root of the equation

$$x^3 - 3x^2 - 3x - 1 = 0.$$

M202

Squared numerators. Prove that if

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} = 1,$$

then

$$\frac{a^2}{b+c} + \frac{b^2}{c+a} + \frac{c^2}{a+b} = 0.$$

M203

Hens and roosters. One day Mrs. Cook bought a hen at the market. This hen laid two eggs and then was cooked for a dinner. As is well known, from each egg either a hen or a rooster can hatch. Each rooster was eaten soon after it hatched, and a hen was eaten only after it had laid two eggs. This process went on for several years, until it ended in a natural way: only roosters remained, and they were eaten. It turned out that the total number of roosters eaten was 1997. What was the total number of hens eaten? (A. Yegorov)

M204

Middle of all chords. Find the location of the midpoints of all the chords drawn in a given circle so that their endpoints lie on different sides of a given straight line intersecting this circle.(I. Sharygin)

M205

Finding the center. A circle on the plane is given. Its center is not shown. Find the center using a compass (without a straightedge) in such a way that the total number of arcs or circles you draw does not exceed six. (V. Panfyorov)

Physics

P201

System in equilibrium. A massless inelastic cord with masses of 1 kg and 3kg attached to its ends is strung over a light pulley. This pulley is set on a shaft with friction, and the force of friction is proportional to the axial load. In this system the acceleration of the larger mass was 2 m/s^2. What is the mass that must be added to the smaller mass to place the system in equilibrium? (S. Varlamov)

P202

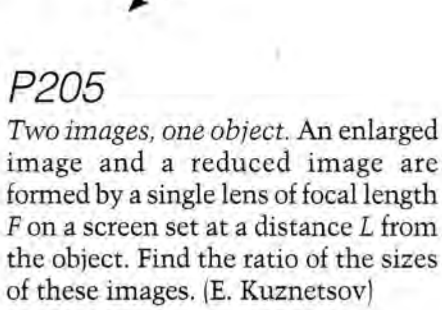
Pucks on ice. A puck of mass M slides on ice with a velocity v_0 and strikes a puck of mass $2M$ at rest. After the impact the first puck stops. The second puck hits a wall and after an elastic rebound hits the first puck head on. Find the velocities of both pucks after the last collision. Note that during a collision a certain fraction of the maximum energy of deformation is transformed into heat. (A. Vargin)

P203

Gas under a piston. One mole of ideal monatomic gas is put under a massive piston in a vertical thermally insulated cylinder at a temperature T_0. The gas is compressed by lowering the piston. After performing work W, the piston is released and assumes a new equilibrium position. Find the temperature T_x for this state. (V. Uzdin)

P204

Collision course. Four military jets are conducting maneuvers. Three of them (1, 2, and 3) follow each other as shown in the figure below; the fourth one (4) flies in a direction at an angle $\theta = 30°$ with the course of planes 1, 2, and 3. None of the speeds designated on the diagram is known. It is known, however, that the pilots in planes 1, 2, and 3 begin to hear plane 4 at the same moment. It is also known that the pilot in plane 1 begins to hear plane 4 when the distance between the two is three times the minimal distance achieved during the approach of plane 4. Is plane 1 flying at supersonic speed? (B. Korsunsky)



P205

Two images, one object. An enlarged image and a reduced image are formed by a single lens of focal length F on a screen set at a distance L from the object. Find the ratio of the sizes of these images. (E. Kuznetsov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 58

Irrationality and irreducibility

It all started with the square root of 2 . . .

by V. A. Oleynikov

THE ANCIENT GREEKS KNEW that

$\boxed{\text{the quantity } \sqrt{2} \text{ is irrational}}$

and could prove it.

In fact, suppose that it's rational. Then we can write it down as an irreducible fraction $\sqrt{2} = a/b$. But this means that

$$2 = \frac{a^2}{b^2}, \quad 2b^2 = a^2.$$

Now it's clear that a is divisible by 2, and a^2 is divisible by 4, so b also is divisible by 2. Thus the fraction a/b is reducible.

The fraction a/b cannot be reducible and irreducible at the same time, so $\sqrt{2}$ is not a fraction. This property made $\sqrt{2}$ an unwelcome guest in the world of numbers, which was ruled by harmony and order, simplicity and perfection.

The number $\sqrt{2}$ owes its existence to the diagonal of a unit square and numerous unsuccessful attempts to measure it by means of rational line segments. These failures troubled the ancient Greeks very much and caused a good deal of intellectual ferment. Reality, embodied in the geometrical figure, was knocking on the door of the beautiful and sublime.

Centuries later, $\sqrt{2}$ claimed its right to be considered a "number" by becoming another reality: the root of the equation $x^2 - 2 = 0$, and it turned out that

$\boxed{\text{the number } \sqrt{2} \text{ is irrational,}}$

because

$\boxed{\text{the polynomial } x^2 - 2 = 0 \text{ is irreducible.}}$

Our aim is to clarify the connection between irrationality and irreducibility.

Irreducibility

An nth-degree polynomial in x

$$P(x) = a_0 + a_1x + \ldots + a_nx^n$$

with integer coefficients

$$a_0, a_1, \ldots, a_n, a_n \neq 0$$

is called *irreducible* if there are no polynomials $L(x)$ and $Q(x)$ with integer coefficients such that

$$P(x) = L(x) \cdot Q(x)$$

and with degrees less than n. Otherwise, $P(x)$ is called *reducible*.

For example, the polynomial $x^3 + x^2 + x + 1$ is reducible (it's equal to $(x + 1)(x^2 + 1)$—check it!), and the polynomial $x^2 + 1$ is irreducible (think why).

The linear binomial $a_0 + a_1x$ is the simplest example of an irreducible polynomial. Its only root $x = -a_0/a_1$ is a *rational* number. However,

$\boxed{\text{an irreducible } n\text{th-degree polynomial, } n \geq 2, \text{ has no rational roots.}}$

This statement is follows from the following more general fact:

THEOREM. *If α is a common root of two polynomials $P(x)$ and $Q(x)$, and one of them—for instance, $Q(x)$—is irreducible, then the polynomial $d \cdot P(x)$, for some integer d, is divisible by $Q(x)$:*

$$d \cdot P(x) = L(x) \cdot Q(x).$$

This theorem bears the name of the famous German mathematician Carl Friedrich Gauss (1777–1855).

It will be proved below. And the property mentioned above is directly implied by it, since no polynomial whose degree n is greater than one can divide a linear binomial evenly.

The possibility of producing many new irrational numbers is now open to us. We can look for them among the roots of irreducible polynomials. The following statement ushers us into the mysterious land of irreducible polynomials:

> EISENSTEIN'S CRITERION. *Suppose that for the given polynomial $P(x)$ there exists a prime p such that the leading coefficient a_n of this polynomial is not divisible by p, all the other coefficients a_k are divisible by p, $k = 0, 1, \ldots, n - 1$, and the constant term a_0 is not divisible by p^2. Then the polynomial $P(x)$ is irreducible.*

During his short life, the German mathematician F. G. M. Eisenstein (1823–1852) suffered much from his own bad luck and his contemporaries' indifference to his work. His ideas were understood only many years after his death.

Proof. Suppose that, on the contrary, there exists a reducible polynomial $P(x)$ with the given properties of its coefficients. It can be represented as the product

$$P(x) = L(x) \cdot Q(x)$$

of the polynomials

$$L(x) = b_0 + b_1x + \ldots + b_lx^l,$$
$$Q(x) = c_0 + c_1x + \ldots + c_mx^m$$

with integer coefficients. The leading coefficients b_l and c_m are nonzero, and we can suppose that $m \geq l \geq 1$. If we add up the coefficients of the same powers of x in this product and compare the results with the coefficients of $P(x)$, we get

$$\begin{aligned} a_0 &= b_0c_0, \\ a_1 &= b_0c_1 + b_1c_0, \\ a_2 &= b_0c_2 + b_1c_1 + b_2c_0, \\ &\vdots \\ a_l &= b_0c_l + b_1c_{l-1} + \cdots + b_lc_0, \\ a_m &= b_0c_m + b_1c_{m-1} + \cdots + b_mc_0, \\ &\vdots \\ a_n &= b_lc_m. \end{aligned}$$

Consider the first of these equalities. We know that the constant term a_0 is divisible by p; this means that either b_0 or c_0 is divisible by p; but it is impossible that both these terms are divisible by p, since a_0 is not divisible by p^2.

Let's assume that b_0 is divisible by p and that c_0 is not. Then we proceed to the second equality: a_1 is divisible by p and b_0c_1 is divisible by p; thus b_1c_0 also is divisible by p. Therefore, b_1 is divisible by p . . .

. . . And so we go on reasoning in this way until we come to the $(l + 1)$st equality (involving the coefficient a_l): a_l is divisible by p, and all of $b_0, \ldots, b_{l-1}$, are divisible by p as well. Therefore, b_lc_0 and hence b_l is divisible by p.

Now we jump directly to the last equality: we conclude that the leading coefficient $a_n = b_lc_m$ is divisible by p, which contradicts the condition of the criterion.

If we assume that in the first equality, c_0 and not b_0 is divisible by p, we'll have to go back to the very beginning, proceed in the same way to the $(m + 1)$st equality (involving the coefficient a_m), and then jump to the last equality.

So we see that the decomposition $P(x) = L(x) \cdot Q(x)$ is impossible, and thus $P(x)$ is an irreducible polynomial. Having established this criterion, we proceed to the next stage.

Irrational radicals

The polynomial $x^2 - 2$ is irreducible, by Eisenstein's criterion (take $p = 2$). Together with $\sqrt{2}$ we obtain the irrational numbers

$$\sqrt[n]{p},$$

where p is an arbitrary prime and $n = 2, 3, \ldots$. All these numbers are roots of the polynomials

$$P(x) = x^n - p,$$

which are irreducible, according to Eisenstein's criterion. The number

$$\sqrt[n]{p_1 \ldots p_k}$$

is irrational if $p_1, \ldots, p_k$ are different primes. This number is a root of the irreducible polynomial

$$P(x) = x^n - p_1 \ldots p_k.$$

To these irrationalities, we can add

$$\sqrt[l]{a + \sqrt[m]{b + \ldots \sqrt[n]{p_1 \cdot \ldots \cdot p_k}}}$$

for all natural $a, b, \ldots$ (try to construct polynomials for these monsters yourself and prove that they are irreducible).

All these examples illustrate how Eisenstein's criterion works. But still, one can't say that it's a big step beyond what the ancient Greeks knew. As a matter of fact, it's possible to prove the irrationality of the last expression, raising it successively to the powers $l, m, \ldots, n$ and then reasoning in the way we did to prove the irrationality of $\sqrt{2}$. The following sum of radicals seems more impressive:

$$\frac{a_1}{b_1}\sqrt[n_1]{p^{m_1}} + \ldots + \frac{a_k}{b_k}\sqrt[n_k]{p^{m_k}}.$$

If all the quotients

$$\frac{m_1}{n_1}, \ldots, \frac{m_k}{n_k}$$

are in lowest terms and different from one another, then the sum is an irrational number. To prove this, let's suppose that it is rational and equal to a/b. Put $N = n_1 \ldots n_k$. Then $\sqrt[N]{p}$ is a root of the polynomial with integer coefficients

$$\frac{a_1B}{b_1}x^{\frac{m_1N}{n_1}} + \ldots + \frac{a_kB}{b_k}x^{\frac{m_kN}{n_k}} - \frac{aB}{b},$$

where $B = b \cdot b_1 \ldots b_k$). The degree of this polynomial is less than N. But, according to our theorem it must be divisible by the irreducible Nth-degree polynomial $x^N - p$, which is impossible.

This sort of activity could be continued successfully (for example, we can combine the two last results to obtain new irrationalities). Success inspires hope, but sometimes it creates illusions. It might seem that if we pile up more and more new radicals and apply the four arithmetic operations to the whole numbers $a, b, \ldots$, we'll get more and more new irrational numbers. The best way to dispel illusions in mathematics is to consider a "special case." One of such special case is the question whether the expression $\sqrt[n]{a^n + b^n}$ is rational for natural

$a, b, (n \geq 3)$. This question is equivalent to Fermat's last theorem:

There are no natural numbers x, y, and z such that $x + y = z$, in which n is a natural number greater than 2.

The great 17th-century French mathematician Pierre de Fermat proposed this problem and left it to us unsolved. Since that time, for over three hundred years, the best (and the worst) mathematicians have tried to solve it.[1] Seemingly simple, this problem has attracted the attention of numerous amateur mathematicians, and many a naive, uneducated soul was swallowed up in this swamp.

In 1908 the German millionaire P. Wolfskehl offered a large monetary award to whoever solved the Fermat's last theorem, thus provoking a growing avalanche of erroneous "solutions." Efforts continued unabated despite the devaluation of the deutsch mark in the 1930s, which lessened the award considerably.

The upshot of these remarks is clear: the mass production of irrational radicals is an apparently promising, but devilishly dangerous occupation.

Variations

Eisenstein's criterion is often insufficient to tell whether the polynomial $P(x)$ is irreducible, because it demands that all the coefficients a_0, a_1, ..., a_{n+1}, except for the leading term a_n, have a common prime divisor p. There are many polynomials—for example,

$$x^2 + 1,\ x^4 + 1,\ x^6 + x^3 + 1$$

—for which no such divisor exists. Nevertheless, the resources of Eisenstein's criterion are far from being exhausted. One merely has to "shake up" the polynomial $P(x)$ by making some substitutions for the variable x.

The question of the irreducibility of the polynomial

$$P(x) = x^2 + 1$$

can be reduced to Eisenstein's criterion in the following way. Suppose that $P(x)$ is reducible. Then $P(x + 1)$ must be reducible, too. But $P(x + 1) = x^2 + 2x + 2$ is irreducible according to Eisenstein's criterion for $p = 2$. The polynomials $x^4 + 1$ and $x^6 + x^3 + 1$ are irreducible for the same reason. A similar transformation allows use to say whether the polynomial

$$P(x) = x^{n-1} + x^{n-2} + \ldots + 1$$

—called the *cyclotomic polynomial*[2]—is reducible.

The polynomial $P(x)$ is reducible if $n = p \cdot k$ is a composite number, because

$$\begin{aligned} P(x) &= \frac{x^n - 1}{x - 1} = \frac{\left(x^p\right)^k - 1}{x - 1} \\ &= \frac{\left(x^p - 1\right)\left(x^{p(k-1)} + x^{p(k-2)} + \ldots + 1\right)}{x - 1} \\ &= \left(x^{p-1} + x^{p-2} + \ldots + 1\right) \\ &\qquad \cdot \left(x^{p(k-1)} + x^{p(k-2)} + \ldots + 1\right), \end{aligned}$$

and it is irreducible if n = p is a prime.

In fact, if $P(x)$ were reducible, then $P(x + 1)$ would be reducible, too. But

$$\begin{aligned} P(x+1) &= \frac{(x+1)^p - 1}{(x+1) - 1} \\ &= x^{p-1} + C_p^1 x^{p-2} + \ldots + C_p^{p-1}. \end{aligned}$$

All the coefficients here[3] are divisible by p, since

$$C_p^k = \frac{p(p-1)\ldots(p-k+1)}{k!} \quad (k < p),$$

and the numerator is divisible by p, while the denominator is not. Moreover, the constant term $C_p^{p-1} = p$ is not divisible by p^2. According to Eisenstein's criterion, $P(x + 1)$ is irreducible, and thus $P(x)$ is irreducible as well.

From this result it follows that the cyclotomic polynomial

$$x^{p-1} + x^{p-2} + \ldots + 1 = 0$$

cannot have any rational root for $p \geq 3$.

It's possible to prove that the cyclotomic polynomial has no real roots either—that all its roots are complex numbers. We can then pose a more difficult question: when are these roots "not too irrational"—that is, when can they be expressed by *quadratic* radicals (obtained by applying to integers the four arithmetic operations and the operation of taking the *square* root)? This question interested the ancient Greek mathematicians, since it's equivalent to the following construction problem: for which p it is possible to construct a regular p-gon with a compass and straightedge?

The young Carl Friedrich Gauss took up this challenge and proved that the roots of the cyclotomic polynomial can be expressed by quadratic radicals if (and only if[4]) p is a *Fermat prime:* $p = 3, 5, 17, 257, \ldots, 2^{2k} + 1, \ldots$. Gauss was very proud of his discovery and even "expressed the desire that a regular 17-gon be engraved on his tombstone."

Here is another, more modest achievement of the great Gauss.

Lemma on primitive polynomials

This lemma will help us prove the theorem formulated above. We call a polynomial

$$P(x) = a_0 + a_1 x + .. + a_n x^n$$

primitive if its coefficients $a_0, a_1, \ldots, a_n$ have no common prime divisors.

Gauss's lemma says that

> *the product of two primitive polynomials is also a primitive polynomial.*

[1]In 1995 the English mathematician Andrew Wiles published a proof of Fermat's last theorem that brings together several disparate areas of modern mathematics.—*Ed.*

[2]The roots of this polynomial are the nth roots of 1 (except for 1 itself), all of which are complex, unless n is even (when one of them is equal to -1). These numbers, on the complex plane, form the vertices of a regular n-gon inscribed in a unit circle, and so divide this circle into n equal arcs. The term "cyclotomic" means "circle splitting" and is used for this reason.

[3]We use the binomial coefficients C_k^i and Newton's binomial theorem (see any standard precalculus text).

[4]This part of the statement was proved by Pierre Laurent Wantzel (1814–1848), a tutor at the Ecole Polytechnique in Paris.

This lemma is used in many theorems of algebra and number theory. If you have followed the proof of Eisenstein's criterion above, you will have no difficulty understanding the proof we give below, because they have a lot in common.

Suppose the opposite is true: there exist two primitive polynomials

$$L(x) = b_0 + b_1x + \ldots + b_lx^l,$$
$$Q(x) = c_0 + c_1x + \ldots + c_mx^m$$

whose product is the polynomial

$$P(x) = a_0 + a_1x + \ldots + a_nx^n$$

that is not primitive. Let p be one of the common divisors of its coefficients. Let b_i be the lowest term of the polynomial $L(x)$ that is not divisible by p, $0 \le i \le l$, and c_j the lowest term of the polynomial $Q(x)$ that is not divisible by p, $0 \le j \le m$. Such coefficients must exist; otherwise the polynomials $L(x)$ and $Q(x)$ would not be primitive. The coefficient of x^{i+j} in the product

$$L(x)Q(x)$$

is equal to

$$\ldots + b_{i-1}c_{j+1} + \ldots + b_{i+1}c_{j-1} + \ldots$$

and is not divisible by p, since all the terms in this sum except b_ic_j are divisible by p. On the other hand, this sum equals a_{i+j}—the coefficient of $P(x)$, and thus must be divisible by p. This contradiction proves the lemma.

And now the end is at hand.

Proof of the theorem

Now we'll prove the theorem offered at the outset: if a polynomial $P(x)$ has common root α with an irreducible polynomial $Q(x)$, then $Q(x)$ divides into $P(x)$ multiplied by some integer $d \ne 0$.

Let's divide the polynomial $P(x)$ by $Q(x)$ "with a remainder":

$$\begin{array}{r|l} & \frac{a_n}{c_m}x^{n-m} + \ldots \\ \hline c_mx^m + \ldots + c_0 & a_nx^n + \ldots + a_0 \\ & a_nx^n + \ldots \\ & \vdots \\ & r_{m-1}x^{m-1} + \ldots. \end{array}$$

We obtain the following equality:

$$P(x) = L_1(x) \cdot Q(x) + R_1(x).$$

The quotient $L_1(x) = (a_n/c_m)x^{n-m} + \ldots$ and the remainder $R_1(x) = r_{m-1}x^{m-1} + \ldots$ will be polynomial functions of x with rational coefficients.

If the remainder of this division were equal to zero (that is, if $P(x)$ were divisible by $Q(x)$), everything would already be clear, because then

$$P(x) = L_1(x) \cdot Q(x) = \frac{1}{d}L(x) \cdot Q(x),$$

where d is the common denominator of the coefficients of $L_1(x)$.

If $R_1(x) \equiv 0$, then its degree is less than that of the divisor $Q(x)$, and the number α, which is the common root of $P(x)$ and $Q(x)$, is a root of the polynomial $R_1(x)$ as well:

$$R_1(\alpha) = P(\alpha) - L_1(\alpha) \cdot Q(\alpha) = 0.$$

Dividing $Q(x)$ by $R_1(x)$, we obtain a new remainder $R_2(x)$ with the same property. Its degree is less than the degree of $R_1(x)$ and so less than the degree of $Q(x)$.

Dividing $Q(x)$ successively by the new remainders $R_1(x), R_2(x), \ldots, R_k(x)$, either we'll arrive at the contradiction

$$R_k(x) \equiv c \ne 0$$

telling us that α is not a root of $R_k(x)$, or we'll find a polynomial $R_k(x)$ that divides into $Q(x)$ without a remainder: $Q(x) = L_k(x) \cdot R_k(x)$. Bringing the rational coefficients of $L_k(x)$ and $R_k(x)$ to a common denominator and removing the common divisors in their numerators, we can rewrite this equality as

$$Q(x) = \frac{a}{b}\left[L(x)R(x)\right],$$

where a/b is an irreducible fraction and $L(x)$, $R(x)$ are irreducible polynomials.

Now it's sufficient to show that the coefficient a/b in the last product is an integer—that is, that $b = \pm1$. Suppose this isn't true. Let p be a prime divisor of the number b. Then it follows from the equality

$$bQ(x) = a[L(x)R(x)]$$

that all the coefficients on the right side are divisible by p; a is not divisible by p, because a/b is an irreducible fraction. Thus all the coefficients of the polynomial $L(x)R(x)$ are divisible by p. But this is impossible, according to the Gauss's lemma.

Therefore, the polynomial

$$Q(x) = [\pm aL(x)]R(x)$$

is reducible, which contradicts the original condition, thus proving the theorem completely.

The ancient Greeks are long gone. The confusion of the Dark Ages has passed. The 19th century—the true "classical age" of mathematics—has slipped away as well. What will be the lasting mathematical monuments of our own century?

A clock wound for all time

The Earth is a timepiece that can measure its own age—almost!

by V. I. Kuznetsov

SEVERAL YEARS AGO, THE *Quantum* article "Physics Fights Frauds" (January/February 1993) described how scientists use carbon dating to determine when very old objects were made. This technique involves measuring the concentration of the radioactive carbon isotope ^{14}C. However, the resolution of this method is limited: the concentration of the ^{14}C isotope is too small in objects whose age is more than 50,000–70,000 years. Sometimes we're interested in events that occurred much earlier. After all, *Homo sapiens* has been around for some 500,000 years, and organic life emerged on Earth more than a billion years ago.

The skeletons and the imprints left by ancient living things in rock can say much about the evolution of life on Earth. The placement of the geological strata can tell us the relative dates of certain events. However, it's much more difficult to find the absolute date, although sometimes it's possible to come up with estimates based on the thickness of sedimentary rocks. But how do we determine the age of the rocks and of the Earth itself?

The first estimates of the Earth's age were made on the assumption that the temperature of the Earth when it was formed was the same as that of the Sun today—about 6,000 K. In this approach, the Earth's age is taken to be equal to the period necessary for the Earth to cool and for a stable crust to form, plus the age of the crust itself. Tackling the problem of the Earth's age, the great English physicist Lord Kelvin assumed that initially the Earth had the temperature of molten rock, and that over time it gradually became cool while radiating heat from the surface into space. Based on his calculations, Lord Kelvin concluded that no more than 100 million years was required for the Earth's surface to cool to the point where it became suitable for plants and animals.

Lord Kelvin made his estimates before the discovery of radioactivity, so he didn't take into account the extra heat released inside the Earth by nuclear reactions. Later scientists came to the conclusion that this "radioactive" heat slowed the rate of cooling significantly. Thus the estimate of the time needed for the crust to form was gradually increased—first to 200 million years and then to 1.5 billion years.

If we know the age of the Earth's crust and add it to the 1.5 billion years it took the Earth's crust to form, the resulting sum will be the Earth's age. But how do we determine the span of time that elapsed from the moment the oldest minerals formed to the present, which is equal to the age of the planet's crust?

For this we need a clock that counts off hundreds of millions and billions of years. This clock must measure "accumulated" time. An example of such a device is the clepsydra—a water clock that counts time by the amount of water dripping from one vessel to another. "I still have much water left"—thus spoke a defendant in a Roman court of law, indicating that he had enough time to make his case. Another example is the hourglass, where sand tracks the passage of time.

Like the clepsydra or the hourglass, the radioactive clock also measures accumulated time. The more atoms of the radioactive isotope that have decayed, the older the mineral. The half-life $T_{1/2}$ of a substance that is suitable for measuring periods of the order of billions of years must also be of this order of magnitude. Only under this condition will a significant number of the

Art by Yury Vashchenko


92
Pa
$^{238}_{92}U \rightarrow 8\,^{4}_{2}He + ^{206}_{82}Pb + 6\,^{0}_{-1}$

radioactive atoms in a mineral be preserved to the present day.

The first type of radioactive decay used to find the age of minerals involved the transformation of uranium isotopes into lead. Natural uranium consists of three kinds of atoms having different masses: the isotopes ^{238}U, ^{235}U, and ^{234}U. The most prevalent isotope of uranium is ^{238}U. It constitutes 99.3% of uranium ore taken from any location.

Let's follow the steps in the decay of the isotope ^{238}U (fig. 1). The atoms of $^{238}_{92}U$ slowly turn into atoms of lead $^{206}_{82}Pb$ and helium $^{4}_{2}He$. Four and a half billion years must pass before half of the initial amount of ^{238}U turns into lead and helium. This process includes 14 radioactive transitions, where each disintegrating uranium atom yields one atom of stable lead and also eight atoms of helium:

$$^{238}_{92}U \rightarrow 8\,^{4}_{2}He + ^{206}_{82}Pb + 6\,^{0}_{-1}e.$$

It's possible to find the age of a piece of dense rock—say, granite—by grinding it in a mill, dissolving the powder in an acid, extracting the lead and uranium, and then determining how many atoms of the isotope ^{206}Pb the mixture contains per atom of ^{238}U. How does that give us the rock's age? Here's how.

Let's denote by N the number of uranium atoms and by N_1 the number of lead atoms ^{206}Pb that were in the piece of rock at the moment of analysis. Then $N_0 = N + N_1$ is the number of uranium atoms at the moment the granite was formed. Here we assume that all the lead atoms in the granite were formed by radioactive decay. According to the law of radioactive decay,

$$\frac{N}{N_0} = \frac{N}{N+N_1} = \left(\frac{1}{2}\right)^{\frac{t}{T_{1/2}}},$$

where t is the age of the granite. Taking the logarithms of both sides of this equation yields a formula for calculating the age of the mineral:

$$t = \frac{T_{1/2}}{\ln 2}\ln\frac{N_0}{N}. \qquad (1)$$

In these calculations we assumed that the sample has no extra lead and does not exchange matter with the surroundings. However, this is not universally true. Indeed, there is an element, radon, that is formed during the decay of uranium. This element belongs to the group of noble gases, and if the rock sample is not dense enough, some of the radon can escape. In this case the number of lead atoms in the sample will be less than that of the uranium atoms that decayed during the time the sample has existed. Under this condition, the age of the rock calculated according to formula (1) will be less than its true age.

On the other hand, the rock might contain lead that was formed simultaneously with the other elements and could have entered the mineral as the Earth's crust was forming. This is called primary lead. The presence of primary lead produces a value for the rock's age that is higher than it should be.

To determine the age correctly, we must be sure that radon was not released from the mineral, and we need to know how to estimate the proportion of primary lead in the sample.

Lead of radioactive origin (radiogenic lead) accumulates not only as the result of the decay of ^{238}U but also due to the radioactive decay of ^{235}U and ^{232}Th. The isotope ^{206}Pb is a descendant of ^{238}U, while the isotopes ^{207}Pb and ^{208}Pb are the final products of the decay of ^{235}U and ^{232}Th, respectively.

Natural lead also contains the light isotope ^{204}Pb, which does not accumulate during the radioactive decay of any natural radioactive element. So where does the "light" lead come from? Only one answer is possible: lead of mass number 204 was formed simultaneously with the Earth's other elements.

The isotopic composition of lead that has no radiogenic admixtures was determined by analyzing iron meteorites. These meteorites have no uranium or thorium, which are the sources of radiogenic lead. The lead in the meteors contains one part ^{204}Pb, 10 parts ^{206}Pb, 10 parts ^{207}Pb, and 29 parts ^{208}Pb.

This isotopic analysis makes it possible to estimate the amount of primary lead from the quantity of ^{204}Pb in a rock. So, if the lead doesn't contain ^{204}Pb or has only a small amount, practically all the lead is radiogenic.

The first measurements of the age

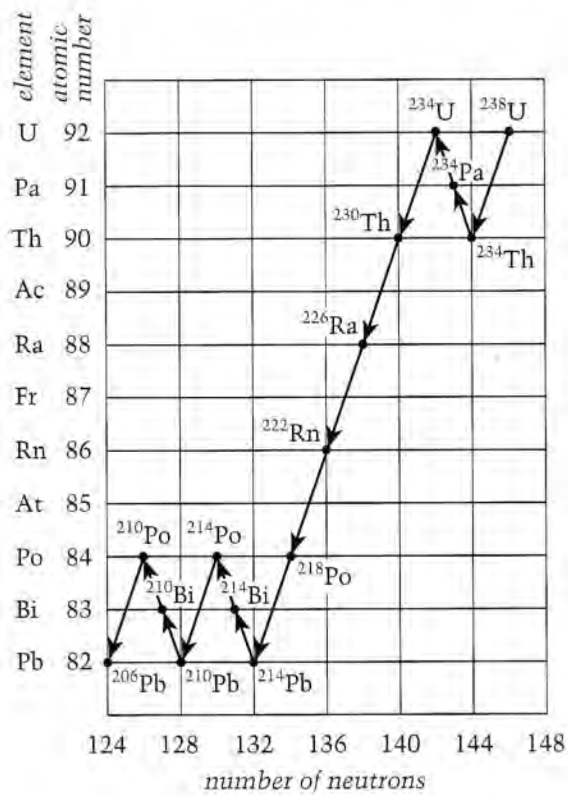


Figure 1

Radioactive decay of uranium. The number of neutrons in a nucleus is plotted on the abscissa and the numbers of protons (the atomic number of an element) along the ordinate. The arrows directed downward mark the process of alpha-decay. In this process the atomic nucleus emits a fast helium nucleus (alpha-particle) thereby losing a pair of protons and two neutrons. The upward arrows designate beta-decay. Here the neutron from the nucleus turns into a proton, resulting in an increase in the atomic number by one.

of minerals and of the Earth itself resulted in far lower values than those accepted by modern science. Perhaps a mistake crept in due to the diffusion of radon. Later, more precise data were obtained not only due to improvements in the lead–uranium method, but also as a result of the development of other techniques. The data are considered reliable when the results obtained by different methods coincide. Fortunately, the uranium isotopes are not the only set of nuclei whose half-life is the same duration as the geologic eras. The table below shows other radioactive isotopes used to verify the age of a mineral.

Natural potassium contains a small amount of the radioactive isotope ^{40}K, whose half-life is 1.3 billion years. Usually the nucleus of ^{40}K emits an electron and turns into calcium.

It's impossible to distinguish between the radiogenic and primary calcium that has accumulated in rocks. However, only 89% of the isotope ^{40}K disintegrates in this way, while 11% of it disintegrates by another way—by *K*-capture, as the physicists say. *K*-capture is a process by which the atomic nucleus captures an orbital electron and becomes the nucleus of an element whose atomic number is less by one. Thus $^{40}_{19}K$ turns into an isotope of argon, $^{40}_{18}Ar$ (fig. 2).

Studies of minerals have shown that in some rocks—for example, in mica—argon has been trapped without leaking out for billions of years. While the Earth's crust was forming, primary argon "boiled out" into the atmosphere, so the minerals must contain only radiogenic argon. Thus

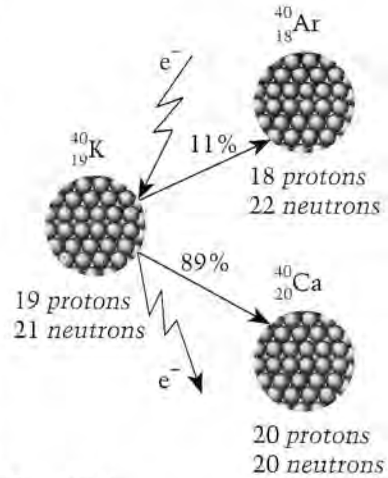


Figure 2

Radioactive decay of the isotope ^{40}K.

by extracting argon from the mica in granite, we can find the age of the granite by potassium decay, and then check the result by the lead–uranium method.

We can determine the half-life of the isotope ^{40}K using a device for detecting beta-particles and an ordinary watch, and we can find the number of atoms N by chemical analysis. The value N_0 equals $N + N_p$ where N_p is the number of ^{40}K atoms that decayed during the entire period of the granite's existence. N_p can be calculated if we know the number of argon atoms in the mica. Let's denote this number by A. Then $N_p = A/0.11$ (because the transformation of potassium into argon makes up 11% of the total number of decays of the radioactive potassium isotope). Thus $N_0 = N + A/0.11$. Inserting this expression into formula (1) we get

$$t = \frac{T_{1/2}}{\ln 2}\ln\left(1 + \frac{A}{0.11N}\right). \qquad (2)$$

Method (isotope)	Measured age (years)	Half-life (years)
Radiocarbon (^{14}C)	100–50,000	5,570
Argon-potassium (^{40}K)	> 100,000	$1.3 \cdot 10^9$
Rubidium-strontium (^{87}Rb)	> 5,000,000	$5.0 \cdot 10^{10}$
Lead-uranium (^{238}U)	> 200,000,000	$4.5 \cdot 10^9$

Of great interest also is the rubidium–strontium radioactive clock. The rubidium isotope ^{87}Rb is radioactive and turns into the strontium isotope ^{87}Sr by emitting beta-particles. Natural rubidium contains 28% of the radioactive isotope. The method for rubidium–strontium dating of minerals is very simple. By chemical analysis one first determines the total amount of rubidium in the sample and then the amount of isotope ^{87}Sr in it. After these procedures, calculations are made according to the formulas for radioactive decay.

All the methods discussed here determine the age of minerals from the moment of their crystallization. However, the products of radioactive decay are kept near the original nuclei only in solid bodies. In molten material the atoms freely intermix, and as the chemical properties of a substance composed of daughter nuclei differ from those of the initial substance, the products of nuclear decay are concentrated elsewhere.

As a rule, the most ancient rocks lie under massive, thick deposits. Only in some regions do they come near the surface. On the Kola Peninsula (in northwest Russia, on the Arctic Ocean) a granite slab was found that solidified and hardened 3.4 billion years ago. This is one of the oldest minerals on Earth. If its age is added to the time needed for a solid crust to form on Earth, the total age of the planet is approximately five billion years. A limitation on this method of determining the Earth's age is that the period of formation of the planet's crust is obtained by calculation, and the initial data are not that reliable. In particular, it's very difficult to take into account the heat dissipated by nuclear fission in the Earth's interior. However, there is another way to find the Earth's age without complicated thermal calculations, one that is based only on radioactive dating.

According to modern views, the meteorites and the Earth are made of the same material and condensed at the same time. The masses of meteorites are small, and so it took far less time for them to cool than was necessary for the Earth. So we can assume that the minerals in meteorites crystallized at the moment the Earth was "created." The age of meteorites can be determined by their lead and uranium content. If the Earth and meteorites were formed simultaneously, the result will give us the Earth's age as well.

When the concentration and isotopic composition of uranium and lead was measured in stone meteorites (which contain uranium, unlike iron meteorites), the age of these heavenly bodies was documented: about 5 billion years. Similar data were also obtained by potassium–argon and rubidium–strontium methods, which indicated that meteorites range in age from 4.3 to 4.8 billion years.

Space research has opened up new prospects for radioactive dating. In the future, space vehicles and probes will bring soil samples from planets in our Solar System to Earth. Scientists will then have a substance that may tell them something about the ages of distant planets.

Samples of lunar soil have already been studied. They also contain radioactive isotopes. The ages of minerals taken from different regions of the Moon proved to be different. This means that the formation of the hard lunar crust took an amount of time comparable to the Moon's age. In some places the lunar matter solidified earlier, in others—later. Here and there the still weak crust was broken and the lava streams filled the hollows. Nevertheless, lunar rocks are extremely old. The youngest have been around for more than 3 billion years, which corresponds to the age of the oldest minerals on Earth. Thus the inner geological life of the Moon stopped in the first 1.5 billion years of its existence. From that time all volcanic activity in the Moon ceased, and this natural satellite of the Earth became a passive celestial body, changing only in response to external events such as solar wind or meteorite bombardment.

Data on the age of planets in the Solar System are essential to research into its origin and history. The question how secondary heavenly bodies formed near primary bodies is key to understanding the processes of creation of the satellite systems of Uranus, Jupiter, and Saturn. Scientists believe that investigations into the origin of the planets' satellites are the most direct route to a general theory explaining the formation of the celestial bodies revolving around the Sun.

Mighty ether has struck out

The outlook wasn't brilliant for Al Michelson that day,
The project he'd been working on just *wouldn't* go his way.
With Morley, he'd been working on this simple apparatus
To indicate the speed of Earth through ether's rigid lattice.

You see, electromagnetism moves through outer space,
And it moves, as had been proven, at an astronomic pace.
And, of course, it was "self-evident," and "every schoolboy knew"
That wave motion needs a medium for it to vibrate through.

The speed at which a wave proceeds depends on the rigidity
Of the stuff it's going through, and hence, extreme rapidity
Requires a medium strange indeed: it must be inelastic,
While still allowing matter through—this stuff must be fantastic!

So even though no evidence had ever been presented,
The existence of this "ether" that the physics world invented
Had so taken scientists by storm that no one had even thought
The idea might be wrong, and it would shortly come to naught.

Well, Michelson and Morley, they believed it like the rest,
And set about to prove its truth, and so conceived a test
Which split a beam of light, and at right angles them aligned,
To show the interference patterns when they recombined.

So Michelson and Morley, they set up a granite slab,
And floated it on mercury they'd gotten from the lab.
They put a light source on the top, some mirrors, and a splitter;
With bated breath, they lit the lamp, their eyes were all a-glitter.

They looked upon the tiny screen for interference fringes
(When a pair of out-of-phase beams on a screen impinges).
But not a single fringe was found, at that or *any* angle
To which they turned their optics bench, their massive stone rectangle.

This history-making failure was truly quite successful,
Although for these two scientists, I'm sure 'twas very stressful.
The lesson that I hope we've learned, as connoisseurs of science,
Is that *truth*, and not *assumptions*, makes with us the best alliance.

—David Arns

David Arns *is a graphics software documentation engineer for Hewlett-Packard in Fort Collins, Colorado, and also operates a small business designing and creating Web sites. In his spare time he dabbles in poetry on scientific themes.*

TOY STORE

Rubik art

A physicist's pastime

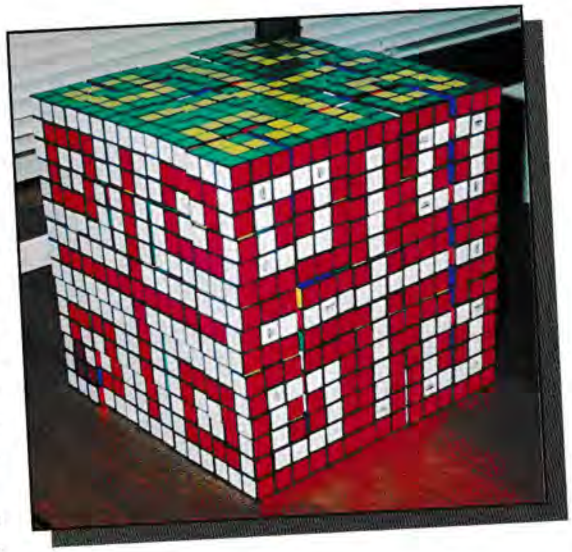
NO, RUBIK'S CUBE HAS NOT spawned a multifaceted monster. The designs in these photographs are the work of Dr. Hana Bizek, a physicist at Argonne National Laboratory, who has taken to building interesting designs out of many copies of Rubik's intriguing toy.

Dr. Bizek's Rubik designs were included in an art exhibition at Argonne. You can visit a virtual version of the exhibition at www.anl.gov/OPA/sciart.

Can you answer the following questions about the designs in the photographs?

1. How many Rubik's Cubes do you need to build one of these designs?

2. Different manufacturers of Rubik-type cubes sometimes arrange the same six colors in different ways on their cubes. In how many ways can this be done? Two colorings are considered distinct if a cube bearing one of them can be moved into a position so that it looks exactly like the other. (Consider mirror-image colorings as distinct).



3. Inside each design are a number of Rubik's cubes that are invisible from the outside. Since they are invisible, it does not matter what they look like, as long as they are of the correct size. Their number depends on the design's overall size. How many such central cubes appear in cubical designs made of 8, 27, 64, and 125 cubes?

4. Assume that the designs in the photographs are constructed out of identical Rubik's cubes, and that they are constructed so that the internal faces (faces that are not part of the large design) that are touching are colored the same. How many colors would a design made from eight solved cubes have?

5. The photographs display designs made from fewer than six colors. Yet they are made from conventional six-color cubes. How is that possible?

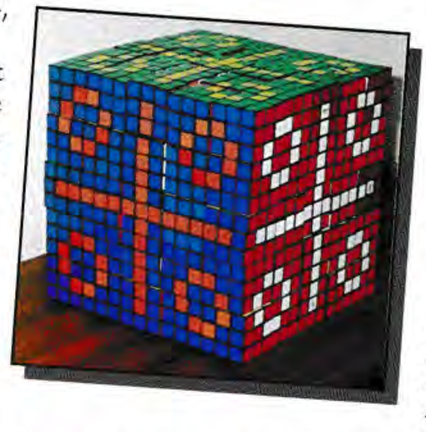


Photo by George Joch/Courtesy of *Argonne News*

Dr. Hana Bizek "twiddling" a Rubik's Cube.

Rubik's Cube has inspired several *Quantum* articles over the years—see, for instance, "The A-maze-ing Rubik's Cube" (September/October 1991), "Portrait of Three Puzzle Graces" (November/December 1991), and "The Last Problem of the Cube" (March/April 1995).

Let's not be dense about it!

("It" being density)

by A. A. Leonovich

"It seems, therefore, that there is no dense matter in the world."—Lucretius

"With a device one can determine when the air is thicker and heavier and when it is thinner and lighter."—Evangelista Torricelli

"Air that is twice as dense is twice as elastic."—Robert Boyle

"The Earth's density is 5.48 times that of water."—Henry Cavendish

"The relationships between pressure, temperature, and density for an ideal gas can be understood if we suppose that the particles move with uniform velocity along straight paths."—James Clerk Maxwell

"AW, COME ON NOW!" THE high school seniors among our readers are saying. "That's child's play, density." But let's not be too cocky—look again at the epigraphs. Although it seems a self-evident and rather modest physical concept, density is ever ready to come to the aid of scientists as they ponder serious questions—the composition of matter; differences in the physical properties of various things; gravitation; motion in fluids; and so on. The list of problem areas can easily be continued, just as we might add the names of many contemporary scientists to the list of famous thinkers above. They study both the microcosm and the structure of the stars, where huge densities can be met; they investigate the far reaches of outer space and the expanding Universe, whose future and destiny dramatically depend on changes in the negligibly small density of matter.

But without venturing so far, we can see how versatile the concept of density is. Indeed, in addition to the density of matter, scientists also use the notions of charge, current, and energy density; there are also such concepts as surface and linear density. So there are many interesting things in our world that are related to density, and here you have a chance to encounter them once again—and renew your respect for this "simple" idea.

Questions and problems

1. What is behind the constant movement of water in a hot-water heating system?

2. Which is heavier: a box filled with small buckshot or a box filled with large buckshot?

3. As they leave the last lock in the Panama Canal, ships drift slowly into the ocean without turning on their engines. What forces push them?

4. A piece of wood floats in water such that three fourths of its volume is submerged. What is the wood's density?

5. A piece of ice floats in a jar filled with water. On the surface of the ice is a wooden ball whose density is less than that of water. Will the water level be different after the ice is melted?

6. A hole is made in the ice in the middle of a large frozen lake. The thickness of the ice is exactly 10 m. What length of rope do you need to fill a pail with water?

7. Can you predict, before a melted substance becomes solid, how its density will change if you have a solid sample of the substance?

8. Which of the two aerometers (devices for measuring the density of liquids) shown in figure 1 should be chosen to record changes in a liquid's density most accurately?

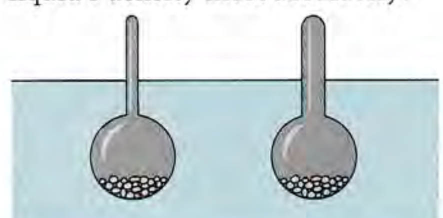
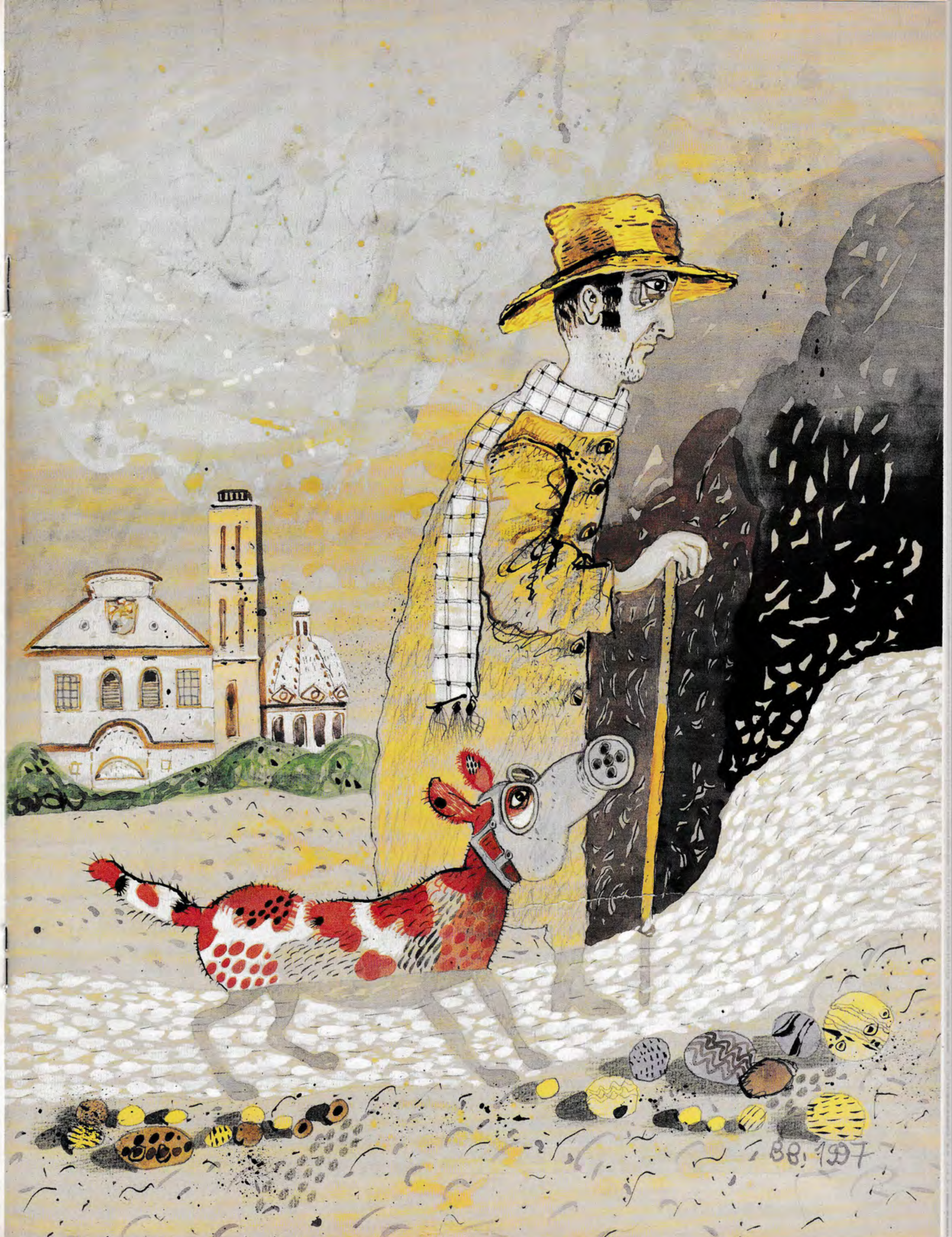


Figure 1

Opposite: entering the "Dog's Cave" (see item on page 34).

Art by Vasily Vlasov


BB, 1997

9. An object is weighed on a precise scale placed under a glass bell jar. Would the reading change if air is pumped out of the bell jar?

10. A object suspended from a spring scale is submerged in a jar of water at room temperature (fig. 2). How will the reading on the scale change if both the water and the object are heated?

11. At the bottom of a jar filled with a fluid (gas or liquid) there lies an object whose density is slightly higher than that of the fluid. Is it possible to lift the object by applying pressure to the fluid?

12. Two volumes of water of the same mass but different temperatures (1°C and 7°C) are mixed. Will the total volume of water change when thermal equilibrium is achieved? Neglect heat exchange with the surroundings.

13. Some water is subjected to increasing pressure. Should one warm it or cool it to keep its volume constant?

14. An empty glass vial floats in water at room temperature. When water is added to the jar, the vial rises. But when some more water is added to the jar, the vial sinks. How can you explain this phenomenon?

15. Plot the graphs of the temperature dependence of the density of an ideal gas during isothermal, isobaric, and isochoric processes.

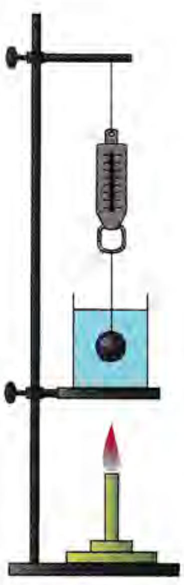


Figure 2

16. Two identical jars are set on a balance scale. One of them is filled with dry air, the other with moist air. The pressure and temperature are the same. Which jar is heavier?

17. How does the lift of a balloon depend on the ambient temperature?

18. Why does a charged conductor covered with dust quickly lose its charge?

19. Two cylindrical carbon electrodes are submerged in a copper sulfate solution. Copper precipitates on the surface of one electrode. Why is the copper layer thickest on the side of the electrode facing the other electrode?

Microexperiment

Try to find the average density of your body. What do you need to do it?

It's interesting that . . .

. . . the ancient Greek physician Hippocrates noted in his writings that rainwater is lighter than other kinds of water. It's remarkable that the ancient Greeks could distinguish between rainwater and well water by their densities, and that they used rainwater to calibrate volumes.

. . . since the 17th century people have determined the density of solid bodies by means of the bilancet, thought to have been invented by Galileo himself. This device, similar to a spring scale, allowed one to compare the weights of objects both in water and in the air.

. . . pondering the existence of the vacuum, Otto von Guericke decided to test experimentally the theory of Descartes that all space is filled with matter. The idea behind these first experiments for producing "emptiness" eventually led to the creation of the vacuum pump.

. . . the originality of Cavendish's experiments to determine the average density of the Earth lay in that they dealt with the gravitational interaction of comparatively small masses under laboratory conditions. Previously all the estimates of this density were based upon the measurements of the plumb line deviation from the vertical line due to the action of a nearby mountain.

. . . in Italy, near Naples, there is a famous cave called the "Dog's Cave." Carbon dioxide gas (which is 1.5 times as dense as the air) is continuously given off in its lower part. This gas spreads along the floor of the cave and slowly seeps out of the cave. A person can safely walk into the cave, but for a dog the foray would be deadly.

. . . the density of amber is close to that of seawater. As a result, amber can be "suspended" in water for dozens of years without dropping to the sea floor and abrading itself against the sand.

. . . the enigmatic anomalous behavior of water when the temperature changes in the range 0–4°C is explained by its quasi-crystal structure. A temperature increase in this range results not only in an increase in the interatomic distances but also in the rearrangement of this structure, which eventually produces a tighter packing of the water molecules.

. . . the very fact that substances have a "critical temperature" demonstrates the absence of a fundamental distinction between gas and liquid (and not only at temperatures higher than the critical point). Indeed, by varying the pressure and temperature, one can turn a liquid into a gas without passing through a boiling phase—that is, in a smooth, continuous way.

. . . if we mentally spread the matter in all the stars in our Galaxy uniformly, the average density of matter will be approximately $5 \cdot 10^{-24}$ g/cm^3.

. . . one ten-thousandth of a second after the big bang (the moment the Universe began to expand), its average density was about 10^{14} g/cm^3—equal to the density of atomic nuclei!

. . . the current value for the average density of the Universe determines how it will evolve further. Either the process of expansion will go on infinitely, or it will be replaced by contraction. It's possible that the matter in the Universe exists in some forms that are difficult to observe, so a precise value for the density of the present-day Universe hasn't been found yet, and the prospects of the Universe's future are open to discussion.

Quantum articles about density

- Yakov Zeldovich, "A Universe of Questions," January/February 1992, pp. 6–11
- William A. Hiscock, "The Inevitability of Black Holes," March/April 1993, pp. 26–29
- V. Mayer, E. Mamayeva, "Two Physics Tricks," March/April 1991, p.35
- A. A. Abrikosov, "The Story of a Dewdrop," September/October 1992, pp. 34–38
- I. I. Mazin, "An Invitation to the Bathhouse," September/October 1990, pp. 20–22
- Albert Stasenko, "From the Edge of the Universe to Tartarus," March/April 1996, pp. 4–8
- Arthur Eisenkraft, Larry 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

ANSWERS, HINTS & SOLUTIONS ON PAGE 61

Grab that chain of thought!

Did an article in this issue of *Quantum* make you think of a related topic? Jot down your thoughts. Then write to us for our editorial guidelines. Scientists and teachers in any country are invited to submit material, but it must be written in colloquial English and at a level appropriate for *Quantum*'s target readership of high school and college students.

Send your inquiries to:

Managing Editor
Quantum
1840 Wilson Boulevard
Arlington VA 22201-3000

PHYSICS CONTEST

Color creation

"She comes in colors everywhere,
she combs her hair,
she's like a rainbow."
—Mick Jagger/Keith Richards

by Arthur Eisenkraft and Larry D. Kirkpatrick

WE LOVE COLORS—THE colors of spring and summer, the colors of butterfly wings and rainbows, the colors of soap bubbles, and the colors from a CD. How are these the same? How are they different? Should we look to the same cause for what appears to be the same effect?

Isaac Newton, in his study of colors, devised some wonderful investigations. Described in his book *Opticks*, published in 1704, he details a series of experiments. In one, he shined light from his window through a prism and observed the colors of the spectrum that had tantalized so many others before him. He then brought these colors back together again with a second prism and saw, as no one before him had, that the white light returned. Newton then surmised that white light is a combination of all the colors of the spectrum. In school we learn the name of Mr. Roy G. Biv to help us remember the order of these colors (**r**ed, **o**range, **y**ellow, **g**reen, **b**lue, **i**ndigo, **v**iolet). These spectral colors are observed through diffraction gratings, through prisms, and in rainbows.

The rainbow is arguably Nature's most beautiful optical display. After a rainfall, the bow of colors can extend from horizon to horizon. The creation of the rainbow involves the physics of refraction and reflection and a geometry first explained by Descartes. The light rays from the sun refract as they enter the raindrops (fig. 1). This refraction causes the different colors of the white light to bend by different amounts, producing a spectrum. Upon hitting the back side of the water droplets, the light is partially reflected back toward the general direction of the sun. These light rays then refract again upon leaving the water droplet. The light rays emerge at many angles depending on where the light enters that rain drop. But each color tends to be concentrated at a special angle. With your back to the sun, the special angle between the shadow of your head and the red light is 42°. If you look up at 42°, you see red light. This also occurs at 42° to the left or right. In fact, it occurs at 42° in every direction. The locus of points that is 42° from the shadow of your head is a cone. This explains why you see a bow of red light across the sky—a bow at 42°. The special angle for orange light is a little less than 42°. The arc of orange light is, therefore, seen just below the red light. Similarly, the arcs for the other spectral colors are unique and together they form the rainbow in Roy G. Biv order from outside to inside.

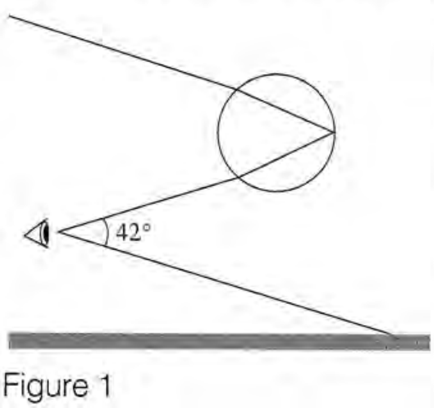


Figure 1

The colors from soap bubbles and oil slicks are not the Roy G. Biv colors of the rainbow. If you have occasion to look at soap bubbles or to notice the colors in oil puddles after a rain, you will recognize the colors as muted reds and blues and not the pure vibrancy of the rainbow colors. The way in which these colors are produced is quite different. We refer to the creation of these colors as due to thin film interference. In thin film interference, light reflected from the top and bottom surfaces of the film interfere, enhancing some colors and diminishing other colors.

Art by Tomas Bunk


ROYAL
SOCIETY
OIL
TOMAS
BUNK

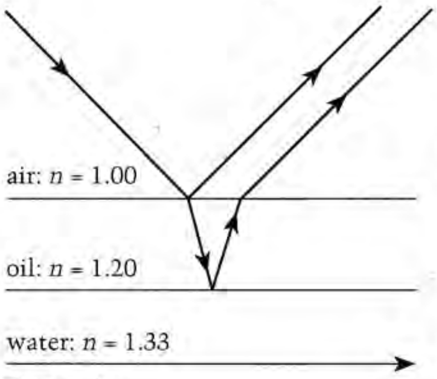


Figure 2

Let's take a closer look at the creation of colors from thin film interference (fig. 2). Imagine the thin layer of oil that rests on a puddle of water. If only red light was shining on the thin layer of oil, some of the light would reflect off the surface and some of the light would refract into the oil layer. Some of the refracted light would then hit the inner surface and some would reflect back into the oil film. Most of this light would then exit the film (undergoing another refraction) and combine with the originally reflected ray of light. If the two light beams had traveled equal distances, they would constructively interfere with one another, and lots of red light would be reflected. But the light ray that traveled into the film and back again traversed an additional distance. If the total distance were equal to one wavelength of red light, then the two rays of red light would, once again, constructively interfere. If the total distance were equal to one-half wavelength of light or $1\frac{1}{2}$ wavelengths of light, the rays would destructively interfere. Destructive interference essentially means that there is no red light reflected.

Let's imagine the more complex situation where white light is shining on the thin film of oil. If the thickness of the film is such that the red light undergoes destructive interference, you would see the complete spectrum minus red light—Oy G. Biv. This light is a muted purple, which we do see in oil slicks and in soap bubbles. If the thickness of the film is such that the violet light undergoes destructive interference, you would see Roy G. Bi, which looks like a muted red.

The actual situation has one more complication that must be taken into account. As many of us have seen, a wave reflecting off a boundary can undergo a phase shift. A crest of a wave can hit a boundary and reflect as a trough. This occurs when the wave is reflected from a stiffer medium or one with a higher index of refraction. A phase shift can be mathematically expressed as a shift of one-half wavelength. To determine the correct thickness of the film for destructive interference, one must calculate the total path difference due to the thickness of the oil and any phase shifts that may occur at the boundaries. Since the index of refraction of oil is 1.20 and the index of refraction of water is 1.33, there is a phase shift of the ray reflected off the oil *and* a phase shift reflected off the water surface.

The phenomenon of thin film interference has industrial applications. One of the problems in building sophisticated lens systems is that the internal reflections can cause stray light in the photograph. By coating the lens with a thin film of magnesium fluoride, the unwanted reflection can be eliminated. Let's assume that the coating of magnesium fluoride has an index of refraction of 1.36 and a 100-nanometer layer is evaporated onto a lens of index of refraction 1.60. Which wavelength of light will not be reflected from the surface? We will assume that the light is incident perpendicular to the surface.

The light must travel through the thin film and back again, a total distance of 200 nm. Since the film has a higher index of refraction than the air ($1.36 > 1.00$), there is a phase shift of one-half wavelength upon reflection. The light which reflects from the film-glass interface also undergoes a phase shift of one-half wavelength, since once again the light is reflecting from a material with a higher index of refraction ($1.60 > 1.36$). The light traveling along path 1 reflected from the first surface and underwent a phase shift of $\frac{1}{2}$ wavelength. The light traveling along path 2 entered the film, traveled 100 nm, reflected off the second surface and underwent a phase shift of $\frac{1}{2}$ wavelength, and then traveled an additional 100 nm to the first surface. But the 100 nm in the film is not like 100 nm in the air. The wavelength of light is shorter in the film by a factor equal to the index of refraction. The light from these two paths will destructively interfere if the total path difference is a multiple of $\frac{1}{2}$ wavelength in the film. For the thinnest coating of magnesium fluoride, the total path length in the film must be equal to $\frac{1}{2}\lambda$ in the film:

$$\tfrac{1}{2}\lambda_{\text{film}} = 200 \text{ nm},$$

or

$$\lambda_{\text{film}} = 400 \text{ nm}.$$

Therefore, the wavelength in air is

$$\lambda = \frac{\lambda_{\text{film}}}{n} = 544 \text{ nm}.$$

We offer two problems this month. One is adapted from a problem first given at the International Physics Olympiad (IPhO) in Czechoslovakia in 1977, and the other is a problem from *Fundamentals of Physics* by Halliday, Resnick, and Walker.

A. White light falls on a soap film at an angle of 30° with the normal. The reflected light displays a predominantly bright green color of wavelength 500 nm. The index of refraction of the liquid is 1.33. (i) What is the minimum thickness of the film? (ii) What color would be seen if the light source fell on the same soap film from the vertical direction

B. A thin film of acetone ($n = 1.25$) coats a thick glass plate ($n = 1.5$). White light is incident normal to the film. In the reflection, fully destructive interference occurs at 600 nm and fully constructive interference at 700 nm. Calculate the minimum thickness of the acetone film.

Please send your solutions to

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

The nature of light

In the November/December issue of *Quantum*, we asked our readers to work with the nonrelativistic Compton effect in which a photon collides with a free electron. Correct solutions were submitted by Robert Marasco (a junior at North Penn High School in Lansdale, Pennsylvania), Timothy Spegar (a graduate student in mechanical engineering at Penn State University), and jointly by André Cury Maiali and Gualater José Biscuola (engineers and physics teachers in Jundiaí, São Paulo, Brazil). For the most part, we will follow Maiali and Biscuola's solution.

A. Knowing that the energy and momentum of a photon are given by $E = hf$ and $p = hf/c$, respectfully, we can write down the equations for conservation of energy and momentum in one dimension:

$$hf = hf' + \frac{1}{2}mv^2, \tag{1}$$

$$\frac{hf}{c} = -\frac{hf'}{c} + mv, \tag{2}$$

where f and f' are the initial and final frequencies of the photons, m is the mass of the electron, and v is the speed of the electron after the collision.

B. We now collect the frequency terms in equations (1) and (2) and square the equations to obtain

$$f^2 - 2ff' + f'^2 = \frac{m^2v^4}{4h^2}, \tag{3}$$

$$f^2 + 2ff' + f'^2 = \frac{m^2v^2c^2}{h^2}. \tag{4}$$

Subtracting equation (3) from equation (4) yields

$$4ff' = \frac{m^2v^2c^2}{h^2}\left(1 - \frac{v^2}{4c^2}\right).$$

Neglecting the last term in the parentheses, we have

$$h^2ff' = \left(\frac{1}{2}mv^2\right)\left(\frac{1}{2}mc^2\right). \tag{5}$$

C. Replace the term in the first parentheses in equation (5) with its equivalent from equation (1) and use the wave relationship $c = \lambda f$ to obtain

$$\frac{h^2c^2}{\lambda\lambda'} = h\left(\frac{c}{\lambda} - \frac{c}{\lambda'}\right)\left(\frac{1}{2}mc^2\right),$$

or

$$\lambda' - \lambda = \frac{2h}{mc}. \tag{6}$$

D. We can get the energies of the X rays from

$$E = hf = \frac{hc}{\lambda} = 2.80 \cdot 10^{-15}\ \text{J}.$$

Because this energy (17.5 keV) is much, much larger than the binding energies of the electrons to their atoms (~10 eV), the electrons can be treated as being free.

The kinetic energy K of the recoil electron can be found from

$$K = h(f - f') = hc\left(\frac{1}{\lambda} - \frac{1}{\lambda'}\right).$$

Using equation (6) and defining $\lambda_C = h/mc$, we obtain

$$K = hf\left(\frac{2\lambda_C}{\lambda + 2\lambda_C}\right) = 1.12\ \text{keV}.$$

Because this is very small compared to the rest energy of the electron (0.511 MeV), the electron can be treated nonrelativistically.

E. We obtain the two-dimensional equation by starting with the two components of the equation for the conservation of momentum.

$$p = p'\cos\theta + p_e\sin\phi, \tag{7}$$

$$0 = p'\sin\theta - p_e\cos\phi, \tag{8}$$

where $p_e = mv$ is the electron's momentum and $p = hf/c = h/\lambda$ and $p' = h/\lambda'$ are the momenta of the photons. In equations (7) and (8), put the electron terms on the left-hand side and the photon terms on the right-hand side, square both equations, and add them together to obtain

$$p_e^2 = p^2 - 2pp'\cos\theta + p'^2. \tag{9}$$

The equation for the conservation of kinetic energy is

$$K = pc - p'c. \tag{10}$$

Divide equation (10) by c, square it, and subtract it from equation (9):

$$p_e^2 - \frac{K^2}{c^2} = 2pp'(1 - \cos\theta). \tag{11}$$

Now let's look at the left-hand side of equation (11) in more conventional terms:

$$m^2v^2 - \frac{m^2v^4}{4c^2} = m^2v^2\left(1 - \frac{v^2}{4c^2}\right)$$

$$\cong m^2v^2 = 2mK = 2mc(p - p'),$$

where we've neglected the term in v^2/c^2. Therefore,

$$mc(p - p') = pp'(1 - \cos\theta).$$

Dividing through by pp' and expressing the momentum in terms of the wavelengths, we get the desired result:

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\theta).$$

Farewell to JCMN

Reflections on the James Cook Mathematical Notes *—in memory of Basil Rennie*

by George Berzsenyi

THE PURPOSE OF THE PRESent column is to pay tribute to the memory of Basil Rennie, who created the *James Cook Mathematical Notes* (*JCMN*) in 1975, as a unique periodical featuring superb mathematical investigations by him and fellow mathematicians around the world. As I am coming close to the end of my term as the author of these columns, I wish to call my readers' attention to the wonderfully conversational style of Basil's superb journal, so that they may find alternative sources for their future investigations. To whet their appetites, I will list below some of the questions posed by Basil in his *JCMN*; to my knowledge, many of them are still not resolved. The references indicate the pages of the issues where these problems appeared; for several other problems and more information about *JCMN*, the reader is referred to my Problems, Puzzles, & Paradoxes column in the Summer 1997 issue of *Consortium*.

Problem 1. Given m things, we want to choose the same number m of k-element subsets so that no two of these have more than one element in common. For each m, what is the largest possible k? The first few seem to be as follows (p. 6137; May 1992):

m	1	2	3	4	5	6	7
k	1	1	2	2	2	2	2

Problem 2. Given a set of n distinct points in the plane, they form $n(n-1)(n-2)/2$ angles, all in the closed interval between 0 and π. What can we say about these angles? One simple thing is that their mean is $60° = \pi/3$. Perhaps we can say that at least one angle $\le A(n)$ and at least one angle $\ge B(n)$, where $A(n)$ and $B(n)$ are given by

n	3	4	5	6
$A(n)$	60°	45°	36°	30°
$B(n)$	60°	90°	108°	120°

How does the table continue? (p. 6090; February 1992)

Problem 3. Two random points in the unit disc (from a uniform probability distribution) give a random line segment (ending at the two points). Find the probability that two such random line segments intersect. (p. 6017; February 1991)

Problem 4. Suppose that three random points in the unit disc are chosen from the distribution with uniform probability density. Calculate the expectation and the variance of the area of the triangle formed by the three points. (p. 5281; October 1990)

Problem 5. An advertisement by a local builders' merchant for "mirror tiles" set me thinking about how hard it is to see one's back. If I tile the walls of my bathroom with these mirror tiles could I look directly (not obliquely) at my back? In a rectangular room it would be possible by standing at any point of the rhombus whose vertices are the midpoints of the sides, or indeed by standing at any point not on either diagonal and facing in the direction of one of the diagonals. But what if the bathroom were triangular? (p. 4128; June 1985)

Much like Paul Erdős (who was one of his frequent contributors), Basil Rennie was constantly probing the boundaries of our mathematical universe. He had an unerring talent for posing problems which would intrigue his readers and friends; we will greatly miss him and his *JCMN*. Several of my own investigations were prompted by his correspondence; I only wish I had learned more from him. He passed away on the 15th of November, 1996.

The first issue of *JCMN* appeared

in 1975, while the last (issue 70) was completed shortly before his death. Issues 1–31 of *JCMN* have been published in three bound volumes; they are probably still available from the Head of the Mathematics Department of James Cook University of North Queensland (Post Office James Cook, North Queensland 4811, Australia). Prior to his retirement, Basil was the Professor and Head there—hence the name of his publication.

We can hope that issues 32–70 will also become available in bound volumes in the near future. I strongly recommend them to my readers. ◘

Magnets, charges, and planets

"Physics has begun to notice that to think about nature is not simply a matter of recording it, but of giving it that form of unity it would not have if it were not contemplated."—Pierre Teilhard de Chardin, Phenomene humain

by Albert Stasenko

WHAT COULD BE MORE ELementary than a straight wire with constant current I? Every student knows that the current generates a magnetic field **B** around the wire. The field lines are circles concentric with the wire lying in planes perpendicular to the wire This can easily be visualized by means of classical iron filings spread on a sheet of cardboard placed perpendicular to the wire. So, what's there to say?

Well, imagine, we placed into this magnetic field a square conducting frame with area $a \times a$, which carries a constant current I_p, (fig. 1). Let the two edges of this frame be parallel to the straight wire and the size of the frame be much less than the distance r to the wire—that is, $a \ll r$. According to figure 1, the magnetic field is perpendicular to all the edges of the frame, so each edge will experience a magnetic force that is proportional to the magnetic field B (generated by the current I), to the length of the edge a, and to the current I_f. Let's examine these forces.

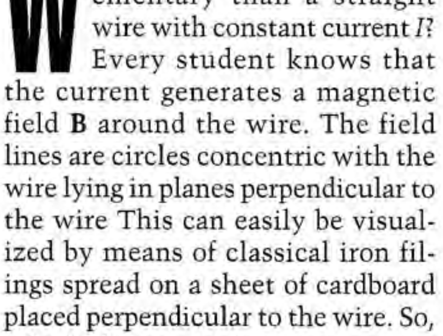


Figure 1

The edge AK located at a distance r from the wire experiences a force perpendicular to the wire and equal to $F(r) = I_f aB(r)$. I'm assuming you know the right-hand rule and that you have already figured out the directions of the vectors **B**, $\mathbf{I}_f$, and **F**. Since the current flowing in the edge CD is directed counter to that in AK, the magnetic field will act on CD in the opposite direction with a force $F(r + a) = I_f aB(r + a)$. The value of this force differs from $F(r)$ because the strength of the magnetic field depends on the distance from the current-carrying wire (and we have a strong suspicion that B decreases as r increases). As for the other forces acting on the edges AD and KC, they counterbalance each other (we assume that the frame is not deformed by the action of all these forces). Thus the net force acting on the frame is equal to the algebraic sum of the two forces acting on the edges AK and CD:

$$\begin{aligned} F &= F(r+a) - F(r) \\ &= I_f a\big(B(r+a) - B(r)\big) \\ &= I_f a^2 \frac{\Delta B}{a}, \end{aligned} \tag{1}$$

where ΔB denotes the change in the magnetic field B over the distance $a \ll r$.

What have we obtained? A small frame carrying a current I_f and lying in the same plane as a straight wire carrying a current I is attracted to this wire by a force proportional to the value of the current flowing in the frame, and also to its area and the rate of change in the magnetic field strength with distance from the current-carrying wire ($\Delta B/a$). To tell the truth, "grown-up" physicists don't use so many words. They use mathematics instead, which begins with some definitions. The product $I_f a^2$ is called the *magnetic moment* $\mathbf{p}_m$ of the current-carrying frame, because the magnetic lines of force produced by a small current-carrying frame look very much like that of the electrostatic field **E** generated by an electric dipole $\mathbf{p}_e$ (only at large distances, of course—that is, when $r \gg a$, which we have assumed from

Art by V. Ivanyuk


МАГНИТ
РАМКА
Электро-
сеть
ПРОВОЛОКА

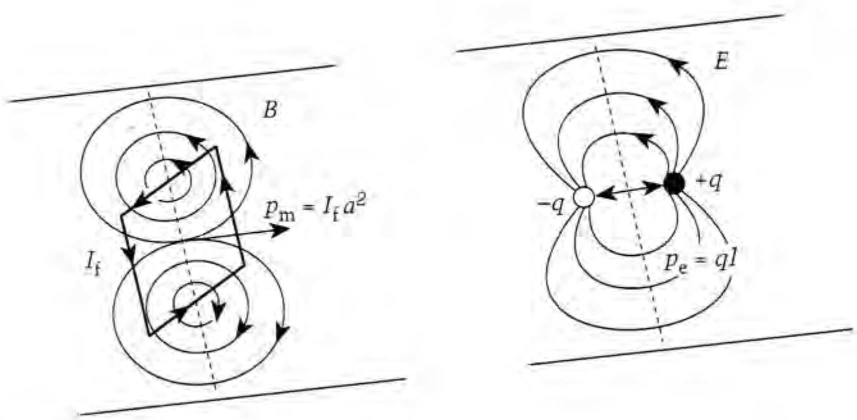


Figure 2

the very beginning). Figure 2 shows schematically the lines of magnetic and electric forces, or in other words, the vector fields **B** and **E**. The ratio $\Delta B/a = \Delta B/\Delta r$ is called the *gradient* of the magnetic field. Now we can express the thought contained in equation (1) in far fewer words, and at the same time guess at some generalizations of the example chosen here (this is a very important moment for physicists, when mathematics helps them greatly).

The point is, the frame can take any shape—circle, triangle, and so on. Also, instead of a frame it may be a very different object—say, a small permanent magnet (the magnetic needle from a Lilliputian compass). The external magnetic field **B** can also be generated not by a current-carrying wire, but by any arbitrary source. In all these various situations, the force affecting a magnet placed in the heterogeneous magnetic field will be proportional to the magnetic moment of a small object (a probe) and to the gradient of the external magnetic field. For example, it is this force that pulls a small permanent magnet into a solenoid connected to a voltage source, and also foils our attempts to extract the magnet from the coil.

Well, let's return to the current that flows in a straight line. It's clear by intuition that its magnetic field decreases somehow with distance. But how? This question is answered by one of Maxwell's laws: the product of the magnetic field $B(r)$ and the circumference $2\pi r$ of a circle concentric with the wire is proportional to the current I generating the field:

$$B(r) \cdot 2\pi r = \mu_0 I.$$

The factor μ_0 (the permeability of free space) is needed to account for the different dimensions on the two sides of the formula (in SI units). It's one of the fundamental physical constants, but we won't pay too much attention to it here—we have other fish to fry. It's more important to recognize that we have formulated a kind of conservation law: however far we move away from an (infinite) straight current-carrying wire, the product $2\pi r \cdot B(r)$ will remain the same. Physicists call this product the "circulation" of the magnetic field along a loop. Thus the magnetic field generated by an infinite straight electric current is inversely proportional to the distance from it:

$$B(r) = \frac{\mu_0 I}{2\pi r}.$$

How can we obtain the value of $\Delta B/\Delta r$? Those who know what a derivative is can differentiate this formula with respect to the radius to immediately get

$$\frac{\Delta B}{\Delta r} = \frac{\mu_0 I}{2\pi}\Delta\left(\frac{1}{r}\right) = -\frac{\mu_0 I}{2\pi r^2}.$$

Those who haven't studied calculus yet can obtain the change in the inverse radius using algebra:

$$\Delta\left(\frac{1}{r}\right) = \frac{1}{r+\Delta r} - \frac{1}{r} = \frac{r - r - \Delta r}{(r+\Delta r)r} = -\frac{\Delta r}{r^2}.$$

In the last term on the right-hand side we neglected the value of Δr in the denominator, because it's very small compared to r itself—remember, we agreed at the very beginning that $a = \Delta r \ll r$.

Now let our small current-carrying frame (or magnet) have mass m and move with a velocity v_0 normal to the straight current I at a distance r_0 from it. How will the frame move?

We know that the frame is affected by a magnetic force that is inversely proportional to the square of the distance to the wire. Where do we meet such forces in physics? Literally everywhere. According to Newton's law of universal gravitation, any two masses are attracted by a force that is inversely proportional to the square of the distance between them:

$$F_N = -G\frac{m_1 m_2}{r^2}.$$

(Of course, Newton's law describes the force acting between two masses regardless of their mutual orientation in space. In our case, the magnetic force always lies in the plane of the current-carrying frame.) According to Coulomb's law, two opposite charges are mutually attracted by a force that is inversely proportional to the square of the distance between them:

$$F_C = -\frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}.$$

So our problem is solved: a current-carrying frame or a permanent magnet will move near the straight current just as massive bodies move near the Sun—either along elliptical trajectories (like the planets) or along parabolas and hyperbolas (like comets) if its initial velocity v_0 is large enough.

Thus, the process we've been examining can be reduced to a physical

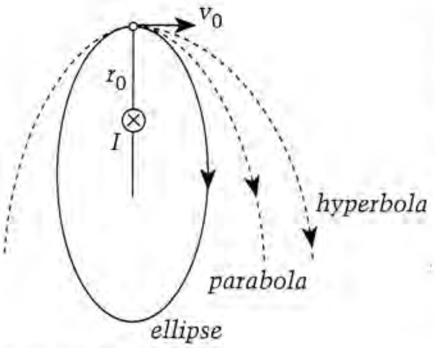


Figure 3

process of an entirely different nature, yet governed by the same mathematical equations (in our case the analogous process is described by Kepler's laws—see "The Fruits of Kepler's Struggle" in the January/February 1992 issue of *Quantum*). The search for physical analogies is a very interesting and practically useful occupation!

Now is the time to exclaim: "How wonderful that the world is so harmonious, so unified!"

But, as usual, after the first paroxysms of joy induced by an interesting finding, we have second thoughts: Did we forget something? Naturally, we did.

First, does the current-carrying frame rotate such that its plane remains normal to the magnetic lines of force at all points along its trajectory? Indeed, the frame is not a point—it has both mass and size. So its inertial characteristics should be taken into account when we consider the rotation of the frame about its axis.

Second, in the general case the distance between the frame and the wire varies, and so does the magnetic flux through the plane of the frame. This variation will generate an electromotive force in the frame, which will change the current flowing in it. The alternating current will in turn induce an emf, which will try to counterbalance the changes in the magnetic field within the frame.

Third, this process will also change the current flowing in the straight wire (the phenomenon of mutual induction).

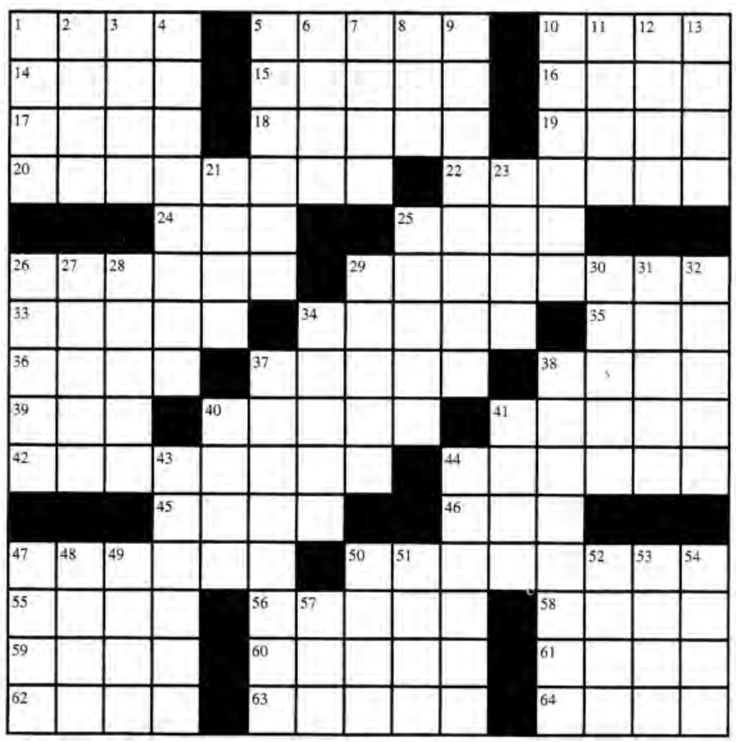
Fourth . . .

Then again, maybe it's better to stop here and say that our theory is correct only when all these effects are very small and can be neglected.

Gridlock!

The grid printed with the March/April CrissCross Science was incorrect. Here is the correct grid. (Now, don't peek at the answers printed in this issue!)

Our apologies for any frustration this may have caused our crossword fans.



AT THE BLACKBOARD II

Adding angles in three dimensions

Taking a theorem for plane figures into the realm of polyhedrons

by A. Shirshov and A. Nikitin

THE THEOREM ON THE SUM of the angles of a plane polygon occupies an important place in Euclidean geometry, and so the question of a spatial version of the theorem arises quite naturally. In this article we'll present the analogous theorem and discuss various subjects associated with it.

The measure of an n-hedral angle

We'll measure plane and polyhedral angles in the following way. Let AOB be an arbitrary angle on the plane. Draw a circle of radius $r > 0$ and center at point O. We'll call *the measure of the angle AOB* the ratio of the arc $A'mB'$ contained within the angle AOB to the length of the entire circle (fig. 1).

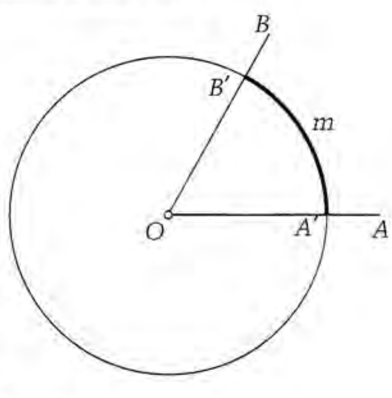


Figure 1

Art by Dmitry Krymov

If this method of measurement is adopted, the measure of a one-degree angle, for example, is $1/360$; the measure of an angle of π radians is $1/2$; and the well-known theorems concerning the sum of the angles of a triangle and the sum of the angles of a convex polygon will read as follows: *the sum of the angles of a triangle is equal to $1/2$, and the sum of the angles of a convex n-gon is equal to $n/2 - 1$.*

Now let's look at an arbitrary n-hedral angle. We'll draw a sphere around its vertex (in the case of a dihedral angle, around an arbitrary point on its edge) and call ratio of the area of the spherical surface contained in the angle to the area of the entire sphere (fig. 2) *the measure of the n-hedral angle.*

If $n > 2$ and $A_1, A_2, \ldots, A_n$ are the points where the edges of the n-hedral angle meet the sphere, then we'll write $|A_1A_2 \ldots A_n|$ for the measure of the n-hedral angle and $|\hat{A}_1|, |\hat{A}_2|, \ldots, |\hat{A}_n|$ for the measures of the corresponding inner dihedral angles $\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_n$ of the n-hedral angle.

Each face of the n-hedral angle intersects the sphere along a great circle, and the figure cut from the sphere by all the faces is called a *spherical polygon* (fig. 2).

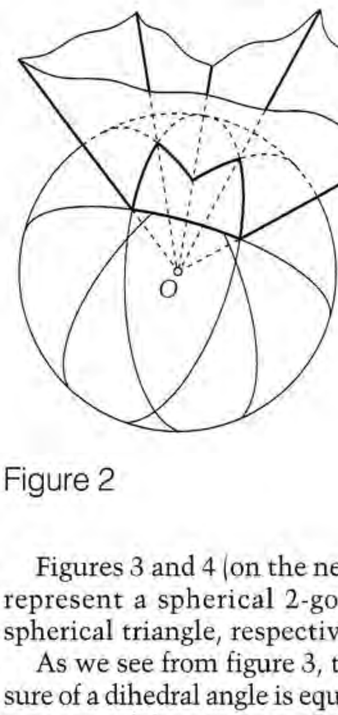


Figure 2

Figures 3 and 4 (on the next page) represent a spherical 2-gon and a spherical triangle, respectively.

As we see from figure 3, the measure of a dihedral angle is equal to the measure of the corresponding linear angle. If we switch to new notation from the well-known equation that expresses the area of a spherical triangle in terms of the radian measure of its angles,[1] we obtain

$$2|ABC| = |\hat{A}| + |\hat{B}| + |\hat{C}| - 1/2, \quad (1)$$

[1] See the article on non-Euclidean geometry, "A Revolution Absorbed," in the January/February issue.—*Ed.*

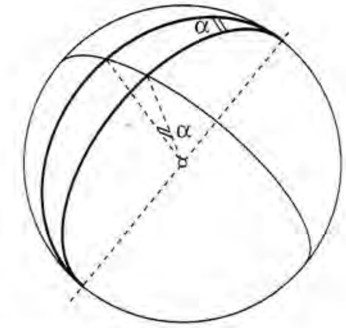


Figure 3

where $\hat{A}$, $\hat{B}$, and $\hat{C}$ are *inner* dihedral angles at the edges OA, OB and OC, respectively, of the trihedral angle.

Indeed, the area of a figure F_1 composed of two spherical triangles ADC and $A'D'C'$ is the same as the area of the 2-gon $\hat{C}$ (we use the same notation for a 2-gon and its corresponding dihedral angle); the area of a figure F_2 composed of the triangles $AD'B$ and $A'DB$ is the same as the area of the 2-gon $\hat{B}$; and a figure F_3 bounded by the semicircles ABA' and $A'CA$ is simply the 2-gon $\hat{A}$. The common part of these three figures $F_1 \cap F_2 \cap F_3 = F_1 \cap F_2 = F_2 \cap F_3 = F_3 \cap F_1$ is the spherical triangle ABC, whose area we are trying to calculate. If we add up the areas of these figures, the result will be three times the area of ABC. On the other hand, if we subtract from this sum the doubled area of triangle ABC, we'll obtain the area of the visible hemisphere (fig. 4). Finally, we have $1/2 = |\hat{A}| + |\hat{B}| + |\hat{C}| - 2|ABC|$—that is, equation (1) is true.

It follows from equation (1), that the sum of the measures of the inner dihedral angles of an arbitrary trihedral angle is half again as large as the doubled measure of the angle.

Now we can find the area of a *convex* n-hedral angle. Keep in mind that any convex n-gon can be decomposed into $n - 2$ triangles. Similarly, any convex n-hedral ($n > 2$) angle can be decomposed into $n - 2$ trihedral angles, and the corresponding spherical n-gon is at the same time decomposed into $n - 2$ spherical triangles (fig. 5). Applying equation (1) to all these triangles, we find

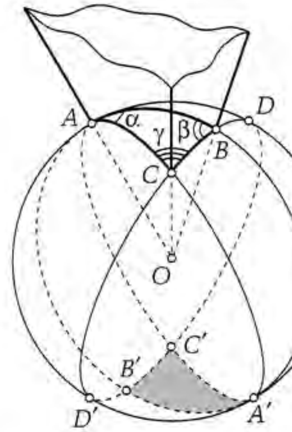


Figure 4

that for an arbitrary spherical n-gon $A_1A_2...A_n$, which corresponds to *some convex* n-hedral angle, the following equation is satisfied:

$$\left|\hat{A}_1\right| + \left|\hat{A}_2\right| + \cdots + \left|\hat{A}_n\right| - 2\left|A_1A_2\ldots A_n\right| = \frac{n}{2} - 1. \tag{2}$$

This means that *the sum of the measures of the inner dihedral angles of an n-hedral angle is n/2 - 1 times larger than doubled measure of the n-hedral angle.*

We'll call the quantity

$$|\hat{A}_1| + |\hat{A}_2| + \ldots + |\hat{A}_n| - 2|A_1A_2\ldots A_n|$$

the "excess" of the polyhedral angle $OA_1A_2...A_n$.

Exercise 1. Use equations (1) and (2) to find (a) the measure of the trihedral angle at a vertex of a cube; (b) the measure of the trihedral angle at a vertex of regular tetrahedron; (c) the measure of the quadrihedral angle at a vertex of octahedron.

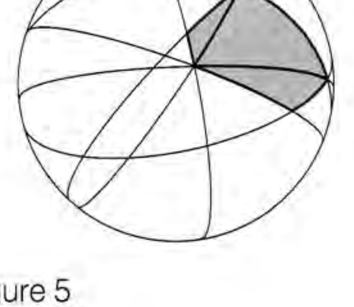


Figure 5

The sum of the angles in a polyhedron

Now let's consider an arbitrary tetrahedron. Denote the values of its trihedral angles as α_i ($i = 1, 2, 3, 4$) and the values of its dihedral angles as β_j ($j = 1, 2, \ldots, 6$). The excess of a trihedral angle is equal to 1/2; thus the sum of the excesses of all the trihedral angles in a tetrahedron is 2. Note that each dihedral angle appears in this sum twice: first at one vertex, adjacent to an edge; and then at another. Thus the following equality is valid:

$$2\sum_{j=1}^{6}\beta_j - 2\sum_{i=1}^{4}\alpha_i = 2.$$

Therefore,

$$\sum_{j=1}^{6}\beta_j - \sum_{i=1}^{4}\alpha_i = 1. \tag{3}$$

Consider an arbitrary convex k-corner pyramid. Let the letter α denote the measure of the k-hedral angle (at the vertex of the pyramid), $\overline{\alpha}$ the sum of the measures of the trihedral angles at its base, β the sum of the measures of all the dihedral angles at its base, and $\overline{\beta}$ the sum of the measures of all the dihedral angles formed by its lateral faces. Now we can calculate the sum of the excesses at all the vertices of the pyramid, as we did above for the tetrahedron.

From equation (2) it follows that

$$\overline{\beta} - 2\alpha = \frac{k}{2} - 1,$$

and from equation (1) we obtain

$$\overline{\beta} + 2\beta - 2\alpha = \frac{1}{2}k.$$

Thus

$$\beta - \overline{\alpha} + \alpha = \frac{1}{2}, \tag{4}$$

$$\beta + \overline{\beta} - (\alpha + \overline{\alpha}) = \frac{k+1}{2} - 1. \tag{5}$$

Thus equations (3) and (5) show that for a tetrahedron, as well as for *any convex k- corner pyramid, the difference between the sum of the*

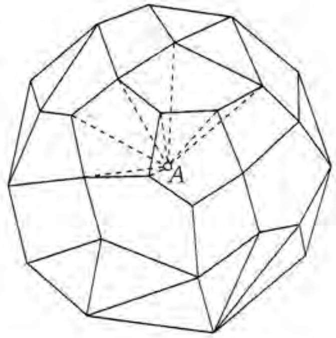


Figure 6

measures of all the dihedral angles and the sum of the measures of all the polyhedral angles at its vertices depends only on the number of faces of the pyramid.

Consider now an arbitrary *convex* polyhedron with n faces (fig. 6). Take an arbitrary point A inside the polyhedron and connect it with all the points on the polyhedron's edges. We obtain a decomposition of the polyhedron into n pyramids.

We note that the sum of the measures of all the polyhedral angles at point A of these pyramids is 1; the sum of the measures of all the trihedral angles at the bases of the pyramids is equal to the sum ϕ of the measures of all the polyhedral angles at the vertices of the polyhedron; and the sum of the measures of all the dihedral angles at the bases of the pyramids is equal to the sum $\bar{\phi}$ of the measures of all the dihedral angles of the polyhedron. This observation, together with equation (4), yields directly the equality

$$1 - \phi + \bar{\phi} = \frac{1}{2}n,$$

or

$$\bar{\phi} - \phi = \frac{n}{2} - 1. \qquad (6)$$

Notice that we need the condition that the polyhedron be convex only to simplify the reasoning. We encourage the reader to make some generalizations in this direction.

The theorem on the generalized sum of the angles of a polyhedron was first proved by Gué in 1783, and further developments were published in 1837 by Brianchon.

One of the wonderful properties of equation (4) is that, unlike its planar analogues, it is valid even in non-Euclidean geometry.

Exercise 2. (a) Find the generalized sum of the angles of the cube and the dodecahedron. (b) Find the sums of the measures of the polyhedral angles at the vertices of these polyhedrons.

Exercise 3. Use equation (1) to calculate the sum of the excesses of the polyhedral angles of a convex polyhedron with E edges and V vertices. (Answer: $E - V$.)

Exercise 4. Use equation (6) and the result from the previous exercise to derive Euler's famous equation that links the number of faces F, the number of vertices V, and the number of edges E of a convex polyhedron: $E - V = F - 2$.

Exercise 5. Give an example of a polyhedron for which $V - E + F = 0$.

Why doesn't the sack slide?

And how does a gynmast "stick" a landing?

by Alexey Chernoutsan

DURING A PHYSICS CLASS A teacher put a matchbox on the table and then placed a glass of water on top of it (fig. 1). "How can one pull the box from under the glass without touching glass? Just tug on it? No, the glass will be dragged along with the box. Pay attention, now!"

With these words the teacher took a heavy ruler, pulled it back, and smacked the matchbox. It flew to a far corner, matches flying, but the glass rested on the table almost in the same place!

Why didn't the glass budge? To avoid misunderstanding, I should note that the teacher struck the box, not the glass. The only horizontal force acting on the glass is the force of friction from the box $F_{fr} = \mu mg$, where μ is the coefficient of friction, m is the mass of the glass, and g is the acceleration due to gravity. Surely, even this force can impart an appreciable speed to the glass, moving the box along the table with a small acceleration. (What would be the upper limit of such an acceleration if the glass is to be prevented from sliding?) The point is not the magnitude of the force, but the fact that after a sharp blow from the ruler the matchbox immediately flew off, so the force of friction acted only during a very brief time. This period was so short that the force of friction had no time to impart any appreciable momentum $\Delta p_x = F_{fr}\Delta t$ to the glass.

This example shows that in analyzing the forces acting on a body or a system of bodies, one must also take into account the duration of their action. For example, at the moment a shell explodes, it is affected by an external force—the force of gravity. In spite of this force, one can assume that the total momentum of the system is conserved. The total momentum of the fragments is equal to the momentum of the shell, because the change in the system's momentum is negligible during the very short time of the explosion.

"Something's wrong here," an inquisitive reader may argue. "Look at a hard ball that bounces on the floor (fig. 2). The duration of the impact is very small, but the change in momentum is quite appreciable: it's $\Delta p_y = mv - (-mv) = 2mv$. What's going on here?"

You're absolutely right. The effect of a force is not always negligible just because it acts for a short time. The effect of a force is evaluated neither by its magnitude nor by the period of its action. The correct way to decide whether a force can be neglected is to estimate the momentum this force would impart to an object if it was the only force acting on the object. When the force is constant, the increase in the object's momentum is equal to $\mathbf{F}\Delta t$ (this magnitude is called the *impulse*); when the force is variable, its impulse is given by the sum $\mathbf{F}_{avg}\Delta t = \Sigma\mathbf{F}_i\Delta t_i$. The effect of this force can be neglected if its value is constant and the period of its action is short. This is what happens when a shell explodes. Nothing of the kind takes place with the force acting on the ball bouncing on the floor. If we decrease the duration of the impact by a factor of 10 (due to the increased hardness of the ball and the floor), the average value of the force will increase by the same factor of 10. As a result, the change in the ball's momentum will still be $2mv$. By convention, such forces are called *impulsive*. The effects of impulsive forces may be substantial even for very brief interactions.

Let's consider firing a cannon as it slides down on an inclined plane

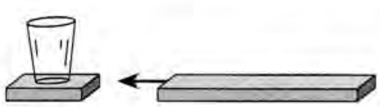


Figure 1

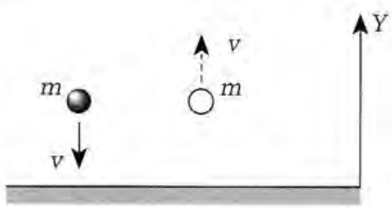
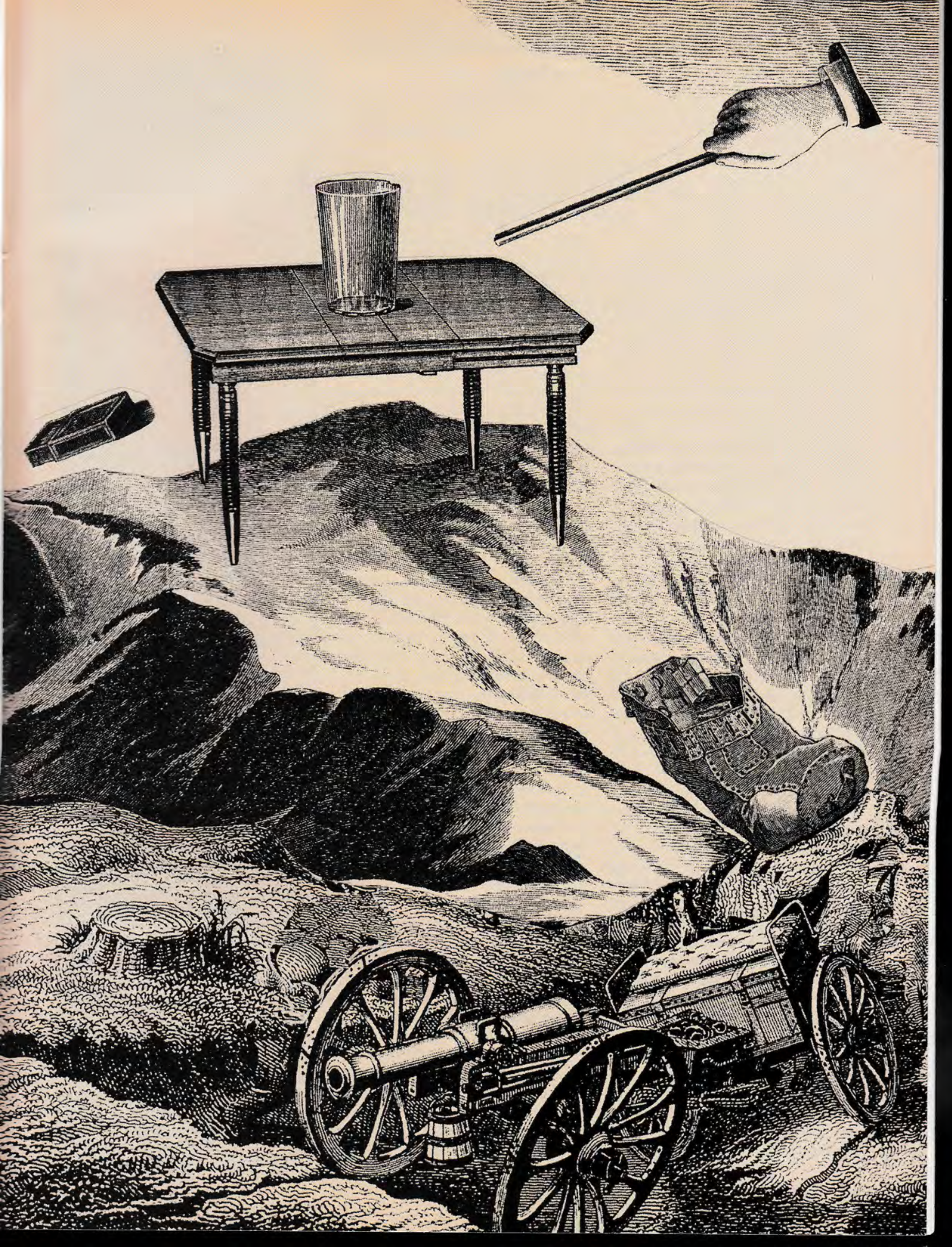


Figure 2

Collage by Vera Khlebnikova



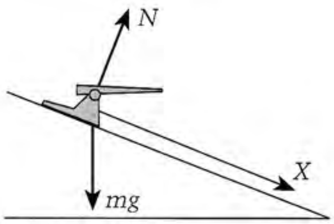


Figure 3

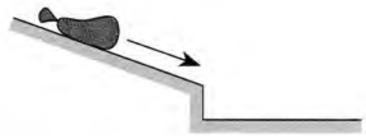


Figure 4

(fig. 3). Can we use the law of conservation of momentum to find the cannon's velocity after the shot? Yes, we can—but first we need to find the direction upon which the projection of the net external force is zero. However, in our case there is no such direction! The system is affected by the force of gravity $m\mathbf{g}$ and the normal force $\mathbf{N}$ due to the plane. If we choose the horizontal axis, we can neglect the force of gravity (its projection is zero), but the projection of the normal force isn't zero, and we can't neglect the impulsive force! If we chose the axis parallel to the inclined plane, the projection of the force of gravity won't be zero. Which is preferable?

What we need to do, actually, is get rid of the impulsive force—that is, the normal force. The impulse of the constant force of gravity can be neglected, because the duration of the cannon shot is very short. On the other hand, the effect of the impulsive normal force doesn't tend to zero and can result in an significant change in the system's momentum. However, since this force is perpendicular to the inclined plane, the momentum of the system "cannon–projectile" won't change if it's projected on the direction of the cannon's motion.

So, during brief processes (bursts, collisions) only impulsive external forces modify the momentum of a body or system of bodies, while the effect of nonimpulsive forces can be neglected.

"Do you understand?" the teacher asks. "Okay, let's look at another example. A sack slides along an inclined chute and falls onto the floor (fig. 4). What happens next? Will the sack stop immediately, or will it slide a little at first under its own momentum?"

"Well, it's pretty clear," a student answers, "the sack is affected by the force of friction in the horizontal direction, and the duration of the impact is small, so the horizontal momentum of the sack won't change—just as in the previous example with the glass. As for the vertical momentum, it will become zero due to the impulsive normal force. So we conclude that, for some time after the fall, the sack will slide."

Something is wrong with this reasoning, and this is felt by everyone who has observed how a sack falls onto the floor. Why? In this example, the force of sliding friction μN is also impulsive—just like the normal force. In the example with the glass, the normal force had a fixed value mg, so the effect of friction was negligible. However, in the case of a sack falling from an inclined chute, both the normal force and the force of friction are impulsive forces, so the change in the horizontal momentum during the impact cannot be neglected. Whether the sack stops or not depends entirely on the coefficient of friction μ: if it's large enough, the velocity of the sack "disappears" entirely during the impact. Try to estimate what value of the coefficient of friction would make such "disappearance" possible.

Similar considerations help explain why gymnasts can land on their feet and stop immediately after jumping from an apparatus (the rest of the explanation has to do with their extraordinary skill!). Try to explain how an object can bounce off a rough floor at some angle other than the angle of incidence, or the mechanics of a "spin serve" in sports. And most important—see if you can find examples on your own and devise problems dealing with impulsive and nonimpulsive forces.

AT THE
BLACKBOARD III

Fair and squared!

What to do when a physics problem has been "reduced" to a math problem involving a quadratic equation

by Boris Korsunsky

WHAT DO YOU DO IF YOU want to solve a physics problem? Well, unless it's a purely conceptual one, you use all the facts to come up with an equation (or several equations). Usually, once you have the equations to play with, you tell yourself: "Okay, the physics is over, let's do some math." Or, as some prominent physicists like to put it, "the physics problem is now reduced to a math problem." But occasionally math gets a little revenge. A very common example involves the quadratic equation. If you have to deal with it as you solve a problem, look out! Once you solve the equation, you have to *interpret* the solution. And this is another *physics* problem—sometimes a tough one.

Let's look at some examples from various areas of physics.

Problem 1. Two cars are traveling along the same road. Car 1 moves at a constant velocity v; car 2 starts from rest with a constant acceleration a. Initially car 2 is a distance d behind car 1. How long (t) will it take car 2 to pass car 1?

This one isn't too bad, is it? Once you write the equation $x(t)$ for each car and set them equal, you end up with something like this:

$$\frac{1}{2}at^2 - vt - d = 0.$$

This equation has, of course, two roots:

$$t = \frac{v \pm \sqrt{v^2 + 2ad}}{a}.$$

Are they both valid solutions? In other words, can the cars meet twice? Not likely. So, which one do we choose? The positive one, of course. And so the answer to the problem is

$$t = \frac{v + \sqrt{v^2 + 2ad}}{a}.$$

The next example is a bit more complicated.

Problem 2. A vertical tube of length l is inserted into mercury to a depth of $l/2$. Then the top end is sealed. Find the height h of the mercury column left in the tube after the tube is pulled out of the mercury. The temperature is constant and the atmospheric pressure is equal to that of a mercury column of height H.

To solve this problem, let's consider the equilibrium of the bottom surface of the column: the pressure "up" must equal the pressure "down." It will be convenient to express the pressure in units of "length": the pressure "up" equals the atmospheric pressure H; the pressure "down" equals the pressure of the air above the column (P') plus that of the column itself (P''). For the air pressure in the tube, using PV = constant, we can write

$$H\frac{l}{2} = P'(l - h),$$

or

$$P' = \frac{H}{2}\frac{l}{l - h}.$$

For the mercury column,

$$P'' = h.$$

The condition $P = P' + P''$ brings us to the equation

$$h^2 - (H + l)h + H\frac{l}{2} = 0,$$

whose roots are

$$h = \frac{H + l}{2} \pm \frac{1}{2}\sqrt{H^2 + l^2}.$$

Now, you're an experienced problem solver and will quickly identify the positive root. The trouble is, they're *both* positive. A second look at the roots, though, reveals that the bigger one is not only greater than zero, it's also greater than l! This leaves us with the answer:

$$h = \frac{H + l}{2} - \frac{1}{2}\sqrt{H^2 + l^2}.$$

These examples show us how to find the root that is not a solution to the problem (of course, sometimes both roots are solutions—see exercise 1 at the end of the article). The next problem is different: it can be solved *without* finding the roots.

Problem 3. The focal length of a converging lens is f. Find the minimum possible distance x between an object and its real image. (Hint: if you aren't familiar with calculus yet, it's okay!)

The lens equation

$$\frac{1}{o}+\frac{1}{x-o}=\frac{1}{f}$$

leads to the quadratic equation with respect to the object distance o:

$$o^2 - ox + fx = 0,$$

the discriminant of which is

$$D = x(x - 4f).$$

Because o exists only for $D > 0$, $x_{\min} = 4f$. Voilà!

In the next example, the algebra gets more complicated, but again, we analyze the discriminant rather than try to find the roots.

Problem 4. A fireworks rocket is tested in the center of a large cylindrical pit of diameter d. After the rocket explodes, the burning pieces are expected to have speeds not exceeding a certain value v. Find the depth h of the pit that will provide safety for observers standing on the very edge of the pit (provided they are careful enough not to fall *into* the pit!).

To solve this one, let's write the equation of motion for the piece that makes it to the top edge of the pit. Suppose the piece takes off at a certain angle θ (fig. 1):

$$x=(v\cos\theta)t=\frac{d}{2},$$
$$y=(v\sin\theta)t-\frac{1}{2}gt^2=h.$$

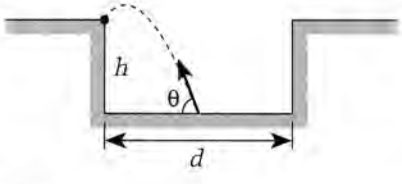


Figure 1

Eliminating the time, we get the following quadratic equation for $\tan\theta$:

$$\tan^2\theta-\frac{4v^2}{gd}\tan\theta+\frac{8hv^2}{gd^2}+1=0.$$

We want this equation *not* to have solutions. That is, we want

$$D=\left(\frac{4v^2}{gd}\right)^2-4\left(\frac{8hv^2}{gd^2}+1\right)<0,$$

or

$$h>\frac{v^4-g^2d^2}{8gv^2}.$$

Looking at the numerator on the right-hand side, we can see that if $v^2 < gd$, then *any* h will do. In other words, the rocket can even be tested on level ground. Otherwise the formula above yields the answer.

The last problem goes back to "root elimination." But this time the procedure isn't at all simple, even though the problem is pretty innocent looking. Let's see how tricky it gets.

Problem 5. A test tube of length l is filled with air at a pressure P and closed with a light movable piston. The test tube is then submerged in water to a depth H (fig. 2) and the piston is released. Find the height h of the air column in the test tube. The density of water is ρ and the atmospheric pressure is P_a.

Using the equilibrium condition for the piston, we have

$$\rho g(H-h)+P_a=P',$$

where P' is the new pressure of the air in the test tube. Also, because

$$PV = \text{constant},$$
$$Pl = P'h.$$

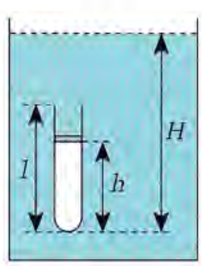


Figure 2

Combining these equations, we get

$$h^2-\left(H+\frac{P_a}{\rho g}\right)h+\frac{Pl}{\rho g}=0.$$

And this is where the trouble begins! First of all, the discriminant of the equation isn't necessarily positive, which makes sense. For example, if $P \gg P_a$ and H is just slightly greater than l, the piston will be pushed out of the beaker. But what if a solution *does* exist? Intuitively, we can expect if to be *the only* one, but the equation gives us two:

$$h_\pm=\frac{1}{2}\left(H+\frac{P_a}{\rho g}\right)\pm\sqrt{\frac{1}{4}\left(H+\frac{P_a}{\rho g}\right)^2-\frac{Pl}{\rho g}}.$$

Both roots are positive, and both must be less than l. How do we choose the "good" one? To answer this question, we must look at the stability of the equilibrium (an important concept that is often overlooked). Consider the graphs of the functions $P' = Pl/h$ and $P' = P_a + \rho g(H - h)$ (fig. 3). The intersection points 1 and 2 correspond to h_- and h_+, respectively. Consider the changes in pressures corresponding to small deviations from the equilibrium positions 1 and 2. Analysis reveals that only point 1 corresponds to a *stable* equilibrium, which is the only real possibility. (Technically, if the piston is somehow brought to a halt at position 2, it will stay there, and solution 2 will be realized.)

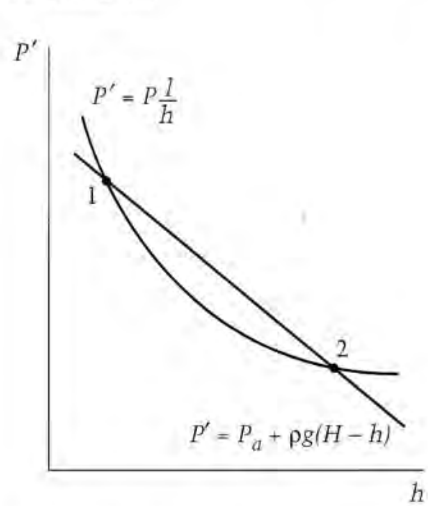


Figure 3

Now that you have some respect for the quadratic equation, try these exercises.

Exercise 1. Two cars begin to move at the same time: car 1 is moving from A to B; car 2 from B to A. Car 1 is moving at a constant velocity v, while car 2 starts at a velocity u and has a constant acceleration a directed from A to B. It's known that the cars met twice, while moving in the *same* direction. Find the range of v that allows this to happen. The distance AB equals l.

Exercise 2. Charges $+Q$ and $-2Q$ are fixed at a distance l from each other. Find the point at which a third charge q will be in equilibrium.

Exercise 3. A rock is thrown out a window, which is set at a height h above the ground. The initial speed of the rock is v. Find the maximum horizontal distance l the rock can travel if the takeoff angle is chosen appropriately. Use an approach similar to that in problem 4. ◘

Boris Korsunsky *teaches at Northfield Mount Hermon School in Northfield, Massachusetts.*

ANSWERS, HINTS & SOLUTIONS ON PAGE 61

What's happening?

Summer study ... competitions ... new books ... ongoing activities ... clubs and associations ... free samples ... contests ... whatever it is, if you think it's of interest to *Quantum* readers, let us know about it! Help us fill Happenings and the Bulletin Board with short news items, firsthand reports, and announcements of upcoming events.

What's on your mind?

Write to us! We want to know what you think of *Quantum*. What do you like the most? What would you like to see more of? And, yes—what *don't* you like about *Quantum*? We want to make it even better, but we need your help.

What's our address?

Quantum
NSTA
1840 Wilson Boulevard
Arlington VA 22201-3000

Bulletin Board

Scholarships for young women

Infusium 23, a producer of hair care products, has established a special awards program aimed at young women in the U.S. "For more than three decades, women have played an increasingly significant part in the American workplace," said Johnna Doyle, Manager of Research and Development at Infusium and Chairperson of the Awards Program. "But girls often need more encouragement to develop a strong interest in fields like science, where role models are generally male."

Infusium 23 will award 23 grants of $1,000 each, specifically earmarked for female high school students to use toward the pursuit of education in science—for college, summer classes, extracurricular programs, etc. The goal of the program will not only be to recognize those young women already interested in scientific fields, but also to encourage other young women to explore the many opportunities available in a field that is often overlooked by females.

In a recent survey conducted by Infusium 23 and targeting thousands of high school students across the country, more than one third of those polled were still undecided as to what career track to take. Of those who had a preference, the top three careers listed by males were computer science (13%), engineering (9%), and medicine/physician (8%). The top three career choices for females were teaching (13%), the arts (12%), and medicine/physician (11%).

Doyle was inspired to pursue her own career when she was one of only two girls to enter a science program sponsored by the Boy Scouts of America. Independent studies have also shown that when females are actually encouraged to enter these so-called "masculine preserves" of science and technology, their test scores in those areas quickly equal, and often overtake, those of boys in the same areas.

To apply for these awards, open to all female high school students, an applicant should provide a statement including

- What career in science she wants to pursue and why;
- Extracurricular activities demonstrating career commitment;
- The person or persons who have influenced her career choice.

These statements should be submitted, along with the applicant's name, address, age, school grade, and name of school, to Infusium 23 Women in Science Awards, 40 West 57th Street, 23rd floor, New York NY 10019. Applications must be postmarked by September 15, 1997.

A rickety cyberbridge

As usual, many of our CyberTeaser contestants showed admirable ingenuity in confronting the most recent problem posted at our Web site (brainteaser B205 in this issue). But, unfortunately, they were often too clever by a half. An answer in the affirmative can be achieved without any contortions, and fortunately many entrants found the way there.

Here are the first ten who submitted a correct answer:

Leonid Borovsky (Brooklyn, New York)
Xi-An Li (Middlebury, Vermont)
Guy Ben Zvi (Ramat Yohanan, Israel)
Masato Kobayashi (Kyoto, Japan)
Clarissa Lee (Perak, Malaysia)
How Yu Khong (Kuala Lumpur, Malaysia)
Jim Grady (Branchburg, New Jersey)
Avner Nevo and Ori Charag (Ramat Yohanan, Israel)
Aaron Manka (Arlington, Virginia)
Jaak Sarv (Tallinn, Estonia)

Each will receive a *Quantum* button and a copy of this issue. Congratulations! All who submitted a correct answer were eligible for a drawing to receive a copy of *Quantum Quandaries*, our collection of brainteasers.

The next CyberTeaser awaits you at www.nsta.org/contest.

Duracell/NSTA competition winners

What useful and entertaining products are missing in catalogs and stores? Six ingenious new ideas have just become realities due to the creativity of American teenagers involved in the 15th Duracell/NSTA Scholarship Competition. Top inventions announced in March included two musical gadgets, two safety tools, a new device for the blind, and a parking space finder.

The 1997 competition awarded over $100,000 in savings bonds to 100 high school inventors and recognized over 700 students nationwide for their innovative ideas.

"While these are prototypes, the technical know-how behind them is sufficient for production of each as a viable product," said Arthur Eisenkraft, judging chair. "These winners showcase the inventive spirit that the Duracell Competition rewards in our youth."

The first-place $20,000 savings bond winner was Ashley Eden, a

CONTINUED ON PAGE 62

Criss × cross science

by David R. Martin

1	2	3	4	■	5	6	7	8	■	9	10	11	12	13
14				■	15				■	16				
17				■	18				■	19				
20				21			■	22	23		■	24		
■	■	■	25			■	26			■	27			
28	29	30			■	31				32		■	■	■
33				■	34			■	35			36	37	38
39			■	40				41			■	42		
43			44			■	45			■	46			
■	■	■	47			48			■	49				
50	51	52		■	53			■	54			■	■	■
55			■	56			■	57				58	59	60
61			62		■	63	64			■	65			
66					■	67				■	68			
69					■	70				■	71			

Across

1 "Who ___ turn to?" Tony Bennett hit
5 Landed
9 Tied one's shoes
14 Cattle
15 Mongoloid or Caucasian, e.g.
16 965,294 (in base 16)
17 Type of gin
18 Physicist Niels ___
19 Bar game
20 Inductance units
22 Small pellets
24 Sleep motion: abbr.
25 Rocky peak
26 OSS's successor
27 10^{12}: pref.
28 ___ catalyst
31 Electron or proton property
33 Child's first word
34 10^5 pascals
35 Ascomycetous fungi
39 ___ agar
40 Two million pounds
42 Hawaiian tree
43 Angle unit
45 Time unit: abbr.
46 Good-hearted
47 Yellow dye
49 Heredity units
50 Focusing device
53 Electrical conductivity unit
54 Distilled coal
55 Arab's robe
56 Wood: comb. form
57 Fluorine or chlorine, e.g.
61 Microwave source
63 Shady mountain side
65 Measuring device
66 699,114 (in base 16)
67 Danish playwright Kaj ___
68 Coup d'___
69 Sortie
70 Type of current
71 Quality: suff.

Down

1 Hyperbolic function
2 Vehicle part
3 Element 10
4 Resistance to acceleration
5 Molecular biologist Werner ___
6 Asian country
7 Fish dermatitis
8 Tb_2O_3
9 Solid-state lamps
10 Abscisic acid: abbr.
11 British novelist John Le ___
12 Diner
13 Knobby spicule
21 Charged particle
23 Heavy particle
26 Like element 24
27 Oolong, e.g.
28 ___ wax (ozocerite)
29 Dalai ___
30 64,957 (in base 16)
31 Unit of heat: abbr.
32 Army leader: abbr.
34 Like computer data
36 Outer body layer
37 Musical sound
38 Arabic letters
40 ___ Baisakhi (dusty squall in Bengal)
41 Element 50
44 "no ___ ands or buts . . ."
46 Petrologen
48 width × height × length
49 Volume unit: abbr.
50 Andean animals
51 965,290 (in base 16)
52 Of the olfactory sense
54 Sticky
56 Long wavelength EM wave
57 Workman
58 Logic circuit
59 Portuguese neurologist Antonio ___ Moniz
60 Meshes of fabric
62 Moray, e.g.
64 Immature shoot

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE MARCH/APRIL PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M201

We can transform the equation into $2x^3 - x^3 - 3x^2 - 3x - 1 = 0$, or $2x^3 = (x+1)^3$, or $x\sqrt[3]{2} = x + 1$. From this we obtain

$$x = \frac{1}{\sqrt[3]{2} - 1}.$$

M202

Let's multiply the first equation by $a + b + c$. Each term simpifies as follows:

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

so the given equation becomes

$$\frac{a^2}{b+c} + \frac{b^2}{a+c} + \frac{c^2}{ab} + a + b + c = a + b + c,$$

which gives the desired result.

M203

We can assume that all the hens that are alive at any given time in the process described lay eggs simultaneously, and that the chicks hatch from all the eggs at the same moment. Let a_k and b_k be the numbers of hens and roosters at the kth stage—that is, $a_1 = 1$, $b_1 = 0$. According to the statement of the problem,

$$a_{k+1} + b_{k+1} = 2a_k. \qquad (1)$$

and $a_n = 0$. Add up equalities (1) for $k = 1$ through n. We get

$$(a_2 + a_3 + \ldots + a_n) + (b_2 + b_3 + \ldots + b_n) = \ldots = 2(a_1 + a_2 + \ldots + a_{n-1}).$$

Since $a_n = 0$ and $a_1 = 1$, we get $b_2 + b_3 + \ldots + b_n = a_1 + a_2 + \ldots + a_{n-1} + 1$. This means that there was one less hen than there were roosters—that is, the number of hens was 1996.

M204

Denote the center of the given circle by O, and let A and B be the points where the line meets the circle. Draw circles with diameters OA and OB (they will pass through the midpoint of AB). The location we seek consists of (1) all points that lie within one and only one of these circles and (2) the point O.

Let's prove this. First, we note that for every point K within the given circle, there is only one chord such that K is its midpoint. This is the chord perpendicular to OK (K must of course be different from O). Draw an arbitrary line through O. Let M_1 and M_2 be the points where this line intersects (for the second time) the circle with diameter OA and the circle with diameter OB, respectively. Consider the situation shown in figure 1. Since OA and OB are diameters of their respective circles, $\angle AM_1O$ and $\angle BM_2O$ are right angles. Now it follows that all the chords whose midpoints belong to the radius OD but not to the segment M_1M_2 lie completely on one side of the line AB. In fact, M_1 is the midpoint of the chord passing through A, and M_2 is the midpoint of the chord passing through B. On the other hand, if a chord meets the segment M_1M_2 in its internal point and is perpendicular to it, then its endpoints lie on the different sides of the line AB. Thus if we forget about O for the time being, the intersection of the point we seek and the line OD consists of the segment M_1M_2. This reasoning is valid for any line, passing through O, even if point O lies between points M_1 and M_2. Finally, we note that point O also satisfies the conditions of the problem.

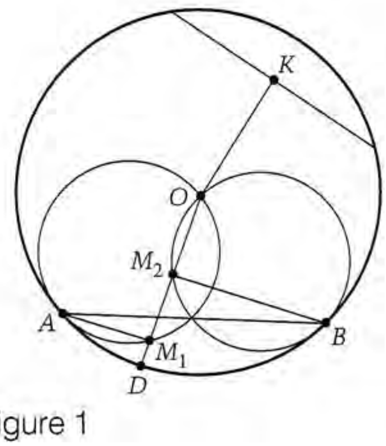
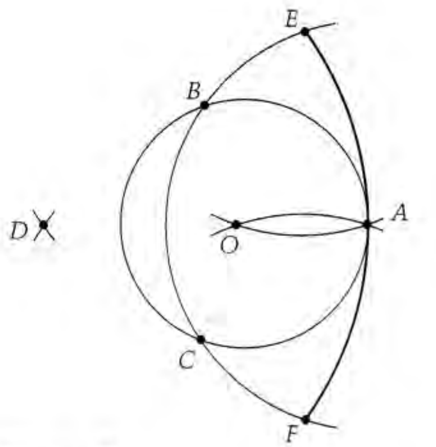


Figure 1

M205

We'll begin by pointing out the necessary construction and then we'll offer a proof of it. Let's take points A and B on the circle such that the distance between them is greater than the radius of the circle[1] and draw a circle passing through B whose center is at A. Let it meet the given circle for the second time at C (we assume that points C and B are different). This is the first circle of our construction. Then we construct point D, symmetric to A with respect to the line BC. For this purpose we will need two more circles; they will be the second and the third circles of our construction. Draw a circle of radius DA whose center is at D. Let this (the fourth) circle meet the first at points E and F. Circles with centers at E and F passing through A—the fifth and the sixth circles of our construction—meet at point O, which is the center

[1]We do not consider that this requires drawing an arc.



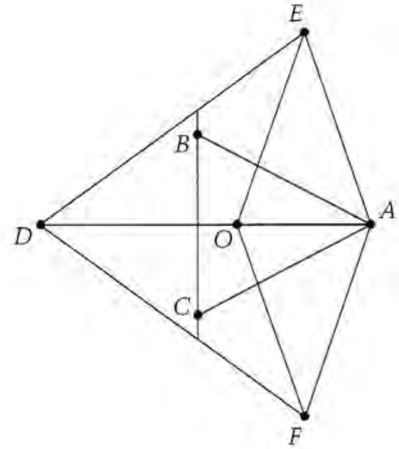


Figure 2

of the given circle.

Proof. It's clear that the point where the fifth and sixth circles intersect lies on the line AD. So it's enough to prove that $OA = R$, where R is the radius of the given circle, which is circumscribed about the triangle ABC. Put $AB = AC = a$ and let h be the length of the altitude drawn to side BC in the triangle. The "extended law of sines"[2] tells us that

$$\frac{AC}{\sin\angle ABC} = 2R,$$

or

$$R = \frac{AC}{2\sin\angle ABC} = \frac{a^2}{2h}. \qquad (1)$$

On the other hand, triangles ADE and AEO are similar, and $AD = DE = 2h$, $AE = AO = a$. Thus

$$\frac{AO}{AE} = \frac{AE}{AD},$$

or equivalently

$$\frac{AO}{a} = \frac{a}{2h}.$$

Therefore,

$$AO = \frac{a^2}{2h}.$$

Comparing this equality with equation (1), we see that $AO = R$. Thus, O is the center of the given circle (fig. 2).

[2]The extended law of sines says that in any triangle ABC, $a/\sin A = b/\sin B = c/\sin C = 2R$, where R is the radius of the circumscribed circle.

Physics

P201

Let's calculate the tension in the cord at both ends of the pulley. Suppose the tension in the cord near mass $M = 3$ kg is T_1. For this load we get

$$Mg - T_1 = Ma,$$

and thus

$$T_1 = M(g - a).$$

Similarly, on the other side of the pulley we have

$$T_2 - mg = ma,$$

and thus

$$T_2 = m(g + a).$$

The difference in the tension on both sides of the pulley counterbalances the friction in the shaft (strictly speaking, one should talk about torques and not forces, but in this case it makes no difference). The friction is proportional to the axial load—that is, to the force $T_1 + T_2$. Thus

$$T_1 - T_2 = k(T_1 + T_2).$$

Although we can find the coefficient k from this equation, it's more practical to obtain the value $(1 - k)/(1 + k)$, since later we'll need just this expression:

$$\frac{1-k}{1+k} = \frac{m(g+a)}{M(g-a)}.$$

Equilibrium is achieved within the entire range of possible "extra" loads—from the minimum (when the downward acceleration of the larger mass is zero) to the maximum (when this mass is lifted uniformly upward). To find the minimum extra load Δm_1 we write

$$\frac{m + \Delta m_1}{M} = \frac{1-k}{1+k},$$

which gives us

$$\Delta m_1 = \frac{2ma}{g-a} = \frac{1}{2}\ \text{kg}.$$

Similarly, for the maximum extra load Δm_2 we have

$$\frac{M}{m + \Delta m_2} = \frac{1-k}{1+k},$$

which gives us

$$\Delta m_2 = \frac{M^2(g-a) - m^2(g+a)}{m(g+a)} = 5\ \text{kg}.$$

P202

This problem can be solved directly by calculating all the successive collisions using the laws of conservation of energy and momentum, taking into account the heat losses. However, the problem has a simple and elegant solution.

Clearly, after the first collision puck 2 has a velocity $v_0/2$. To calculate the result of the collision of the pucks after the recoil from the side wall, it will be convenient to switch to the moving reference system, which travels with a velocity $v_0/2$ from the side wall. In this reference frame puck 2 is again at rest (as before the first collision), while puck 1 again approaches it—with half the velocity, however. After the second impact puck 1 stops, just as after the first impact, and puck 2 acquires velocity $v_0/4$. In the laboratory frame of reference the velocity of puck 1 is $v_0/2$ and is directed away from the side wall, while the velocity of puck 2 is $v_0/4$ and is also directed away from the wall.

P203

The work W performed on the system is spent in changing both the internal energy of the gas ΔU and the potential energy of the piston ΔE_p:

$$W = \Delta U + \Delta E_p.$$

For one mole of monatomic ideal gas, the change in the energy is given by the formula

$$\Delta U = \frac{3}{2}R(T_x - T_0).$$

The change in the piston's potential energy can be found in this way: it is equal to the work needed to move the piston quasi-statically from the initial state to the final one. While this is happening, the external force at any moment must be equal to the force of gravity mg acting on the piston. Since the piston is at equilibrium in the initial and final states, the force of gravity equals the gas pressure P times the area of the piston S (we neglect the pressure of the outside air).

Denoting the change in the piston's height by h, we get

$$\Delta E_p = mg\Delta h = PS\Delta h = P\Delta V,$$

where ΔV is the change in the volume of the gas. Using the ideal gas law for one mole of gas, we have

$$\Delta E_p = P\Delta V = R\,(T_x - T_0).$$

From this we obtain

$$W = \frac{3}{2}R(T_x - T_0) + R(T_x - T_0) = \frac{5}{2}R(T_x - T_0),$$

and consequently

$$T_x = T_0 + \frac{2}{5}\frac{W}{R}.$$

P204

First let's find v_4. The only way the sound waves produced by plane 4 reach planes 1, 2, and 3 at the same moment is if plane 4 is moving at a supersonic speed ($v_4 > c$, where c is the speed of sound). Then, knowing that the shock wave forms an angle $\sin\theta = v_4/c$ and that the wave front must be parallel to the course, trigonometry allows us to find $v_4 = c/\sin\theta$ (see figure 3).

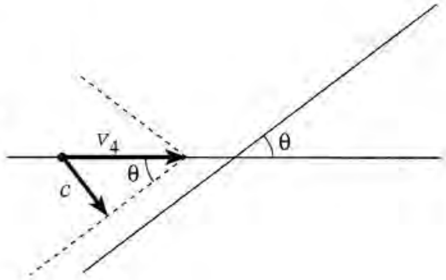


Figure 3

Now let's consider planes 1 and 4. In the reference frame of plane 1, plane 4 approaches at a velocity $v_4 - v_1$. In the triangle in figure 4, we can see that $\sin\alpha = 1/3$. The law of sines yields

$$\frac{v_1}{\sin(\theta - \alpha)} = \frac{v_4}{\sin\alpha},$$

or

$$\frac{v_1}{\sin(\theta - \alpha)} = \frac{c}{\sin\theta\sin\alpha}.$$

Solving this equation and substituting numerical values for the trigonometric functions of θ and α, we obtain the answer:

$$v = 1.1c.$$

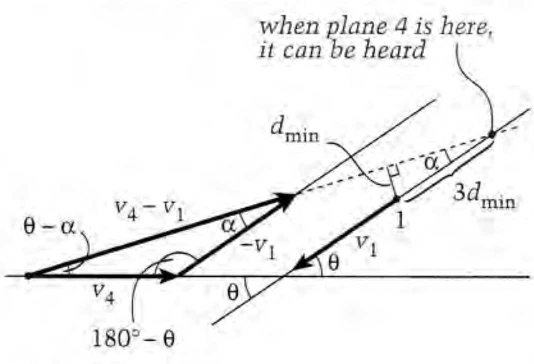


Figure 4

P205

Let the height of the object be h. Then the image's height is $H = mh$, so the ratio of the sizes of the images is

$$\frac{H_1}{H_2} = \frac{m_1 h}{m_2 h} = \frac{i_1/o_1}{i_2/o_2}.$$

Now we need to find o_1, o_2, i_1, and i_2. According to the lens formula,

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{F},$$

By the statement of the problem,

$$o + i = L.$$

Eliminating o from the above equations yields the square equation

$$i^2 - iL + FL = 0,$$

from which we get

$$i_{1,2} = \frac{L}{2} \pm \sqrt{\frac{L^2}{4} - FL}.$$

In addition, the principle of reversibility of optical rays results in $o_1 = i_2$ and $o_2 = i_1$. Therefore,

$$\frac{H_1}{H_2} = \frac{i_1^2}{i_2^2} = \left(\frac{L/2 + \sqrt{L^2/4 - FL}}{L/2 - \sqrt{L^2/4 - FL}}\right)^2.$$

Brainteasers

B201

Look at figure 5.

B202

The number will decrease by a factor of 50. In fact, $1/1996 = 0.0005\ldots$. If the first digit (5) were omitted, we would obtain the number

$$\left(\frac{1}{1996} - \frac{1}{2000}\right)\cdot 10 = \frac{1}{1996}\cdot\frac{1}{50}.$$

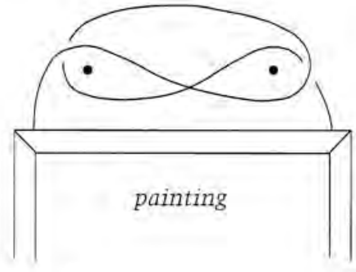


Figure 5

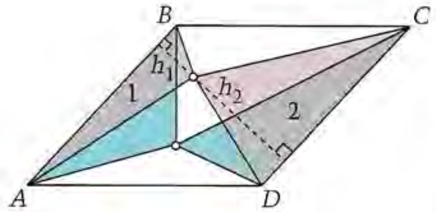


Figure 6

B203

The oil lies on top of the vinegar in the bottle. If the bottle is turned over, the vinegar will still lie beneath the oil, but it will be in the neck of the bottle. So you can pour out some oil from the bottle in the upright position. Then turn the bottle over, stopping the bottle with your finger, and release the vinegar in a controlled stream.

B204

If we add triangles 1 and 2 to the red ones (fig. 6), the total area will be equal to half the area of the parallelogram. (Indeed, $S_{ABM} + S_{CDM} = {}^1/_2 h_1 \cdot AB + {}^1/_2 h_2 \cdot CD = {}^1/_2 AB(h_1 + h_2) = {}^1/_2 AB \cdot h = {}^1/_2 S_{ABCD}$, since $AB = CD$.) The same thing happens if triangles 1 and 2 are added to the blue ones.

B205

First the son and daughter cross the bridge together. (It takes them 3 minutes). Then one of them—for instance, the son—returns to his parents. (Add 1 minute.) The father and mother cross the bridge together (10 minutes). The daughter comes back (3 minutes). The son and daughter cross the bridge again (3 minutes). Thus, the total time is $3 + 1 + 10 + 3 + 3 = 20$ minutes.

Kaleidoscope

1. The difference in the densities of water heated to various temperatures.

2. The volume occupied by the buckshot doesn't depend on its radius, so both boxes have the same mass.

3. The freshwater in the lock has a lower density than the saltwater in the ocean. The lock is opened when the hydraulic pressure is equal on both sides, which means that the level of freshwater is higher than that of the seawater. Water flows out of the lock and carries the ship with it.

4. Three quarters of the water's density.

5. It will not.

6. In the middle of a large lake, the ice doesn't rest on the lake bed but floats on the water. Since the ratio of the densities of ice and water is 0.9, the same proportion (0.9) of the ice's thickness is submerged in water. So the distance from the top surface of the ice to the water (and thus the rope's length) is 1 meter.

7. Throw a solid test piece into the melt. If it floats on the surface, the density will decrease when the melt hardens; if it sinks, the density will increase.

8. The aerometer with the thinner scaling tube.

9. The balance reading will increase if the mean density of the object is less than the density of the weights. In the opposite case it will decrease, and it will not change at all when the densities are equal.

10. If the object and the water expand equally when heated, the reading on the scale will not change. If the object expands less than the water, the reading will increase, and vice versa.

11. If the object is compressed less than the fluid when pressure is applied, then at some pressure its density will become less than that of the fluid, so the object will rise to the surface.

12. The volume will decrease.

13. If the initial temperature of the water was less than 4°C, it must be cooled; but when the opposite is true, the water must be heated.

14. Initially, water that is cooler than the water in the vessel is added; then water that is warmer is added.

15. See figure 7, where I is the isotherm, II is the isobar, and III is the isochore.

16. The jar with dry air is heavier.

17. The lift is proportional to the difference in the densities of the air

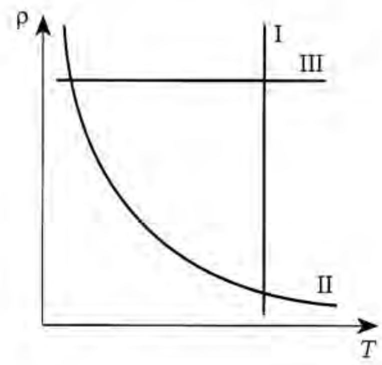


Figure 7

and the gas in the balloon. Since the density difference (and the density itself) is inversely proportional to temperature, the lift is greater in cooler air.

18. There are projections on the dust particles where the charge density is greatest and from which the electric charge quickly "flows off."

19. The current density is greatest in this area.

Microexperiment

Divide the mass, measured with a balance, by the volume, determined by the amount of water displaced in, say, a bathtub. Compare the resulting density with that of water.

Fair and squared!

1. The conditions of the problem will be satisfied if the equation

$$vt = l - ut + \frac{at^2}{2}$$

has two roots, both greater than u/a. Calculations yield

$$\sqrt{2al} - u < v < \frac{al}{u} - \frac{u}{2}.$$

2. Intuitively, charge q must be placed as shown in figure 8. Then, from Coulomb's law,

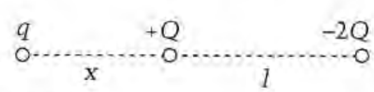


Figure 8

$$\frac{qQ}{x^2} = \frac{2qQ}{(x+l)^2},$$

or

$$x^2 - 2lx - l^2 = 0,$$
$$x = l\left(1 \pm \sqrt{2}\right).$$

The negative sign corresponds to placing charge q between $+Q$ and $-2Q$; of course, no equilibrium is possible in this case. So the answer is

$$x = l\left(1 + \sqrt{2}\right).$$

3. The equations of motion are

$$x = (v\cos\theta)t,$$
$$y = h + (v\sin\theta)t - \frac{gt^2}{2}.$$

When the rock hits the ground, $x = l$, $y = 0$. Combining the equations, we obtain a quadratic equation for $\tan\theta$ (see problem 1). Solving the equation, we get

$$\tan\theta = \frac{v^2}{gl}\left(1 \pm \frac{2gh}{v^2} - \frac{(gl)^2}{v^4}\right).$$

The answer exists if

$$(1 + 2gh/v^2 - (gl)^2/v^4) \geq 0,$$

or if

$$l \leq (v/g)\sqrt{v^2 + 2gh}.$$

The maximum possible l is thus

$$l = (v/g)\sqrt{v^2 + 2gh}.$$

Corrections

The Gallery Q in the last issue contained a misprint. The marine chronometer was invented, and George Washington was three years old, in 1735 (not 1835).

* * * * *

Several readers wrote to object to a passage in Lev Tarasov's "The Green Flash" in the January/February issue. Professor Andrew T. Young at San Diego State University writes:

> One cannot say the Sun is 2 degrees below the horizon [p. 39] because of the light-travel time; if such an argument were true, large stellar systems like globular clusters and galaxies, across which the light-time is many years, would be smeared out around a great circle of the sky by the Earth's rotation! Apart from a very small displacement due to the aberration of light, the Sun and stars really are about where they appear to be, because there is a continuous stream of light from these celestial sources to our eyes. The light-travel time has no effect on apparent position, except for the very small displacements of the objects themselves during the time of light propagation. The diurnal motion is just the reflex of the Earth's rotation, so it produces no such effect as you describe.

Prof. Young provided the Web address of his own green flash page (www.isc.tamu.edu/~astro/research/sandiego.html) and pointed us to another (www.bishop.hawaii.org/bishop/planet/Greenflash.html). He also took us gently to task for referring to John William Strutt as "Sir John Rayleigh." He was, of course, Lord Rayleigh.

BULLETIN BOARD CONTINUED FROM PAGE 56

16-year-old junior at Montgomery Blair High School in Silver Spring, Maryland. Eden invented Harmony Helper, a practical and entertaining musical training machine that teaches people to sing harmony. Three lights indicate whether a singer has produced a tone that is sharp, flat, or on pitch. Harmony Helper offers three octaves, and when its users can't get the right note, it features a "hint" button that, when pressed, plays it correctly.

Ashley, her parents, and her sponsoring teacher, Doris Sandoval, were guests of Duracell at the 45th annual National Science Teachers Association convention in New Orleans. She was honored along with the five second-place winners at an April 2 ceremony and participated in an April 3 invention workshop and birthday party for the competition.

Second-place awards were given to Scott Fulford, a junior at the Colorado Academy in Denver, for the Sensational Metronome; Hilde Anne Heremans, a senior at Detroit Country Day in Troy, Michigan, for the Compass for the Blind; Michael Kennedy, a senior at Fox Lane High School in Bedford, New York, for ClearView Goggles; Seung-Joo Lee, a sophomore at the Academy for the Advancement of Science and Technology in Hackensack, New Jersey, for TAPS—The Available Parking Spaces; and Leonard Shtargot, a senior at San Mateo High School who lives in Foster City, California, for the Portable 60-Hz Power Line Detector. Each is the winner of a \$10,000 bond and was honored in New Orleans along with their parents and teacher-sponsors.

In its fifteenth year, the Duracell/NSTA Scholarship Competition also announced ten third-place winners, who each receives a \$1,000 savings bond; 25 fourth-place winners, who received \$200 bonds; and 59 fifth-place winners, who received \$100 bonds.

Index of Advertisers

COWCULATIONS

Slipping silage

A corny caper

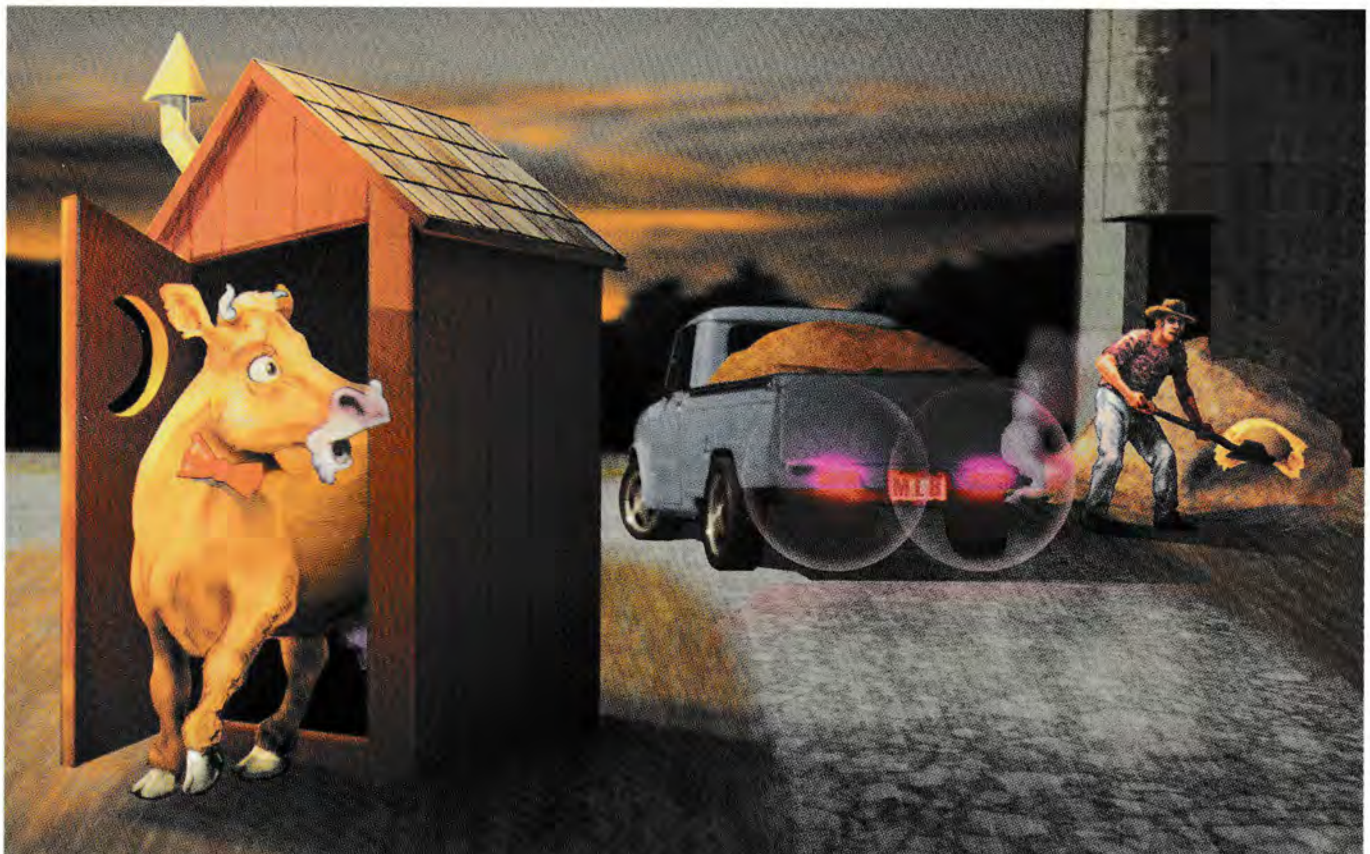
by Dr. Mu

WELCOME BACK TO COWCULATIONS, the column devoted to problems best solved with a computer algorithm.

A concrete silo stands tall next to Farmer Paul's barn. It was built in 1910 for the princely sum of 85 dollars, following plans published in the *Farmers' Institute Bulletin*. It has been used ever since to store our winter feed corn. Corn is an ideal silage since it is a source of energy, fiber, and protein. Without it, we cows would go dry in the winter and Farmer Paul would lose his monthly milk check. Keeping the silage safe and dry is a high priority around here. That's why we are determined to solve a most distressing corny caper.

As he does every year, Farmer Paul put up his feed corn last October in a huge silo capable of holding up to 120,000 pounds. In late fall, he began taking out a daily feed allotment of 300 pounds. Unknown to us, our silo was being broken into during the night and a fixed proportion removed. Each night, the silage thief stole exactly $(1/n)$th of the corn remaining in our silo. Oddly enough, this was always an integer number of pounds. Don't ask me to explain how I know this—cows are psychic.

Each day, Farmer Paul took out another 300 pounds for our feed, unaware of the nightly skimming. After five days and nights, I discovered the culprit quite by accident and turned him in to Farmer Paul.



Art by Mark Brenneman

We were all relieved.

At this point, Special Agent Mark from Barn & Silo Insurance entered the case. His investigation establishing the following sequence of events.

Each day Farmer Paul took 300 pounds for feed; each night the thief stole $(1/n)$th of the silage left in the silo (n is the same for each night). This was repeated for a total of five days and nights. The thief always stole an integer number of pounds. After the theft was stopped, Farmer Paul still had enough silage left to feed us for 210 days. Thank goodness!

COW 4. In order to determine the proper monetary settlement, Special Agent Mark needs to determine exactly how many pounds of silage were stolen.

Okay, cowhands, it's time to get up, get out, and start cowculating. COW 4 is waiting for you. Send your cowculation to drmu@cs.uwp.edu. To view all previous ruminations, take a peek at http://usaco.uwp.edu/cowculations.

Yes, you may write a better rhyme,
But can you do it every time?
Solve the COW and then we'll see
Just how clever you can be!

—Dr. Mu

Solution to COW 3

Last time we posed the following milk bottle problem: find an efficient algorithm for cowculating the number of ways of pouring 10 gallons of milk into bottles of the following sizes: two-gallon, gallon, half-gallon, quart, pint, and half-pint.

Whenever I have a big problem to solve I like to lie down, get comfortable, and think on it for a spell. Rushing out to write code before I've ruminated a bit is a big mistake. Don't make it.

Before we get started, let's introduce some notation to keep track of things in an orderly way. First there is the different bottle sizes we are using. It's easiest to express them in terms of how many half-pints each holds. Our half-pint, pint, quart, half-gallon, gallon, and two-gallon sizes are translated into a list of half-pints:

```
size={1,2,4,8,16,32};
```

Each element of the size list can be referenced with an index. So **size[[m]]** is the mth element in the list. For example:

```
size[[3]]
4
```

Now we define a two-dimensional array **Whey[m,n]**, which equals the number of ways you can distribute n half-pints of milk using any subset of the first m bottle sizes **Size[[1]]**, **Size[[2]]**, ..., **Size[[m]]**. Thus, if $m = 2$ and $n = 10$, **Whey[2,10]** is the number of ways 10 half-pints of milk can be distributed into any combination of half-pint or pint bottles. This number can be broken down into two cases. If we decide to fill a pint bottle, there are 8 half-pints left, which can be distributed into pints and half-pints in **Whey[2,8]** ways. If we don't use any pint bottles, the 10 units of milk can be distributed into half-pints in **Whey[1,10]** ways. Thus we arrive at the all important relationship:

```
Whey[2,10] = Whey[1,10] +
Whey[2,8]
```

But the same argument applies for any m and n. Ruminate on this one:

```
Whey[m_,n_]:= Whey[m-1, n]
+ Whey[m, n-size[[m]]]
```

Notice when you use one bottle of **size[[m]]**, the number of half-pints of milk left to distribute is **n-size[[m]]**. Got it? Thus we've transferred the problem of finding **Whey[m,n]** into two subproblems. But these subproblems can be pushed back to more subproblems, using the same relationship, until finally we are down to three simple cases.

1. As we all know, if we have only half-pint bottles to put the milk in, this can be done in exactly one way. This is expressed in Mathematica™ as

```
Whey[1,n_]=1
```

2. If you have a bottle of **size[[m]]** and this is exactly how much milk you have left (n = **size[[m]]**), then if you use at least one bottle of **size[[m]]**, it can be distributed in only one way—all in that bottle. Now there is no milk left to distribute. Thus when the milk is down to zero, we assign the value 1. This is expressed in Mathematica as

```
Whey[m_,0]=1
```

3. Finally, it's impossible to use a quart bottle when you have less than a quart of milk left to distribute. Of course, this applies generally to all bottle sizes whenever n < **size[[m]]**. This is expressed in Mathematica as

```
Whey[m_,n_] = 0 /; n<0
```

Combining these commands, we have the basic recursive Mathematica solution:

```
Clear[Whey]
Size={1,2,4,8,16,32};
Whey[m_,n_]:=0   /; n<0
Whey[m_,0]=1;
Whey[1,n_]=1;
Whey[m_,n_]:=Whey[m,n]=Whey[m-
1,n]+Whey[m,n-Size[[m]]]
```

So let's see how many ways we can distribute 10 gallons (160 half-pints) using all 6 bottle sizes:

```
Whey[6,160]
64350
```

Okay, now look at how the number changes as we increase the bottle sizes allowed from 1 to all 6:

```
Table[Whey[n,160],{n,1,6}]
{1, 81, 1681, 12341, 38841,
64350}
```

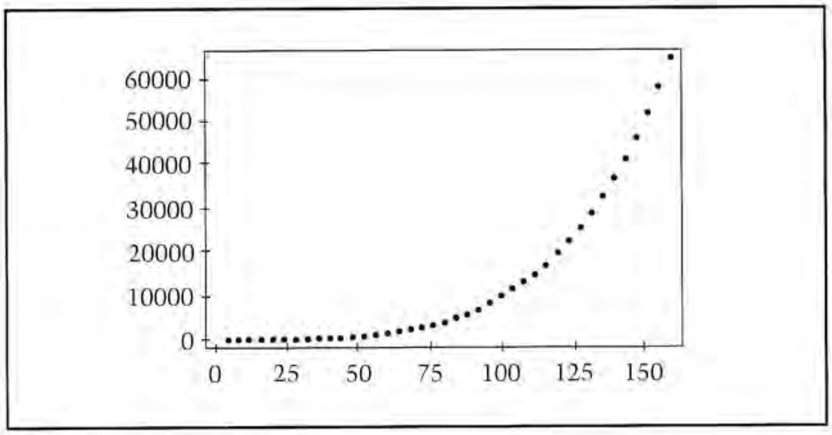
Finally, let's see what happens using all six bottle sizes by increasing the amount of milk from 1 gallon to 10 gallons, in jumps of a quart:

```
Wheys=Table[{n,Whey[6,n]},
{n,4,160,4}]
{{4, 4}, {8, 10}, {12, 20},
{16, 36}, {20, 60}, {24,
94}, {28, 140}, {32, 202},
{36, 284}, {40, 390}, {44,
524}, {48, 692}, {52, 900},
{56, 1154}, {60, 1460},
{64, 1827}, {68, 2264},
{72, 2780}, {76, 3384},
{80, 4088}, {84, 4904},
{88, 5844}, {92, 6920},
```

```
{96, 8148}, {100, 9544},
{104, 11124}, {108, 12904},
{112, 14904}, {116, 17144},
{120, 19644}, {124, 22424},
{128, 25509}, {132, 28924},
{136, 32694}, {140, 36844},
{144, 41404}, {148, 46404},
{152, 51874}, {156, 57844},
{160, 64350}}
```

The graph on the right shows all the wheys. It was created by the command

```
ListPlot[Wheys, Frame->True]
```



Postscript. Notice that this solution can easily be changed for farmers who live in other countries and use different bottle sizes. Just change the size list to match the containers you use. And it can be used for other applications. For example, if size = {1,5,10,25,50,100}, which represents the number of pennies in each U.S. coin from a penny to a silver dollar, then **Whey[6,1000]** is the number of ways you can make change for 10 dollars using U.S. coins. The answer is 2,103,596.

Come to the SILO!

As you can see, Mathematica is more than just the largest collection of mathematical functions ever assembled in one package for doing Mathematics. It is also a powerful symbolic programming language that can be used to solve your most personal barnyard problem, provided you know how it works. If you would like to learn how to cowculate in Mathematica, join me on the Internet at the Mathematica SILO (Summer Internet Learning Opportunity). During one week in July, I will ruminate on Mathematica fundamentals between milkings. You'll need to have access to the Internet and a willingness to do some mental chores. You don't need any prior Mathematica knowledge or even the software itself to participate. If you'd like join the herd at the SILO this summer, send an e-mail message to drmu@cs.uwp.edu.