QUANTUM

NOVEMBER/DECEMBER 1995

$5.95

Springer

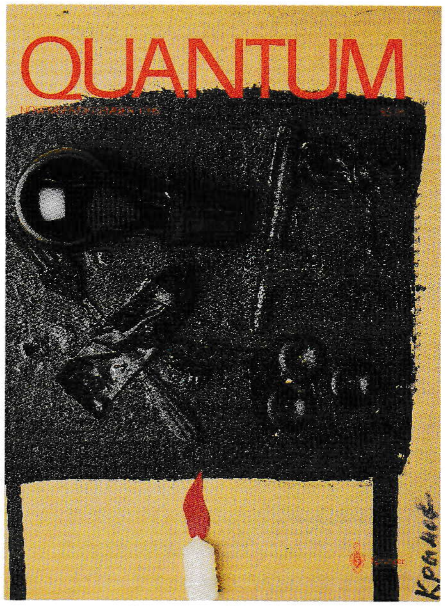
A View of the Mountain Pass Called the Notch of the White Mountains (Crawford Notch) (1839)
by Thomas Cole

IF THOMAS COLE (1801–1848), ONE OF THE FOUNDERS of the Hudson Valley school of American painting, had rendered this scene 30 years later, he might have included the cog railway that was built from Crawford Notch to Mt. Washington, about eight miles (13 km) to the northeast. Mt. Washington is the highest peak in the Presidential Range of the White Mountains and is noted for its extreme weather conditions—winds of 231 mph (372 km/s) were recorded there in 1934. Much of the surrounding area now lies within the White Mountain National Forest. Crawford Notch itself is a state park, and much of the natural beauty Cole found there has been preserved to the present day.

The nearby peak depicted here, towering over a tranquil landscape, is not being battered by high winds. But Cole has offered us a view of a different sort of meteorological phenomenon. He was obviously struck by the way the clouds have massed on one side of the peak, while the sky is clear on the other. Is this a common occurrence in the mountains? Why does it happen? Turn to "Smoky Mountain" on page 38 for some answers.

QUANTUM

NOVEMBER/DECEMBER 1995 VOLUME 6, NUMBER 2



Cover art by Dmitry Krymov

Our cover gives a whole new twist to the notion of a "candlelight dinner"! The fare isn't particularly appetizing—a razor blade, a light bulb, a pen . . . And besides, everything seems to have been burnt to a crisp!

What's being cooked up here? Maybe we have it backwards, and the objects were already black before they were subjected to the flame. Clearly there's more here than meets the eye (which we can assume is actually a small percentage of the incident light.) The article that begins on page 4 will certainly shed some visible radiation on the matter.

Indexed in *Magazine Article Summaries, Academic Abstracts, Academic Search, Vocational Search, MasterFILE,* and *General Science Source*

FEATURES

DEPARTMENTS

Bridging the gap

Between classical and quantum mechanics, between teaching and being taught

IN MY FIRST YEAR OF HIGH school teaching, back in 1957, I had a 14-year-old student who was remarkable in his abilities. He came to me frustrated because he could not understand some of the math in electronics. I agreed to help him. I gave him a college algebra text (Richardson, Prentice-Hall) and asked him to work every problem in the book. He did so, coming back to me often to discuss problems. Then I gave him a book on calculus and analytic geometry (George Thomas, Addison-Wesley) and asked him to do the same thing. He did! Within a short time he had prepared his first mathematical paper (a general trinomial theorem), and when he was 15 years old he was working at Midwest Research Institute doing mathematical modeling of the chest cavity using potential theory—the ultimate goal was a better method of interpreting electrocardiograms.

The student's name was Michael C. Mackey, an only child. After some time I realized he did not belong in high school, so we talked his parents into allowing him to quit high school before graduating. He did so and attended the University of Kansas for his undergraduate work and completed his Ph.D. in biophysics at the University of Washington in Seattle. He is now a professor of biophysics at McGill University. Mike has published extensively, and I have three of those publications: *Ion Transport through Biological Membranes*; *From Clocks to Chaos: The Rhythms of Life* (with Leon Glass); and a very special article published in *Reviews of Modern Physics* (Vol. 61, No. 4, October 1989, pp. 981–1015), "The Dynamic Origin of Increasing Entropy."

What makes this article so special? That will take some explaining.

Looking for connections

When I was first trying to learn quantum mechanics, I found myself constantly faced with questions for which no answers were likely to be provided by my professors. I wanted to *understand* quantum mechanics, and I had just completed the very solid grounding in classical course work in mechanics (Goldstein), electrodynamics (Jackson), thermodynamics (Callen), and statistical mechanics (the classical portions of Huang). I was particularly disturbed by this assertion: "It is quite clear that no deductive reasoning can lead us to [the Schrödinger wave equation]. Like all equations of mathematical physics, it must be postulated and its only justification lies in the success of the comparison of its predictions with the experimental result" (*Quantum Mechanics*, Vol. I, Messiah, p. 61). My professors were impatient with questions about the fundamental assumptions underlying the quantum mechanics. They wanted to get on with applying it to all of the various problems for which it was so enormously successful. Like other graduate students, I therefore dutifully pursued those applications in solid-state physics and in atomic and nuclear physics. I also found the enormous utility of quantum mechanics in properly explaining chemistry.

It was only years later, when I was no longer a student and my interest was again piqued by those fundamental questions, that I returned to examine the underlying hypotheses of quantum mechanics. Regrettably, this was also years after my mind was at its peak in terms of mathematical knowledge and skill, and I found it very difficult to work through the details of various mathematical derivations—like those you probably encounter in *Quantum*, but, I assure you, at a much higher level of abstraction.

My first interest was the simple matter of Planck's constant. A classical treatment of blackbody radiation leads to a disastrously wrong distribution (sometimes referred to as the ultraviolet catastrophe). Planck was able to derive the correct distribution only by postulating that the energy exchange between matter and radiation can occur not continuously as required by classical theory, but discontinuously, in quanta of size $h\nu$, where h was an empirically determined constant (which we now call Planck's constant). What was this h, and where does it come from? Can it somehow be derived from other considerations? Or is it some kind of universal constant like the G in the law of universal gravitation?

I was fascinated to find that Shin'ichiro Tomonaga (who shared the Nobel Prize in physics in 1965 with Richard Feynman and Julian Schwinger for developing quantum electrodynamics) had, in his own mind, raised and answered some of my questions in his text on quantum mechanics. I was particularly intrigued by the classical phenomenon of *adiabatic invariance*, in which he provides first a simple and, later in the text, an elaborate and general proof that establishes the fact that the energy E and the frequency ν are proportional, and that the ratio E/ν equals the adiabatic constant. Here's an example: if you pass the string of a simple pendulum over a pulley and very slowly pull the string, shortening the length of the swinging pendulum, the frequency increases and the amplitude, and therefore the energy, increase as well (since the energy is proportional to the square of the amplitude). But the ratio of energy to frequency of this pendulum remains constant. The value of the constant doesn't come out of the derivation, since it results as a constant of integration. This adiabatic invariance, so long as the changes imposed are done infinitely slowly, is always correct (Ehrenfest's adiabatic hypothesis), but if the change is sudden or over a short time, there is a transition, and the new state differs from the initial state by $h\nu$.

The other area that gave me a better connection between quantum mechanics and classical mechanics was Hamilton–Jacobi theory (Goldstein, *Classical Mechanics*, see especially pp. 307–314). This theory can make the solution of complicated problems in mechanics much easier, but there is an inverse relationship between making the solution simple to do and the mathematical sophistication required for that simple solution. The idea of Hamilton–Jacobi

CONTINUED ON PAGE 46

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 National Science Teachers Association is an organization of science education professionals and has as its purpose the stimulation, improvement, and coordination of science teaching and learning.

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 6 (6 issues) will be published in 1995–1996. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau of the Russian Academy of Sciences, as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000, telephone (703) 243-7100. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York NY 10010-7858.

Second class 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.

Subscription Information:

North America: Student rate: $15; Personal rate (nonstudent): $20; Institutional rate: $39; Single Issue Price: $5.95. 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$28 including surface delivery charges; institutional rate is US$39; airmail delivery is US$13 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.

Less heat and more light

Is the answer to high-efficiency illumination hiding inside the "ideal black body"?

by Y. Amstislavsky

THE STUDY OF HOW HEATED bodies give off light (heat or thermal radiation) played an important role in the development of physics. Suffice it to say that it was the study of thermal radiation that marked the beginning of the quantum era. One of the basic laws of this phenomenon was formulated by the German physicist Gustav Kirchhoff in 1859. And this law is the subject of our story.

Can black be bright?

First, a word about "black." In physics this notion has to do with the properties of physical bodies that allow them to absorb incident radiation (in the visible or any other region of the electromagnetic spectrum). The blacker the body, the more incident radiation it absorbs. The *ideal black body* (which we'll abbreviate IBB) absorbs all the incident radiation—and it does this for all regions of the spectrum. The opposite of "black," as we all know, is "white." The more a body reflects, the less it absorbs and the less black it appears.

Now, a word about "bright" bodies. A bright body is assumed to radiate appreciably (in the visible or any other region of the spectrum). The more the body radiates, the brighter it is. The opposite of "bright" is "dark." A bright body radiates a lot, while a dark one radiates only a little.

So "black" and "bright" (as well as "black" and "dark") are notions of a different "order." They describe different properties of bodies. An important question is whether these properties are related to each other. If such a relationship exists and is universally applicable, we could use our knowledge of a body's absorption characteristics to predict how that body will radiate under various conditions.

Common experience tells us that bodies radiate light differently at various temperatures. We need only recall how drastically the visible radiation of an electric lamp's filament changes with increasing current—from a barely noticeable red glow at T = 800 K to a dazzling white incandescence at T = 2,800 K. And you've undoubtedly seen (perhaps without really noticing) the transformations of a smoking flame, caused by changes in a body's temperature. Heated to high temperatures (of the order of 1,800 K), the "black" carbon particles (soot) glow brightly and together produce yellow tongues of flame (here black becomes bright), but the selfsame particles of soot, which haven't burned completely and have had time to cool, produce jet-black tongues of soot (here black becomes dark). Clearly we must compare the absorption and the radiation characteristics of various bodies at the same temperature.

In our everyday life we usually observe bodies at room temperature, and in so doing we often see black ones (in the visible range of the spectrum or even beyond it). It may be a black cloth, a piece of coal, something covered with soot, a bird's plumage, the opening of a cave or burrow, a nest tucked in the face of a cliff, and so on. Comparing a black body with nonblack bodies nearby, we see that it is dark, while the bodies nearby are much brighter. So we might subconsciously come to the conclusion that black is always dark, and the blacker something is, the darker. But we would be mistaken.

This paradox can be easily explained when two circumstances are taken into account. First, we always compare different bodies illuminated by electric light or daylight (nobody does it in a dark room or in the dead of night), and the light perceived by our eyes is by no means the thermal radiation of the bodies themselves, but rather the reflected light of some high-temperature source—the Sun or an electric lamp. Unlike black bodies, nonblack bodies dissipate this incident radiation strongly, and so they look like bright objects. Second, we usually observe the objects around us with the naked eye. This wonderful device is indeed sensitive, but only to visible light—that is, to a very small range of electromagnetic radiation. In this region of the spectrum, the amount of radiation given off by any object at room temperature is practically nil!

Analysis of the relation between the radiation and the absorption properties of bodies led Kirchhoff to an important conclusion, now known as Kirchhoff's law. It can be expressed as follows: *the more radiation a body absorbs at a given temperature, the more it radiates (that is, the brighter a black body is).*

To write this law in mathematical notation, let's formulate the concepts of absorptivity and intensity. The absorptivity $A_{\lambda,T}$ is the fraction of the incident radiation of wavelength λ that is absorbed by a given body at absolute temperature T. The absorbtivity is a dimensionless value varying from 0 to 1 for different bodies. For the absolutely white body,[1] $A = 0$, and for the IBB it is 1. The intensity per unit wavelength $I_{\lambda,T}$ is the power radiated by a given body in a unit interval of wavelength[2] near λ from a unit area of the body's surface at a temperature T.

Imagine that we have a number of

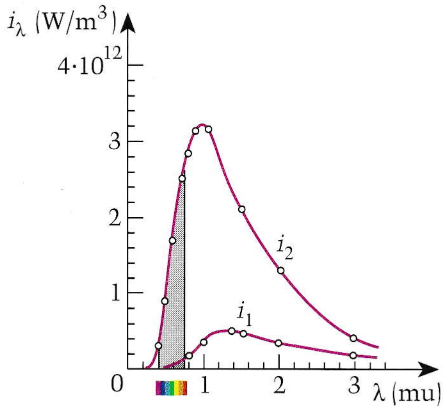


Figure 1

bodies, the IBB and ideal white body among them, with different $A_{\lambda,T}$. Let all the bodies be heated to the same rather high temperature T. Kirchhoff's law tells us that these bodies will radiate differently: the brightest will be the IBB, and its opposite will be absolutely dark. Let's denote by special symbols the absorptivity and intensity per unit wavelength of the IBB: $A_{\lambda,T}^{\text{IBB}} = a = 1$, $I_{\lambda,T}^{\text{IBB}} = i_{\lambda,T}$. It's very important that the radiation of the IBB not only is the most intense at a given temperature, but in addition is characterized by a strictly determined spectral composition. In other words, $i_{\lambda,T}$ is a universal function of λ and T. In view of the aforementioned formulation and these definitions, Kirchhoff's law can be written as

$$I_{\lambda,T} = A_{\lambda,T} i_{\lambda,T}. \tag{1}$$

Many attempts were made to find the function $i_{\lambda,T}$ theoretically, and finally the German physicist Max Planck found it in 1900:

$$i_{\lambda,T} = \frac{2\pi hc^2}{\lambda^5\left(e^{hc/\lambda kT} - 1\right)}, \tag{2}$$

where $c = 3 \cdot 10^8$ m/s is the speed of light in a vacuum, $h = 6.62 \cdot 10^{-34}$ J · s is Planck's constant, and $k = 1.38 \cdot 10^{-23}$ J/K is the Boltzmann constant.

When the temperature T of a body is held constant, the functions A, I, and i depend only on λ. In this case they are denoted by A_λ, I_λ, and i_λ, respectively. Figure 1 shows a graph of Planck's functions for two temperatures, T_1 = 2,000 K and T_2 = 3,000 K.

Can red become blue?

The most convenient model of the IBB is a small opening in a closed container made of a refractory or fireproof material. The body's temperature is changed by an electrical, gas, or other sort of heating device. The shape of the container is of no consequence.

Let's consider an IBB model in the form of a hollow sphere made of a refractory metal with a black inner surface (fig. 2). A pencil of light falls on the opening S from outside, passes into the container, and quickly becomes very weak after many reflections inside the sphere. Consequently, such a beam practically does not leave the sphere. This is true for any spectral interval and for any temperature. Thus, the hole S in our model behaves just like the IBB. It may seem that, absorbing everything, the opening radiates nothing. But it isn't so. True, the opening returns none of the original radiation. But, absorbing all the incoming radiation and "processing" it completely, the IBB generates its own radiation corresponding to the given temperature T. Kirchhoff's law says that, compared to the radiation of any other body heated to the same temperature T, the radiation of the opening S will be the strongest. Kirchhoff himself was the first to think of using a hole in a closed container as an ideal black radiator (in 1859). However, it wasn't until many years later that experimental studies of thermal radiation with this model of the IBB became traditional.

Let's look at three examples that illustrate Kirchhoff's law.

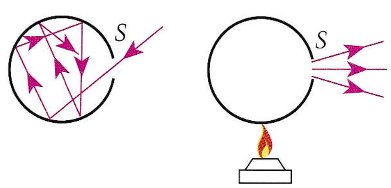


Figure 2

[1]Or for the perfect reflecting body (the ideal mirror). The absolutely white surface dissipates all the incoming rays homogeneously in all directions, while the ideal mirror reflects them according to the angle of incidence.

[2]To avoid confusion, keep in mind that "in a unit interval of wavelength" doesn't mean $\Delta\lambda$ is equal to, say, 1 m. The point is that the radiating power is proportional to $\Delta\lambda$—that is to say, $\Delta E/(\Delta S\Delta t) = I_{\lambda,T}\Delta\lambda$, where ΔE is the energy radiated in the wavelength interval $\Delta\lambda$ from an area ΔS in time Δt.

1. Suppose the outer surface of the sphere (fig. 2) is polished brilliantly and has the properties of a mirror across a wide spectral range. What will an observer see, looking at the side of the sphere where the hole S is located, in the following two cases: (a) at room temperature (T = 300 K) and (b) at the temperature of white incandescence (T = 3,000 K)?

(a) To exclude the masking effect of the initial radiation, we'll conduct the experiment in a darkened room. However, at room temperature we can't see anything with the naked eye, because according to equation (2), Planck's function i_λ at T = 300 K is negligible in the visible range of the spectrum ($0.4 \cdot 10^{-6}$ m $\le \lambda \le 0.76 \cdot 10^{-6}$ m). Nevertheless, a body heated to T = 300 K does radiate, and its radiation is maximal in the spectral interval where the function i_λ is maximal—that is, at the wavelength $\lambda_m = 10 \cdot 10^{-6}$ m (10 µm), which corresponds to the infrared region of the spectrum. Based on Kirchhoff's law, we can conclude that the opening S should "glow" brightly in the infrared range against the dark background of the rest of the sphere's surface. However, such a picture could be "seen" only by eyes that are highly sensitive to the 10-µm region of the spectrum or equipped with a suitable converter of infrared rays into visible light.

(b) When a body is heated to white incandescence, the function i_λ increases sharply in the visible range. As follows from equation (2), when T changes from T_1 to T_2, the intensity per unit wavelength of the IBB in the given spectral region λ increases by a factor of

$$\frac{i_{\lambda,T_2}}{i_{\lambda,T_1}} = \exp\left[\frac{hc}{\lambda k}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)\right].$$

(This follows directly from equation (2) when we take into account that $\exp(hc/\lambda kT) \gg 1$.) For the middle of the visible range ($\lambda = 0.55 \cdot 10^{-6}$ m) at T_2 = 3,000 K and T_1 = 300 K ($T_2/T_1 = 10$), we get

$$\frac{i_{\lambda,T_2}}{i_{\lambda,T_1}} = e^{78} = 10^{34}.$$

Clearly no converter is necessary in this case, and we'll see with the naked eye a bright opening S against the dark background of the body's surface ("the blacker, the brighter!").

2. Imagine that we have a colored refractory mineral that looks red when light passes through it (fig. 3a) because it absorbs virtually all of the light-blue, blue, and violet rays (for which $A_\lambda = 1$), but is translucent for the orange-red part of the spectrum ($A_\lambda = 0$). For the sake of definiteness, we'll assume also that the region of intense absorption practically does not shift along the spectrum as the temperature varies. Will this stone shine when heated to, say, T = 3,000 K? If so, will the radiance be colored? How?

At T = 3,000 K, Planck's function i_λ in the visible range is high enough, and according to Kirchhoff's law the heated mineral must radiate strongly at the absorption regions of the spectrum, where $A_\lambda = 1$:

$$I_\lambda = A_\lambda i_\lambda.$$

As the strong absorption relates only to the short-wave region of the visible spectrum, this radiance should be colored light-blue and blue (fig. 3b).

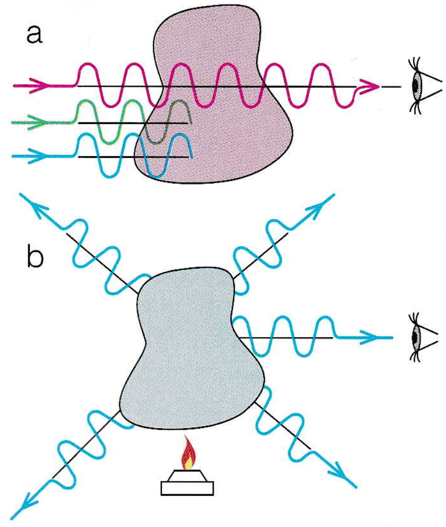


Figure 3

The American physicist Robert Wood performed an interesting experiment with fused quartz at the beginning of the 20th century. Fused quartz is transparent in the visible range of the spectrum, where absorption is practically absent—that is, $A_\lambda = 0$. So a column of fused quartz heated on a gas burner remains dark notwithstanding the high temperature. Wood "colored" the fused quartz by adding some neodymium oxide and prepared a homogeneous melt. Solutions of rare elements (including neodymium) are known to have narrow spectral regions of strong absorption in the visible range. So it wasn't unexpected that the colored fused quartz acquired some absorption bands. For neodymium they are located in the red, orange, and green regions of the visible range. If a column of such a melt is heated by the same burner, it will radiate strongly. Wood studied this radiation with a spectroscope and found that it indeed consisted of bright bands in the red, orange, and green regions.

3. Some crystals are known to absorb light polarized differently in various planes. This means that the parameter A_λ is larger for light oscillations in one direction and smaller for oscillations in the other (for the same values λ and T). A classic example of such a crystal is Iceland spar[3]—a one-axis colored (green

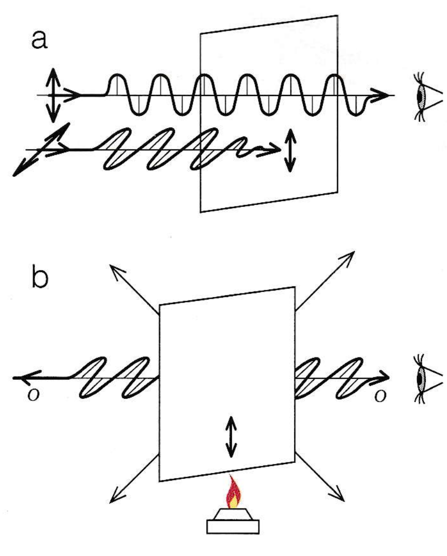


Figure 4

[3]See "A Polarizer in the Shadows" in the January/February 1994 issue of *Quantum*.—*Ed.*

most often) refractory crystal, which predominantly absorbs the ordinary oscillations (fig. 4a). A crystal of Iceland spar several millimeters thick is a natural polarizer: it transmits light with oscillations lying practically in the same plane. The problem is: how is the radiation from a heated crystal of Iceland spar polarized?

The parameter A_λ differs for ordinary (O) and extraordinary (E) oscillations. So, O- and E-oscillations will radiate differently as well. Kirchhoff's law says that the radiation from heated Iceland spar must be partially polarized with a marked prevalence of O-oscillations (fig. 4b). And this is what Kirchhoff himself observed during a qualitative investigation of the radiation from heated Iceland spar in 1859. Later quantitative studies at the beginning of the 20th century confirmed the relationship $A^E/A^O = I^E/I^O$.

Can 3 W equal 100 W?

One of the most important applications of heated bodies is using them as light sources. Think of the filaments in incandescent light bulbs, which continue to be a major source of artificial illumination. Turning now to practical applications of Kirchhoff's law, we should note that a very small fraction of the total radiant energy of the IBB corresponds to the visible range of the electromagnetic spectrum (the shaded portion of the graph in figure 1)—only 0.3% at T = 2,000 K. At T = 3,000 K this value increases to 3%, but it is still rather small. For a source made of tungsten, as we'll see below, the situation is a little better. However, because of inevitable additional losses due to heat conductivity and heat convection, the actual efficiency of modern tungsten filament lamps does not exceed 2–3%. This means that in the best case, about 97% of the supplied power goes "out the window."

Now let's turn our imaginations loose. Assume that a refractory, electrically conductive material is found (with a melting temperature T = 3,000 K) that absorbs light strongly in the visible range (A_λ = 1) at high temperatures and absorbs practically nothing in other regions (A_λ = 0). Kirchhoff's law predicts that filament lamps made of such a material would save huge amounts of energy. Indeed, the illuminating power of such a lamp consuming several watts would correspond to that of a modern 100-W lamp.

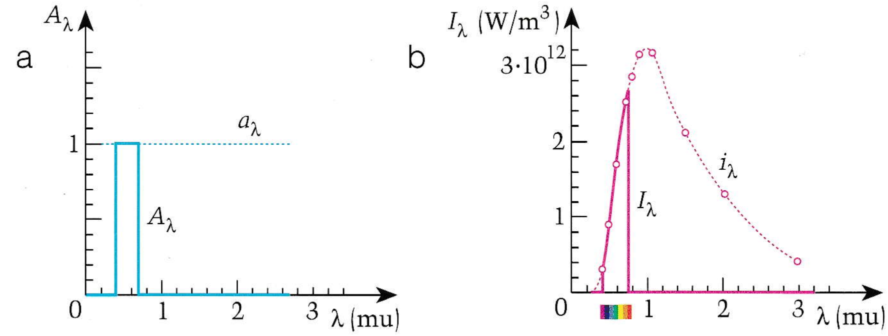


Figure 5

One problem that naturally arises in this context is to determine what form the function A_λ must take to give the maximum possible illuminating power when one of the following conditions is met: (1) the brightness of the source is maximum at a given temperature T; (2) the brightness is maximum and the color composition of the emitted light is the most comfortable for the human eye.

Let's denote these sources by S_1 and S_2. The functions A_λ we are interested in are shown in figures 5a and 6a, and the corresponding intensities are given in figures 5b and 6b. According to Kirchhoff's law, both sources are economically ideal since they emit only visible light (A_λ = 0 outside the visible range of the spectrum). But source S_1 is brighter at a given temperature, because it emits light as an IBB. However, this light has a red tint caused by the prevalence of warm colors—this is because I_λ increases sharply with λ. On the other hand, the warm colors are inhibited in the light emitted by source S_2 by the decrease of A_λ in the long-wave portion of the spectrum, so the energy distribution is similar to that of sunlight. The light emitted by source S_2 is "whiter" and thus more comfortable for the human eye. (Note that the function A_λ in figure 6a has been chosen in the simplest way—perhaps the reader can propose a better source. Some knowledge of visual perception and biophysics will help in this regard.)

Clearly the invention of lamps like S_1 or S_2 would bring revolutionary engineering changes in their wake. But to make such a lamp, a preliminary revolution in technology is necessary—one must know how to produce materials with the theoretically necessary characteristics: light absorption, melting temperature, and so on.

Unfortunately, the thermal light sources that currently exist differ greatly from the ideal supereconomical sources S_1 and S_2. Yet we can find some elements in common.

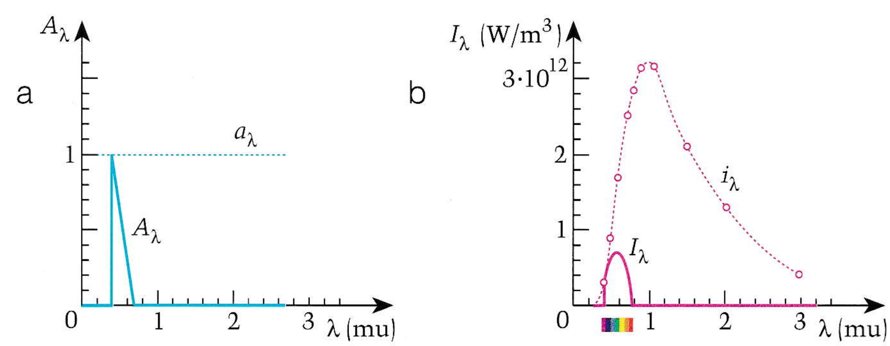


Figure 6

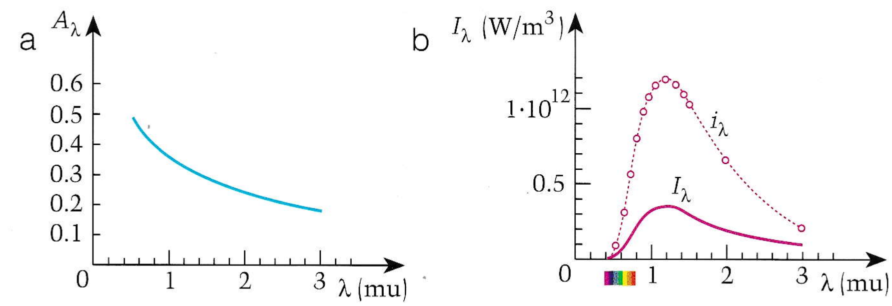


Figure 7

It's worthwhile to recall in this context the good old Auer burner. At the beginning of the 20th century this burner became widely known as a source of visible light and mid-range infrared rays, but now it's primarily of historical interest. Nevertheless, the properties of this light source are rather curious. The basic component of the burner is a mantle, heated by gas flame to 1,800 K. The grid is made of thorium oxide with a small amount of cerium oxide (0.75–2.5%). Cerium oxide provides the strong absorption (practically like that of the IBB) in the entire visible range of the spectrum and absorption of mid-range infrared waves comparable to the IBB, with almost no absorption in the near infrared region. Thus, the most energetic part of the IBB spectrum at T = 1,800 K is absent in the radiant light of the Auer burner, and so this burner radiates visible light and mid-range infrared rays strongly (almost like the IBB), while its total radiation is low.

Now let's turn to incandescent lamps. The basic material used as the working element of the modern filament lamp is tungsten. The A_λ and I_λ curves for tungsten heated to T = 2,450 K are shown in figure 7. We can see that at the same temperature tungsten radiates far less than the IBB across the entire spectrum. Also, it emits half as much in the visible range, which is not so good, of course, for a light source. However, in the infrared region tungsten radiates at 1/3 to 1/5 the output of the IBB, which is a good feature in a light source, because it translates to a drastic decrease in the amount of power consumed. Summing up, we can say that the tungsten filament lamp, while not as bright a source of light as the IBB heated to the same temperature, is somewhat more comfortable to the eye and certainly more economical. However, its efficiency is very low—as we noted above, it doesn't exceed 3%. Increasing filament lamp efficiency by even a few percentage points while retaining its advantageous features would be a significant achievement in lighting design.

In closing, I leave you with an easy question to check your undestanding. Figure 8 shows A_λ and I_λ curves for a certain heated body. What can you say about its radiation?

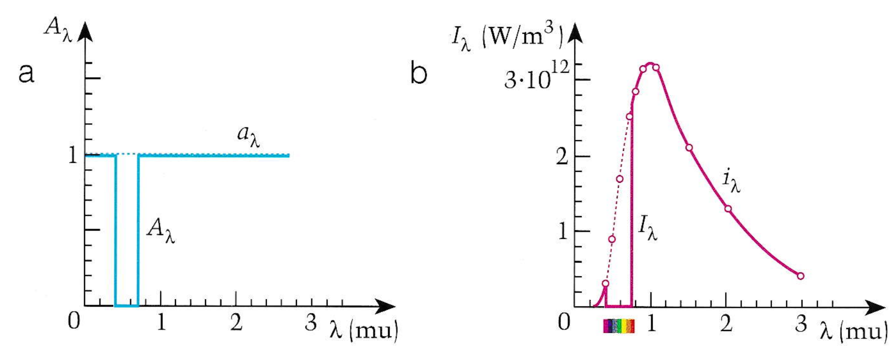


Figure 8

THE International Physics OLYMPIAD

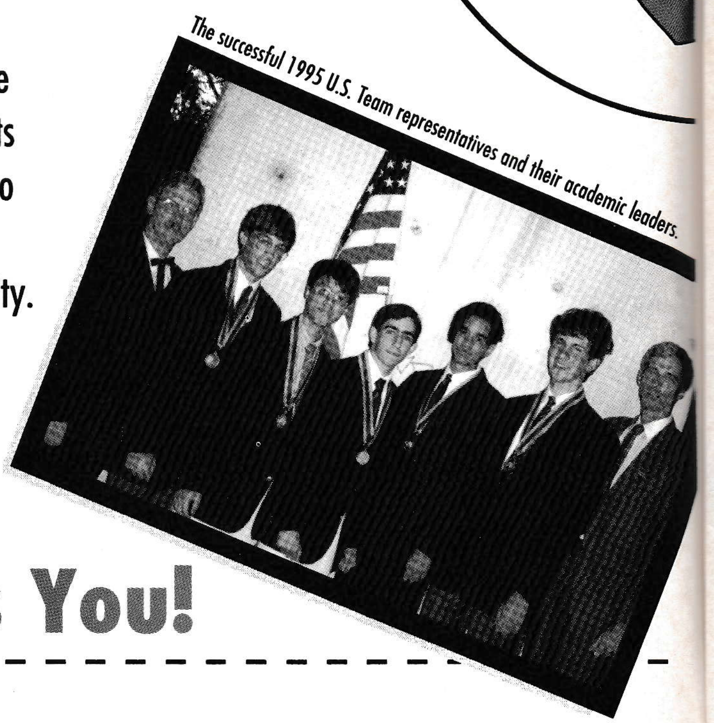
AMERICAN INSTITUTE OF PHYSICS

It's what the World Series is to baseball and the Academy Awards are to acting. It's the International Physics Olympiad (IPhO)–a global competition open to physics students in grades 9-12.

Each year the IPhO brings together qualified physics students from over 50 countries around the world to compete for recognition and medals. The U.S. Physics Team, organized by the American Association of Physics Teachers (AAPT) and the American Institute of Physics (AIP), has earned 31 medals since joining the IPhO in 1986.

In 1995 the U.S. Team placed second in the world at the Olympiad in Canberra, Australia–the team's best finish to date, acquiring four gold medals and one silver medal.

Not only do team members have the opportunity to compete with students from around the world, but they also experience a different culture and how it relates to the world community.



The successful 1995 U.S. Team representatives and their academic leaders.

We want you on our team! For more information, call AAPT at 301-209-3344, or return the form below.

The U.S. Team Needs You!

Application Request Form

For a complete packet of IPhO information and a U.S. Team application (deadline January 19, 1996), please fill out this form and return it to AAPT. The 1996 IPhO will be held in Oslo, Norway.

Name ______________________ Student ____ Teacher ____

School ______________________

Address (School or Home) ______________________

____

Phone ______________________

FAX ______________________

Return completed form to: AAPT, U.S. Physics Team, One Physics Ellipse, College Park, MD 20740-3845 ◆ FAX: 301-209-0845

Just for the fun of it!

B156

Two equations. Solve two simultaneous equations in integer unknowns T, W, O:

$$T - W - O = T \div W \div O = 2.$$

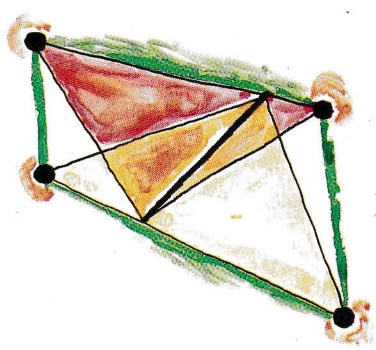
(Y. Alenkov)

B157

Financial dealings. The little tycoon Johnny says to his fellow capitalist Annie, "If I add 7 dollars to 3/5 of my funds, I'll have as much money as you do." To which Annie replies, "So you have only 3 dollars more than me." How much money does each have? (N. Antonovich)

B158

Experiment in the abyss. Two very long copper tubes are immersed in the ocean. One tube is hermetically sealed at both ends, the other has one end open. What will happen to the tubes in the depths of the ocean?



B159

Playing with triangles. Two isosceles right triangles are placed one on the other so that the vertices of each of their right angles lie on the hypotenuse of the other triangle (see the figure at left). Their other four vertices form a quadrilateral. Prove that its area is divided in half by the segment joining the right angles. (V. Proizvolov)

B160

Don't complete it! Two players take turns coloring squares in a 4×4 grid, one at a time. As soon as a player forms a completely colored 2×2 square, he or she loses. Who can force a win: the player who begins or the second player? (S. Tokarev)

ANSWERS, HINTS & SOLUTIONS ON PAGE 61

Art by Pavel Chernusky

Signals, graphs, and kings on a torus

Assuring clear communication by a variety of transmitters

by A. Futer

IN THE MORNING the phone rang. I leaped up out of bed, grabbed the receiver, and yelled:

"Hello!"

The receiver replied:

"Why are you grunting?"

"What do you mean, 'grunting'? I'm not grunting," I said.

"Stop grunting! Speak human!" Mishka cried.

"But I am speaking human! Where do you get this 'grunting'?"

"Come on, stop fooling around! Anyway, I just can't believe you'd bring a piglet to your room."

—Nikolay Nosov, "The Telephone"

WHEN TWO PEOPLE HAVE an unhurried conversation, the words of either of them are usually absolutely clear to the other one. They rarely ask each other to repeat or clarify an incomprehensible word. But if the same conversation is conducted by telephone, each person may find some of the words hard to catch, especially if the line is noisy because of a bad connection. Sometimes the missing words lead to a loss of com-

prehension. And if we try to transmit by phone arbitrary sequences of letters, like *rptfffskllaaaam...*, rather than intelligible words, we'll see that some letters often get mixed up—for instance, *f* and *s*, *m* and *n*, and so on.

Practically any means of communication—telephone, radio, binoculars—corrupts the signal. There are various ways to combat this phenomenon, but they all impair the efficiency of communication. For instance, instead of separate letters sent by telephone, we can send entire words that begin with that letter—for example, sending *b* as *Bill*, *g* as *Gregory*, and so on. But then instead of one letter we'd have to send 4 or 7 letters. Alternatively, every signal can be repeated many times, which was the case when the first photographs of the far side of the Moon were transmitted to Earth in 1959. But the time it takes to do this increases by the same factor.

In this article we'll consider one economical method of noise control.[1] The idea behind this method is quite simple: since we know which signals can be mixed up, we can send only one signal from each such group and ignore all the rest. For instance, of the three sounds *f*, *s*, and *x*, we can send only one by phone—say, *s*.

[1]This problem was first explored by the prominent American mathematician Claude Shannon.

Error graphs

Any message usually consists of separate "elementary" signals: words (in a telephone conversation), letters (in semaphore), or other signs (the dot and dash in telegraphy). These elementary signals constitute the set S, called the *input alphabet*. Let's picture each signal from S as a small circle and join every two circles whose corresponding signals can be confused with each other during transmission. In general, if you have a set of points and some of them are joined by lines, you have a *graph*. The points are called the graph's *nodes*, and the lines between them are the graph's *edges*. Two nodes, v and w, joined by an edge are said to be *adjacent*—this is denoted $v \sim w$. The graph on the set S constructed above is called the error graph of the transmitter in question.

By way of example, consider an electric clock with the minute hand changing its position in leaps—as soon as a full minute ends, the hand jumps to the next division on the dial. If the clock is far from us, we can't exactly determine the position of the minute hand. But suppose our error isn't too great—say, not greater than a minute. Then the input alphabet will comprise 60 elements—60 possible positions of the minute hand (each of which corresponds to a certain point on the dial's circumference)—and the error graph G will be a regular 60-gon: every point is joined to its two neighbors.

Now let's try to understand what should be done to correctly identify the greatest number of different readings of the minute hand. Adapt the clock so as to make its minute hand leap over every two minute divisions—that is, so as to indicate only an even number of minutes. Then we'll never mix up any two indications of the clock, because we don't allow for the errors of two or more minutes. So if we agree to confine the set of transmitted signals only to 30 even numbers from 0 to 58, they will always be correctly received. On the other hand, thirty is obviously the maximum number of faultlessly distinguishable signals: any 31 nodes in our graph G will

Art by Sergey Ivanov

contain at least two adjacent nodes (prove this!).

In this example, we've constructed a subset M of the graph's nodes with the following property: *any two nodes from M are not joined by an edge.* Any subset with this property is called an *independent set* of nodes. In our example, any set of even indications of the minute hand—say, {2, 8, 34, 52, 56}—or the set of indications divisible by five—{0, 5, 10, ..., 55}—are independent, as are many others.

If the number of nodes in an independent set M is the greatest among all independent sets in a graph, we call this set M a *greatest independent set* (GIS), and its number of nodes $\alpha(G)$ is called the *independence number* (IN) of the graph.

The graph G considered above has many independent sets, but only two greatest independent sets (the set of all even numbers from 0 through 58 and the set of odd numbers from 1 through 59), so $\alpha(G) = 30$.

If G is the error graph of a certain transmitting device T, then $\alpha(G)$ is the maximum number of signals that can be transmitted through this device without mixing them up. Therefore $\alpha(G)$ is also called the *throughput* of the device T.

Problems

1. Find the GIS and IN for the graphs in figure 1.

2. A number n of points are plotted around a circle. Each of them is connected to $2k$ points (consisting of the k nearest points on either side). Find the independence number of this graph.

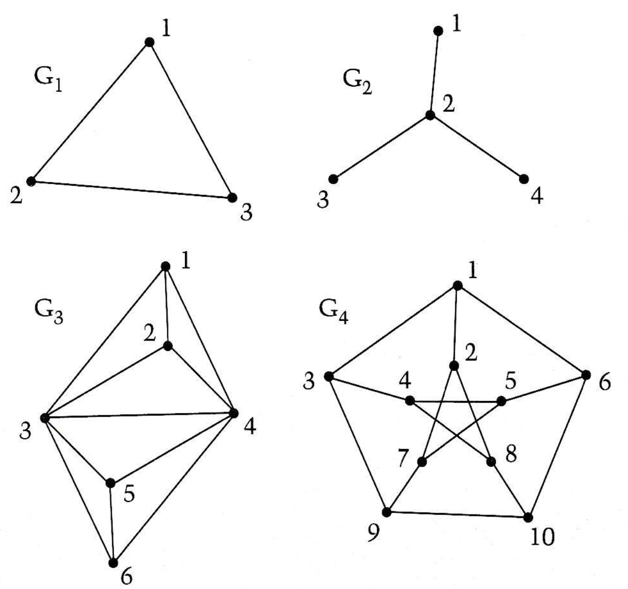


Figure 1

An alphabet squared

Any transmitter T can be described in terms of its input alphabet S, the error graph G, and the independence number of the graph G (or the throughput of T) $\alpha(G)$. But what should we do if the transmitter T is given and fixed once and for all, but its throughput is insufficient for us? For instance, what if T is a telegraph, which can send only dots and dashes, but we want to send letters? The solution is no secret—it's the well-known Morse code. We have to send packs of several signs (dots and dashes) through T and consider any such pack a single signal.

Let's first try to send pairs of successive signals from S through the same device T and see by how much this will increase our possibilities. We can assume that now we've got a new transmitter T^2 whose input alphabet S^2 consists of two-letter signals (v_1, v_2), where v_1 and v_2 are elements of S (for convenience, from here on we'll refer to the elements of any alphabet as "letters").

Let's try to determine the throughput of the transmitter T^2. To this end we must draw its error graph, denoted by G^2. When can a signal (v_1, v_2) be confused with (w_1, w_2)? Clearly, if (and only if) one of these conditions is satisfied:

(a) $v_1 = w_1$, $v_2 \sim w_2$;
(b) $v_1 \sim w_2$, $v_2 = w_2$;
(c) $v_1 \sim w_1$, $v_2 \sim w_2$.

So two nodes (v_1, v_2) and (w_1, w_2) of the graph G^2 are joined by an edge if the graph G has edges v_1w_1 (or $v_1 = w_1$) and v_2w_2 (or $v_2 = w_2$).

For instance, if a graph G has only two vertices joined by an edge, then G^2 is a square with diagonals (fig. 2).

Problem 3. Draw the square graph G^2 for each of the graphs G in figure 3.

So G^2 can be thought of as a "fiber" graph, in which every vertical or horizontal "fiber" is the graph G, and each edge ab of G corresponds to a square with diagonals in G^2.

Now let's try to find an independent set in G^2. If M is such a set in graph G, then no elements of M are mixed up with one another when transmitted by the device T. Then none of the letters from M is corrupted when pairs of these letters are transmitted, so in the alphabet S^2 the pairs (a, b), with $a \in M$, $b \in M$, aren't mixed up. The set of pairs (a, b), where $a \in M$, $b \in M$, is denoted by M^2. Obviously, the number of elements in M^2 is p^2, where p is the number of elements in M. Therefore, the throughput of A^2 is not less than the square of the throughput of the device A—that is, in terms of graphs,

$$\alpha(G^2) \geq (\alpha(G))^2.$$

However, it should be noted that this increase in the throughput isn't gratis—it's achieved at the expense of the transmission speed falling by a half.

In actual telegraphy, letters are transmitted somewhat differently. There are three, rather than two, elementary symbols in the telegraph alphabet (Morse code)—dot, dash, and *blank* (an increased interval between signals). So it's not necessary to use the same number of signs (dots and dashes) for all letters—it's reasonable to denote more frequent letters by shorter strings of symbols, reserving longer strings for rarer

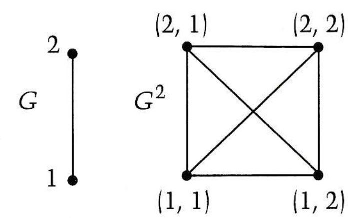


Figure 2

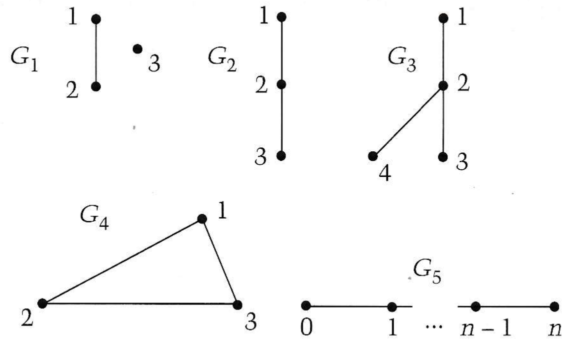


Figure 3

letters. For instance, there are 33 letters in the Russian alphabet. If they were all to be represented by the same number of symbols, we'd have to use 6-symbol sets (because there are only $2^5 = 32$ sets of five dots and dashes). In fact, no more than four symbols suffice to denote almost all Russian letters: $2^1 + 2^2 + 2^3 + 2^4 = 2 + 4 + 8 + 16 = 30$. Of the remaining three letters, two are simply identified with certain pairs of letters out of the thirty, and one letter—the rarest one—has a 5-symbol code. The sets of more than four symbols are used for figures and punctuation marks.[2]

Kings on a torus

Let's make the "linear" graph G_5 in figure 3 a little more complicated by gluing together its endpoints 0 and n. This gives an n-gon $0\ 1\ 2 \ldots (n-1)$. Denote it by P_n. The square P_n^2 of this graph is obtained from the square of the linear graph (which is a grid of $n \times n$ squares with diagonals) by gluing together its opposite sides. But this operation turns a square (a normal, genuine square) into a torus! (See the Kaleidoscope in the March/April 1994 issue of *Quantum*.) So P_n^2 can be viewed as a grid on a torus consisting of n^2 "unit squares" with diagonals (the red grid in figure 4).

The vertices of these squares can be identified with the centers of the squares of an $n \times n$ toroidal chessboard (shown in black in figure 4). Then adjacent nodes of P_n^2 will correspond to chessboard squares connected by a chess king's moves. So the independence number for P_n^2 is equal to the maximum number of kings not attacking one another on the $n \times n$ toroidal chessboard. This number is studied in problem M156 in this issue of *Quantum*.

To get a clearer idea of the structure of GIS for the graph P_n^2, solve the following problems.

Problems

4. Let M be the greatest independent set for P_n^2 (or the toroidal chessboard). Prove that (a) if $n = 2a$, then M can be chosen in the form M_1^2, where M_1 is any GIS for P_n, and $\alpha(P_{2s}^2) = s^2$; (b) if $n = 4s + 1$, then there are exactly s nodes (kings) from M in each horizontal or vertical file of the board; (c) if $n = 4s + 3$, then each vertical or horizontal file contains either s or $s + 1$ elements of M.

5. Define a *cyclic shift* of the graph P_n^2 on a torus as a "translation modulo n," which takes any point (x, y) into $(x + s, y + t) \pmod n$ with certain fixed s and t—that is, into the point (x', y'), where x' and y' are the remainders of $x + s$ and $y + t$ when divided by n. Prove that any GIS for P_5^2 can be reduced to the form shown in figure 5 (on the next page) by way of an appropriate cyclic shift and, perhaps, a line reflection of the board.

Up the dimensionality ladder

Let's develop our transmitter further: consider the device T^k for an arbitrary natural k. That is, we'll use the transmitter T to send packs of k letters at a time from the original input alphabet S. By analogy with the case $k = 2$, we can construct the error graph G^k of the transmitter T^k. The set of its nodes is the alphabet S^k consisting of all k-letter sets $(v_1, v_2, \ldots, v_k)$, where all $v_i \in S$. It's also clear which nodes in this set are adjacent—that is, which signals $(v_1, \ldots, v_k)$ and $(w_1, \ldots, w_k)$ can get mixed up. This happens when the letters in each "coordinate" can be confused, which means that for any i $(1 \le i \le k)$ one of the two conditions listed above must be satisfied: either $v_i = w_i$ or $v_i \sim w_i$ (if $v_i = w_i$ for all i, the two sets coincide).

It's rather difficult to find the throughput of the transmitter T^k exactly in the general case, but we can estimate it from below.

Problem 6. Prove that $\alpha(G^k) \ge (\alpha(G))^k$.

As for further improvements . . . Even in one of the simplest cases,

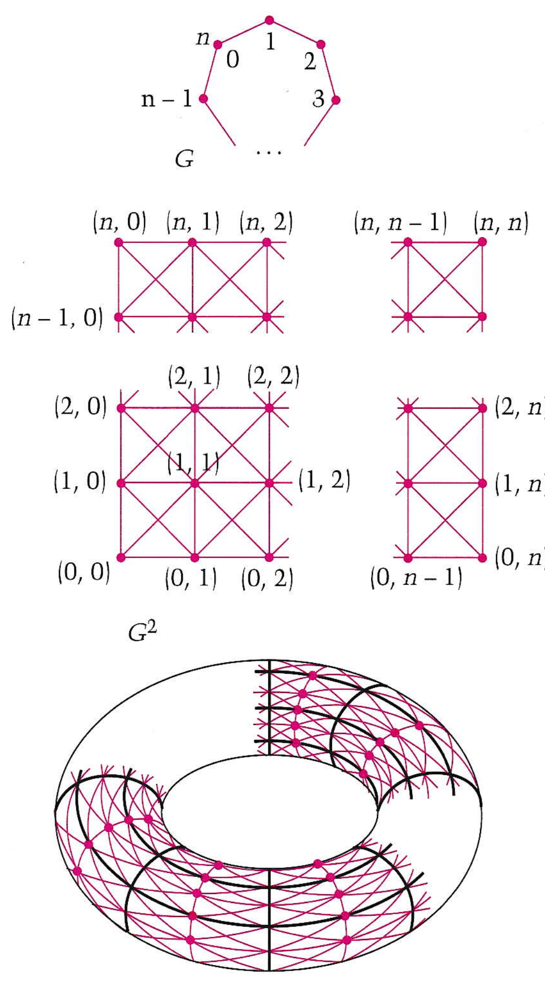


Figure 4

---

[2]"Uniform" codes, with equally many symbols per letter, have their own advantages and are also used in telegraphy, perhaps even more often than Morse code.—*Ed.*

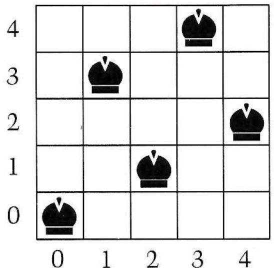


Figure 5

where the error graph is an n-gon P_n, we don't know much about $\alpha(P_n^k)$. Now I'll summarize the results obtained up to now.[3]

A square or a chessboard with its opposite sides glued together is a *two-dimensional torus*. Similarly, the kth power of the n-gon (the graph P_n^k) can be called a *k-dimensional torus* of size n. The nodes of the graph P_n^k can be represented as sets of k integers $(x_1, x_2, \ldots, x_k)$ where each x_i varies from 0 to $n - 1$.

Problem 7. Find the number of nodes in P_n^k.

By definition, two sets $(x_1, x_2, \ldots, x_k)$ and $(y_1, y_2, \ldots, y_k)$ are adjacent in the graph P_n^k if each pair of coordinates (x_i, y_i) are "neighbors" in P_n—that is, differ by no more than 1 (mod n): for any i, $|x_i - y_i| \in \{0, 1\}$ (mod n). For instance, in figure 6 you can see all the neighbors of the set (0, 1, 2) for the case $k = 3$, $n = 5$.

Problems

8. Find the number of neighbors of each node on the k-dimensional torus.

9. Prove that $\alpha(P_n^k) \le [(n/2)^k]$, where $[x]$ is the integer part of a number x.

10. Find $\alpha(P_n^k)$ for an even n.

For odd values of n the following inequality holds:

$$\left(\frac{n-1}{2}\right)^k \le \alpha\left(P_n^k\right) \le \left[\left(\frac{n}{2}\right)^k\right],$$

and sometimes even the exact value of $\alpha(P_n^k)$ can be found.

THEOREM. *If $n - 1$ is divisible by 2^k, $k \ge 1$, then*

$$\alpha\left(P_n^k\right) = \frac{n-1}{2^k} \cdot n^{k-1}.$$

A complete proof of this theorem is rather long, so I'll only present an independent set with $\alpha(P_n^k)$ elements. Take an arbitrary set of values for the $k - 1$ first coordinates $(x_1, x_2, \ldots, x_{k-1})$, $0 \le x_i \le n - 1$, and compute $r = (n - 1)/2^k$ values for the kth coordinate using the formula

$$x_k = 2l + r(2^{k-1}x_1 + 2^{k-2}x_2 + \ldots + 2^1 x_{k-1}) \pmod{n},$$

where l is any integer from 0 to $r - 1$. This gives $n^{k-1} \cdot r$ nodes $(x_1, \ldots, x_k)$ of P_n^k that form an independent set.

If you draw the set of nodes whose coordinates satisfy the last formula for the case $k = 2$, you'll get the answer to problem M156 for the case of $n - 1$ divisible by four.

These results allow us to write out a table of the values of $\alpha(P_n^k)$ (fig. 7). Some of the entries in the table have been filled out "theoretically"; the rest were obtained using computers. You see that the proportion of blank spaces becomes greater as n and k grow.

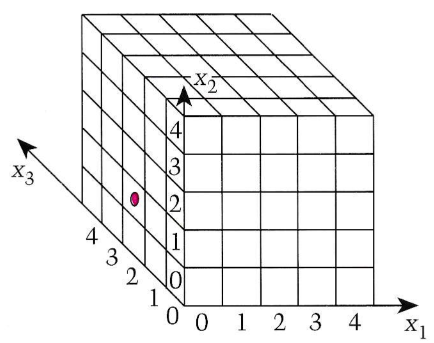
Perhaps some of our readers will manage to fill the gaps in this table and solve the problem of calculating the throughput of transmitters.

ANSWERS, HINTS & SOLUTIONS ON PAGE 61

$n \backslash k$	1	2	3	4	5
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	2	4	8	16	32
5	2	5	10	25	
6	3	9	37	81	243
7	3	10	33		
8	4	16	64	256	1024
9	4	18	81		
10	5	25	125	625	3125
11	5	27			
12	6	26	216	1296	7776
13	6	39			
14	7	49	343	2401	16807
15	7	52			

Figure 7

[3]That is, at the time of the original publication of this article in *Kvant*, about 15 years ago. It's very likely that many of the blank spaces in the table of results given here have been filled in by now.—*Ed.*



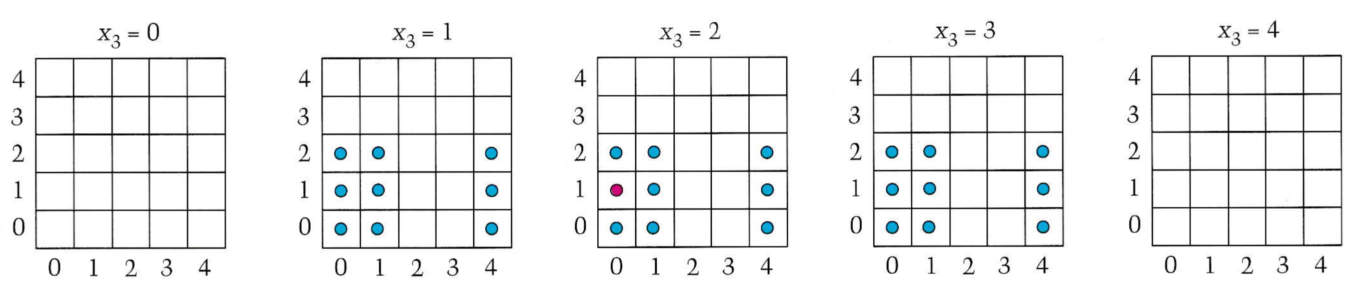


Figure 6

The importance of studying the physics of sound insulation

A detective story told in the words of Dr. John Watson

by Roman Y. Vinokur

I TRY TO TAKE IT EASY, BUT day in and day out I have to hear a tremendous rumble when a tram passes by my house. My nerves are on edge, and besides, my office is not the best place for curing the sick. I have to insulate my unfortunate patients from the traffic noise entering the room from the street below. The thick brick walls of the house reflect all the sound, but the window lets most of it through, even if it is shut. To my deep regret, I was not a diligent student when I studied physics at the university. That branch of science holds the keys to many nettlesome problems and mysteries.

Actually, why not take the advice of the talkative glass-cutter who called on me last night? I didn't invite him—he came of his own accord, just to improve the sound insulation of my window. He spent a great deal of time standing at the window. I believe he left rather late, when the windows across the street fell dark and my neighbors went to bed. He recommended double glazing rather than the single pane now installed. Two panes, each 3 mm thick, with an air gap of 15 mm . . .

"I'm sorry to interrupt, but such a recommendation is worthless," said Sherlock Holmes in an even tone. "The thickness of the air gap, in your case, needs to be much more than 15 mm. Trust me."

I turned my head to look at him. Holmes was sitting with his newspaper at my study table. I had forgotten he was there, because he seemed deeply involved in his reading. He noticed my questioning look and shook his head with a kind smile.

"No, my dear Watson, I am not a mind reader. I am just able to observe and analyze. By deduction, I arrive at the correct conclusion.

"First of all," he continued, "I knew that the constant noise coming into your room drives you crazy. When the tram passed a few minutes ago, I saw that you were very upset. Then, you began to poke around the window, not merely looking through it but touching it. At the same time, I found in yesterday's newspaper—utterly by chance—the address of the glass-cutter's shop. You underlined it and wrote a note about 'double glazing to reduce noise: 3 + 15 + 3.' By my lights, this solution[1] is not correct. That's all there is to it."

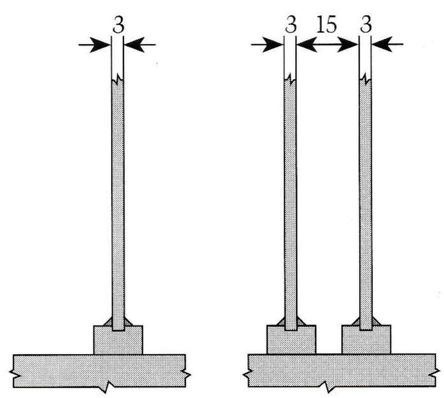


Figure 1

"Yes," I replied, "that sounds reasonable. The glass-cutter came to my place last night and urged me to do as he advised. But why did you reach such a sudden conclusion that his recommendation was faulty?"

"I had taken it upon myself to help you reduce the noise in your

[1]See figure 1.—*Ed.*

Art by Sergey Ivanov


S. Ivanov

office. I have read some books on acoustics and mechanics, and now I am quite conversant with the current thinking on sound insulation. The same does not hold for the glass-cutter, in my opinion."

My friend's words encouraged me greatly. Holmes stood and walked to the window. I watched him. He seemed rather enthusiastic.

"This problem seems difficult at first, but you will soon get the hang of things. Have you ever tried your hand at differential calculus or vector analysis? No? Well, perhaps logarithms? Fine, let's leave it at that. And what do you know about sound?"

"I feel that my experience in this field is close to nil," I replied uneasily. "Sound is . . . sound is wave motion. It occurs when the particles of air, for instance, are set in motion by an oscillating body—what the technical people call the 'sound source'—and the movement gradually spreads to particles further from the source. The most elementary type of sound wave," I rattled on, warming to my subject, "is one whose particles vibrate at a single frequency. It's called a 'harmonic wave.'"

Holmes praised my answer. Despite its sketchiness, he was being quite sincere. I know that he is not keen on making a fool of somebody. He just wanted to see how much I already knew about acoustics.

"You could think of a sound wave as being like a bullet," he began, "and a wall as being like . . . a wall. The image doesn't simplify the phenomenon of sound transmission through a wall, but it's a crude mechanical analogy to simplify the explanation. It goes without saying that any analogy leads one astray if one pushes it too far. Nevertheless, analogies are useful when we seek to illuminate complicated phenomena.

"If a bullet strikes a wall," Holmes went on, "three things can happen. One, the bullet might be deflected (or, we might say, 'reflected'). Two, it might embed itself (that is, be 'absorbed') in the wall. Three, it might pierce the wall (in other words, the bullet is 'transmitted' through it). A similar situation arises when a sound wave comes up against a partition (say, a window or a wall), but the result is more exotic. Both reflected and transmitted waves form simultaneously, and on top of that, acoustic energy is absorbed by the partition. Do you find this scenario strange, my friend?"

"Well," I searched for a reply, "it's certainly not obvious to me. First of all, the idea of a wall being punctured by a sound flying through the air strikes me as highly improbable. Otherwise the walls around us would be full of holes, but I don't see any such evidence . . ."

"Precisely! A world where sound waves could pierce a wall would not be a very secure one, now would it? No, things happen a bit differently. The incident and reflected waves merely shake the wall. The wall vibrates and thereby radiates sound in two directions. The radiated waves are less intense than the wave that struck it.

"By the way," Holmes continued, "this phenomenon is not unique. Experiments of this sort are actually quite common. When you strike the prongs of your tuning fork, they vibrate and act on the surrounding air particles. To make a long story short, they radiate sound waves. The role of the incident wave, in this case, is played by whatever you use to strike the tuning fork.[2] Does this make sense to you?"

"Not only does it make sense," I said, "I find it quite interesting. You have a knack for teaching people, Holmes. Now, what I'd like to know is, what kind of walls provide the best sound insulation. I find this issue more pressing than any theory, really."

. . . I don't think I'd be able to get by without Holmes in my life. A few days ago, he left the hospital, and he still looked rather weak. His health is improving, thank God. He had tried to defend a woman against three young ruffians who were assaulting her. Within moments, he looked more dead than alive. The thugs proved to be too much for him. Suddenly, a young man showed up and turned the tables on the three attackers. His blows broke their jaws and noses and knocked them to the ground. This young gent might well be the best boxer in all of London, but to everyone's surprise, he ran off just as the police arrived at the scene of the crime. They found four men lying unconscious and a woman who seemed more excited than frightened. She admitted, though, that she was not the only person who did battle with the hooligans . . .

"Watson! You seem lost in thought. Have you been following what I've been saying?" He looked at me narrowly.

"Yes, of course, Holmes! Please, go on. I'm all ears."

"Let's consider the simplest case of airborne sound transmission through a single wall.[3] The pressure acting on the left side of the wall is uniformly distributed over its surface and varies harmonically in time as $p \sin(\omega t)$, where ω is the angular frequency, t is time, and p is the amplitude.

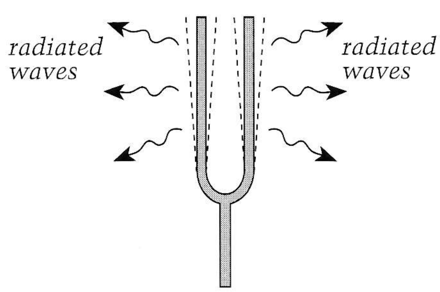


Figure 2

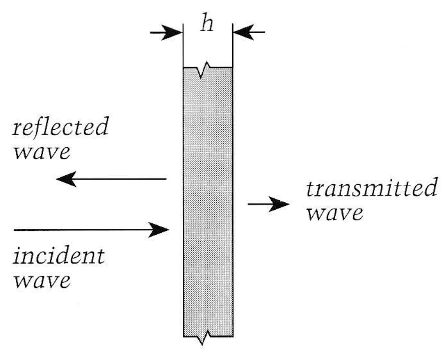


Figure 3

[2]See figure 2.—*Ed.*

[3]See figure 3.—*Ed.*

"What is the pressure acting on the wall? It's approximately twice the sound pressure of the incident wave. Why? The transmitted wave propagating to the right from the wall is weak compared to the incident and reflected waves, which have about the same amplitude. This is not particularly surprising. In the case of an elastic collision, for instance, a light ball bounces off a massive obstacle that is initially motionless with approximately the same speed it had before the collision. So the pressure exerted on the wall is about twice the sound pressure of the incident wave.

"Now, using Newton's second law, we can find the acceleration of the wall:

$$a = \frac{2pS\sin(\omega t)}{M}, \qquad (1)$$

where M and S are the mass and area of the wall, respectively. For harmonic motion,

$$|v| = \frac{|a|}{\omega}, \qquad (2)$$

where $|v|$ and $|a|$ are the amplitudes of the vibrational velocity and acceleration, respectively. Substituting equation (2) into equation (1), and after a little algebra, we finally arrive at

$$|v| = \frac{2p}{m\omega}, \qquad (3)$$

where $m = M/S$ is known as the surface density of the wall and can be expressed in kg/m^2. This parameter is very important for insulating single walls against airborne sound, but the insulation also depends on the elastic and dissipative properties of the walls. Nevertheless, surface density is still the most important parameter. The more massive a single wall, the greater its sound insulation.

"Do you have any questions, Watson? Please, stop me if it seems to you that I'm talking rot. Personally, I can't stand to take someone at his word without probing the matter or doing my own investigation . . ."

"Don't worry, Holmes, I follow. Actually, I'm quite pleased with myself that I'm understanding you. The only thing I'd like to know is the correlation between the surface density and the acoustical insulation of a wall. I think you're aware . . ."

"Yes, certainly. I like your questions! We're successful, you see, because we pull together. A reasonable question is sometimes more important than the appropriate answer. Researchers have come up with approximate relationships to correlate the sound insulation of a single wall with its surface density. For example, the walls of my building are made of brick and are 0.75 m thick. The density of a brick is about 1,600 kg/m^3, so the surface density of the walls is $m = (0.75\,\text{m})(1{,}600\,\text{kg/m}^3) = 1{,}200\,\text{kg/m}^2$. In this case, my neighbors hardly hear a thing when I take target practice with my pistols. If m = 500 kg/m^2, you cannot hear a loud conversation on the other side of the wall. If m = 200 kg/m^2, you can hear it as noise, but quiet conversation is still inaudible. If $m \leqslant 20$ kg/m^2, you can hear even a quiet conversation.

"Naturally, such rules of thumb are applicable in practice only if there are no holes or other openings in the wall.

"One example of a material with low sound insulation is a thin plywood board. This is because of its low surface density . . ."

"I see your point," I interrupted. "The density of glass is 2,500 kg/m^3, so the surface density of a 3-mm pane is only 7.5 kg/m^2 . . . Good heavens! This is depressing. Is it even possible to improve the sound insulation of windows?"

"Calm down, Watson! Yes, it is," Holmes said, smiling. "Double partitions provide comparatively high sound insulation. The same is true of people: two heads are better than one, isn't that so?"

"In my experience, it depends on the heads. And besides, sometimes people don't get along with each other. But let's stick to the point. Physics and psychology are two different things . . ."

"You have a point. In fact, its validity goes beyond psychology. It holds for physics as well, despite what you say. Two similar forces can act in opposition. This situation is not all that uncommon, even in sound insulation. The point is that double partitions afford good sound insulation only if the air gap between the two sheets of material is thick enough. The lower the masses of the two sheets, the wider the minimum air gap. Otherwise, you may get poorer results than you would with a single partition of the same surface density."

"I'm sorry, Holmes, but you've lost me there. Maybe I'm out of shape, mentally . . . As I understand it, the air gap is good for heat insulation. A double pane protects a room against cold much better than a single pane. Naturally, the amount of heat insulation increases with the thickness of the air gap—the wider the gap, the higher the insulation level . . ."

"I fully agree, Watson," Holmes interjected. "However, one of the misconceptions inherent in many proposed methods of sound insulation is the erroneous assumption that methods effective for heat insulation are also effective for sound insulation. Sometimes the analogy seems very close, but it may be misleading. Suffice it to say that your overcoat, which protects you against the cold air, is ineffective from the standpoint of sound insulation. Its surface density is not high enough. . . . Now let us talk about resonance, which plays an important role in sound insulation by multiple partitions."

"My dear Holmes, I'm sorry to interrupt you, but I fear the problem posed has proven too complex for me. I am not a scientist."

"Don't give up, my friend!" Holmes exclaimed. "Buck up and carry on for a few minutes more. We have almost reached our objective! I will simplify matters by means of a vivid example: a tram."

"A tram! Some friend you are. I'd prefer to hear neither its infernal rumble nor anything about it. You know it is the bane of my existence!"

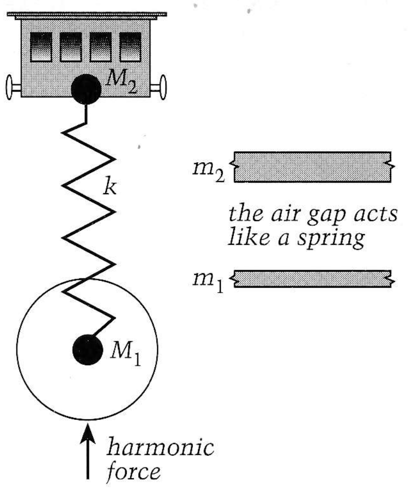


Figure 4

"I trust in your courage, Watson! Structurally, a tram can be envisioned as a simple mechanical system consisting of two masses and a spring. Here, let me sketch it for you.[4] The lower mass stands for the chassis, and the upper one simulates the coach with passengers and driver. The spring plays the role of the tram's suspension system. The mechanical structure shown here has a resonant frequency determined by the equation

$$f_r = \frac{1}{2\pi}\sqrt{\left(\frac{1}{M_1}+\frac{1}{M_2}\right)k},$$

where k is the spring constant and M_1 and M_2 are the masses. If a harmonic force acts on the first mass, the amplitude of the vibration of the second mass depends fundamentally on the frequency f ($f = \omega/2\pi$). If $f \ll f_r$, the spring does not protect the second mass against the vibrational motion caused by the force acting on the first mass. If $f \gg f_r$, the vibration isolation is significant. This is why a suspension system is used. The amplitude of the vibration of the second mass is comparatively small and decreases still further with an increase in the frequency.

"Now, a dangerous situation occurs if $f = f_r$ (that is, if the frequency of the induced vibration equals the resonant frequency). The vibration of the second mass may be greatly amplified, especially if the internal damping of the spring is low."

"Thank you, Holmes! It's a rather straightforward and, I believe, true analogy of sound transmission through a double partition. The masses are the individual partitions, and the spring is . . . the air gap, because air is resilient!"

"Bravo, Watson! You got it yourself. Yes, you are absolutely right. The resonant frequency of a double partition is given by a similar equation:

$$f_r = \frac{1}{2\pi}\sqrt{\left(\frac{1}{m_1}+\frac{1}{m_2}\right)\frac{\gamma P}{d}}, \qquad (4)$$

where m_1 and m_2 are the surface densities of the partitions, d is the width of the air gap, P is the atmospheric pressure, and $\gamma = 1.4$ is the ratio of the specific heat at constant pressure to that at constant volume. Let's calculate the resonant frequency of the double pane recommended by your glass-cutter. Substituting $m_1 = m_2 = 7.5$ kg/m^2 (in accordance with your previous estimate, Watson), $P = 10^5$ Pa, $d = 0.015$ m, and $\pi = 3.14$ into our equation (4) . . . we get

$$f_r \cong 250 \text{ Hz}.$$

Now we must evaluate the most 'noisy' frequency of tram noise."

"But Holmes, we have no device for measuring such a thing. As far as I know, it could be done with Helmholtz resonators."[5]

"Worry not, my dear Watson. The best device available is my own ears. You know me to be a good musician, correct? To my ears, the tram noise is high, in the frequency range of 100 to 400 Hz. The average frequency of this range is 250 Hz. So I was right, Watson. If you did what the glass-cutter recommended, you would create a 'resonant' window!"

"That would have been a disaster. Thank you, Holmes, for setting me straight. You have been a great help—now I can calculate the correct air gap. If it were 150 mm (a tenfold increase), the resonant frequency, as follows from our equation, would equal approximately 80 Hz. This would suffice to achieve the favorable condition: $f \gg f_r$. It's so simple, Holmes!"

"It's as easy as pie, if you know what you are doing. Your glass-cutter, however, proved a little wet behind the ears, which is strange: as a rule, such specialists know the ropes because their lack of formal education is compensated by their practical experience . . ."

Holmes fell silent. He walked up to the window and, after a long pause, asked me a few questions about the glass-cutter (his appearance and manners). Shortly thereafter, he left.

He called on me at my office the next morning. There was something puzzling in all this. Holmes saw the question in my eyes and, putting his fingertips together (as is his wont) and his elbows upon his knees, he explained the situation.

"You should be on guard," he began ominously. "Your glass-cutter is a remarkable man who has been to Oxford. He has a sharp mind, he is a great boxer, and in addition, he is of noble birth. Four years ago, however, he was found guilty of burglary. Nobody understood why he did it. Three years later, he escaped from custody and, at present, he is wanted by the police."

"You speak of danger, Holmes. But there is no need to worry over my safety. I regret to say my business brings in little money. I am not a rich man. If that person is truly intelligent, he will not choose me as a victim. Besides, he could have attacked me a few days ago. I don't see your point, my dear Holmes."

"As far as you are concerned, I agree with you entirely, Watson. In this affair, you are free and clear. His target is an apartment in the building across the street. Your neighbor who lives over there is a skilled jeweler,

[4]See figure 4.—*Ed.*

[5]Nowadays, special sound analyzers or sound meters with frequency filters could be used. Because of their great expense, however, one would not expect to find them in every home!—*Ed.*

David Polack. Are you acquainted with him?"

"He is a patient of mine. A kind and intelligent man. As I understand it, he made a mark for himself cutting and selling diamonds. He is quite a marksman, I hear, though he never served in the army. One more thing—his ears are rather sensitive. So he suffers from the noise even more than I!"

"Very good, Watson! I would like to mention as well that your neighbor does not believe in safes. He hides his most valuable diamonds in 'secret' places in his rooms. Sometimes he forgets to close the curtains, and in the evenings, when the lights are on in his rooms, he can be observed from your window. I saw him hiding his diamonds just yesterday. Your so-called glass-cutter did the same the day before yesterday."

"I see. It all seems plausible enough." I patted my empty jacket pocket. "Will my army revolver come in handy?"

"No, no, Watson! On no account will you fire a shot. We'll try to help our 'glass-cutter' mend his ways. Let us hope we succeed . . ."

The guest came in a few minutes later. He was so agitated, he didn't seem to recognize the room. He didn't even see me. He addressed Holmes as if he were his last hope.

"I read in the morning paper that Mr. Holmes knows something about a young woman named Lillian Wilson. I am looking for her because . . ."

"Because you are her brother, Ronald Wilson," said Holmes in a friendly manner. "Am I right?"

Our guest whirled around. I was standing by the door, my hand in my pocket. Wilson laughed, but without mirth. He turned back to Holmes.

"Congratulations! Now I recognize you. You're the gentleman I saved. And this is the thanks I get! Good work, Mr. Policeman."

"First of all, Mr. Wilson, I am not a policeman. Second, if you wish to leave, you may. However, it would not be beneficial to you for a number of reasons. So calm down and listen to me carefully. Your sister Lillian is married and lives in the United States of America, at West Chester in the state of Pennsylvania. Her husband is a wealthy young doctor. She is happy. He is happy as well.

"Now, about the jeweler across the way. I would not recommend that you irritate Mr. Polack. He has sensitive ears, and he is a sharpshooter. It would not be wise to visit him uninvited, now would it, sir?"

Our guest was flabbergasted. "You are a wizard, sir!"

"No," Holmes replied, "I am not a wizard. My name is Sherlock Holmes. This is my best friend and associate, Dr. Watson, before whom you may speak as freely as before me. But you seem to be ill at ease. You could do with a cup of coffee, couldn't you?"

Wilson was happy to learn that his lovely younger sister was well and happy. He told us that when his parents died, they inherited nothing but debts. The family jewels were taken and sold at auction. The same was true of their house and furniture. Wilson worked hard to make a living. He and his sister seemed to be surviving against all odds, but he was driven to get the family jewels back, come hell or high water. And so he found himself on the opposite side of the law—he had broken into the house of a banker who had bought most of the Wilson family jewels. He ended up in prison and lost touch with his sister. He escaped and, not long before, read that some of the family jewels had been sold to Mr. Polack.

"I'm glad to know that my sister is safe and sound," Wilson said with feeling. "As for me," he added bitterly, "I'm a criminal."

"Cheer up, young man!" Holmes exclaimed. "The young lady you saved from the ruffians is the daughter of that banker. She recognized you. A woman's memory is an amazing thing sometimes. She promised me that she would talk with her father. I believe your case will be reviewed and you will be found innocent.

"By the way," he added, "she would like to meet you. As far as I can recall, she is the type that turns heads, but that's not my business."

"Thank you very much, Mr. Holmes. I appreciate your friendly support. But I am in hot water. It's rather difficult to find work at present. And beside, I have been in prison . . ."

"Well, in that case, I shall introduce you to Mr. Polack. Yes, the very same! He is going to South Africa in search of diamond deposits, and a companion such as yourself would be a great boon to him. My recommendation is sufficient in this case."

A year has passed, and we received a letter from Wilson. His trip with David Polack was a resounding success. He has become a rich man. In his letter he wrote that his new house is in a quiet rural place.

"Dear Mr. Holmes and Dr. Watson," he wrote, "I found that David is really a great shooter. Not only that, he studied acoustics, and we discussed many problems, including those of sound insulation. Now I know my 'advice' to you, Dr. Watson, was wrong, because of the resonance. On the other hand, that error changed my life. By the way, I believe the best sound insulation is to go away from all sources of noise (the farther, the better). I invite you to see my quiet house, me, and my wife. Sincerely, Ronald Wilson."

"Very well," I said, "let's go to his place next month. It would be convenient for me."

"I regret to say, dear Watson, that I cannot take you with me."

I was taken aback. "Why not, Holmes?"

"Simply because you can't stand noise."

He opened his newspaper. It reported that the government had decided to extend a rail line to Wilson's vicinity, near his new home. This stretch of railway is considered a matter of priority for economic reasons, so it must be built in the shortest possible time . . . Oh, noise, noise! You are everywhere!

Challenges in physics and math

Math

M156

Politics on the chessboard. What greatest number of chess kings can be put on a toroidal $n \times n$ chessboard so as not to attack one another? (A toroidal chessboard is obtained by gluing together the opposite edges of the ordinary chessboard. A king on the torus always attacks eight squares—see figure 1.) (A. Tolpygo)

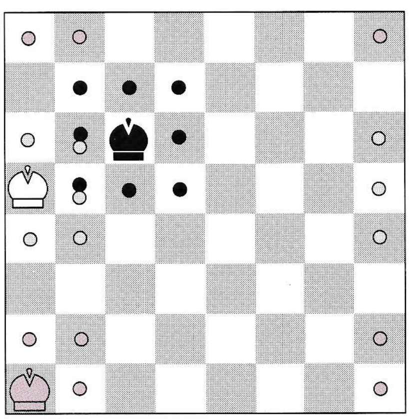


Figure 1

M157

Sparse products. Prove that for any sequence of positive numbers a_n the integer parts of the square roots of the numbers

$$b_n = (a_1 + a_2 + \cdots + a_n)\left(\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}\right)$$

are all different. (L. Kurlyandchik)

M158

A bunch of circles and lines. Circles S_1 and S_2 touch each other externally at point F. Line l touches S_1 and S_2 at points A and B, respectively, and the line parallel to l and tangent to S_2 at C meets S_1 at points D and E. Prove that (a) points A, F, C are on the same line; (b) the common chord of the circumcircles of triangles ABC and BDE passes through F. (A. Kalinin)

M159

Period of recollection. At the vertices of a regular n-gon are placed m chips ($m > n$). A pair of chips at the same vertex is moved to the vertices next to it—one chip to each of the adjacent vertices. Then another pair is separated in the same way, and so on. After a number of such moves, the numbers of chips at each vertex are restored to their initial values. Prove that this number of moves is a multiple of n. (I. Rubanov)

M160

Positive positions. Each square of an infinite square grid on the plane has a real number written in it. Two figures consisting of a finite number of grid squares are considered. The figures are allowed to be shifted parallel to grid lines by any integer number of squares. Prove that if for any shift of the first figure the sum of the numbers it covers is positive, then there exists a shift of the second figure such that the sum of the numbers it covers is positive. (B. Ginzburg, I. Solovyov)

Physics

P156

Spinning bobbin. The end of a thread wound around a bobbin is passed over a nail in a wall (fig. 2).

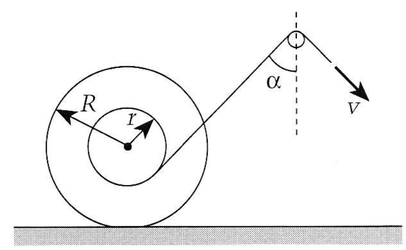


Figure 2

The thread is pulled with a constant velocity $\mathbf{v}$. What is the velocity of the bobbin's center at the moment the thread makes an angle α with the vertical? The outside radius of the bobbin is R, the inside radius is r, and assume that the bobbin rolls without slipping. (S. Krotov)

P157

Nitrogen bubble. A soap bubble is inflated with nitrogen at room temperature. At what diameter will the bubble start to ascend? The surface tension of the soap solution is $\sigma = 0.04$ N/m. Neglect the mass of the soap film. (A. Sheronov)

P158

Aircraft design. An inventor designs the following "aircraft." The upper surface of a large flat plate is kept at a constant temperature of 0°C, while the temperature of the bottom surface is 100°C. The inventor asserts that such a plate will be suspended in the air like a dirigible. Explain the phenomenon. Estimate the lift on such a plate (order of magnitude) if it is 1 m^2 in area. The air temperature is 20°C.

P159

Current and frame. A square frame made of wire with a diameter d_0 is placed near a long straight wire carrying electric current I_0 (fig. 3). When this current is turned off, a momentum p_0 is imparted to the frame. What is the direction of this momentum? What momentum would be imparted to the frame if the initial current in the wire were $I_1 = 3I_0$ and the diameter of the wire $d_1 = 2d_0$? (V. Mozhayev)

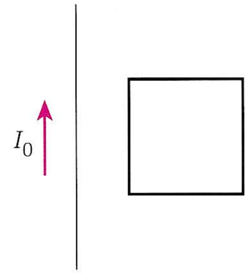
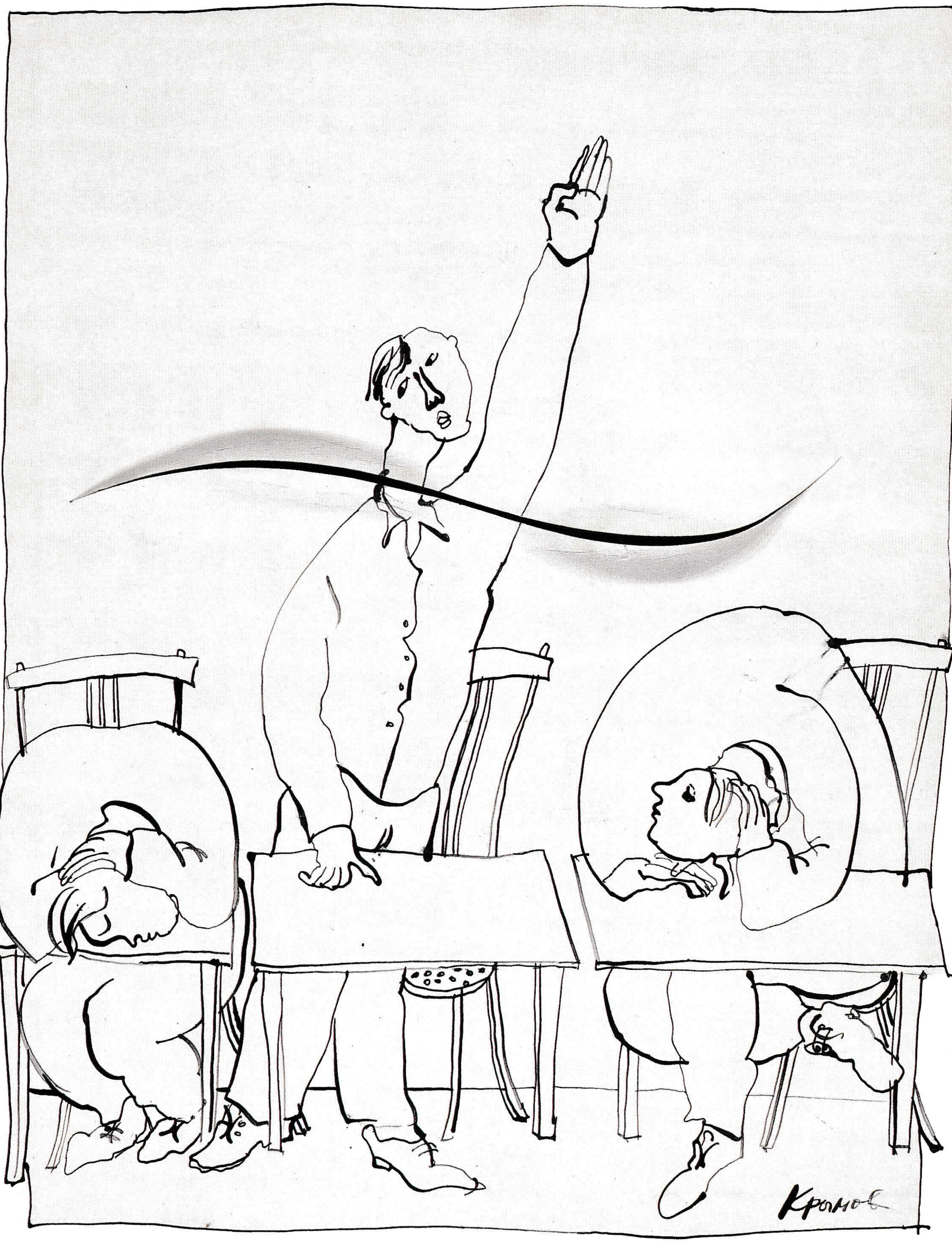


Figure 3

P160

Compound lens. A bar of width h = 5 mm is cut from a convex lens of diameter d = 5 cm and focal length F = 50 cm. The resulting parts are moved close to each other. A point source of light S is placed at a distance l = 75 cm from the compound lens along the axis of symmetry. At what maximum distance from this lens can optical interference be observed?

ANSWERS, HINTS & SOLUTIONS ON PAGE 57



The mean value of a function

An arithmetic concept is stretched and applied in unusual places

by Yury Ionin and Alexander Plotkin

YOU UNDOUBTEDLY KNOW what the arithmetic mean of finitely many numbers is. In this article we'll extend this notion to functions defined on a segment, circle, and sphere and show its somewhat unexpected applications to geometry. In particular, we'll explain how to measure length by measuring width.

To read this article, you will have to know a little bit about the geometry of vectors. You will need to know how to add vectors and how to find their length (absolute value). We will also be dealing with the projection of a vector onto a line, but we'll define this for ourselves. We will also use certain terminology and notation from the integral calculus, but readers who don't know any calculus at all will be able to understand in other ways what is meant. Only the last two or three problems we discuss involve a direct knowledge of the calculus.

On a finite set

At an annual meeting of the Parents Fund Committee of a certain Business Mathematics School, the Chairman addressed the audience thus: "It is the honorable duty of every parent to make a greater-than-average contribution!" Apparently the Chairman wasn't up on his math. Let's think it through.

The average, or arithmetic mean, of n numbers $x_1, x_2, \ldots, x_n$, according to the usual definition, is the number $M = M(x_1, x_2, \ldots, x_n)$ given by the equation

$$M(x_1, x_2, \ldots, x_n) = \frac{x_1 + x_2 + \cdots x_n}{n}. \quad (1)$$

Heeding the Chairman's call, the parents would have to simultaneously satisfy the inequalities $x_1 > M$, $x_2 > M$, ..., $x_n > M$. Adding them up and dividing by n we get a contradiction to equation (1).

We note the following important properties of the arithmetic mean:

1. $M(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) = M(x_1, x_2, \ldots, x_n) + M(y_1, y_2, \ldots, y_n)$;
2. $M(\alpha x_1, \alpha x_2, \ldots, \alpha x_n) = \alpha M(x_1, x_2, \ldots, x_n)$;
3. $\min(x_1, x_2, \ldots, x_n) \le M(x_1, x_2, \ldots, x_n) \le \max(x_1, x_2, \ldots, x_n)$.

Readers are urged to find their own verbal descriptions of these properties.

Exercises

1. Prove properties 1–3 above.

2. For any integer $n > 1$, prove that the inequality $\binom{2n}{n} > 2^{2n-1}/n$ (where $\binom{2n}{n} = (2n)!/(n!)^2$ is a binomial coefficient). This can be done by induction, but also by using property 3.

Property 3 is often used to establish that one of the numbers $x_1, x_2, \ldots, x_n$ is greater than a given d: it suffices to check that the average $M(x_1, x_2, \ldots, x_n) > d$. Look, for instance at the following problem.[1]

Problem 1. There are fifty correctly running watches lying on a round table. Prove that at a certain moment the sum of the distances from the table's center to the tips of their minute hands will be greater than the sum of the distances from the table's center to the watches' centers.

Solution. Denote by $f(t)$ the sum of the distances from the center O of the table to the minute hands' tips at a moment t (hours). Let d be the sum of the distances from O to the centers of the watches. We have to prove that $f(t) > d$ at a certain time t. We'll show that $M(f(t_0), f(t_0 + 1/2)) > d$ for a certain t_0, which will mean that one of the moments t_0, $t_0 + 1/2$ is the one we're looking for.

Denote by O_i, A_i, and B_i the center of the ith watch, the tip of its minute hand at a moment t, and the same tip a half-hour later (at the moment $t + 1/2$), respectively. Since all the watches are correct, there is

[1]Proposed by S. Fomin at the 10th All-Union Mathematical Olympiad (Dushanbe, Tadjikistan, 1976). The idea can also be traced in the solution to M160 in this issue.

Art by Dmitry Krymov

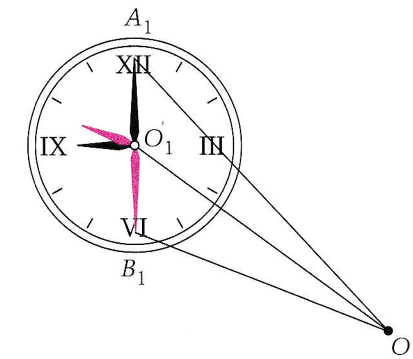


Figure 1

a moment t_0 at which the points O, A_1, and B_1 are not on the same line. Consider the triangle OA_1B_1 and its median OO_1 (fig. 1).

Exercise 3. Prove that a median in a triangle is less than the half-sum of the sides drawn from the same vertex.

By exercise 3, $OO_1 < (OA_1 + OB_1)/2$ at the moment t_0. For all the other watches $OO_i \le (OA_i + OB_i)/2$ (be sure you understand why we must replace the strict equality with "$\le$"). Adding these 50 inequalities and dividing by n, we get $M(f(t_0), f(t_0 + 1/2)) > d$.

On a segment

In problem 1 it sufficed to estimate the arithmetic mean of *two* values of a function. Is it possible to define the average of *all* values? To approach such a definition, let's give some geometric meaning to the average of n positive numbers $x_1, x_2, \ldots, x_n$.

Divide a segment $[a, b]$ into n equal pieces of length $(b - a)/n$ and construct a "bar chart" consisting of the rectangles with these pieces as bases and the numbers $x_1, x_2, \ldots, x_n$ as heights (fig. 2). The area of this figure equals $M(x_1, x_2, \ldots, x_n) \cdot (b - a)$. So $M(x_1, x_2, \ldots, x_n)$ is the height of the rectangle with the base $[a, b]$ equal in area to our bar chart.

Now consider a continuous non-

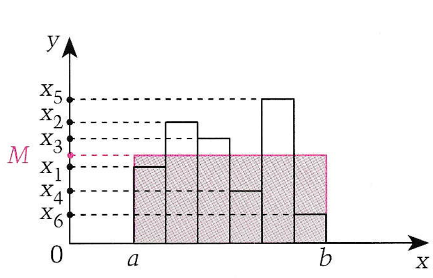


Figure 2

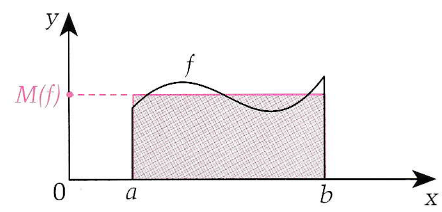


Figure 3

negative function f on a segment $[a, b]$. By its *mean value* on $[a, b]$ we'll mean the height of the rectangle with base $[a, b]$ whose area is equal to the area of the "curvilinear trapezoid" bounded by the lines $x = a$ and $x = b$, the x-axis, and the graph of f (fig. 3). This area can be expressed as the integral $\int_a^b f(x)dx$.

[*Note (for readers who are not familiar with integrals)*: Perhaps you've thrown up your hands at this point and are ready to quit. Hang in there! Throughout this article the formula $\int_a^b f(x)dx$ can simply be understood as a notation for the area described above, where the integral sign $\int$ means "area," the letters a, b, and f designate its boundaries, and dx indicates the argument (x) of the function f to distinguish it from other parameters on which f may depend. The actual integration begins when we have to compute this area, but we'll need only one numerical result below, and you can just take it for granted.—*Ed.*]

So the mean value of f on $[a, b]$ is *defined* as

$$M(f) = \frac{1}{b-a}\int_a^b f(x)dx. \qquad (2)$$

(This definition is good for *any* function on $[a, b]$ for which the right side of equality (2) makes sense, but we'll deal only with continuous nonnegative functions.)

The mean value thus defined has properties similar to properties 1–3:

1′. $M(f + g) = M(f) + M(g)$;
2′. $M(\alpha f) = \alpha M(f)$;
3′. $\min\limits_{a\le x\le b} f(x) \le M(f) \le \max\limits_{a\le x\le b} f(x)$.

Let's explain the first of these properties in terms of areas. (A rigorous proof is based on a rigorous definition of the integral.) Consider the area between the graphs of f and $f + g$ (both functions f and g are assumed to be positive—see figure 4a). It can be approximated with any desired precision by a set of bars as shown in the figure. Shift all the bars down on the x-axis (fig. 4b). Then they'll form a "bar chart" approximating the area under the graph of $g(x)$ (because their heights are equal to the values of the difference $(f(x) + g(x)) - f(x) = g(x)$ at the corresponding points). Since the approximations can be made as precise as we wish, the area between the graphs of f and $f + g$ is the same as under the graph of g—that is,

$$\int_a^b \big(f(x)+g(x)\big)dx - \int_a^b f(x)dx = \int_a^b g(x)dx,$$

which is equivalent to property 1′.

Property 2′ is proved similarly. And property 3′ is clear from figure 5 (the area under f lies between the areas of the rectangles $ABCD$ and $ABFE$—that is, between $(b - a)\min f$ and $(b - a)\max f$).

Exercise 4. From properties 1′–3′

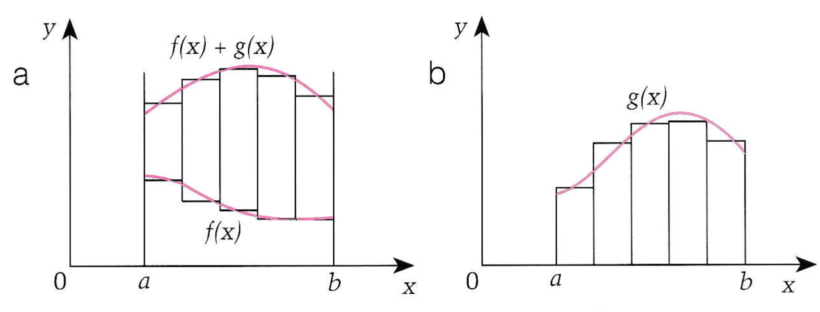


Figure 4

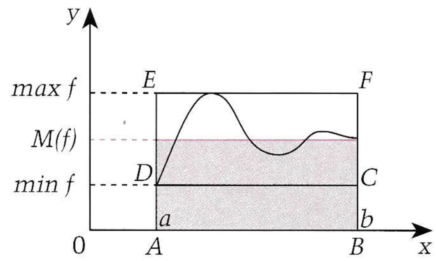


Figure 5

derive the following:

4′. If $f(x) \le g(x)$ for all $x \in [a, b]$, then $M(f) \le M(g)$.

In the next problem the mean value of a function is applied in a setting that doesn't seem to bear even the slightest relation to it.

Problem 2. The sum of four vectors **a**, **b**, **c**, **d** on the plane is zero. Prove the inequality

$$|\mathbf{a}| + |\mathbf{b}| + |\mathbf{c}| + |\mathbf{d}| \ge |\mathbf{a} + \mathbf{d}| + |\mathbf{b} + \mathbf{d}| + |\mathbf{c} + \mathbf{d}| \quad (3)$$

(here $|\mathbf{a}|$ is the length of vector **a**, and so on).

The solution is based on the following one-dimensional version of this problem, which is left to the reader as an exercise.

Exercise 5. For any real numbers a, b, c, d with zero sum, prove that

$$|a| + |b| + |c| + |d| \ge |a + d| + |b + d| + |c + d|. \quad (4)$$

Solution to problem 2. Consider the projections of the vectors **a**, **b**, **c**, **d** on an arbitrary axis (directed line). For our purposes, we define the projection of a vector $\overrightarrow{AB}$ on a (directed) axis l as the number $\pm A'B'$, where A' and B' are the projections of A and B on l (fig. 6) and the sign is chosen according to whether the vector $\overrightarrow{A'B'}$ is directed with (+) or

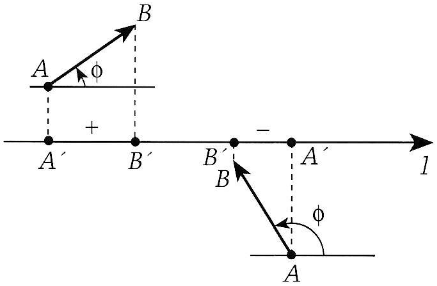


Figure 6

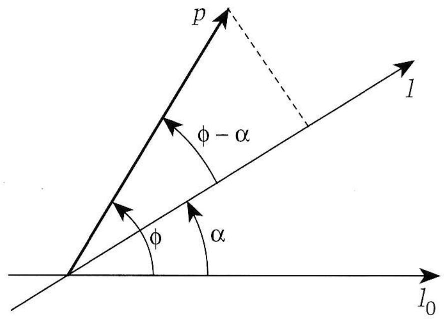


Figure 7

against (−) the axis—in other words, it's equal to $AB \cos \phi$, where ϕ is the angle between l and $\overrightarrow{AB}$.

By exercise 5, if we replace the vectors in expression (3) with their projections on any given axis, we'll get a true inequality (because the sum of the projections is equal to the projection of the sum—that is, zero). This suggests that we might prove inequality (3) by considering all possible projections of the given vectors.

Fix an axis l_0 and for any vector **p** define the function $p(\alpha)$ as the projection of **p** on an axis l that makes an angle α with l_0 (fig. 7). If the angle between l_0 and **p** is ϕ, then $p(\alpha) = |\mathbf{p}| \cos(\phi - \alpha)$. Now consider the mean value $M(|p|)$ of the function $\alpha \to |p(\alpha)|$ on the segment $[0, 2\pi]$. By property 2′, $M(|p|) = |\mathbf{p}| M(|\cos(\phi - \alpha)|)$, and

$$M\left(\left|\cos(\phi - \alpha)\right|\right) = \frac{1}{2\pi}\int_0^{2\pi} \left|\cos(\phi - \alpha)\right| d\alpha.$$

It's clear from figure 8 that the area under the graph of $|\cos(\phi - \alpha)|$ on the segment $[0, 2\pi]$ is equal to the area under the graph of $|\cos(\pi/2 - \alpha)| = |\sin \alpha|$ on $[0, 2\pi]$ (or twice the area

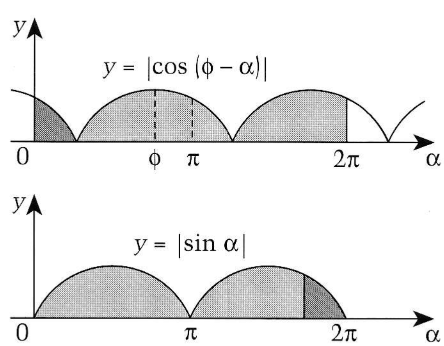


Figure 8

under an arc of the standard sinusoid). So

$$M\left(\left|\cos(\phi - \alpha)\right|\right) = \frac{1}{2\pi}\int_0^{2\pi} \left|\cos\left(\frac{\pi}{2} - \alpha\right)\right| d\alpha = M\left(\left|\sin \alpha\right|\right).$$

The integral calculus tells us that this area is in fact equal to 4, but the really important thing for us here is that it doesn't depend on ϕ and is positive. So the mean value of the function $p(\alpha)$ is proportional to the length of **p** with a constant positive coefficient k:

$$M(|\mathbf{p}|) = k|\mathbf{p}| \quad (5)$$

(using calculus, we can find that $k = M(|\sin \alpha|) = 2/\pi$). And the basic property of this function used in our argument was that the length of the segment $[0, 2\pi]$ over which it was averaged is its period.

Now we're ready to finish the solution.

As we mentioned above, for any $\alpha \in [0, 2\pi]$ the sum of the numbers $a(\alpha)$, $b(\alpha)$, $c(\alpha)$, $d(\alpha)$ is zero. Therefore,

$$|a(\alpha)| + |b(\alpha)| + |c(\alpha)| + |d(\alpha)| \ge |a(\alpha) + d(\alpha)| + |b(\alpha) + d(\alpha)| + |c(\alpha) + d(\alpha)|.$$

Using properties 1′ and 4′ (see example 4) to take the mean value of both sides and expressing the mean values according to expression (5), we get

$$k|\mathbf{a}| + k|\mathbf{b}| + k|\mathbf{c}| + k|\mathbf{d}| \ge k|\mathbf{a} + \mathbf{d}| + k|\mathbf{b} + \mathbf{d}| + k|\mathbf{c} + \mathbf{d}|.$$

All that remains is to cancel the factor k out.

On the circle and sphere

Problem 3. Prove inequality (3) of problem 2 for any four vectors **a**, **b**, **c**, **d** in space with a sum of zero.

It would be nice if we could extend our "plane reasoning" to space. To do this we have to somehow define the mean absolute value of projections of a vector on axes of all directions in space. We'll see below that this will amount to averaging a function on a sphere. So let's begin with a simpler but similar case of

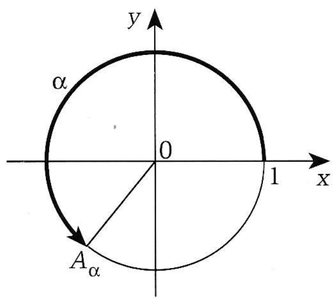


Figure 9

functions on the unit circle (circle of radius 1).

Let's take a function f whose domain is the set of points on the unit circle and whose range is a set of real numbers. For any real α define $\bar{f}(\alpha) = f(A_\alpha)$, where A_α is the point on the unit circle obtained by rotating the point $E(1,0)$ through an angle of radian measure α about the origin (fig. 9). This allows us to extend an arbitrary function f, defined on the unit circle, to a function $\bar{f}$ defined on the entire real axis. Clearly $\bar{f}$ is a periodic function with period 2π. Now we can define the *mean value of f on the unit circle* as the mean value of $\bar{f}$ on the segment $[0, 2\pi]$. Recall that the crucial point in the solution to problem 2 above was the fact that the mean $M(|p|)$ was independent of the direction of vector **p**. This independence is a particular case of the following general property of the mean values of functions on the unit circle.

5′. Let f and g be two functions on the unit circle that differ by the rotation through a certain angle ϕ about the circle's center—that is, $g(A) = f(r^\phi(A))$ for any point A on the circle, where r^ϕ is this rotation. Then $M(g) = M(f)$.

Exercise 6. Prove this property.

Proceeding to problem 3, consider the unit sphere centered at the origin O. It is possible to define the mean value for functions on this sphere so that properties 1′–5′ are satisfied! (Of course, r^ϕ in property 5′ should now by understood as the rotation of the sphere through ϕ about a certain *axis* passing through O.) We cannot explain here how this is done, because an accurate definition involves integration over the sphere.

However, let's assume that the mean value $M(f)$ of a function on the sphere with properties 1′–5′ is somehow defined. That is, for this discussion we can take the existence of a mean value $M(f)$ of a function obeying properties 1′–5′ as an "axiom." It in fact turns out that the properties 1′–5′ define the mean value of a function uniquely.

Pick a point A on the unit sphere. Consider any vector **p** and define an auxiliary function $p(A)$ on the sphere as the projection of **p** on the axis OA (with the same direction as vector $\overrightarrow{OA}$. Then the mean value $M(|p|)$ of the function $A \to |p(A)|$ on our sphere satisfies equation (5) for a certain fixed $k \neq 0$ and any vector **p**. To prove this, it suffices to show that $M(|p|) = M(|q|)$ whenever $|\mathbf{p}| = |\mathbf{q}|$.

Exercise 7. Prove that this is indeed sufficient. (Use property 2′.)

Suppose $|\mathbf{p}| = |\mathbf{q}|$. Then there exists a rotation r of space about some axis through the origin O that takes the ray OQ, where $\overrightarrow{OQ} = \mathbf{q}$, to the ray OP, where $\overrightarrow{OP} = \mathbf{p}$. It's clear that $q(A) = p(r(A))$ for any point A on the sphere. But then, in view of property 5′, $M(|q|) = M(|p|)$.

Now the solution to problem 3 can be completed by repeating the end of the solution to problem 2 above.

Length in terms of width

The same idea can tell us something about the perimeter of a convex polygon in the plane.

Let $a_1, a_2, \ldots, a_n$ be its successive sides. Fix an axis l_0. Denote by $W(\alpha)$ the "width" of the polygon in the direction of an axis l_α that makes an angle α with l_0 (fig. 10). It turns out

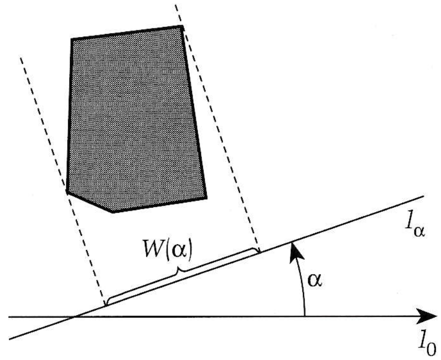


Figure 10

that if we know this width in any direction—that is, if we can calculate the function $\alpha \to W(\alpha)$—then we can find the perimeter P of the polygon. Let's see why this is so.

Denote by $a_i(\alpha)$ the length of the projection of side a_i on the axis l_α.

Exercise 8. Prove that $W(\alpha) = ½(a_1(\alpha) + a_2(\alpha) + \ldots + a_n(\alpha))$.

In the solution to problem 2, it was shown that the mean value of the function $\alpha \to a_i(\alpha)$ is proportional to the length $|a_i|$ of the side a_i. We've also mentioned that the proportionality coefficient $k = 2/\pi$. From exercise 8 and properties 1′ and 2′ it follows that the mean value $M(W)$ equals half the sum of the mean values of the functions $\alpha \to a_i(\alpha)$. Therefore,

$$M(W) = \frac{1}{2}\left(\frac{2}{\pi}|a_1| + \frac{2}{\pi}|a_2| + \cdots + \frac{2}{\pi}|a_n|\right)$$

$$= \frac{1}{\pi}P.$$

Recalling our definition of the mean value of a function (equation (2)), we finally get

$$P = \pi \cdot M(W) = \pi \cdot \frac{1}{2\pi}\int_0^{2\pi} W(\alpha)d\alpha$$

$$= \frac{1}{2}\int_0^{2\pi} W(\alpha)d\alpha.$$

Thus, the length of a polygonal path (formed by the sides of our polygon) can be expressed in terms of the function W.

Exercise 9. Prove that if the lengths of all the sides and diagonals of a convex polygon are less than d, then its perimeter is less than πd.

Our expression for the perimeter is valid for any plane closed convex curve. The method described above for calculating the length if the width (in each direction) is known was proposed by the well-known Polish mathematician H. Steinhaus in 1930.

CONTINUED ON PAGE 37

Graphs an

Connecting with ancestors and

THE TERM "GRAPH" (FROM the Greek γραφω—"I write") is used in mathematics basically in two senses. The more "classical" meaning relates to curves in the coordinate plane that represent functions or equations in two variables. But here we'll talk about the other meaning of "graph": in discrete mathematics and topology it means a diagram showing a number of points, some of which are joined by lines (not necessarily straight). Diagrams of transportation systems, such as subway lines or networks of railroads depicted on geographical maps, are typical examples of graphs (fig. 1).[1]

The points of a graph are called its nodes or vertices, and the lines joining them are its edges. The really essential thing about a graph is the way in which its edges connect its nodes. If there is a one-to-one correspondence between the sets of nodes of two graphs that associates any two nodes connected in one graph to connected nodes in the other graph, then these graphs are considered the same.

Figure 2 shows another example of a graph—the "ascending" genealogical tree of the great writer Count Leo Nikolayevich Tolstoy (in Russian, "graf Lev Nikolayevich Tolstoy").[2] Here the vertices of the graph are the members of this famous Russian noble family, and the lines between them indicate the parent–child relationship. The term "tree" in graph theory means a graph without cycles—that is, a graph whose edges do not form any closed paths. In other words, on a tree you can't start from a node, travel along a number of edges, and arrive back at the initial node. Genealogical trees are trees in this sense, too, if there are no marriages between relatives in the family.

It's easy to understand that a graph-tree can always be depicted on the plane so that its edges do not intersect one another. The same property is true for the graphs formed by the vertices and edges of a convex polyhedron. Examples of such graphs for all five regular polyhedrons are shown in figure 3.

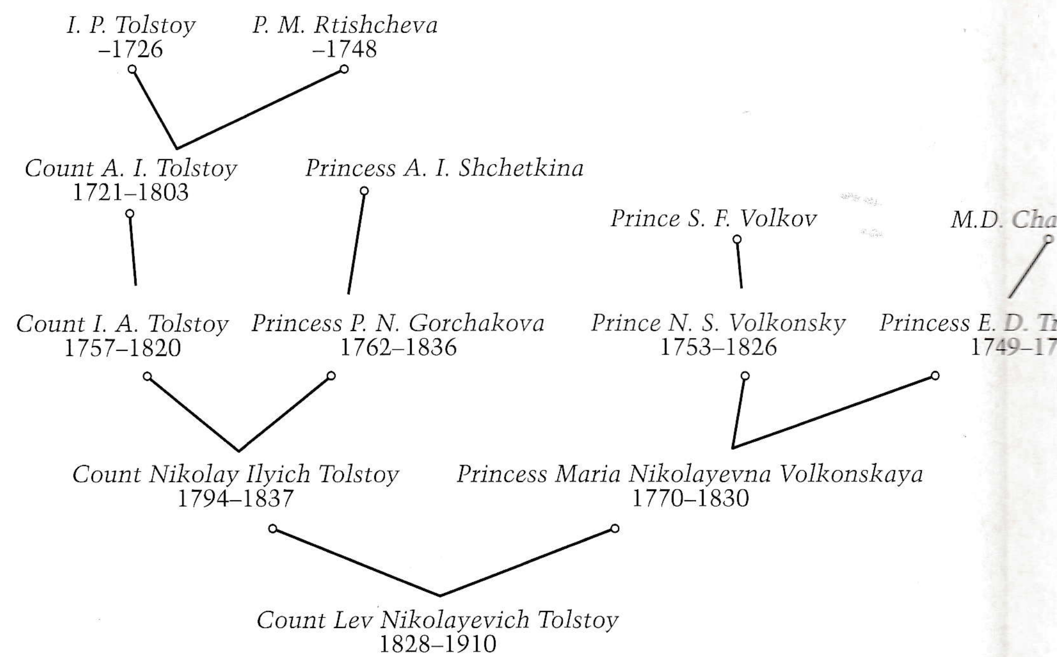


Figure 2

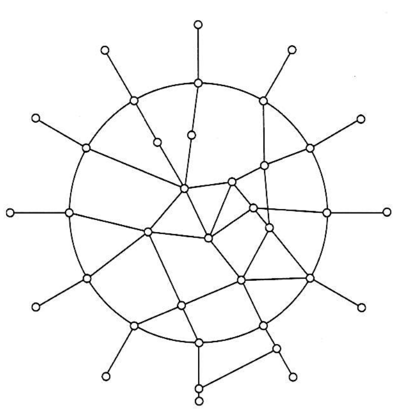


Figure 1

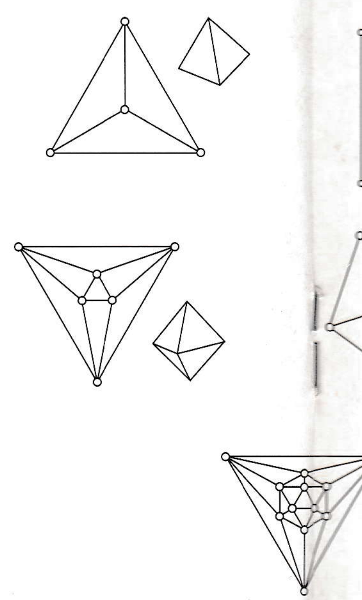


Figure 3

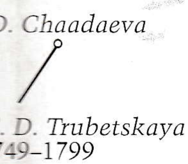
---

[1] See also "Signals, Graphs, and Kings on a Torus" in this issue.—*Ed.*

[2] The same title was used in German as well. Recall, for instance, the *Graf Zeppelin*—the dirigible that inaugurated transatlantic passenger service. It was named after its designer, Ferdinand, Graf von Zeppelin.—*Ed.*

and grafs

and avoiding one's neighbors



One of these graphs—that of the tetrahedron—has the property that any two of its nodes are connected by an edge. A graph with this property is called *complete*. Another example of a complete graph—for the case of five nodes—is shown in figure 4. Try to draw it so as to get rid of intersections.

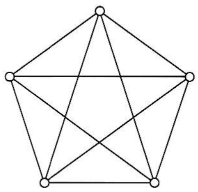
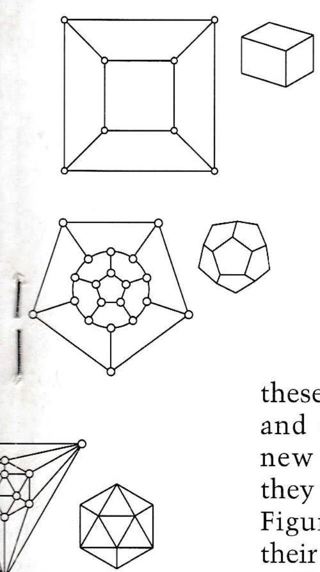


Figure 4

Have you tried? And what was the result? No need to answer—I knew in advance what you'd say! This graph can't be placed on the plane without self-intersection. This is as impossible as to fulfil the intentions of the three persons once described by Lewis Carroll.



They lived in three houses and there were three wells nearby—one with water, another with butter, and the third with jam. They used to go to the wells by the pathways shown in figure 5. But one day these persons quarrelled and decided to create new pathways so that they wouldn't intersect. Figure 6 shows one of their attempts to do this.

The graphs in figures 5 and 6 proved to be of

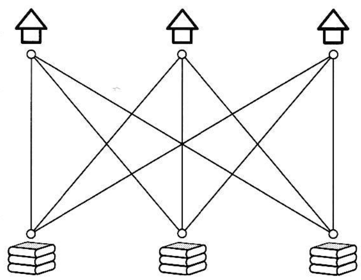


Figure 5

decisive significance in determining whether a given graph is planar—that is, whether it can be drawn on the plane without self-intersection.

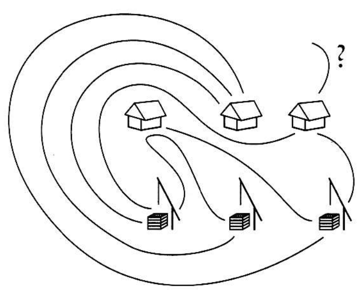


Figure 6

The Polish mathematician Kazimierz Kuratowsky and the Russian mathematician Lev Pontryagin independently proved that a graph is planar if and only if it contains neither of these two graphs (the complete 5-node graph and "houses-and-wells") as a subgraph.[3]

We find graphs in the block diagrams of computer programs and in network models for construction work, where the edges correspond to different kinds of work and their arrangement shows the sequence in which the tasks must be completed—that is, which parts of the entire process must be finished before other parts can begin (fig. 7).

Graphs are very helpful in solving puzzles. In our Toy Store articles we have used them many times (see, for instance, "A Manual for the Mathematical Gambler" in the previous issue or "The A-maze-ing Rubik's Cube" in the September/October 1991 issue).

Graph theory is a part of topology as well as combinatorics. Its topological nature stems from the fact that the properties of a graph in itself are independent of the position of its nodes and the shape of its edges. And the "language" of graphs turned out to be very well suited for formulating combinatorial problems, which made graph theory a powerful tool in combinatorics.

—Anatoly Savin

[3]In more recent literature, this graph is called the "utility" graph, since it illustrates the problem of connecting three houses to sources of, say, water, gas, and electricity, without any lines crossing.—*Ed.*

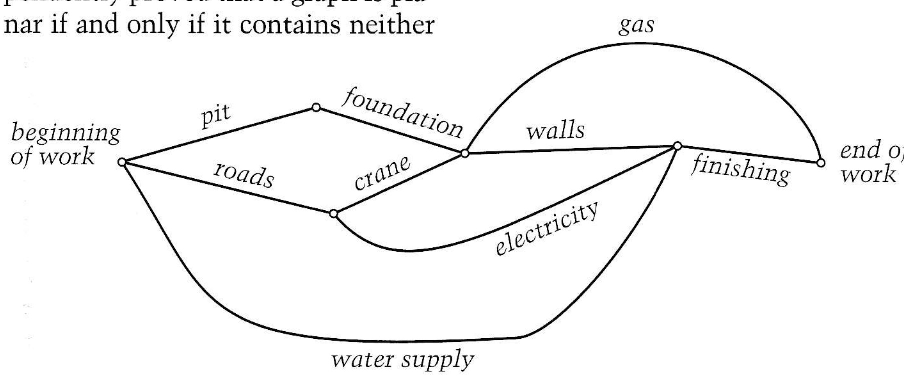


Figure 7

PHYSICS CONTEST

Gravitational redshift

"Impulses from what scarce was matter
Bounced off a shallow platter
Into the realm of number pure . . ."
—Howard Nemerov, "Druidic Rimes"

by Arthur Eisenkraft and Larry D. Kirkpatrick

AS WE VIEW THE STATE trooper ticketing a speeding motorist, our mind returns to those lectures in physics class where Doppler shifts and radar filled the board. The same type of radar that catches speeding motorists records the speed of pitches in baseball and serves in tennis. At this year's US Open Championship, Monica Seles and Steffi Graf both served at speeds in excess of 100 mph, while Pete Sampras recorded serves faster than 120 mph.

The classic example of the Doppler shift is the change in pitch of a train whistle. As the train approaches us, our ears record a higher pitch, which drops as the train passes us. By measuring the change in pitch of the whistle, we can determine the speed of the train.

Doppler shifts occur for all types of wave. Because atoms and ions near the surfaces of stars emit spectral lines characteristic of each element, Doppler shifts are a very important tool in astronomy. For instance, the shift in the frequency of the emission spectrum from a star (or galaxy) tells us how fast the star is approaching or receding from Earth.

Because electromagnetic radiation can propagate through a vacuum, the Doppler shift formula for electromagnetic radiation is simpler than for sound waves. The size of the shift depends only on the relative velocity v of the source and the detector, rather than the velocities of each of these relative to the air. If we use the notation $\beta = v/c$, where c is the speed of light and β is positive when the distance between the source and the detector is increasing, the shifted frequency f is given by

$$\frac{f}{f_0} = \sqrt{\frac{1-\beta}{1+\beta}},$$

where f_0 is the frequency emitted by the source. When $|\beta| \ll 1$, we can use an even simpler expression that can be obtained by approximating $(1 \pm \beta)^n \cong 1 \pm n\beta$. Therefore,

$$\frac{f}{f_0} = (1-\beta)^{1/2}(1+\beta)^{-1/2}$$

$$\cong \left(1-\frac{\beta}{2}\right)\left(1-\frac{\beta}{2}\right)$$

$$\cong 1-\beta,$$

where we've dropped the term in β^2. (The formula used by the police is actually $1 - 2\beta$, because the radar wave is shifted twice. You can think of your car as the first detector and the police car as the second.)

Edwin Hubble showed that most stars (and galaxies) are receding from Earth, which led to the theory that the universe began with a big bang. Because of the expansion, the frequencies of the spectral lines from the stars are shifted to lower values—that is, the light is redshifted. However, this not the only redshift that occurs. A photon leaving a star is also redshifted as it rises in the gravitational field of the star. The gravitational redshift for our sun is too small to be detected accurately, but the redshifts of photons leaving white dwarfs can be measured and are equivalent to the redshifts corresponding to speeds around 20 km/s.

One of the problems given at the XXVI International Physics Olympiad in July (see Happenings on page 53 for a report) combines the effects of these two types of redshift. We use part of this problem for this month's column.

Art by Tomas Bunk


SPEED LIMIT 299,000,000 m/SEC
PHOTON
TOM BUNK

Although gravitational redshifts are normally calculated using Einstein's theory of general relativity, we can develop a feeling for the effect by performing a semiclassical calculation. A photon with a frequency f has an effective inertial mass determined by its energy

$$mc^2 = hf.$$

Let's assume that the photon's gravitational mass is the same as its inertial mass and that the photon is emitted from the surface of a star. As the photon travels upward, it loses energy in the form mc^2 as it gains gravitational potential energy.

A. Show that the frequency shift Δf of the photon at an infinite distance from the star is

$$\frac{\Delta f}{f} = -\frac{GM}{Rc^2},$$

where G is the gravitational constant and M and R are the mass and radius of the star.

Let's now imagine launching a probe to a distant star to measure both the mass and radius of the star. Photons emitted from He^+ ions on the surface of the star are monitored through resonant absorption by He^+ ions in the probe. Resonance absorption only occurs if the ions in the probe are given a velocity toward the star that compensates for the gravitational redshift. As the probe approaches the star radially, the velocity v of the He^+ ions relative to the star is measured as a function of the radial distance d from the surface of the star. The experimental data are given in the following table:

$\beta = v/c$ $(\times 10^{-5})$	d $(\times 10^8$ m$)$
3.352	38.90
3.279	19.98
3.195	13.32
3.077	8.99
2.955	6.67

B. Utilize these data to determine the mass and radius of the star.

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 and their authors will receive special certificates from *Quantum*.

The first photon

Our contest problem in the May/June 1995 issue described our experimental determination of the particle properties of light. The problem was well received, and excellent responses came from Christopher Rybak (a junior at the Prairie School in Racine, Wisconsin), Jonathan Devor (who represented Israel in the International Physics Olympiad in Australia this past summer), Aaron Manka (a student in Huntsville, Alabama), and Lori Sonderegger and Arthur Hovey (a student and teacher, respectively, at Amity Regional High School in Woodbridge, Connecticut).

The first part of the problem required readers to interpret data from a photoelectric experiment. The equation relating the kinetic energy of the electron and the frequency of the light is

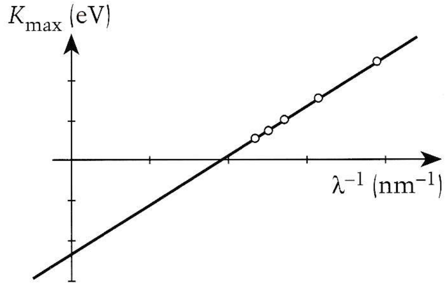
$$K_{\max} = h\nu - \phi.$$

We can substitute $\nu = c/\lambda$, where c is the speed of light in a vacuum and λ is the wavelength. This yields

$$K_{\max} = \frac{hc}{\lambda} - \phi.$$

If we plot $K_{\max}$ on the y-axis and $1/\lambda$ on the x-axis, the resulting graph should be linear; the slope is hc, and the y-intercept is the negative of the work function ϕ (see the figure below). The slope of this line is $\Delta K_{\max}/\Delta(1/\lambda)$, which is equal to 1,242 eV · nm. Therefore,

$$hc = 1{,}242 \text{ eV} \cdot \text{nm},$$



where $c = 3 \cdot 10^8$ m/s $= 3 \cdot 10^{17}$ nm/s and $h = 4.14 \cdot 10^{-15}$ eV · s. The work function for lithium is equal to 2.32 eV.

Part B asked for a proof that a free electron cannot completely absorb a photon. We can assume that absorption can take place, apply conservation of momentum and energy, and then arrive at a paradox.

Momentum conservation requires

$$\frac{h\nu}{c} = mv.$$

Energy conservation requires

$$h\nu = \tfrac{1}{2}mv^2.$$

Dividing these equations requires that the electron's velocity be equal to twice the speed of light—an impossibility.

Part C(i) asked how far away a 50-W bulb would need to be placed so that a human eye, sensitive to single photons, would detect an average of one photon per second.

Given a wavelength of 500 nm for the light, the individual photons will have a corresponding energy of $hc/\lambda = 3.97 \cdot 10^{-19}$ J. Therefore, a 50-W light bulb emits on average $1.26 \cdot 10^{20}$ photons/s. These photons spread out isotropically and will intercept the human eye. Since the pupil has a given diameter d of 0.5 cm, its area is $\pi(d/2)^2 = 0.0625\pi$ cm^2.

The ratio of the surface area of the eye to the surface area of the light sphere is equal to the ratio of the number of photons hitting the eye to the total number of photons. The distance at which one photon hits the eye every second is

$$\frac{0.0625\pi \text{ cm}^2}{4\pi R^2} = \frac{1 \text{ photon/s}}{1.26 \cdot 10^{20} \text{ photons/s}}.$$

Thus,

$$R = 14{,}000 \text{ km!}$$

Part C(ii) asked for the distance at which this light source would have to be placed if the density of photons were to be 1 photon/cm^3 on average. If we imagine a spherical shell

1 cm thick, we can assume it is made of 1-cm^3 cubes. The density stated requires one photon to be in the cube at any time. Since the photons travel at $3 \cdot 10^{10}$ cm/s, they will each traverse the cube in $3.33 \cdot 10^{-11}$ s. Therefore, $3 \cdot 10^{10}$ photons will have to travel through the cube in 1 s for the density to be 1 photon/cm^3.

This is identical to having an area density at the shell of $3 \cdot 10^{10}$ photons/cm^2/s. Applying the same procedure as in part C(i), we can solve for the distance R:

$$\frac{1.26 \cdot 10^{20} \text{ photons/s}}{4\pi R^2} = 3 \cdot 10^{10} \text{ photons/cm}^2\text{/s},$$

and

$$R = 183 \text{ m},$$

which is considerably smaller than the previous answer.

"MEAN VALUE" CONTINUED FROM PAGE 30

The length of the sum

Problem 4. The sum of the lengths of vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ on the plane is equal to 1. Prove that we can choose a number of these vectors such that the length of their sum is not less than $1/\pi$.

Solve this problem following the plan we give below.

Define the *pseudoprojection* of vector $\mathbf{p}$ on axis l as the ordinary projection if the angle between $\mathbf{p}$ and l is acute, and zero otherwise. Fix a certain axis l_0. If the angle between $\mathbf{p}$ and l_0 is ϕ, then the pseudoprojection of $\mathbf{p}$ on axis l_α that makes an angle α with l_0 can be written as $|\mathbf{p}|g(\phi - \alpha)$, where

$$g(x) = \begin{cases} \cos x, & \text{if } \cos x > 0, \\ 0, & \text{if } \cos x \le 0. \end{cases}$$

Exercise 10 (for readers who know some calculus). Check that

$$\int_0^{2\pi} g(\phi - \alpha)d\alpha = 2$$

for any ϕ. (If you don't know any calculus, you can assume this fact and still work on the last few problems.)

Now denote by $f(\alpha)$ the sum of the pseudoprojections of vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ on l_α.

Exercises

11. Prove that the mean value of f on the segment $[0, 2\pi]$ equals $1/\pi$.

12. Prove that the sum of pseudoprojections of the given vectors on a certain axis l_α is not less than $1/\pi$.

13. Prove the statement of problem 4.

The constant $1/\pi$ in problem 4 cannot be replaced by any greater number. This becomes clear if you take a sufficiently large n and the vectors drawn along the sides of a regular n-gon.

For vectors in space, the statement of problem 4 turns out to be true even if the constant $1/\pi$ is replaced by $1/4$.

ANSWERS, HINTS & SOLUTIONS ON PAGE 62

AT THE BLACKBOARD

Smoky mountain

Why the air is always warmer on the leeward side

by Ivan Vorobyov

> "It'll be a lovely day tomorrow," I said.
>
> The captain made no reply and pointed to a tall mountain which rose up directly before us.
>
> "What's that?" I asked.
>
> "Gud-Gora."
>
> "Oh yes, what about it?"
>
> "Look at the way it's smoking."
>
> Gud-Gora was indeed smoking. Light wisps of cloud crept along its sides, and on its summit lay a cloud so black that it seemed a blot on the dark sky.
>
> —Mikhail Lermontov, *A Hero of Our Time*

When crossing a mountain ridge in the direction the wind is blowing—that is, from windward to leeward—travelers immediately notice a change in the weather. As they climb the peak, they find themselves in cloudy conditions, if not rain or snow, but beyond the ridge the weather is fine—not a cloud in the sky, and the wind is gentle, warm, and dry. (Such a wind is called a foehn.) Why does such a drastic change in the weather occur? And is the air "warmed"? To answer these questions, let's see what happens when a high mountain ridge stands in the way of a wet wind.[1]

After colliding with the mountain, the air stream climbs up the slope. In so doing, any given portion of the air mass moves to a region of lower pressure. This results in an increase in the volume occupied by this air mass. As a first approximation we can disregard the heat exchange between the particular portions of the air mass and consider the air expansion to be adiabatic. In this case the work is performed at the expense of the air's internal energy, which causes a drop in the air temperature.

Now let's see what happens with the humidity as the air ascends the mountain. We know that the saturated vapor pressure decreases with temperature. The greater the altitude, the more moisture condenses, resulting in the appearance of a great number of small droplets suspended in the air—that is, a cloud or fog.

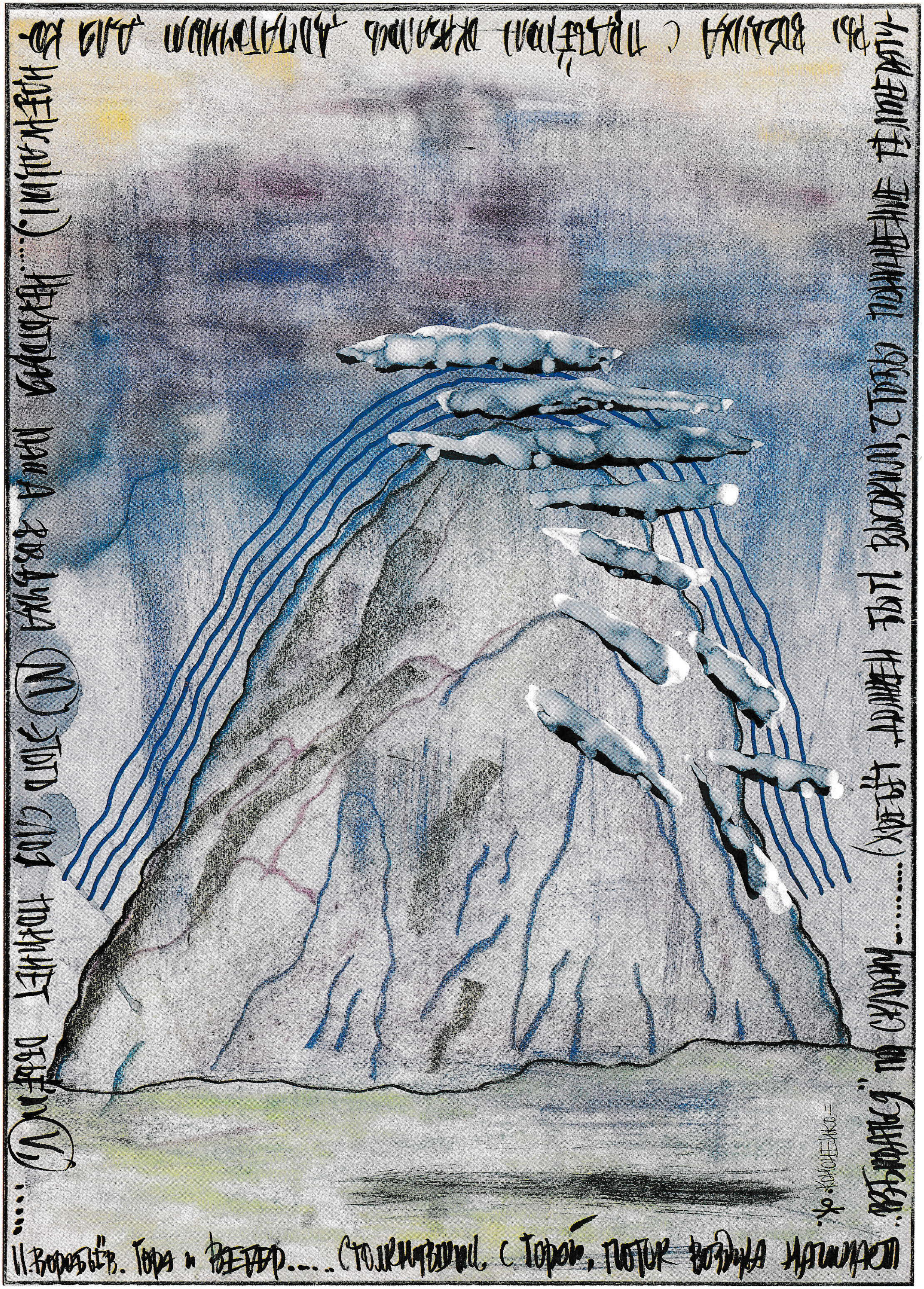
The process of condensation proceeds with the release of the latent heat of vaporization. This energy is by no means negligible: at 18°C each kilogram of vapor condensed into water releases about $2.5 \cdot 10^6$ J. Due to this release of energy, the temperature of the rising humid air decreases less quickly than it would for dry air. If wind-borne clouds crossed the ridge without losing a single drop, then during the descent along the leeward slope, as the temperature increases, the drops of water would evaporate, which would require the same amount of energy as was previously released. At the foot of the leeward side of the mountain, the air would have the same temperature and humidity that it had at the foot of the windward side.

However, if the wind is wet and the ridge is high enough, a great amount of water will drop out as fog, rain, or snow. The air descending on leeward side will be somewhat drier, and its temperature during its descent will increase more quickly than it decreased during its ascent as wet air. (Of course, the ridge must be high enough to produce a temperature decrease sufficient for condensation.) So, at the same altitude the air temperature is higher on the leeward side than on the windward side.

Now let's estimate the temperature difference quantitatively. We'll look at a curved layer of air with sides parallel to the wind and ends located on opposite sides of the mountain ridge. In a short interval of time a certain mass M of this layer will leave a volume V_1 on the windward side at the foot of the mountain, where the temperature is T_1 and the pressure is P_1. As for the leeward side, the same air mass will descend and occupy a new volume V_2 at a temperature T_2 and pressure P_2. (We assume here that at any fixed point the air temperature, pressure, and velocity do not vary with time. In particular, this

[1]See also "Cloud Formulations" in the January/February 1995 issue.—*Ed.*

Art by Yury Kononenko


11. Воробьёв. Гора и ветер..... столкнувшись с горой, поток воздуха начинает
„взбираться" по склону........(хребет должен быть высоким, чтобы понижение температу-
ры воздуха с подъёмом оказалось достаточным для ко-
нденсации.)...... Некоторая масса воздуха (M) этого слоя покинет объём (V) .....

means that in equal time intervals the same amount of air passes through any cross section of the layer.)

The internal energy of one mole of monatomic gas at temperature T is $(3/2)RT$. Air consists predominantly of diatomic molecules of oxygen and nitrogen. For a diatomic gas, the internal energy per mole is $(5/2)RT$. The change in the internal energy of our layer of air is

$$\Delta U = \frac{5}{2}\frac{M}{\mu}R(T_2 - T_1),$$

where μ is the molar mass of air and M/μ is the number of moles of air.

Now let's calculate the work performed against the forces of external pressure. Coming down on the leeward side, the wind performs positive work P_2V_2 by displacing air that previously occupied a volume V_2 at a pressure P_2. On the windward side, the air leaves a volume V_1 at a pressure P_1, performing the negative work $-P_1V_1$. The total work performed is

$$W = P_2V_2 - P_1V_1.$$

Using the ideal gas equation $PV = (M/\mu)RT$, we express the amount of work performed in terms of the temperatures T_2 and T_1:

$$W = \frac{M}{\mu}R(T_2 - T_1).$$

(Since the volumes V_1 and V_2 are at the same altitude, the potential energy of the air is the same in both the initial and the final states.)

We'll assume that the mountain is rather high, so during the ascent on the windward side, almost all the moisture condenses and falls as rain. If the mass of the precipitation is Δm, the heat Q liberated during condensation is $L\Delta m$ (where L is the heat of vaporization). It is this heat Q that causes the change in the air's internal energy ΔU and performs the work W:

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

Let's estimate the value of Q. The specific heat of vaporization varies weakly with temperature, so we'll consider it to be constant. Let the humidity of the air on the windward side correspond to the partial vapor pressure p at pressure P_1. The air of mass M and volume V_1 carries a mass of water vapor equal to Δm. According to the equation of state,

$$PV_1 = \frac{\Delta m}{\mu_0}RT_1, \quad P_1V_1 = \frac{M}{\mu}RT_1,$$

where μ_0 is the molar mass of water. From this we obtain

$$\Delta m = M\frac{\mu_0 P}{\mu P_1}.$$

Thus,

$$Q = LM\frac{\mu_0 P}{\mu P_1}.$$

In our previous reasoning we considered the entire mass M of air occupying the volume V_1 to cross the ridge—that is, we neglected the change in mass due to the rainfall. The expression obtained for Δm shows that this simplification is acceptable: $\mu_0 = 18$ g/mole, $\mu = 29$ g/mole, $P \ll P_1$ (for example, if the humidity is 50%, the temperature is 18°C, and the pressure $P_1 = 10^5$ Pa, the vapor pressure is $P \cong 0.01 \cdot 10^5$ Pa); therefore, $\Delta m \ll 1$.

Now let's write out the expression we obtained earlier, $Q = \Delta U + W$, in its final form:

$$LM\frac{\mu_0 P}{\mu P_1} = \frac{5}{2}\frac{M}{\mu}R(T_2 - T_1) + \frac{M}{\mu}R(T_2 - T_1),$$

which simplifies to

$$L\mu_0\frac{P}{P_1} = \frac{7}{2}R(T_2 - T_1).$$

This yields the formula for the temperature increase:

$$T_2 - T_1 = \frac{2}{7}\frac{L\mu_0 P}{RP_1}.$$

Inserting the numerical values $P_1 = 10^5$ Pa, $P = 0.01 \cdot 10^5$ Pa, $L = 2.5 \cdot 10^6$ J/kg, $R = 8.3$ J/mole · K), and $\mu_0 = 18$ g/mole into this formula shows that $T_2 - T_1 \cong 15$°C! That's the difference in the air temperature on the leeward and windward sides of a high mountain!

Highlanders and mountain climbers can readily confirm this estimate.

MATH INVESTIGATIONS

Lost in a forest

A problem area initiated by the late Richard E. Bellman

by George Berzsenyi

THE FOLLOWING PROBLEM was posed by Richard Bellman, one of the most outstanding applied mathematicians of our era, as problem 6 on page 133 of his 1957 classic *Dynamic Programming*, a RAND Corporation Research Study, published by Princeton University Press:

> *We are lost in a forest whose shape and dimensions are precisely known to us. How can we get out in the shortest time?*

Bellman, who not only coined the term "dynamic programming" but was also the major contributor to that field, was mainly interested in a mathematical formulation of this problem, so as to apply the theory of dynamic programming to its general solution. Unfortunately, as far as I know, no general methods were developed for the solution of this problem, and there are only a few scattered results in the literature concerning special cases. For example, the problem was resolved for the case of a forest occupying the region between two parallel lines (by O. Gross—see problem 7 on the same page of Bellman's book). The optimal solution was also found (by J. R. Isbell—see pp. 357–59 of the 1957 volume of the *Naval Research Logistics Quarterly*), in the case where the forest occupies a half-plane and one knows how far one traveled from its boundary. This problem, along with the case of a circular forest, was featured in part 2 of volume II of the *USSR Olympiad Problem Book*, which was, unfortunately, never translated into English. Even volume I, which was seemingly translated twice, was unavailable until its recent publication (ISBN 0-486-27709-7) by Dover Publications, Inc.—I highly recommend it to my readers.

The only other reference I found to Bellman's problem is an article by Gábor Tóth in Hungary's high school mathematics journal, *Kőzépiskolai Matematikai Lapok*, which recently celebrated its 100th anniversary. The article appeared in 1982 (pp. 53–55, volume 65), when Tóth was still a high school student. In addition to verifying Isbell's result and dealing with the circular forest, Tóth also managed to find the shortest paths for forests shaped like regular polygons with an even number of sides, and for rectangles whose longer side is less than $\sqrt{3}$ times its shorter side. My initial challenge to my readers is to resolve all of the aforementioned special cases.

It seems the answer is not known even for equilateral triangles. Moreover, nothing seems to be known about three-dimensional extensions of the problem. So there is a lot of ground yet to be covered—especially by one who is truly lost in a forest!

The purpose of this column is to direct the attention of *Quantum*'s readers to interesting problems in the literature that deserve to be generalized and could lead to independent research and/or science projects in mathematics. Students who succeed in unraveling the phenomena presented are encouraged to communicate their results to the author either directly or through *Quantum*, which will distribute among them valuable book prizes and/or free subscriptions.

Feedback

I am happy to report wonderful progress on the problems featured in my May/June 1995 column in *Quantum*. One of my readers, Professor Les Reed of Southwest Missouri State University, managed to prove that the value of $G(3,k)$, the maximum value of the greatest common divisor of $n^3 + k$ and $(n + 1)^3 + k$, is $27k^2 + 1$ if k is even, and $(27k^2 + 1)/4$ if k is odd. This result was also obtained by Carl Bosley, a high school student, during his stay at our NSF-supported Young Scholars Summer Program.

In response to another question, Stavros Sainidis, who is a civil engineer in Greece, proved that if m is even and if $n = 2^{m-1}k$, then $2^mk + 1$ is a common divisor of $n^m + k$ and $(n + 1)^m + k$. He also showed that for odd m, $2^m - 1$ divides both $(2^m - 3)^m + 1$ and $(2^m - 2)^m + 1$. One should be able to extend his second result to the case of arbitrary k, and hence prove that $G(m, k)$ is not equal to 1 for any values of m and k. Unfortunately, I haven't yet managed to do so.


SAE40 SAE30 10W40 SAE
SAE10
S. Ivanov

A viscous river runs through it

A study of the engine-saving properties of motor oil

by Henry D. Schreiber

YOU CHECK THE OIL IN YOUR car's engine. The dipstick indicates the oil level is low. Off you go to buy a quart of oil. Shelves are full of Pennzoil, Quaker State, Valvoline, Amoco, and other brands as well as generic brands. Each brand has selections of 10W40, 5W30, and SAE30, among others. The various sales promotions extol the merits of each: this one provides "outstanding resistance against viscosity breakdown," another one has "excellent fluidity at low temperature and maintains protection at high temperature," and another is "multiviscosity and stands up to hot engine operating conditions." Which one are you to buy? Does it matter what brand or type or grade of oil? What do the numbers mean? Do you need to just add more oil or change the oil? Maybe a check of the owner's manual will help . . . Maybe you should have gone to a service station and let them do it . . . Arrgh!

Perhaps it would first help to know why oil is needed in car engines. Oil's primary function is lubrication—to circulate through the engine to various rubbing surfaces and to form a film between these surfaces to keep them from touching. The oil has to be thin enough to flow into these locations when the engine is cold but thick enough to stay between the surfaces when the engine is hot. However, the oil cannot be too thin or it will leak from the clearances, resulting in surfaces that grind; nor too thick, requiring the engine to exert excessive power to overcome the drag between the rubbing surfaces. Other functions of motor oil include transfer of heat from the hot engine parts to the oil pan, removal of wear particles from the rubbing surfaces, and cushioning the shock of combustion.

Oil in the engine needs to be replaced periodically because it becomes ineffective in its primary role—that is, lubrication. Its properties, especially thickness, change upon prolonged exposure to the operating temperature of the engine. According to its design, a particular engine requires oil of a certain thickness; thus, changing the oil means first getting oil of the specified thickness, or "grade."

Art by Sergey Ivanov

Viscosity of liquids

"Slow as molasses in winter" refers to the rate of flow of molasses at low temperatures. It flows slower than water, and much slower when cold. Flow properties are a function of the type of liquid as well as its temperature. To quantify flow, we use the term *viscosity*. Viscosity is a fluid's internal friction, which makes it resist a tendency to flow. Viscous or thick liquids have a cohesive and sticky consistency. Whether a liquid has high or low viscosity at room temperature results from its molecular structure: are the molecules able to move past one another easily, are they intertwined, or are they attracted to each other?

Consider a liquid between two plates as shown in figure 1. The bottom plate is stationary, while the top plate is moving at a velocity v_x. The layer of liquid next to the moving plate also moves at velocity v_x, and that next to the stationary plate has zero velocity. Between these two extremes there is a velocity gradient that changes linearly from top to bottom—that is, perpendicular to the direction of the top plate. This rate of change of velocity with distance dv_x/dy is measured perpendicular to the direction of liquid flow. How fast this gradient drops off depends on the viscosity of the liquid, which in turn depends on the molecular attractions of the liquid layers. In order to move this top plate relative to the stationary bottom plate, a shearing force F must be applied to the system. The magnitude of this force can be related

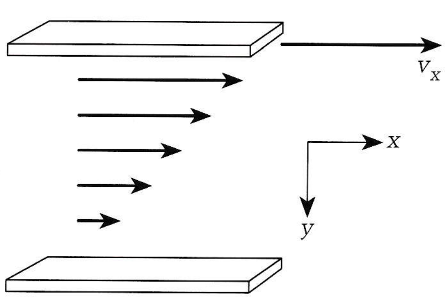


Figure 1

to the velocity gradient by

$$F = -\eta \frac{dv_x}{dy}, \qquad (1)$$

where η is the index of viscosity (or just the viscosity) and can be viewed as a proportionality factor analogous to the coefficient of friction for the motion of objects. The relationship is negative because v_x decreases with the distance away from the plate with the applied force. An η of 1 poise (0.1 Pa · s) is obtained when a shearing force of 1 dyne/cm^2 results in a gradient dv_x/dy of 1 cm/s per centimeter perpendicular to the shearing plane.[1] Most liquids have viscosities that are measured in centipoise (0.01 poise).

Although equation (1) defines the viscosity η, experimental measurements of the applied force and velocity gradient are not easy to obtain in the laboratory. Usually one measures the viscosity indirectly by monitoring a property of liquid flow that depends on the viscosity. For example, classical methods measure the time it takes for a liquid to flow through a capillary tube or out of an orifice in a container bottom, the torque needed to turn a crank in a liquid, or the time for a ball to drop through a column of a liquid.

To understand the effect of temperature on the flow of liquids, consider the fact that vacancies (defects) exist in the structure of liquids. That is, open space exists between adjacent molecules making up the liquid. Molecules are continually moving in and out of the vacancies to permit the liquid to flow, but such movement requires energy for the molecules to enter or leave the vacancies. This energy is called the activation energy E. Because more energy is available at high temperature, the molecules constituting the liquid can flow more easily at the high temperature. Accordingly, the viscosities of most liquids decrease with increasing temperature. Mathematically, the fluidity of a liquid is related to the absolute temperature T by an exponential Arrhenius function

$$\text{fluidity} = \frac{1}{\eta} = Ae^{-E/RT}, \qquad (2)$$

where A is a constant specific to that liquid and R is the ideal gas constant (8.31 J/mole · K). If the natural logarithm is taken of both sides of this equation and the terms rearranged, we obtain

$$\ln \eta = \frac{E}{RT} - \ln A. \qquad (3)$$

A plot of the natural logarithm of the viscosity versus the reciprocal temperature is linear with a slope equal to the activation energy for flow divided by the ideal gas constant.

Viscosity of oil

Viscosity is oil's most important property in lubricating a car's engine. The ideal oil has low viscosity at low temperatures so that it flows between the engine's rubbing surfaces when cold, even though liquids have a tendency to thicken at low temperatures. On the other hand, the ideal oil does not have too low a viscosity at high temperatures—otherwise, the layers between the rubbing surfaces are too thin. All oils thin as they get hot, as shown by equation (2). Exposure to the high temperature of an operating engine tends to break down the oil with time. In a sense, the oil burns or oxidizes in the hot engine, which leads to a different molecular structure for the liquid. The oil thickens or becomes more viscous under this long-term exposure to high temperature, no longer providing the correct lubrication.

All motor oils are graded according to an SAE (Society of Automotive Engineers) number that provides a viscosity range. The oils are classified by their performance at a low temperature (T = 0°F) followed by the letter W, which at one time indicated a winter grade of oil, and at a high temperature (T = 210°F). Figure 2 provides these viscosity ranges for the different grades of oil. Note that the ranges for the different grades of oil are quite wide, allowing variability of motor oils within each classification. The grading scale provides that SAE30 oil falls only into that one classification at the operating temperature of 210°F; 20W20 fits into the 20 weight classification at both hot and cold temperatures; while 10W40 meets different classifications at the two extremes of temperature. Oils classified as 10W40 and 5W30 are examples of multigrade oils, for which the effect of temperature on viscosity is quite different than straight oils such as SAE30. This effect is shown schematically in figure 3.

Experimental measurement of viscosity

Industry measures the viscosity of oil with the Saybolt viscometer, which determines the time required for the oil to flow through an orifice at the bottom of a standard container. However, the falling-ball method can more readily be adapted to simple

Temperature (°F)	SAE number	Viscosity (η)
0	5W	–1,200 cp
	10W	1,200–2,400 cp
	20W	2,400–9,600 cp
210	20	5.7–9.6 cs
	30	9.6–12.9 cs
	40	12.9–16.8 cs
	50	16.8–22.7 cs

units: cp = centipoise, cs = centistokes
cp = cs × density of oil at temperature

Figure 2

[1] A dyne is the force unit in the centimeter–gram–second system. One dyne = 10^{-5} N.

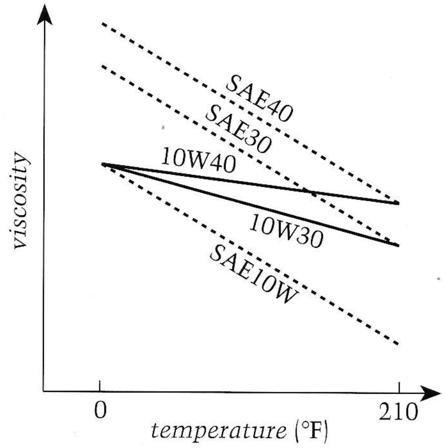


Figure 3

laboratory equipment. The goal of this experiment is to determine and compare the temperature dependence of different brands or different grades of motor oil.

Suppose a spherical ball is dropped through a column of a liquid as shown in figure 4. The force causing the ball to settle in a liquid is the ball's effective mass times its acceleration due to gravity g. The effective ball mass is the mass of the ball minus the mass of the liquid that it displaces. If r and ρ are the radius and density of the ball, respectively, and ρ_0 is the density of the liquid, then

$$\text{effective mass} = \frac{4}{3}\pi r^3(\rho - \rho_0), \quad (4)$$

which is simply the spherical volume times the difference in densities. Then the downward force of the ball is

$$F_{\text{down}} = \frac{4}{3}\pi r^3(\rho - \rho_0)g. \quad (5)$$

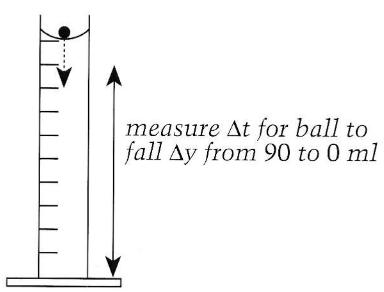


Figure 4

An inherent assumption is that the radius of the ball is small enough with respect to the column radius so there are no turbulence or edge effects.

The force retarding the motion of the falling ball equals the frictional coefficient f of the liquid times the velocity v of the ball:

$$F_{\text{up}} = fv. \quad (6)$$

This frictional coefficient is related to the liquid's viscosity, and in fact Stokes showed for spheres that

$$f = 6\pi\eta r. \quad (7)$$

But the velocity of the falling ball is simply dy/dt, or on average $\Delta y/\Delta t$, so that

$$F_{\text{up}} = 6\pi\eta r\frac{\Delta y}{\Delta t}. \quad (8)$$

Equating up and down forces when the ball achieves a constant settling rate, we obtain

$$6\pi\eta r\frac{\Delta y}{\Delta t} = \frac{4}{3}\pi r^3(\rho - \rho_0)g, \quad (9)$$

or

$$\eta = \frac{2}{9}r^2(\rho - \rho_0)g\frac{\Delta t}{\Delta y}. \quad (10)$$

However, to measure the absolute viscosity accurately, we need to know the radius of the ball and the densities of the ball and liquid rather exactly. Usually the viscosity is measured with respect to a standard (a liquid of about the same density) in the same apparatus with a ball of the same dimensions and density. In such a case, we could simply say

$$\eta = k\Delta t, \quad (11)$$

where Δt is the time for the ball to fall a certain distance and k is the proportionality constant, which is determined by measuring the time required for the ball to fall the same distance in a standard liquid of known viscosity:

$$k = \frac{\eta_s}{\Delta t_s}, \quad (12)$$

where the subscript s indicates the standard liquid. A good approximation for the viscosity of oil is obtained by measuring times for a ball to fall a specified distance in the oil with respect to a reference liquid. The relative viscosity is proportional to the ratio of the times:

$$\frac{\eta}{\eta_s} = \frac{\Delta t}{\Delta t_s}. \quad (13)$$

In order to determine the viscosities more accurately, we also have to consider density differences between the reference liquid and oil.

The experimental apparatus is quite easy to construct, as shown in figure 4. Fill a 100-ml graduated cylinder with the liquid to the 100-ml mark. With a pair of tweezers place a plastic bead or a Teflon ball (about 3–4 mm in diameter) at the center of the surface. Time (to the nearest 0.1 sec) the bead as it falls from the 90-ml mark to the bottom. (The first 10 ml is a sufficient distance to achieve a constant settling rate.) The standard must be relatively viscous in order to give measurable times and a good reference for motor oil. A cooking oil for which the viscosity can be found in the *Handbook of Chemistry and Physics*, or a viscous liquid such as glycerol or ethylene glycol, is a good standard.

Take several measurements using a sample of motor oil at room temperature to get a good average time. The motor oil can be gently warmed in a container of hot water and subsequently cooled in a container of ice water. The reference liquid need only be measured at room temperature.

These data can be used in equation (13) to provide values of the oil's viscosity as a function of the temperature. Further, the natural logarithm of the viscosity can be plotted against the reciprocal of the temperature (in kelvins) according to equation (3) to obtain the activation energy for the oil as well as the viscosities at the reference temperatures (0°F and 210°F). An example of some typical experimental results is

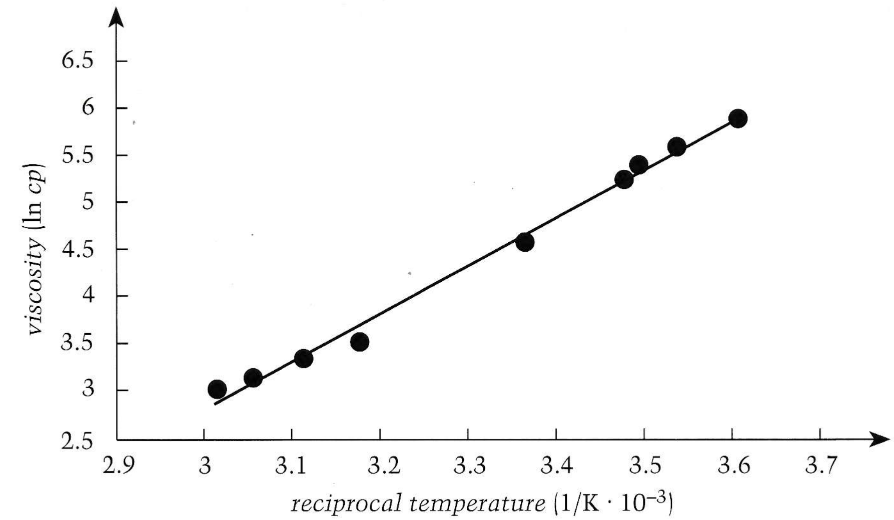


Figure 5

shown in figure 5. In your experiment you might be able to determine the differences among several brands of motor oil of the same grade—for example, 10W40 from Amoco, Pennzoil, Quaker State, and other manufacturers. Or you can ascertain the differences between the various grades, such as SAE30, 10W30, and 10W40, of the same brand.

Characteristics of motor oil

Advertisements promote individual brands of motor oil and their resistance to changes from their claimed viscosities. In order to be classified as a certain grade, the oil only has to hit the right ranges of viscosity at one or two temperatures, and these ranges are quite wide. About 20% of the volume of the motor oil may be additives to inhibit oxidation of the oil so that it retains its viscosity, to inhibit rust and corrosion of the rubbing surfaces, to clean deposits within the engine, to keep particles dispersed in the oil until filtered, and to neutralize acids produced during combustion. Such additives reinforce the oil's properties and also provide new beneficial properties. But the additives may not be compatible from one brand to another, so when you add oil to a car engine, use the same brand and grade to minimize the chance of conflicting or countereffective additives.

As oil circulates in the engine, the oil deteriorates due to exposure to the high temperature of the engine. The hot oil reacts with oxygen, breaks down, and forms oxidized deposits in the engine. Dirt, coolant, and sludge from various sources also contaminate the oil over time. When the oil no longer possesses the proper viscosity, it needs to be changed before it loses its effectiveness. The development of synthetic motor oils has resulted in more stable oils that are less subject to oxidation. As a result they don't thicken under long-term high-temperature use. Research in this area has also produced lighter oils, which translates into less engine friction.

Here are a few more tips. Always use the motor oil grade specified in the owner's manual for that engine. Use the same brand of oil if you are adding to what's already present in the engine. Unfortunately, there is no viscometer or other instrument to determine the oil's viscosity while it is in the engine to see whether the oil has deteriorated enough to be changed. We now rely on the rather general prescription: change your oil every few thousand miles. Perhaps one of our readers can devise a monitoring sensor or gauge to replace this rather unscientific rule of thumb!

Henry D. Schreiber *is a professor of chemistry, specializing in teaching general as well as physical chemistry, at the Virginia Military Institute in Lexington, Virginia.*

"PUBLISHER'S PAGE" CONTINUED FROM PAGE 3

theory is to find a transformation of the position and momenta (the six coordinates in phase-space) at some time t to the initial time $t = 0$. Then the latter becomes the solution for the system. When this theory is applied in a certain way, you find that there can be wave fronts of the principal function $S(q, P, t)$ propagating in phase-space much like a water wave moves across the surface of water. These wave fronts are mathematical descriptions of the mechanical particle under consideration. By means of mathematical derivations, with no reference to quantum principles, this analysis leads to an equation $(\nabla W)^2 = 2m(E - V)$, where $S = W - Et$ defines W. This equation is identical to the *eikonal equation of geometrical optics*. In the final application to optics, it can be shown that E and ν must be proportional, leading exactly to the Schrödinger wave equation, except that the constant h isn't determined.

Filling this gap gave me a great sense of satisfaction that there was a clear connection between classical physics and quantum physics, and that the constant h, determined by Planck empirically, was just as reasonable to me as finding G in the Cavendish experiment to provide the constant in the law of universal gravitation, which Newton didn't express when he developed his law.

Still I wondered about the origins of h, and I wondered about the fundamental aspects of quantum behavior. Instead of using hypotheses in an axiomatic way, could there be some kind of logical deduction leading to quantum mechanics?

Enter my former student, Mike Mackey.

The student as teacher

Mike came to see me once several years ago, and I discussed with him my curiosity about the underlying ideas of quantum mechanics. We talked about "hidden variable theory" and about the peculiar connection to thermodynamics in the adiabatic invariant. I expressed dismay that my mathematical abilities were never good enough, and that what I did know was now so rusty I could not seriously pursue something like this. Knowing his remarkable mathematical ability I urged him at least to consider some of these matters theoretically. "The Dynamic Origin of Increasing Entropy" was the result of his efforts.

Mike's paper addressed the fundamental question of why entropy approaches a maximum value in a system, since all of the laws of physics are formulated as reversible dynamical systems. What Mike found suggests that "all currently formulated physical laws may not be at the foundation of the thermodynamic behavior we observe daily." He further observed that either "physical laws are incorrect and that more appropriate formulations in terms of irreversible semidynamical systems await discovery," or "other phenomena may mask the operation of these reversible systems so they appear to be irreversible to the observer."

What this paper said to me was that *if* the second law of thermodynamics is a correct and universal law, *then* there must be "hidden variables" requiring new laws of physics, or there must be other phenomena with which we are unfamiliar or whose mechanisms are unknown. In either case these conclusions offer exciting opportunities for our best young minds to create new knowledge as theoreticians, or to make new discoveries as experimentalists.

I was frustrated and saddened by my inability to follow Mike's mathematics in the paper, like *Frobenius-Perron operators* (a Markov operator) or the *forward Kolmogorov equation*. But I was cheered by the realization that the intensely curious young man I first encountered when he was 14 years old now did have that remarkable knowledge and could create such beautiful work. This is surely one of the most precious rewards of teaching!

—Bill G. Aldridge

Georg Cantor

"No one shall expel us from the paradise that Cantor has created for us."—David Hilbert

by Vladimir Tikhomirov

MARCH 3, 1995, WAS THE 150th anniversary of the birth of Georg Cantor. The outstanding Russian topologist Pavel Sergeyevich Alexandrov said: "I don't think there was a mathematician in the second half of the 19th century who had a greater impact on the development of mathematical science than the creator of abstract set theory, Georg Cantor." This opinion won't be shared by everyone—after all, the great Henri Poincaré lived and worked at the same time. But one cannot deny the tremendous, incomparable influence of Cantor's work on all the mathematics that followed. He enriched our science with fundamental new concepts, profound results, substantial theories, fruitful methods . . .

Cantor's ideas were received initially with a healthy dose of scepticism, then—by many mathematicians—with admiration. Later they were criticized, and the repercussions of this criticism are still to be heard. But here's the opinion of one of the greatest mathematicians of all time, David Hilbert: "I think that [Cantor's theory of sets] is the highest expression of human genius and one of the greatest achievements of human spiritual activity." And some time later, when the paradoxes of set theory had plunged many thinkers into doubts about its significance, Hilbert uttered the words quoted in the subtitle above.

So what was Cantor's contribution to mathematics? Let me start simply with an inventory. Essentially, Cantor brought forth the idea of building an entire body of mathematics on the basis of set theory. He introduced the fundamental notions of the theory of sets and topology, laid the foundations for the theory of sets itself, created one of the constructions of the real number system, invented the diagonal method, proved that the continuum is uncountable and spaces of different dimensionality have the same cardinality, proved the existence of transcendental numbers, constructed Cantor's perfect set and Cantor's staircase, proved the nonexistence of the "highest" cardinality, proved the fundamental uniqueness theorem for a trigonometric series, and posed the continuum problem.

Cantor's contribution was so basic and fundamental that its main concepts can be explained to just about anyone. (In his speech at the Paris congress of mathematicians, where he formulated his famous set of problems, Hilbert said, "A mathematical theory can be considered perfect only if . . . it can be explained to just about anyone." All of Cantor's creations bear this sign of perfection.)

I'll try to tell you about almost all of Cantor's most substantial achievements (and in sufficient detail, too) in a few pages of a magazine intended primarily for high school students. This article will, I hope, attest to the fact that the depth of mathematical results need not always be measured by the length of the text and the difficulty of the proofs!

Now, let's move on and review Cantor's legacy. We'll begin with his most fundamental achievement.

Sets and the structure of mathematics

It is Georg Cantor who deserves the credit for introducing the notion of the "set" (or "collection") in mathematics. It belongs to the category of primary undefined notions. It can only be interpreted, explained, and illustrated by examples.

A set, wrote Cantor, is a collection of definite, distinguishable objects of perception or thought conceived as a whole. P. S. Alexandrov and A. N. Kolmogorov write in one of their books: "For example, we can speak of the set of all persons in a given room, or the set of geese swimming in a pond." You can easily continue this list of possible sets.

Kolmogorov said, "At the basis of all of mathematics lies the pure theory of sets." This statement reflects a view of the structure of mathematics that was professed by many mathematicians of his generation. Basically, this ideology is a brainchild of Cantor. It was distinctly formulated by Hilbert and Weyl, and later consolidated and developed in the fundamental work of the group of French mathematicians united under the nom de plume of Nicolas Bourbaki.

In the essay "The Architecture of Mathematics," Bourbaki writes: "The intrinsic evolution of mathematical science . . . strengthened the unity of its parts. . . . The end result was the trend usually called the 'axiomatic method.'" (Many prominent mathematicians reject Bourbaki's assertion and believe that the axiomatic method is a dead end in the history of mathematics. But this interesting subject lies outside the purview of an article dedicated to Georg Cantor.) According to Bourbaki, mathematics falls into *structures* of ever increasing complexity—that is, sets supplied with algebraic operations, systems of subsets, and so on.

This notion takes us back to the main character in our story, Cantor, who developed the theory of one of the most widely known structures—the real number system.

The theory of real numbers

From the axiomatic standpoint, the system of real numbers is a *complete, ordered field*. An ordered field is a set with two operations—addition $(a + b)$ and multiplication (ab)—and an order relation $(a < b)$ that obey all the basic properties (axioms) we learn about in school: $a + b = b + a$; $a(b + c) = ab + ac$; "if $a < b$ and $b < c$, then $a < c$"; "the sum of any two positive numbers is positive"; and so on. Most authors use fewer than 20 such axioms, and they can be chosen in various ways. The term "complete" means that this structure satisfies one other very important axiom (a topological one)—the *completeness axiom*.

In Cantor's time the formulation of one or another completeness axiom constituted the theory of the real number system per se, because the aforementioned algebraic properties of addition, multiplication, and order were implied. Such axioms were introduced (or introduced in effect) by all the eminent mathematicians who laid the rigorous foundations of calculus—Cauchy, Bolzano, Dedekind, Weierstrass, and Cantor.

THE CANTOR COMPLETENESS AXIOM (axiom of nested intervals). *Any sequence of nested segments whose lengths approach zero has a unique common point.*

Now we have the complete list of axioms for the real numbers. It can be shown that a set defined by this system of axioms is, in essence, unique (up to isomorphism). This means that any two objects satisfying all these axioms can be placed in a one-to-one correspondence that preserves the two algebraic operations on them and the order. The object (unique in the stated sense) thus described is called the real number system, or the real line.

Cantor faced the problem of constructing the real number system when he tried to apply his set theory to a particular question in algebra and number theory—the problem of the existence of nonalgebraic numbers (we'll return to this later).

It should also be noted that Cantor gave another definition of real numbers—as classes of equivalent fundamental sequences. In fact, he coined the term "fundamental sequence."

But let's proceed to the next topic.

Georg Cantor was born in 1845 to the family of a German diplomat in St. Petersburg. He went to school in Berlin and studied mathematics in Zurich, Göttingen, and Berlin, where in 1867 he received his master's degree. After that he settled in Halle and taught mathematics at the local university until 1913. From 1879 he was the head of the mathematics department at the university. His very productive career was cut short in 1897, when he fell seriously ill. He died in Halle in 1918.

The basics of set theory

The most important notion in the theory of sets is undoubtedly Cantor's notion of the *cardinality* (or cardinal number) of a set. It generalizes the notion of the number of elements in a set. Two sets are called *equivalent* if there exists a one-to-one correspondence between them. So the cardinality of a set is defined as the common feature of all sets that are equivalent to one another. Cantor wrote: "The cardinality of a set is what is left in our mind after we abstract ourselves from the qualitative nature of its elements and their order." The smallest infinite cardinality is that of the natural numbers. Cantor designated it as $\aleph_0$ ("aleph null"). Any set of cardinality $\aleph_0$ is called *countable*—its elements can be enumerated. The cardinality of the set of real numbers forming the segment $I = [0, 1]$ is called the cardinality of the continuum and is sometimes denoted by c.

Cantor proved the following theorems that form the core of set theory. He showed that the integers are countable by listing them as 0, 1, –1, 2, –2, 3, –3, ... Any set that can be arranged in a list like this is certainly countable, since we can say that the first element corresponds to the natural number 1, the second to 2, the third to 3, and so on.

THEOREM 1. *The set of pairs of natural numbers is countable.*

Indeed, any natural number is uniquely factored into the product of a power of two and an odd integer. So the formula $(m, n) \leftrightarrow 2^{m-1}(2n-1)$ defines a one-to-one correspondence between the pairs (m, n) (with natural m, n) and the set of all natural numbers.

Another method of enumeration is illustrated in figure 1.

COROLLARY. *A set consisting of the union of countably many countable sets is countable.*

As we've noted, a countable set can be arranged as an infinite "list." So think of the first countable set as written down in a horizontal list, and second written down as a list under it, the third under the second, and so on. The diagram will look something like figure 1, and the enumeration illustrated there works once more.

THEOREM 2. *The continuum is uncountable.*

Proof. Let's suppose that the continuum can be enumerated. Each point of the segment I is representable as an infinite decimal fraction (for uniqueness, finite decimal fractions will be written with a periodic 9—for instance, $0.25 = 0.24(9) = 0.2499...9...$).

If the numbers in the segment I are countable, then let's list them vertically:

$$\begin{aligned} x_1 &= 0.x_{11}x_{12}\ldots x_{1n} \\ x_2 &= 0.x_{21}x_{22}\ldots x_{2n} \\ &\vdots \qquad \vdots \end{aligned}$$

where x_{ij} is one of the digits 0 through 9 and none of these infinite decimal fractions ends in a solid row of zeros starting from a certain place. Consider the decimal $\alpha = 0.y_1y_2\ldots y_n\ldots$, where $y_i = 1$, if $x_{ii} \neq 1$, and $y_i = 2$, if $x_{ii} = 1$. (It is this construction that was termed the *Cantor diagonal process.*) By our assumption, the number α must have a certain number in our enumeration—say, $\alpha = x_N$. But then $\alpha = 0.x_{N1}x_{N2}\ldots x_{NN}\ldots$, which is impossible because, by construction, the Nth digit in α is $y_N \neq x_{NN}$. This contradiction proves the theorem.

The diagonal process has become an important and widely used tool in mathematical proofs. For instance, proofs of two remarkable theorems—Sousline's theorem about the existence of a new class of sets (A-sets) and Gödel's incompleteness theorem—are based on the Cantor diagonal.

Now we'll derive an important corollary to this theorem. A real number is called *algebraic* if it is a root of a polynomial with integer coefficients. One example is the number $\sqrt{2}$—the root of the equation $x^2 - 2 = 0$. Nonalgebraic numbers are called *transcendental.*

THEOREM 3. *Transcendental numbers exist.*

A polynomial of zero degree (with integer "coefficient") is an integer, and this set of polynomials is certainly countable. A polynomial of degree 1 looks like $ax + b$ and is specified by the ordered pair (a, b) of integers. So the set of first-degree polynomials is countable (by theorem 2).

A polynomial of degree 2 looks like $ax^2 + bx + c$ and so is specified by the ordered triple (a, b, c) of integers. Or, we can say it is specified by a number a and a first-degree polynomial $bx + c$. So the set of second-degree polynomials is specified by ordered pairs, each of whose elements (an integer and a first-degree polynomial) come from countable sets. We can thus arrange them as in figure 1, and the scheme in that figure shows that they are countable.

In the same way, we find that the polynomials with integer coefficients of degrees 3, 4, 5, ... are all countable. So the set of polynomials with integer coefficients is a union of countably many countable sets and is therefore (by the corollary to theorem 1) countable.

Cantor's own proof of this theorem was somewhat different. Suppose the algebraic numbers were countable and we could arrange them in a list. Take the first number. Find a segment that doesn't contain it. Take the second number. Find a segment inside the first segment that doesn't contain the second algebraic number. Then proceed in the same way to construct a sequence of nested segments whose lengths approach zero. By the Cantor axiom they have a common point, which must be a transcendental number.

In the same way (without the diagonal process) the uncountability of the continuum can be proved.

The first "explicit" construction for a transcendental number was given by Liouville in 1844. He showed that the number $10^{-1} + 10^{-2!} + 10^{-3!} + \ldots$ is transcendental. Many mathematicians erroneously contrasted Liouville's "constructive" solution with Cantor's "pure existence" theorem. However, this is a false contrast. Cantor's method is perfectly constructive. It is possible to make a computer program that will "calculate" a transcendental number step by step. Details can be found in the article "Georg Cantor and Transcendental Numbers" by Robert Gray (*American Mathematical Monthly*, November 1994, 819–32).

The next theorem created a sensation at the time. Isn't it obvious that there are "more" points in a square than in a segment? For a long time Cantor also thought that the cardinality of a square was greater than that

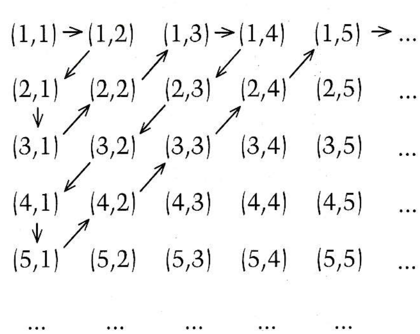


Figure 1

A sketch of the diagram drawn on the memorial plaque dedicated to Cantor in Halle.

of a segment. But later, to his great astonishment, he discovered that this isn't so. He expressed his shock in a letter to Dedekind. The letter—written, naturally, in German—has a French exclamation: "Je le vois, mais je ne crois pas!" (I see it, but I don't believe it!).

THEOREM 4. *The set I of the points of the segment $[0, 1]$ is equivalent to the set I^2 of the points of the square $\{(x, y) \mid 0 \le x \le 1,\ 0 \le y \le 1\}$.*

Proof. Split the decimal notation of a number in I into blocks consisting, by definition, of a significant (nonzero) digit with all zeros immediately preceding it. (We adhere to the same agreement about the representation of finite decimals as in theorem 2.) For instance, the number 0.0032050007... is partitioned into the blocks {003}{2}{05}{0007}.... Blocks will be labeled b_i and c_i. This establishes a one-to-one correspondence between decimal fractions and sequences of blocks. Take a pair (x, y) from I^2. Suppose its first coordinate $x = 0.x_1x_2\ldots$ is associated with the sequence of blocks $0.b_1b_2\ldots$, and the second coordinate $y = 0.y_1y_2\ldots$ with the sequence of blocks $0.c_1c_2\ldots$. Now consider the mixed sequence of blocks $0.b_1c_1b_2c_2\ldots$, which corresponds to a unique number z in I. Conversely, for any z in I the inverse process of "separating blocks" allows us to find the corresponding point (x, y) in the square I^2. Thus, we have a one-to-one correspondence between the square and the segment, which completes the proof.

THEOREM 5. *The cardinality of the set of all subsets of any given set is greater than the cardinality of the given set.*

First of all, I have to explain what is meant by the phrase "the cardinality of set Y is greater than the cardinality of X." It means that X is equivalent to a subset of Y, but no subset of X is equivalent to Y. (Cantor very much wanted to prove that if X is equivalent to a subset of Y and Y is equivalent to a subset of X, then X and Y are equivalent. He found a proof; but by that time the theorem had already been proved by Felix Bernstein. Now it's called F. Bernstein's theorem.)

Now let's prove theorem 5. Let X be the given set and S the set of all its subsets. Set X is equivalent to the set of all one-point subsets of X, which is a subset of S. So it remains to show that S is not equivalent to X.[1]

Suppose the two sets are equivalent. Then there is a one-to-one correspondence between elements of X and the subsets of X. Denote by $A(x)$ the subset corresponding to element x. Now x is an element, and $A(x)$ is a set, so either x is an element of $A(x)$ or x is not an element of $A(x)$. Let's call x a "good" element if $x \in A(x)$, and let's call x a "bad" element otherwise. Then each element of X is either good or bad.

Now we consider the subset of X consisting of all the "bad" elements of X. This subset must correspond to some element ξ. Is this element good or bad? Well, if it's good, then $\xi \in A(\xi)$, which says that ξ is bad. So ξ cannot be good. Can it be bad? Well, if it's bad, then $\xi \in A(\xi)$. But this says that ξ is good! So our assumption that there is a one-to-one correspondence between elements of X and subsets of X leads to a contradiction.

Let's look at a few more notions introduced by Cantor and some of his constructions.

Topology and constructions

Cantor laid the foundation of general topology. He defined its most important notions for the case of a straight line, but it will be more convenient for us to consider them in the most general setting.

Let X be a set and τ a certain system of its subsets. The pair (X, τ) is called a *topological space* if the intersection of a finite number and the union of any number of sets from τ all belong to τ. The sets from τ are called *open* sets, and any open set that contains the point $x \in X$ is called a neighborhood of x. (The concept of topological space crystallized at the beginning of this century and took its final shape in the work of Hausdorff.)

[1]Here F. Bernstein's theorem is implicitly used.

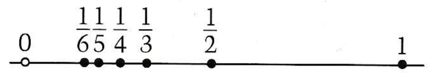


Figure 2

The set $A = \{1/n \mid n \ge 1\}$, whose limit point is 0. The derivative set consists of 0 alone; $A \cup \{0\}$ is closed.

Here are the definitions of some basic topological concepts.

A point x in a topological space is a *limit point* for set A (fig. 2) if any neighborhood of x contains a point of A distinct from x. A set that contains all its limit points is *closed*. The set of all limit points of A is the *derivative set* of A. A set that coincides with the set of its limit points is a *perfect set*. The set consisting of the points of A and its limit points is the *closure* of A. A set such that the intersection of its closure with some open set coincides with that set is said to be *everywhere dense* in it. And a set that is not everywhere dense in any open set is said to be *nowhere dense*.

Mathematicians master these concepts at the very beginning of their mathematical education, and this can create the illusion that they've always been there. But no, they all came from one person—Georg Cantor.

Now I'll describe two remarkable Cantor constructions.[2]

Take the unit segment I and cut out the middle interval of length 1/3 (in other words, the interval (1/3, 2/3)). Two segments [0, 1/3] and [2/3, 1] will remain. Subject each of them to a similar operation (deleting of the middle third). This will leave four segments, and we'll do the same with each of them. This process is continued forever. What is left in the end? It turns out that the leftover is both big and small. The remaining set is perfect, nowhere dense, and has the cardinality of the continuum. On the other hand, it can be covered with finitely many

[2]These constructions have already appeared in *Quantum*—see "Smale's Horseshoe" in the May/June 1995 issue and "Bushels of Pairs" in the November/December 1993 issue for additional discussion and applications.

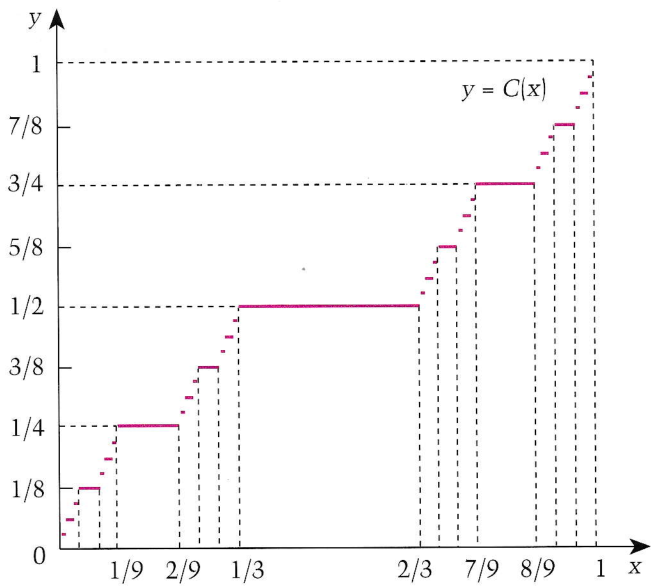


Figure 3

The first four steps in constructing the Cantor set. C(x) is Cantor's staircase. The sum of the lengths of all the deleted segments is equal to the length of the entire segment: 1/3 + 2/9 + 4/27 + ... = (1/3)/(1 – 2/3) = 1.

intervals of arbitrarily small total length (sets with this property are said to be of measure zero). The remarkable set thus constructed is called the *Cantor set.*

For some time such sets had been considered freaks that could not arise in classical calculus. But it turned out that this isn't the case. Such sets emerged in the most classical problems of mechanics and even in the works of Poincaré himself! (Nowadays many are convinced that "freaks" are all that one encounters in "real life." But that's another subject, one that I can't touch upon here.)

On the next two intervals ("of the second rank") we set $C(x) = 1/4$ and $C(x) = 3/4$ (see figure 3); on the intervals of the third rank the function takes the values 1/8, 3/8, 5/8, 7/8; and so on. This defines our function on all the deleted intervals. In the Cantor set itself, the function is extended so as to preserve continuity. The function $C(x)$ is called *Cantor's staircase.* It's a monotonic function and (if you know this concept from calculus) has a derivative on all deleted intervals, whose total length is one (in fact, the derivative is zero inside each deleted interval).

Of course, we're far from exhausting the legacy of this extraordinary scholar. (One theory we haven't even touched is the theory of transfinite numbers. You can read about it in any book on set theory.) Nonetheless, I think this article has covered the most significant areas of Cantor's work.

Concluding remarks

I'd like to advise you, if you haven't done it yet, to look through Cantor's volume on set theory. I want to quote P. S. Alexandrov again. He wrote that Cantor's works were among the very first mathematical texts he read in his youth. They made an indelible impression on him. Alexandrov expressed the hope that these works would be enthusiastically studied "and would help bring forth young people [among its readers] who have an interest in and talent for mathematics." My sentiments exactly.

Georg Cantor's life was tragic in many respects. At the very beginning of his research he posed the *continuum problem*: he wanted very badly to prove that there are no intermediate cardinalities between $\aleph_0$ and the cardinality of the continuum. A number of times he thought he'd reached his long-sought goal, only to realize he was mistaken. The solution of the continuum problem (the first in the list of problems posed by David Hilbert as a challenge to twentieth-century mathematicians) by Gödel and Cohen is one of the most important achievements of our century. It turned out that within the framework of the existing axioms of set theory, this problem can be neither proved nor refuted.

Several factors, including the intensity of his work regimen, the failure of many mathematicians to see the significance of his results, and his inability to solve the continuum problem, brought Cantor to a state of severe depression. He died in a neurological clinic in Halle on January 6, 1918. Let this be a lesson to young scientists—take care of yourself!

Cantor didn't live to see the time when his ideas would become universally recognized. This occurred in the 1920s. The upsurge of mathematics after World War I, the development of topology and functional analysis, and, in general, the revision of the essence of mathematics itself in the works of Hilbert, Weyl, Bourbaki, and many others—all these were consequences of the Cantor revolution.

"Can-do" competitors in Canberra

US Physics Team wins four golds and a silver

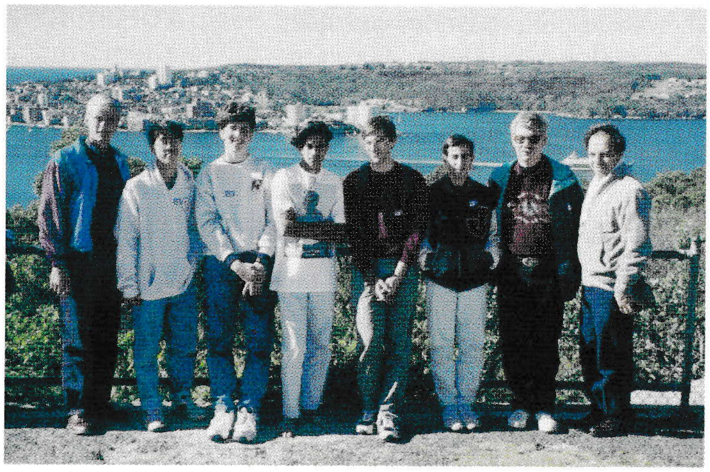
by Larry D. Kirkpatrick

THE US PHYSICS TEAM RECEIVED four gold medals and a silver medal at the XXVI International Physics Olympiad held in Canberra, Australia, in July. This is the first time in its ten-year history that all five members of the US Physics Team have won a medal. The Chinese team was the only team to win more gold medals, winning five for the second straight year. The Iranian team won two golds and three silvers, while the teams from Germany and Russia each earned two golds, two silvers, and a bronze. Great Britain took home two golds and three bronzes. The remaining gold medals were earned by students from Hungary, Italy, the Netherlands, South Korea, Turkey, and Vietnam. Australia, the host country, won two silvers and three bronzes for its best medal count ever, and the Canadians garnered two silvers and two honorable mentions.

In terms of the total points earned by the five members of a team, the US students placed second among the 51 nations at the Olympiad. This was an improvement over their third-place finish last year in Beijing, China.

The US Physics Team was led by gold medal winner Rhiju Das, who attained the second highest score. Rhiju graduated from the Oklahoma School of Science and Mathematics in Oklahoma City, where he studied with Xifan Liu. Paul Lujan placed in the top ten in the world and will be a senior next year at Lowell High School in San Francisco, where Richard Shapiro is the physics instructor. Ben Rahn won a gold medal after graduating from Thomas Jefferson High School of Science and Technology in Alexandria, Virginia. He studied physics with John Dell, as did his schoolmate Jooh Pahk, who will return to Thomas Jefferson for his senior year and may compete for another gold medal next year. The silver medalist is Daniel Phillips, who graduated from Concord Carlisle High School in Massachusetts. His physics teacher was William Barnes. The three seniors are now attending Harvard University.

The success of this year's team can be attributed to four factors: (1) the group learning and friendly competition at the training camp held at the University of Maryland; (2) the very intense studying by the five team members during the five weeks between the training camp and the Olympiad (they are asked to solve all of the problems from the previous 25 Olympiads); (3) the three-day training camp held at California State Polytechnic University–Pomona just before leaving the US; and (4) the intensity of the team members in their group study during the trip.



Dwight Neuenschwander, Jooh Pahk, Paul Lujan, Rhiju Das, Daniel Phillips, Ben Rahn, Larry Kirkpatrick, and Bernard Khoury enjoy a sightseeing tour of Sydney.

Australia

It is a 14-hour flight to Sydney from Los Angeles, one of the longest nonstop routes in the world. The team spent three days in Sydney adjusting to the time change and thoroughly enjoying an introduction to the "land down under." Upon arriving at 6 am, the team cleared customs, collected luggage, and traveled to the hotel to leave off the luggage. A late breakfast and a two-hour walk around Darling Harbor got us back to the hotel in time to claim our rooms and take hot showers. Then it was off for a harbor cruise to the zoo, so that we didn't fall asleep and could use the sunshine to help reset our biological clocks. Shortly after dinner, it was off to bed and a very long night's sleep, solving most of the jet-lag problem. The next two days were occupied with some sightseeing mixed in with study sessions.

Then it was off to Canberra for the Olympiad. The Olympiad organizers consisted of Prof. Rod Jory (normally director of the Australian Physics Team) and approximately 90 of his students and former team members. Even the coaches of the Australian Team were former team members. The organization was very professional and the hospitality well beyond anything that could have been expected. The Australians made all of us feel very welcome and special. And it was very interesting to encounter the new flora and fauna. It was a strange experience having your noon-day shadow point south, seeing many new constellations in the night sky, and having winter in July.

The problems are the heart of the Olympiad competition and were very well prepared by teams of physicists from the major cities of Australia. Several of these problems will provide inspiration for Physics Contest problems in *Quantum* during this next year. (The first appears in this issue.)

The 1995 Team

The other members of the US Physics Team (with their physics teachers and high schools) are: Matthew Ahart, North Hollywood, California (John Feulner, Harvard–Westlake School), James Belk, Endicott, New York (Mitchell Johnson, Union-Endicott High School), Franz Boas, La Jolla, California (Martin Teachworth, La Jolla High School), Michael Emanuel, New Rochelle, New York (Anthony Soldano, New Rochelle High School), Chris Holleman, Durham, North Carolina (Hugh Haskill, North Carolina School of Science and Mathematics), Yoon-Ho Lee, Wallingford Connecticut (Lawrence Stowe, Choate Rosemary Hall), Chen Ling, Cleveland Heights, Ohio (Robert Quail, Cleveland Heights High School), Edward Miller, New Orleans, Louisiana (Tony Asdourian, Isidore Newman School), Vivek Mohta, Northville, Michigan (Robert Sharrar, Northville High School), Chris Norris, Andover, Massachusetts (J. Peter Watt, Phillips Academy), Mark Oyama, Honolulu, Hawaii (Carey Inouye, Iolani School), Brian Patt, Birmingham, Michigan (James Bedor, Seaholm High School), Casey Rothschild, Northfield, Massachusetts (Boris Korsunsky, Northfield Mt. Herman High School), Ari Turner, Los Alamos, New Mexico (Julia Wangler, Los Alamos High School), and Daniel Wesley, Rosemont, Pennsylvania (Robert Schwartz, Harriton High School).

Rhiju Das, who had the second-highest score in the competition, carries the US flag during the closing ceremonies.

The US Physics Team is coached by Larry Kirkpatrick—academic director (Montana State University), Dwight Neuenschwander—senior coach (Southern Nazarene University, now with the American Institute of Physics), Ted Vittitoe—senior coach (retired from Libertyville High School in Illinois), Hugh Haskill—coach (North Carolina School of Science and Mathematics), and Mary Mogge—coach (California State Polytechnic University–Pomona). The US Physics Team is organized by the American Association of Physics Teachers under the direction of Bernard Khoury and with the invaluable assistance of Maria Elena Khoury and her staff. Financial support is organized by the American Institute of Physics with help from its member societies.

The next International Physics Olympiad will be held in Oslo, Norway, from June 30 to July 5, 1996. Teachers who have not received application materials by mid-December should contact Maria Elena Khoury at AAPT (301 209-3344).

Larry D. Kirkpatrick *is a professor of physics at Montana State University and a field editor for* Quantum. *He has completed his eighth and final year as academic director of the US Physics Team.*

Bulletin Board

The Society of Physics Students . . .

. . . wants *you*! The Society of Physics Students (SPS) is the professional society that engages all college and university students who are interested in physics. Membership is through the local chapter. Chapters are found in nearly 700 colleges and universities. Although some high school students join SPS through a nearby chapter, and many graduate students are SPS members, the majority of the members are undergraduates.

The purpose of the SPS is to help transform students from course-takers into contributing members of the community of professionals. Through undergraduate research, presenting papers at regional meetings, "taking physics on the road" in science outreach programs in local schools (K–12), chapter colloquiums with guest speakers, tours, and social events, the SPS chapter sharpens the student's professional development, communication skills, and leadership qualities.

The only requirement for membership is interest in physics. Students majoring in physics, engineering, chemistry, mathematics, computer science, pre-med, geology, and other fields will be found within the SPS. The annual dues are $13.

Each member receives the *SPS Newsletter*, the magazine *Physics Today*, and *The Journal of Undergraduate Research in Physics*. In addition, by being an SPS member, the student has the option of membership at reduced rates in many other more specialized physics professional societies and reduced subscription rates to various magazines and journals.

The physics honor society, Sigma Pi Sigma, is housed within the SPS. Students who have demonstrated high standards of scholarship in physics are eligible for election into the local Sigma Pi Sigma chapter. The SPS and Sigma Pi Sigma governing board is the SPS Council, which consists of 18 elected student members and 18 elected faculty members. Sigma Pi Sigma, founded in 1921, is a member of the Association of College Honor Societies.

For more information about the Society of Physics Students, contact the SPS National Office, One Physics Ellipse, College Park MD 20740. The phone number is 301 209-3007, and the e-mail address is sps@aip.org.

—Dwight E. Neuenschwander,
SPS Director

International Olympiad in Informatics

Four high school students representing the United States won medals in the 1995 International Olympiad in Informatics, held in Eindhoven, the Netherlands, from June 26 to July 3. The US team competed against 210 students from 51 countries.

Russell Cox, 16 (Delbarton High School, New Providence, New Jersey), ranked 11th in the contest and won a gold medal. Hubert Chen, 18 (Upper Dublin High School, Fort Washington, Pennsylvania), ranked 23rd and received a silver medal. Daniel Adkins, 15 (McKinley High School, Baton Rouge, Louisiana), ranked 68th, and Valentin Spitkovsky, 17 (Lafayette High School, Williamsburg, Virginia), ranked 100th, which earned them bronze medals. Erika Hoffeld, 17 (Montgomery Blair High School, Silver Spring, Maryland), also competed for the US team.

In the medal rankings, the USA team ranked 6th behind Czechoslovakia, China, Russia, Hungary, and Romania.

Don Piele, professor of mathematics at the University of Wisconsin–Parkside and leader of the United States team, reported that the Eastern European countries are extremely strong in algorithmic computer problem solving because of the many competitions held in this region. "They are consistently in the top ranked teams at IOI," Piele noted. "To be ranked in the top six with these teams is a great achievement for our team. We continue to improve each year as we develop our training program at UW–Parkside."

During the IOI competition, students compete on two separate days where they are given five hours to use logic, mathematics, and computer programming skills to create original computer programs that solve three difficult problems.

Since 1992, when the US first entered the IOI, participating US teams have won four gold medals, four silver medals, and six bronze medals and have always ranked in the top 10 countries.

Rob Kolstad (deputy team leader), president of Berkeley Software Design, Inc., Colorado Springs, Colorado, accompanied the team to the international event.

Financial backing for the United States of America Computing Olympiad (USACO) is provided by USENIX, a UNIX user group headquartered in Berkeley, California, and an anonymous donor.

The USACO home page (http://usaco.uwp.edu) on the World Wide Web contains information and pictures of all USACO competitions including this year's IOI. For more information on the competition, contact Don Piele at (414) 595-2231.

CyberTeaser winners

The following visitors to *Quantum*'s home page on the World Wide Web were the first ten persons to electronically submit a correct answer to the CyberTeaser they found there (brainteaser B157 in this issue):

Leonid Borovsky (Brooklyn, New York)
Chris Rybak (Racine, Wisconsin)
Daniel E. Sealey (Freeland, Michigan)
Steve VanDeBogart (Reno, Nevada)
Scott Bilker (Barnegat, New Jersey)
Brian Platt (Woods Cross, Utah)
Douglas E. Norton (Havertown, Pennsylvania)
Galina Yakovenko (University Park, Maryland)
John-David Rusk (Dahlonega, Georgia)
Richard Krueger (Sherwood Park, Alberta)

Galina was our youngest winner (she's eight years old), and no one knows who the oldest was. Each winner will receive a *Quantum* button and a copy of this issue of the magazine. Congratulations!

We received correct answers from many others, and we thank them for taking part in our contest. The new CyberTeaser is posted soon after each issue of *Quantum* goes to press. Maybe it's there already—give it a try! Go to http://www.nsta.org/quantum and follow the links.

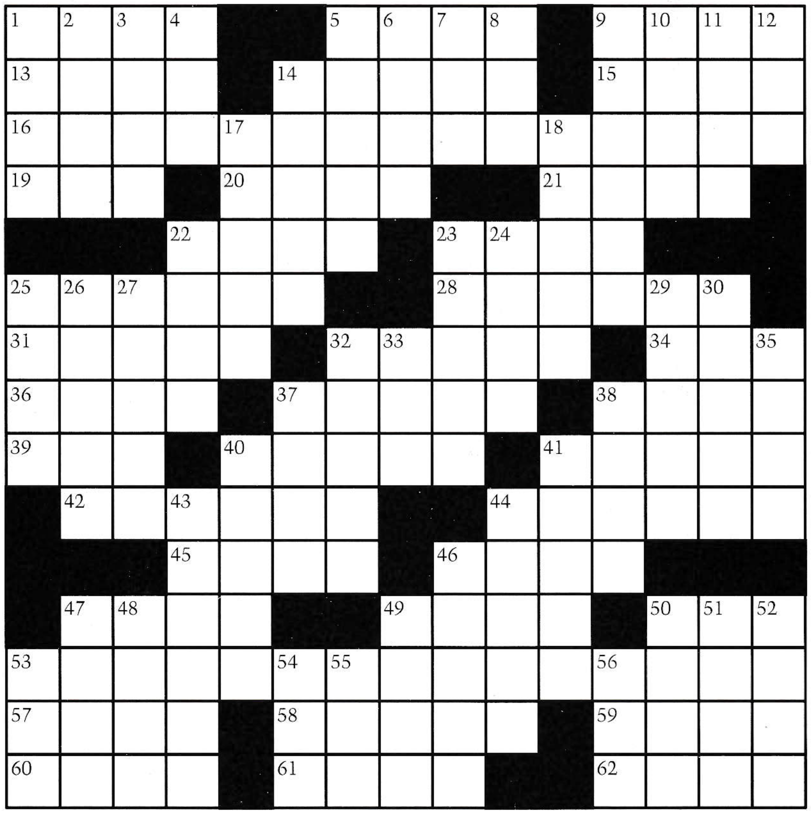
What's happening?

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.

fax:	703 522-6091
e-mail:	quantum@nsta.org
snail mail:	*Quantum* 1840 Wilson Blvd. Arlington VA 22201-3000

Criss × cross science

by David R. Martin



Across

1 Pigment
5 Astronomer George ___ (1868–1938)
9 God of war
13 Swedish botanist ___ Afzelius (1750–1837)
14 Threadlike structure
15 ___ number (ratio of speeds)
16 Study of certain solids
19 Sense organ
20 Type of current
21 10^{-18}: pref.
22 Sailing
23 Bustle
25 ___ law of refraction
28 Fabry–Perot ___
31 Leg bone
32 Living room item
34 Wager
36 ___ process (of making German butter)
37 Vaporizes
38 Berry with a hard rind
39 Retina cell
40 Liquid foods
41 Monochromatic radiation source
42 Polytetrafluoroethene
44 Units of volume
45 For fear that
46 Microbiologist ___ Dubos (1901–1982)
47 Indian
49 44,714 (in base 16)
50 Intrinsic FORTRAN function
53 Afterglow
57 Hyperbolic function
58 Engines' air controller
59 Ten: pref.
60 Engaged
61 Medulla oblongata to midbrain connector
62 Surgeon ___ Wangensteen (1898–1981)

Down

1 Frilly threads
2 ___ martini (2 wds.)
3 Comedian Danny ___
4 Printer's measures
5 Poet ___ Doolittle (1886–1961)
6 Competently
7 German explorer ___ Frobenius (1873–1938)
8 Unit of energy
9 Ammonium nitrate and trinitrotoluene
10 Engrossed
11 Reverberation
12 Bashful
14 Wanes
17 Unit of magnetic flux
18 Ship detector
22 Inter ___ (amongst other things: L.)
23 Fresh water ducks
24 Elevator man
25 Large plasma ball
26 Egyptian
27 966,622 (in base 16)
29 Fat
30 8.686 decibels
32 Enumerate
33 Waist
35 Rocky hills
37 Sounds from an audience
38 Top of the head
40 Unconscious state
41 Linear accelerator
43 Corporeal
44 Rental contract
46 Stinks
47 Chinese dynasty
48 Biologist ___ Harrison (1870–1959)
49 French sociologist Raymond ___ (1905–1983)
50 Afresh
51 48,334 (in base 16)
52 Connery or Penn
53 Plastic ingredient: abbr.
54 Type of sphere packing: abbr.
55 Exclamation of triumph
56 Former name of Tokyo

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE SEPTEMBER/OCTOBER PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M156

On the torus the centers of eight squares attacked by a king always form a square measuring 2×2 with sides parallel to those of the chessboard (see figure 1). Two kings are not attacking each other if and only if their surrounding 2×2 squares do not overlap by more than a line segment. So our problem reduces to the following question: *what is the greatest number N of 2×2 squares described above that can be placed on the $n \times n$ toroidal chessboard so as not to overlap with one another?*

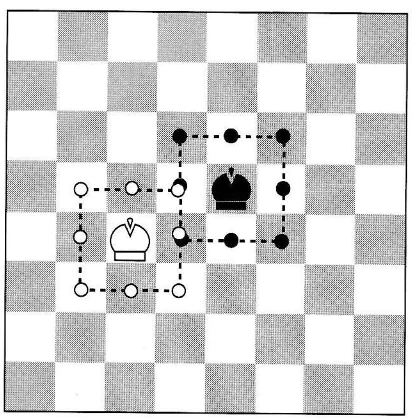


Figure 1

For convenience, let's shift all the 2×2 squares up and to the right by a half diagonal of a unit chessboard square (fig. 2) so each of them will

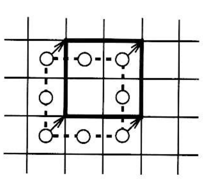


Figure 2

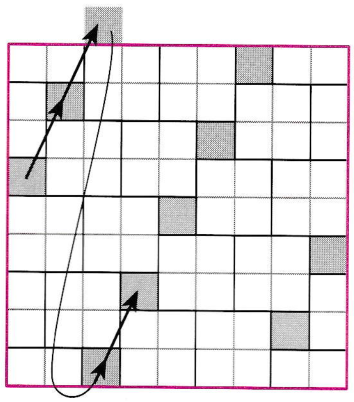


Figure 3

cover exactly four chessboard squares. Now it's obvious that in the case of an even $n = 2s$ the chessboard can simply be covered by s^2 nonoverlapping 2×2 squares, so in this case $N = s^2 = n^2/4$.

If n is odd, any arrangement of 2×2 squares will leave at least one unit square in each horizontal file uncovered, so the number of the covered squares is no greater than $n(n-1)$, and $N \le N_0 = [(n^2 - n)/4]$ (the square brackets denote the integer part of a number). We'll prove that $N = N_0$; that is, we can always place N_0 non-overlapping 2×2 squares on our chessboard.

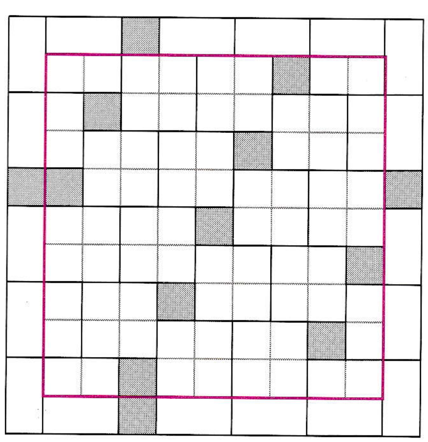


Figure 4

Again consider two cases.

Suppose $n = 4k + 1$. Then it follows that $N_0 = kn$. Let's leave uncovered the n squares obtained from one another by the chess knight's move in a fixed direction (fig. 3—don't forget that we're on a torus!). It's not difficult to see that the remaining area can be split into 2×2 squares. The required arrangement of kings can be obtained by placing them, say, at the left bottom corners of these squares.

Finally, if $n = 4k + 3$, we can show that $N_0 = (k+1)(n-1)$. This many 2×2 squares can be arranged as follows: separate a one-square-thick frame around the chessboard, fill its interior square of size $4k + 1$ as above, and then fill the frame leaving four of its squares uncovered as shown in figure 4.

M157

It will suffice to prove a somewhat stronger assertion: each term of the sequence $\{\sqrt{b_n}\}$ is greater than the previous term by no less than 1:

$$\sqrt{b_{n+1}} \ge \sqrt{b_n} + 1. \qquad (1)$$

Using the notation

$$a_1 + \cdots + a_n = a,$$

$$\frac{1}{a_1} + \cdots + \frac{1}{a_n} = c,$$

$$a_{n+1} = x > 0,$$

we have to prove that for any $x > 0$,

$$\sqrt{(a+x)\left(c+\frac{1}{x}\right)} \ge \sqrt{ac} + 1.$$

Using the fact that the geometric mean of two numbers never exceeds their arithmetic mean, we get $a/x + cx \geq 2\sqrt{ac}$, which leads to

$$(a+x)\left(c+\frac{1}{x}\right) \geq ac+1+2\sqrt{ac}$$
$$=\left(\sqrt{ac}+1\right)^2.$$

Now inequality (1) follows by taking the square root of both sides.

M158

(a) This statement is true for any tangent l to circle S_1 (not necessarily touching the second circle—see figure 5). One of the simplest ways to

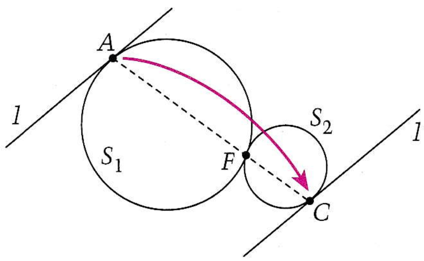


Figure 5

prove it is to use the fact that two tangent circles are homothetic to each other relative to their point of contact. Here, the dilation relative to F that takes circle S_1 to circle S_2 takes the tangent l to S_1 to the parallel tangent l' to S_2; so it takes the point of contact A on the first circle into the point of contact C on the second circle. But this means that A and C lie on the same line through F.

(b) Notice first that $\angle ABC = \angle BFC = 90°$ (see figure 6), because B and C are diametrically opposite on the circle S_2. So the circumcenter O of $\triangle ABC$ is the midpoint

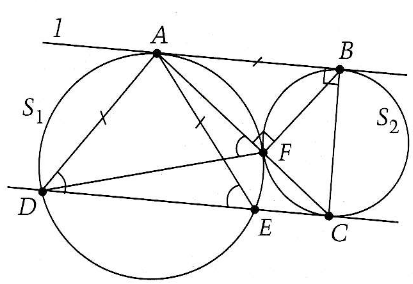


Figure 6

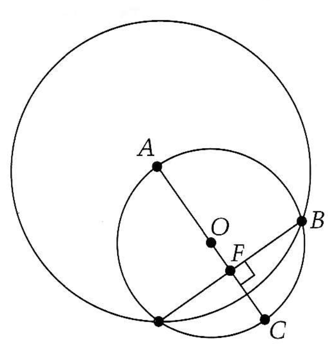


Figure 7

of AC ($\angle ABC = 90°$) and, as we'll show below, the circumcenter of $\triangle BDE$ is point A. It follows that BF is the perpendicular dropped from a common point B of the two circumcircles on the line AC through their centers. But this just means that F lies on the common chord (fig. 7).

So it remains to prove that A is the circumcenter of $\triangle BDE$, or that $AE = AD = AB$. The first of these equalities is easy: ED is a chord of S_1 parallel to the tangent at A, so $\overset{\frown}{AD} = \overset{\frown}{AE}$. As for the second, notice that the right triangles ABF and ACB are similar (they have a common acute angle at A). So $AF/AB = AB/AC$, or

$$AB^2 = AF \cdot AC$$

(actually, this is a particular case of the well-known theorem on a tangent and secant drawn to a circle from the same point). We now show that $\triangle AFD$ and $\triangle ADC$ are also similar. Indeed, they have a common angle at A, and $\angle AFD$ equals $\angle AED$ (they both intercept $\overset{\frown}{AD}$), which in turn equals $\angle ADE$ (as noted above). From this similarity we get $AF/AD = AD/AC$, or

$$AD^2 = AF \cdot AC = AB^2,$$

completing the solution. (V. Dubrovsky)

M159

Let's number the vertices of the n-gon 1, 2, ..., n in, say, the clockwise direction. Denote by a_i the number of moves made from the ith vertex. Then the number of chips that leave this vertex is $2a_i$ and the number of chips that arrive there is $a_{i-1} + a_{i+1}$. So we have

$$a_1 = \frac{a_n + a_2}{2},$$
$$a_2 = \frac{a_1 + a_3}{2},$$
$$\vdots$$
$$a_n = \frac{a_{n-1} + a_1}{2}.$$

Assume for definiteness that a_1 is the largest of the numbers a_i. Then the equality $a_1 = (a_n + a_2)/2$ is possible only if $a_n = a_2 = a_1$. Now the equality $a_2 = (a_1 + a_3)/2$ gives $a_3 = a_1$, and so on. Thus, $a_1 = a_2 = \ldots = a_n$, so the total number of moves is na_1.

M160

Fix certain positions P and Q of the first and second figures on the grid and the center O of any grid square. Let $P_1, \ldots, P_n$ and $Q_1, \ldots, Q_m$ be the centers of the squares constituting figures P and Q, respectively. Consider the sum of mn numbers written at the endpoints of the vectors $\overrightarrow{OP_i} + \overrightarrow{OQ_j}$ drawn from O ($i = 1, \ldots, n$; $j = 1, \ldots, m$)—some numbers in the sum may be repeated. This sum is positive, because it can be obtained by adding the sums in the m copies of figure P (perhaps overlapping—see figure 8) obtained by the shifts through the vectors $\overrightarrow{OQ_1}, \ldots, \overrightarrow{OQ_m}$. But it can also be obtained by adding the sums in the n shifts of Q by $\overrightarrow{OP_1}, \ldots, \overrightarrow{OP_n}$. Therefore, at least one of the sums in the copies of Q is positive. (N. Vasilyev)

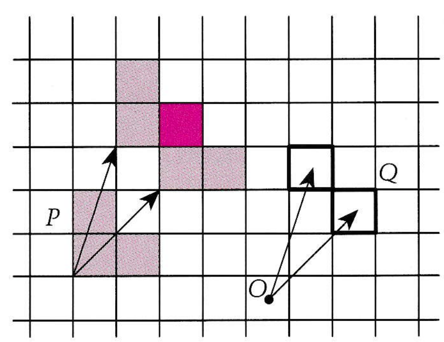


Figure 8

Physics

P156

If the thread is pulled as shown in figure 9, the bobbin rolls to the right and rotates clockwise about its axis.

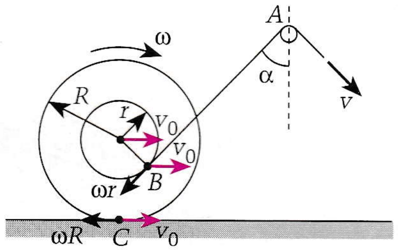


Figure 9

For point B the sum of the projections of the velocity $\mathbf{v}_0$ of the translational motion and the linear velocity due to the rotation with angular velocity ω along the direction of the thread is

$$v_0 \sin\alpha - \omega r = v,$$

where v_0 is the magnitude of the velocity we seek. As there is no slipping of the bobbin, the sum of the corresponding velocities for point C is zero:

$$v_0 - \omega R = 0.$$

The above equations yield

$$v_0 = v\frac{R}{R\sin\alpha - r}.$$

Clearly, when $R\sin\alpha = r$ (which corresponds to the case when points A, B, and C are on the same line), the formula obtained for v_0 is no longer valid. It should also be noted that this formula describes the bobbin's motion both to the right (if point B is located to the left of AC and $R\sin\alpha > r$), and to the left (when point B lies to the right of AC and $R\sin\alpha < r$).

P157

Neglecting the mass of the soap film means that a soap bubble will start to rise when the density of the nitrogen inside it equals that of the surrounding air:

$$\rho_N = \rho_a.$$

The equations of state for nitrogen and air yield

$$\rho_a = \frac{P_0\mu_a}{RT}$$

and

$$\rho_N = \frac{(P_0 + 8\sigma\, d)\mu_N}{RT}.$$

Here P_0 is the atmospheric pressure, μ_N and μ_a are the molar masses of nitrogen and air, R is the gas constant, T is room temperature, and d is the diameter of the bubble. Equating the densities yields

$$d = \frac{8\mu_N\sigma}{P_0(\mu_a - \mu_N)}.$$

For $P_0 = 10^5$ Pa, $\mu_N = 28$ g/mol and $\mu_a = 29$ g/mol, the diameter we seek is

$$d \cong 9 \cdot 10^{-5}\ \text{m}.$$

P158

The air molecules striking the plate have a root-mean-square speed

$$v_{rms} = \sqrt{\frac{3kT}{m}},$$

where T is the air temperature and m is the mass of an air molecule. The average value of the projection of the molecular velocities onto the OY-axis set perpendicular to the plate is

$$\bar{v}_y \cong \frac{v_{rms}}{\sqrt{3}} = \sqrt{\frac{kT}{m}}.$$

After the molecules collide with the plate, their "temperature" becomes equal to that of the plate. So, at the bottom surface of the plate the velocity projection after collision is

$$\bar{v}_{yb} = \sqrt{\frac{kT_b}{m}},$$

while that at the top surface is

$$\bar{v}_{yt} = \sqrt{\frac{kT_t}{m}}.$$

As a result, the projection p_y of the momentum of a molecule colliding with the bottom surface changes by

$$\Delta p_{yb} = m\left(\sqrt{\frac{kT}{m}} + \sqrt{\frac{kT_b}{m}}\right)$$

and that on the top surface by

$$\Delta p_{yt} = m\left(\sqrt{\frac{kT}{m}} + \sqrt{\frac{kT_t}{m}}\right).$$

In a time interval Δt each surface is struck by molecules located at a distance less than $\bar{v}_y\Delta t$. The number of such molecules is $N = n\bar{v}_y s\Delta t$, where n is the number of molecules in a unit volume and s is the area of the plate.

In accordance with Newton's second and third laws, the plate is affected vertically by forces having magnitudes equal to the changes in the momentum projections of the molecules per unit time. So, the bottom surface experiences an upward force (see figure 10)

$$F_b = \frac{N\Delta p_{yb}}{\Delta t} = mn\bar{v}_y s\left(\sqrt{\frac{kT_b}{m}} + \sqrt{\frac{kT}{m}}\right),$$

and the top surface experiences a downward force

$$F_t = \frac{N\Delta p_{yt}}{\Delta t} = mn\bar{v}_y s\left(\sqrt{\frac{kT_t}{m}} + \sqrt{\frac{kT}{m}}\right).$$

Since $T_b > T_t$, the resulting force R is directed upward and its magnitude is

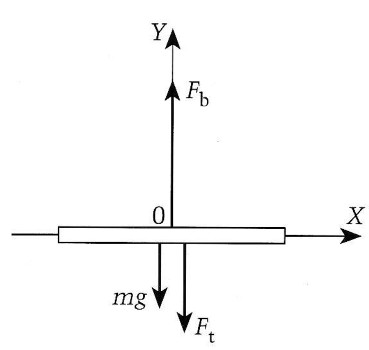


Figure 10

$$R = F_b - F_t$$
$$= mn\bar{v}_y s\left(\sqrt{\frac{kT_b}{m}} - \sqrt{\frac{kT_t}{m}}\right)$$
$$= nsk\sqrt{T}\left(\sqrt{T_b} - \sqrt{T_t}\right).$$

The number of air molecules in a volume V at pressure P and temperature T is $N_V = \nu N_A$, where ν is the number of moles of air in a volume V and N_A is Avogadro's number. On the other hand, $\nu = M/\mu = PV/RT$, so the number of molecules in a unit volume is

$$n = \frac{P}{RT}N_A.$$

Thus,

$$R = \frac{PN_A sk}{RT}\sqrt{T}\left(\sqrt{T_b} - \sqrt{T_t}\right)$$
$$= \frac{Ps}{\sqrt{T}}\left(\sqrt{T_b} - \sqrt{T_t}\right).$$

Inserting the numerical data into this formula (and assuming $P = 10^5$ N/m^2) yields

$$R \cong 1.5 \cdot 10^4 \text{ N}.$$

P159

As the current decreases, the frame is infused with a varying magnetic flux $\Phi(t)$ that is proportional to the electric current $I(t)$ in the wire:

$$\Phi(t) \sim I(t).$$

The resulting emf produces a current in the frame:

$$i(t) = -\frac{1}{R}\frac{\Delta\Phi(t)}{\Delta t},$$

where R is the resistance of the frame. Since $R \sim 1/d^2$,

$$i(t) \sim \frac{\Delta I(t)}{\Delta t}d^2.$$

The left side of the frame is affected by the magnetic field of the straight wire with an Amperian force (see figure 11)

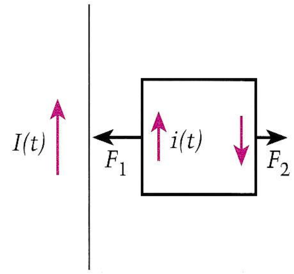


Figure 11

$$F_1 \sim I(t)i(t) \sim I(t)\frac{\Delta I(t)}{\Delta t}d^2$$
$$\sim \frac{\Delta\left(I^2(t)\right)}{\Delta t}d^2.$$

Clearly, a similar force F_2 acts on the right side of the frame. The resulting force is

$$F = F_1 - F_2 \sim \frac{\Delta\left(I^2(t)\right)}{\Delta t}d^2$$

and is directed to the left.

In an infinitely small time interval Δt the frame is acted on by an impulse

$$\Delta p = F\Delta t \sim \Delta(I^2(t))d^2.$$

The total impulse imparted to the frame during the change of electric current in the wire from the initial value I down to zero is

$$p \sim I^2 d^2.$$

In the first case this impulse is equal to p_0 and is directed to the left. In the second case the frame receives the impulse

$$p_1 = \left(\frac{I_1 d_1}{I_0 d_0}\right)^2 p_0 = 36p_0.$$

P160

Each part of the compound lens forms its own image of the source S. These images (S_1 and S_2 in figure 12) are located at a distance H from the system's axis. The rays passing through both parts of the lens and forming the images overlap, and in the overlapping region one can see the interference. We must find the boundaries of the overlapping region—that is, the distance x in figure 12.

First let's find the positions of the images S_1 and S_2—that is, the distance f from the lens and the distance H from the axis. Using the lens formula and taking into account that the focal length of both parts of the compound lens is F, we get

$$f = \frac{Fl}{l - F} = 3F.$$

The optical axes of the new lenses are at a distance $h/2$ from the system's axis and are parallel to this axis. This means that the source S is at a distance $h/2$ from the optical axis of any lens, and its image is formed at distance $H - h/2$. The formula for the linear magnification of the lens is

$$\frac{H - h/2}{h/2} = \frac{f}{l},$$

from which we get

$$H = \frac{h(l + f)}{2l} = \frac{h(l + 3F)}{2l} = 1.5h.$$

Now it's easy to find x. The similarity

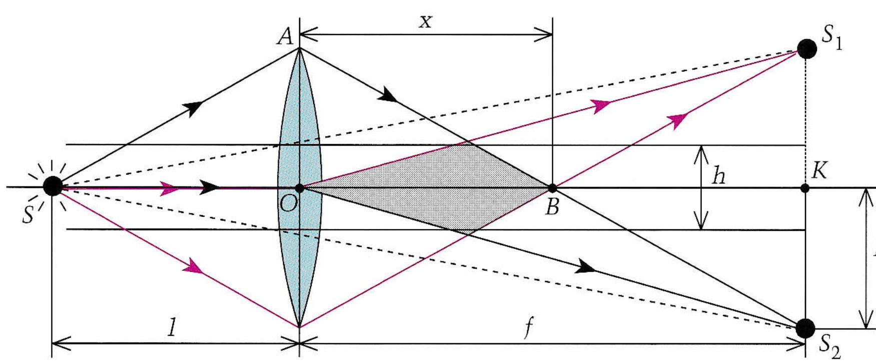


Figure 12

of triangles AOB and BS_2K result in

$$\frac{AO}{S_2K} = \frac{OB}{BK}$$

and

$$\frac{(d-h)/2}{H} = \frac{x}{f-x},$$

from which we get

$$x = \frac{f(d-h)}{d-h+2H} = \frac{3F(d-h)}{d+2h}$$
$$= 112 \text{ cm}.$$

Brainteasers

B156

From the second equation, $T = 2WO$. Substituting into the first we get $2WO - W - O = 2$, or

$$W(O-1) + O(W-1) = 2.$$

Now, neither O nor W can be zero (since the problem statement asks us to divide by them) or negative (they're digits), so the two addends on the left of the displayed equation are nonnegative. If either is equal to 1, then $O = 1$ or $W = 1$. If $O = 1$, then $W = 3$ and $T = 6$. If $W = 1$, then $O = 3$ and $T = 6$. It's not hard to see that the two addends on the left of the displayed equation cannot both be 1 (and if either were greater their sum couldn't be 1). Thus, the only two solutions are (6, 3, 1) and (6, 1, 3).

B157

If we add 7 and then 3 dollars to 3/5 of Johnny's funds, we get his entire holdings. So 2/5 of his total equals 10 dollars. It follows that Johnny has 25 dollars and Annie has 22 dollars.

B158

At great depths the tube with both ends sealed will be flattened by the tremendous pressure of the water. The open tube, however, will not be deformed.

B159

In figure 13, the area of the quadrilateral $AECF$ is equal to $\frac{1}{2}AC \cdot EF \cdot \sin\alpha$ and the area of $EBDC$ is $\frac{1}{2}BC \cdot ED \cdot \sin\beta$. Now we notice that $AC = BC$, $EF = ED$ and $\sin\alpha = \sin\beta$, because $\alpha + \beta = 360° - (\angle ACB + \angle FED) = 180°$.

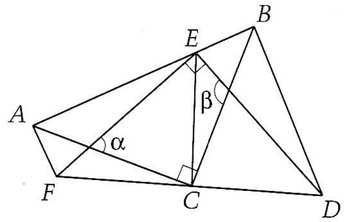


Figure 13

B160

The second player can win by applying the following "symmetric strategy": the grid is divided in half (say, horizontally) and each move of the first player is exactly repeated in the other half. In other words, the second player should always color the square with the same number (in the numbering shown in figure 14) as that of the last square colored by the first player. This strategy ensures a win for the second player, because if this player completes a colored 2 × 2 square with numbers a, b, c, d, then the other four unit squares with these numbers would have to be already colored. So seven of these eight unit squares would have been colored after the previous move by the first player, and they necessarily contain a 2 × 2 square.

Another winning strategy is based on the following approach. Imagine that the entire grid is divided into four 2 × 2 squares. If each of these squares has a colored unit square not on a diagonal of the grid (which the second player can easily ensure in the first few moves), and the total number of colored squares is less then 12, then one more unit square can be colored in accordance with the rules of the game. (Check this!) So the game will continue until the second player colors the twelfth square. Then the first player will be unable to make the next move, because it will necessarily complete one of the four corner 2 × 2 squares. (S. Tokarev, V. Dubrovsky)

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

Figure 14

Kings on a torus

(Answers supplied by the editors)

1. $\alpha(G_1) = 1$, GIS = {1}, {2}, or {3}; $\alpha(G_2) = 3$, GIS = {1, 3, 4}; $\alpha(G_3) = 2$, GIS = {i, j}, where $i \in$ {1, 2}, j = {5, 6}; $\alpha(G_4) = 4$, GIS = {3, 6, 7, 8} (or four other similar sets).

2. $[n/(k+1)]$.

4. These statements follow from the solution to math challenge M156 in this issue.

5. By problem 4b each file of the board must contain exactly one king. Any GIS can be cyclically shifted so as to bring one of the kings to the central square (2, 2) of the board (see figure 5 in the article). This will leave only 12 possible squares for the remaining four kings: the three squares (0, 0), (0, 1), and (1, 0), and three similar triples at the other corners of the board. Obviously there must be exactly one king in each of these triples of squares, and it's easy to see that the square (0, 0) must be free (because the king in (0, 0) attacks or is in the same file with all the squares in the top left and bottom right triples). We can assume that there is a king in (0, 1); otherwise apply reflection in the (0, 0)–(4, 4) diagonal. Then the last three kings are positioned uniquely: (1, 4), (3, 0), and (4, 3). It remains to shift this arrangement one square down.

6. Any different sets $(v_1, \ldots, v_k)$ with all coordinates from a fixed GIS for the given graph G are nonadjacent in G^k. Since each v_i takes $\alpha(G)$ values, the total number of such sets is $(\alpha(G))^k$.

7. n^k.

8. $3^k - 1$.

9. The graph P_n^k can be viewed as

a k-dimensional grid cube consisting of n^k unit k-dimensional cubes: the vertices of these cubes are the nodes of the graph and two nodes are adjacent if and only if they are vertices of the same unit cube. In the case $k = 2$ this interpretation was discussed in the article; the case $k = 3$ is also easy to imagine. High-dimensional cases are more difficult to visualize, but you can take a more formal approach: the 2^k vertices of any one unit cube can be written as $(v_1 + h_1, v_2 + h_2, \ldots, v_k + h_k)$, where $(v_1, v_2, \ldots, v_k)$ is fixed and $h_i = 0$ or 1 for all $1 \le i \le k$). Each node of P_n^k is a common vertex of 2^k unit cubes surrounding it; two nodes are nonadjacent if and only if the sets of unit cubes surrounding them are disjoint. So an independent set cannot have more than $n^k/2^k$ nodes.

10. For $n = 2s$, $\alpha(P_n^k) = s^k$. In this case the k-dimensional "cube neighborhoods" of nodes considered in the previous solution (the sets of 2^k unit cubes surrounding a node) can be chosen so as to fill the big $n \times n \times \ldots \times n$ cube without gaps. Their centers form the largest independent set consisting of s^k nodes.

Toy Store

1. One twisted ring.
2. Two linked, twisted rings.
3. (a) Three interlaced rings; (b) a three-ring chain.
4. Two linked rings—one large, the other small, neither twisted.
6. A three-ring chain, with no rings twisted.
7. A five-ring chain. But if you vary the instructions, intentionally or otherwise, you'll get some other configurations of rings. Some of these are interesting and useful in further experiments.

Mean value

(Answers supplied by the editors)

1. These properties are not difficult to prove by means of the definition of M and some algebraic manipulation.

2. For any n, we can show that

$$\binom{2n}{0} + \binom{2n}{1} + \cdots + \binom{2n}{2n} = 2^{2n}$$

(for example, by writing out the expansions of $(x + y)^{2n}$ and setting $x = y = 1$). So the number $2^{2n-1}/n$ is the mean value of all the binomial coefficients $\binom{2n}{k}$. Comparing the ratio

$$\frac{\binom{2n}{k}}{\binom{2n}{k+1}} = \frac{k+1}{2n-k}$$

with 1, we find that $\binom{2n}{n}$ is the greatest of these coefficients, because this ratio is greater than 1 for $0 \le k \le n$ and greater than 1 for $n \le k \le 2n$. It remains to use property 3.

3. Let AM be a median in a triangle ABC. Using the construction in figure 16, we get $2AM = AD < AB + BD = AB + AC$.

4. $M(g) = M(f) + M(g - f) \ge M(f)$, because $g - f \ge 0$.

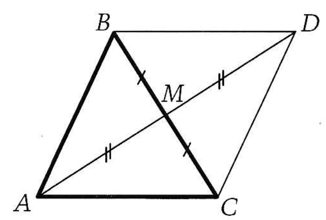


Figure 16

5. If any of the numbers a, b, c, d are 0, the proof is easy (in fact, the inequality of the statement becomes an equality). So we can assume that none of them is 0. Then at least two of the numbers a, b, c—say, a and b—are of the same sign. Without loss of generality, we can assume this sign is positive (otherwise, we can reverse the signs of all the numbers). Substituting $d = -(a + b + c)$, we rewrite the inequality in question as $|c| + |a + b + c| \ge |b + c| + |a + c|$, or

$$|a + (b + c)| - |b + c| \ge |a + c| - |c|.$$

Consider the function $f(x) = |a + x| - |x|$. Plotting its graph (see figure 17), we see that $f(x)$ is nondecreasing. Since $b + c \ge c$, $f(b + c) \ge f(c)$. But this is exactly our reworked inequality.

6. If we construct $\bar{f}$ and $\bar{g}$ as in the paragraph just preceding the statement of this problem, we find that the graph of $\bar{g}$ is a shifted graph of the 2π-periodic function . So we can again use the argument given in

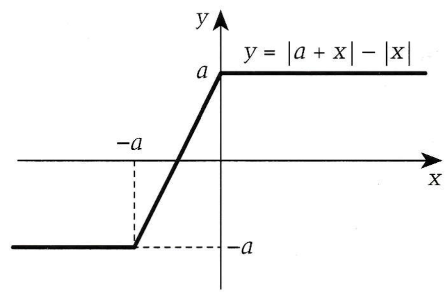


Figure 17

Index of Advertisers

the article for the case of the sinusoid.

7. Take an arbitrary unit vector **u** from the unit sphere, and let $k = M(|u|)$. Then for any **p** we have $\mathbf{p} = |\mathbf{p}| \cdot M(M(|v|))$, where **v** is the unit vector with the same direction as **p**. But $M(|v|) = M(|u|) = k$.

8. The projection of the given polygon on the line l_α is a segment I. Any perpendicular to I drawn through a point inside I meets the boundary of the polygon exactly twice (because the polygon is convex). So the segment I will be covered by the projections of the polygon's sides in exactly two layers at every point, which means that its length is the half-sum of the lengths of all sides' projections.

9. If segment I is a projection of the polygon (see the preceding solution), then the endpoints of I are the projections of certain vertices. So I is the projection of a side or diagonal and has a length of no greater than d. In other words, $W(\alpha) \le d$ for all α, $M(W) \le d$, and so the perimeter, equal to $\pi M(W)$, does not exceed πd.

10. The value of $\int_0^{2\pi} g(\phi - \alpha)d\alpha$

does not depend on ϕ (see solution 6), and for $\phi = \pi/2$ it takes the form $\int_0^{\pi} \sin\alpha\, d\alpha = 2$. The value of this integral is 2.

11. For any vector **a**, look at the function $\mathbf{a} \to \alpha(\mathbf{a})$, where $\alpha(\mathbf{a})$ is the pseudoprojection of **a** on α (as in the text). The mean value of this function is

$$\frac{1}{2\pi}\int_0^{2\pi} ag(\phi - \alpha)d\alpha = \frac{a}{\pi}.$$

Now, the function f is the sum of the functions $\alpha(\mathbf{a}_i)$ for our given vectors. The conclusion follows from property 1′ and the fact that the lengths of our given vectors add up to 1.

12. The sum in question is a continuous function of α, $0 \le \alpha \le 2\pi$. So it attains its maximum value S (this is proved in a calculus course, but is not hard to guess without a formal proof). But S is no less than the mean value of this function, which is equal to $1/\pi$ by exercise 11.

13. Choose the axis l_α from exercise 12 and all the vectors $\mathbf{a}_i$ that make acute angles with l_α. Their sum is not less than its projection on l_α, which is equal to the sum of their projections on l_α or to the sum of the pseudoprojections of *all* vectors $\mathbf{a}_i$ on α. By the choice of l_α, it is greater than $1/\pi$.

TOY STORE

Hands-on topology

The Möbius strip and linked rings

by Boris Kordemsky

THE SURFACE OF A RING such as we wear on our finger has two sides (fig. 1). One side touches the finger, the other faces out. These sides have two borders (that is, two edges), each of which is a circle. If a bug decides to travel from the outer side to the inner side, it will inevitably have to cross one of the borders.

It's easy to make a simple model with a completely different set of properties—a one-sided surface (fig. 2), as opposed to the ring-shaped, two-sided surface described above. The first to describe this surface was August Ferdinand Möbius (in 1863).[1] Give a half-turn twist to one end of a rectangular paper strip and glue it to the other end. You'll get a model of a surface that doesn't have two sides ("inner" and "outer"). And—just like that—you've created a Möbius strip.

To convince yourself that the Möbius strip has only one side, take a marker and draw a line on the strip without lifting the marker from the surface or crossing the edge. When you come back to the starting point, you'll see that the line travels along the entire surface of the strip, even though you never crossed the line that seems to separate the "two sides."

Now get a few sheets of flexible paper (newsprint will do), tape or glue, and scissors. It's time to do some practical exercises with the Möbius strip and other models made of rectangular paper strips.

Let's begin with two rather well-known tricks.

Experiment 1. What do you get when an ordinary (nontwisted) paper ring is cut along its midline? Clearly, two narrower rings, each with the same circumference as the original ring. But if you cut the Möbius strip along its entire midline, you'll obtain _____.

Do it and see what happens!

Experiment 2. Make another Möbius strip, rather wide this time, and cut it with scissors, keeping the slit 1/3 of the way between the two (apparent) edges. The thing you've created is _____.

You'll get a similar result if, before gluing the ends to form the ring, you twist one of its ends by 360° and then slit the model along its midline.

Now let's engage in some more advanced paper surgery.

Experiment 3. Slice the ends of the strip as shown in figure 3. (a) Tape or glue together *A* and *D*. Pass *B* under *A* and *E* over *D*, and tape *B* and *E* together. Then pass *C* under *B* and over *A*. Pass *F* over *E* and under *D* (note that *C* and *F* are treated somewhat differently here). Tape *C* and *F* together. Tape all the ends directly,

Figure 1

[1]A biographical note about Möbius, as well as a lot of interesting things about the topology of two-dimensional surfaces, can be found in the article "Flexible in the Face of Adversity" in the September/October 1990 issue of *Quantum*.—*Ed.*

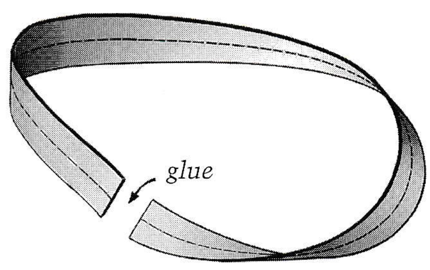


Figure 2

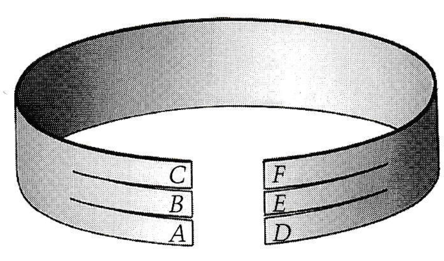


Figure 3

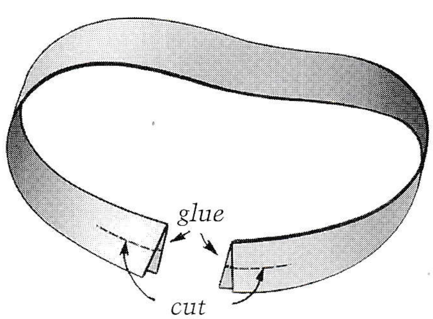


Figure 4

without twists. Now continue all the cuts you've begun all along the strip. You get _____. (b) Tape or glue together *A* and *D*. Pass *B* under *A* and *E* over *D*, and tape *B* and *E* together. Pass *F* over *E* and under *D*, but this time tape it directly to *C*, without doing anything with *C*. As before, tape all the ends directly, without twists. Then, after the cuts are made, you'll have _____.

Experiment 4. Prepare another strip sliced at the ends as before (fig. 3). Half-twist the end *E* (its top side away from you) and glue it to *C*. Half-twist the end *F* the same way and glue it to *B*. Pass *A* under *B* and glue it to *D* without twisting. Now continue all the cuts all the way along the model to obtain _____.

Experiment 5. (This idea comes from M. Brooke and J. Madachy.) Consider the following curious problem: make a paper-strip model from which a link of *n* rings (that is, an ordinary chain) can be obtained by a single unbroken closed cut.

The secret here is to prefold the strip lengthwise before cutting it and to slice and glue together the ends of the folded strip in a special way.

Begin with a strip creased once (fig. 4). Make one full (360°) twist of one end and tape the ends together crease to crease. Then slit this

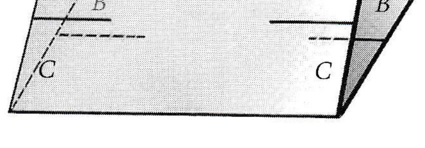


Figure 5

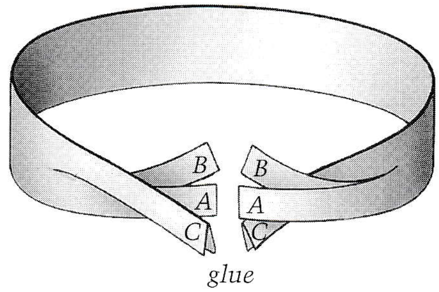


Figure 6

double layered band along its midline: the result is three twisted linked rings (each pair of rings is linked together).

Experiment 6. Crease the strip and cut it at the ends as shown in figure 5. (The strip is shortened for clarity, but should actually be as long as the one in figure 4.) Without twisting the ends, glue them together as shown in figure 6. Note that *B* remains behind *A*, and that *C* goes over both on the left and under both on the right. Continue the cut along the entire double-layered strip. If you do it exactly as I told you, you'll get _____.

Experiment 7. To obtain a chain of rings by a single closed cut, take a sufficiently wide strip and give it an accordion crease (fig. 7—again, the strip has been shortened). Slice the ends as shown in the figure, bend the folded strip into a ring, and glue together the pairs of ends labeled the same in the following order: ends *C*, *D*, and *E* directly (no twist); ends *B* after passing one of them under the rings *CC* and *DD*; the ends *A* after passing one of them under the rings *CC* and *EE*. Now complete the cut through all layers to obtain _____.

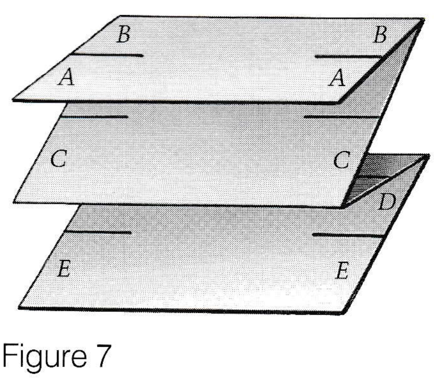


Figure 7

The answers to the questions and experiments above are given on page 62. But I don't want to supply hints to the following six experiments. You'll have to search for ways to tackle these problems on your own. I set only one condition: you must make do with a single slit around the model, constructed from a prefolded paper strip whose ends were properly sliced, interleaved, and glued together.

Experiments

8. Make a chain of four rings.

9. Make two separate chains of two and three rings.

10. Make a three-ring chain with the fourth ring hooked through the middle ring of the chain.

11. Make a nine-ring chain.

12. Make three separate chains of three rings each (out of one paper strip!).

13. Make two separate chains—one of four rings, the other of five rings (out of one strip).

ANSWERS, HINTS & SOLUTIONS ON PAGE 62