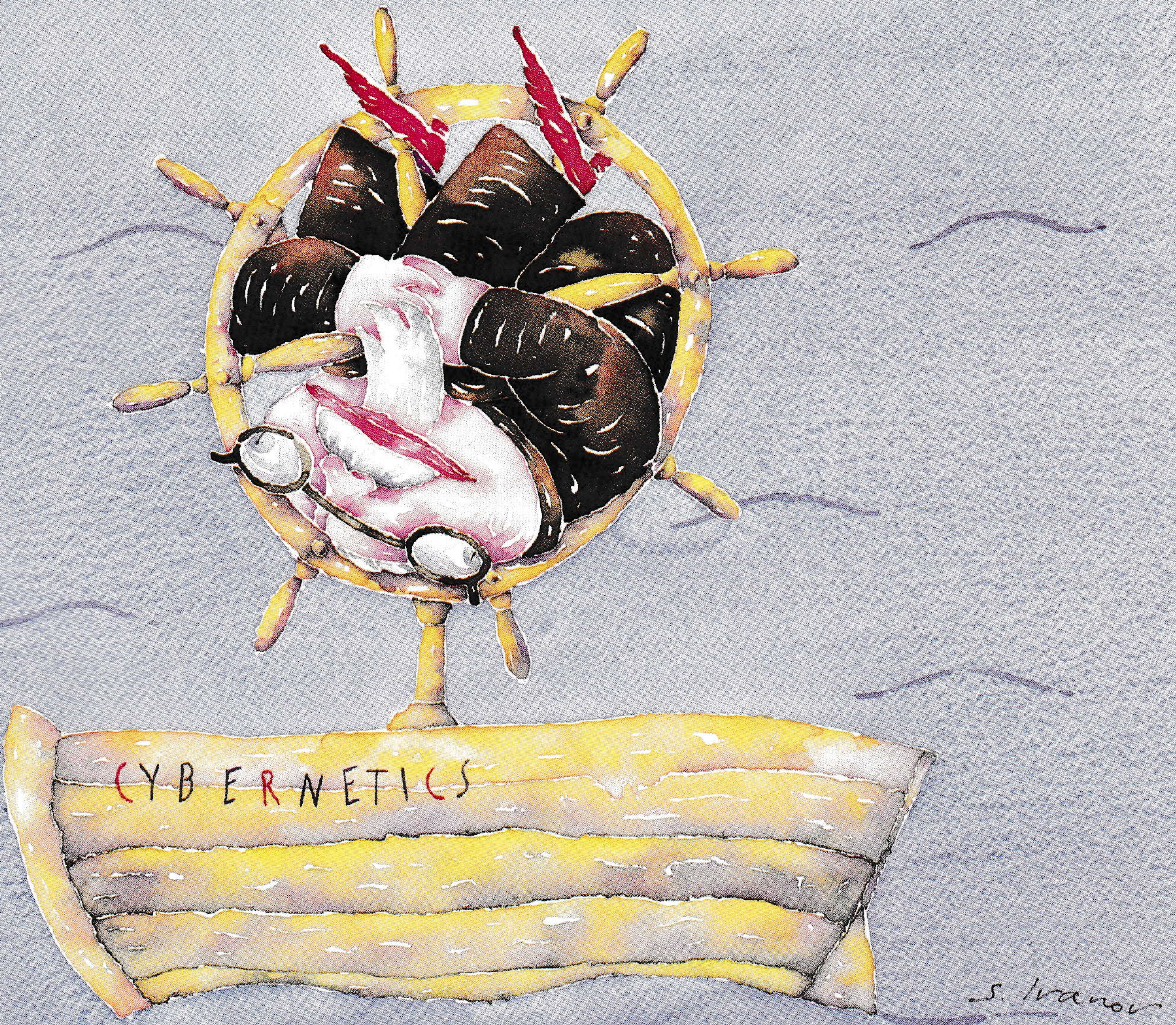
QUANTUM

NOVEMBER/DECEMBER 1994 $5.95



SPRINGER INTERNATIONAL

GALLERY Q

National Museum of American Art, Washington DC/Art Resource, NY

African Jungle Picture:
If the Ladies Had Knew the Snakes Wouldn't Bite Them They Wouldn't Have Hurt the Snakes;
If the Snakes Had Knew the Ladies Wouldn't Hurt Them They Wouldn't Have Bit the Ladies
(1989) by Thornton Dial

HOW IS IT POSSIBLE FOR TWO SPECIES TO INhabit the same ecological niche and yet misunderstand each other so fundamentally? Let's assume that snake flanks are not a delicacy in these ladies' cuisine. We know that snakes are not in the habit of eating people. The irony here, then, is that a nonpredator–nonprey relationship plays out as a *double* predator–prey relationship: each participant predator and prey both!

Or perhaps we should say: attacker and attacked. For no good comes of this encounter. No one comes away with a full belly. In the worst case, no one is left alive.

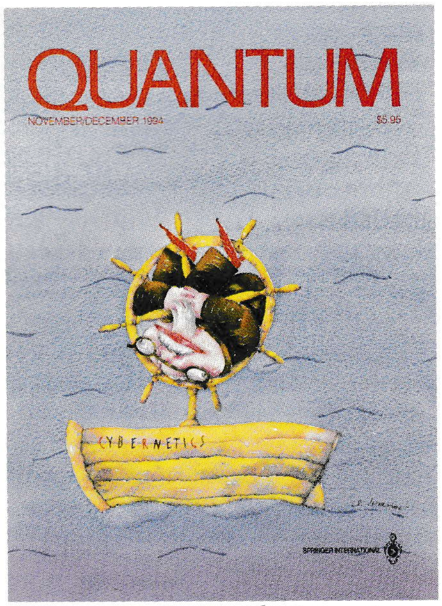
We are left to wonder how such a pointless tragedy could have been avoided. When the snakes reared their heads in self-defense, why did the ladies interpret that as aggressive? When the ladies screamed, why did the snakes not understand the ladies were afraid and simply slither away? In the immortal words of the prison guard in the classic movie *Cool Hand Luke*, "What we have here is a failure to communicate."

Then again, maybe the ladies and the snakes *were* communicating, but the noise arising from their instinctive mutual fears drowned out the messages. Such things have been known to happen, even among members of the same species.

This topsy-turvy vision of predators and prey is righted in the article that begins on page 15. And questions of communication theory will arise in later installments of the essay on Norbert Wiener (founder of the science of cybernetics), which begins on page 47.

QUANTUM

NOVEMBER/DECEMBER 1994 VOLUME 5, NUMBER 2



Cover art by Sergey Ivanov

The relaxed captain of the good ship *Cybernetics* is Norbert Wiener, who coined the term for the study of control and communication in organisms, automatic machines, and organizations. (The word comes from the Greek for "helmsman.") He was a major force in the development of this interdisciplinary science, and his work on Brownian motion, potential theory, and generalized harmonic analysis places him among the great mathematicians of the century. His three hundred publications range through philosophy, quantum mechanics, neurology, and religion, in addition to mathematics, and include book reviews, science fiction, and (rumor has it) pseudonymous detective novels.

Turn to page 47 for the first installment of a commemorative essay on the life and work of this innovative thinker.

(By the way, the chromatic binary code in the ship's name contains a second message the artist hadn't intended. Can you guess what it is? See page 22.)

FEATURES

DEPARTMENTS

PUBLISHER'S PAGE

Throwing the book at them

Maybe we should be throwing it out

ALMOST WITHOUT EXCEPTION, the science textbooks currently in use are seriously defective. No wonder students decide they can't learn science, or that it's boring.

What's wrong? First of all, the level of abstraction in these books is unevenly distributed, and often it's clumped at the very beginning. Second, the material is sequenced according to the logic of the discipline, not according to how we learn science. Third, the books tell us what the authors know, but not how they know it. Fourth, and most importantly, the books fail to distinguish between empirical science, on the one hand, and theories and models, on the other.

Open any 9th or 10th grade biology text—a book used by students before they've taken chemistry. In chapter 3 or 4 you'll encounter structural organic chemistry that is a condensation of what you would find in an 11th grade chemistry book. Physics textbooks typically *begin* with one of the most *abstract* of physical concepts: vectors. In addition to a disregard for how and when to introduce abstractions, these books are an unnourishing byproduct of the "layer cake" approach to science education in this country. Rather than study *all* of these subjects *every* year in an integrated way, American kids are fed disjointed textbooks, one per grade.

Have you ever noticed that a book will give you a technical term and then explain what it means? It should rather be providing the experience (or appealing to vicarious experience) first, *then* giving the name. Experience first, names later! But find a book that actually does this.

And what about evidence? When you studied photosynthesis, your textbook probably gave an elaborate explanation of the process, including the light and dark reactions, the two photosystems, and several dozen names of organelles and molecules, including complex organic molecules. But how many books describe the *evidence* for the structure or function of the various parts of the tissues or cells associated with photosynthesis? Have you ever noticed that the formula in some books uses six carbon dioxide molecules plus six water molecules to produce a carbohydrate molecule and six water molecules? Other books show twelve water molecules going in and six coming out. Now, why in the world would nature take in twelve molecules of water, only to give six back again as products? The answer is crucial: the *only way* the process could split oxygen from water (instead of from carbon dioxide) is by using twelve water molecules. How do we *know* the oxygen comes from the water instead of from the carbon dioxide? What is the line of *evidence* that demonstrates this fact? It's not enough to learn *what* happens—we need to know *how* it happens, and how we *know* it happens in a particular way.

The most serious problem with existing textbooks is the failure to distinguish between (1) the observations and laws of empirical science and (2) the theories or models we use to account for that empirical knowledge. For example, light is often called "light waves," and the books say the light waves do this or that. What about water waves? We don't say they "do" anything. Certain aspects of water can be described in terms of waves, but there are others where the wave analogy falls short. For example, water flow requires a different description. Likewise, light also behaves like a particle. Some phenomena (for instance, the photoelectric effect) require a particle model of light.

Some authors seem to enjoy creating the impression that there is some kind of paradox in nature because "light must be both a wave and a particle." This confusion is caused by imposing the model on the phenomenon. Light is neither a particle nor a wave. Light is light! The paradox lies in our inability to adequately describe or explain it. Nature speaks with one voice—it harbors no paradoxes. What we need is a model that can explain both sets of phenomena. (Actually, a more sophisticated use of a statistical version of quantum, or photon, theory can accomplish this unification.)

We must be able to distinguish between a definition ($D = m/v$) and a law (PV = constant). One is made up to represent a concept (in this case, density), while the other describes a relationship between two concepts expressed symbolically. Observations and measurements represent our appeal to nature, and the results are sometimes summarized in the form of empirical laws. When we want to explain those observations or measurements, we create a theory or model.

For example, when we want to explain Boyle's law, Charles's law, and Guy-Lussac's law, we create a theory called the kinetic theory of gases. We apply Newton's laws of motion to a set of particles, and we work under a certain set of assumptions. When we're finished, we find, to our surprise, that the temperature of such a gas is equivalent to the average kinetic energy of the molecules. Yet books often *define* temperature as the average kinetic energy of molecules—that is, a concept has been defined in terms of a theoretical prediction. (Actually, the temperature is *not* the average kinetic energy of molecules. At very low temperatures—below the lambda point for liquid helium—the temperature is more nearly connected with the orientation of atoms and nuclei rather than with the motion of atoms or molecules.) In a similar vein, Stephen J. Gould has written eloquently in a recent issue of *Natural History* on the distinction between the *facts* of evolution and the *theory* of natural selection.

One of the most enjoyable things about learning science is trying to explain what you've observed. You start with an educated guess (hypothesis) and, by testing it, you begin to create a more elaborate framework (a theory or model). This is the creative part of science, and it's a shame to deprive students of that experience, or to deride their initial efforts because they may not match the most current, "correct" explanation held by scientists. Those same scientists went through as many of the false starts as the novice student who is trying out his or her first hypotheses.

If you're a teacher, refuse to use textbooks that don't measure up and make sure your teaching matches how we actually learn science. If you're a student, insist on understanding the how's and why's, and pay attention to the evidence. Be skeptical of textbooks and of unsupported assertions. There is only one authority in science. It is neither your textbook nor your teacher. It is nature itself, and you appeal to that authority only through careful observation and measurement of natural phenomena.

—Bill G. Aldridge

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 5 (6 issues) will be published in 1994–1995. *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.

Advertising:

Advertising Representatives: (Washington) Paul Kuntzler (202) 328-5800; (New York) Bob Vrooman (212) 460-1700; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, Germany, telephone (0) 30-82 07-1, telex 185 411.

Second class postage paid at New York, NY, and additional mailing offices. **Postmaster:** send address changes to: *Quantum* Springer-Verlag New York, Inc., Journal Fulfillment Services Department, P. O. Box 2485, Secaucus NJ 07096-2485.

Subscription Information:

North America: Student rate: $15; Personal rate (nonstudent): $20; Institutional rate: $34; 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: Subscription rates in U.S. currency as above (all rates calculated in DM at the exchange rate current at the time of purchase) plus postage and handling. SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Air mail delivery to all other countries is available upon request. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, Germany.


B.B.94.

Strolling to Chebyshev's theorem

It isn't really our goal, but somehow we'll get there

by Victor Ufnarovsky

IT'S A SPECIAL DELIGHT TO walk "following your nose": you wander aimlessly, and all of a sudden you encounter something absolutely unexpected, something that could never enter your head when you started your walk. Let me invite you to take such a walk along one intricate mathematical path.

We'll set off from the well-known criterion for divisibility by 9: *a number n is divisible by 9 if and only if the sum of its digits is divisible by 9.* Or, to put it even more imposingly: *if the sum $\sigma(n)$ of the digits of a number n is subtracted from n, then the result is always divisible by 9.*

But first let's take a seat and talk about notation. The choice of notation, however strange it may seem, determines a lot. That's why mathematicians are usually rather conservative in this regard. For instance, the notation n will tell any mathematician that we're talking about integers, most probably positive integers. (By the way, what did *we* mean by n in the preceding paragraph?) The notation σ wasn't picked out of thin air either: from time immemorial the Greek letter "sigma," and its capital version Σ as well, have been used to denote summation. Our sum is small, so we'll use the small letter.

Let's look at notation from another angle. You probably know that the Arabs write from right to left rather than from left to right, as we do. But did you ever consider the fact that we ourselves treat numbers in the Arabic way—from right to left? Go ahead, don't believe me—after all, we write down numbers from left to right. True enough. But how do we add them? What digit do we begin with: the first or the last? And what about multiplication? Just try it the other way around! In fact, it would be more convenient to write digits from right to left, but there's nothing we can do about it: this custom has become second nature to us. To get around this habit, let's write numbers in another form—not as a string of digits, but as an expansion in powers of ten—say, not 234, but $4 + 3 \cdot 10 + 2 \cdot 100$. Our habit doesn't rebel against it, so let's write our number n in this way:

$$n = a_0 + a_1 \cdot + a_2 \cdot 10^2 + \ldots + a_k \cdot 10^k,$$

where a_0 is the last digit of n, a_1 is the next to last, . . . , and a_k the first, so that the total number of digits is $k + 1$ and their sum is

$$\sigma(n) = a_0 + a_1 + a_2 + \ldots + a_k.$$

Now it costs nothing to prove our statement:

$$\begin{aligned} n - \sigma(n) &= (a_0 - a_0) + a_1(10 - 1) \\ &\quad + a_2(10^2 - 1) + \ldots + a_k(10^k - 1) \\ &= 0 + 9 + 9a_1\ 99a_2 + \ldots + 99\ldots9a_k, \end{aligned}$$

which is, of course, divisible by 9. (By the way, how many nines are in the last coefficient?)

Now that we've admired the results of our labor to our hearts' content, let's continue our walk. What else can we obtain just as easily, without exerting ourselves too much? We can change either the problem, its proof, or the notation. Let's not fiddle with the proof. Can we change the problem? Sure, it's not hard to think up and prove a criterion for divisibility by 11—we only have to consider the alternating sum $a_0 - a_1 + a_2 - a_3 + \ldots$. And if we dig into it deeper, we'll get something like a universal divisibility criterion (prove it as an exercise):

Let m be a natural number and let $p_1, p_2, \ldots, p_k$ be the remainders of the numbers $10, 10^2, \ldots, 10^k$ upon division by m. Then the number

$$n - (a_0 + a_1p_1 + a_2p_2 + \ldots + a_kp_k)$$

is divisible by m.

For $m = 9$ we get the result proved above. (And what about 11?) For $m = 7$ we get the sequence of remainders 3, 2, 6, 4, 5, 1, 3, 2, and so on, periodically. Therefore, the remainder of 1994 when divided by 7 equals the remainder of $4 + 9 \cdot 3 + 9 \cdot 2 + 1 \cdot 6 = 55$, which in turn equals the remainder of $5 + 5 \cdot 3 = 20$, which in turn . . . But perhaps at this point we can stop and say that it's 6. Not too interesting . . .

Let's walk another way and try to change notation. How? Those who

Art by Vasily Vlasov

are even slightly familiar with programming will at once suggest that we use another number system—for instance, binary notation:

$$n = b_0 + b_1 \cdot 2 + b_2 \cdot 2^2 + \ldots + b_m \cdot 2^m,$$

where the coefficients b_i are ones and zeros—for example, $25 = 1 + 0 \cdot 2 + 0 \cdot 4 + 1 \cdot 8 + 1 \cdot 16$. We can also compute the sum of binary digits $\sigma_2(n) = b_0 + b_1 + \ldots + b_m$. (The index 2, of course, alludes to the base of the number system, so the sigma we used above was σ_{10}.) The corresponding theorem reads: *The number $n - \sigma_2(n)$ is divisible by . . .* Indeed, by what? In the decimal system the corresponding difference was divisible by $9 = 10 - 1$, so here it must be divisible by $2 - 1 = 1$. This fact is true but not very valuable. What if we try the general case of p-nary notation with an arbitrary base p? We write

$$n = a_0 + a_1 p + a_2 p^2 + \ldots + a_k p^k,$$

where $a_i < p$, and denote

$$\sigma_p(n) = a_0 + a_1 + a_2 + \ldots + a_k.$$

Now we arrive at a beautiful theorem.

Theorem 1. *The number $n - \sigma_p(n)$ is divisible by $p - 1$.*

Let's prove it. Our reasoning remains basically the same:

$$n - \sigma_p(n) = (a_0 - a_0) + a_1(p - 1) + a_2(p^2 - 1) + \ldots + a_k(p^k - 1)$$

is certainly a multiple of $p - 1$, because for any positive integer m, $p^m - 1 = (p - 1)(p^{m-1} + p^{m-2} + \ldots + 1)$.

For instance, in the octal system the number $n = 124$ is divisible by 7 (because $n - \sigma_8(n) = n - 7$ is divisible by 7). Care to check? Here goes:

$$4 + 2 \cdot 8 + 1 \cdot 64 = 84,$$

which is indeed divisible by 7. So we've created a new criterion for divisibility by 7. It's a pity we're not used to the octal notation.

Where else could we go? What else could be derived from divisibility? What if we . . . actually divided? Actually, that's a very good question: what is the quotient

$$\delta_p(n) = \frac{n - \sigma_p(n)}{p - 1}$$

equal to? Interesting . . . Why not begin simply with $p = 2$—at least we won't have to do any work to divide here. For starters, we'll fill in the following table:

n_{10}	n_2	$\sigma_2(n)$	$\delta(n) = n - \sigma_2(n)$
0	0	0	0
1	1	1	0
2	10	1	1
3	11	2	1
4	100	1	3
5	101	2	3
6	110	2	4
7	111	3	4
8	1000	1	7
9	1001	2	7
10	1010	2	8
11	1011	3	8
12	1100	2	10

What do we see? The number $\delta(n)$ doesn't change for odd values of n, but does change whenever n is even. By how much? Aha! Simply by the number of zeros at the end of the binary notation of n. And this number, as is clear, is the greatest power of two that is a factor of n.[1] For example, $n = 12$ is divisible at most by $4 = 2^2$, and we make a step of magnitude 2 from 8 ($n = 11$) to 10 ($n = 12$). So we can say that $\delta(n)$ "counts" how many powers of two are contained in the numbers $1, 2, \ldots, n$, or, we ought to say, in the product $1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$, which is called *n factorial* and denoted by $n!$. So, it looks as if the following fact is true: *$\delta(n) = n - \sigma(n)$ is the maximum exponent of a power of two that is a divisor of $n!$.* How about its proof?

The technique of induction naturally springs to mind. For small n everything is clear from the table. Let's take the "inductive step." Suppose we already know that our statement is true for $n - 1$; that is, $(n - 1) - \sigma_2(n - 1)$ is the greatest power of two that is a divisor of $(n - 1)!$. Consider the number n. To obtain $n!$ we must multiply $(n - 1)!$ by n. In so doing we increase the number of powers of two in $(n - 1)!$ by that in n—that is, by the number of zeros at the end of the binary notation of n. And what happens to the difference $(n - 1) - \sigma_2(n - 1)$? The minuend $n - 1$ increases by 1. How does the subtrahend $\sigma_2(n - 1)$ change?

Suppose the binary notation of n has k zeros ($k \geq 0$) at the end: $n = \ldots 100\ldots0$. Then the binary notation of $n - 1$ ends with exactly k ones: $n - 1 = \ldots 011\ldots1$ (it may consist of nothing but these ones). Therefore, the number of ones has decreased by $k - 1$, and the total change equals $1 - (-(k - 1)) = k$. This completes the induction proof.

What next? Well, it would be interesting to check if this is true in general—that is, whether the number

$$\delta_p(n) = \frac{n - \sigma_p(n)}{p - 1}$$

is the maximum exponent of a power of p that is a divisor of $n!$.

Exercise 1. Prove this for $p = 3$.

Alas, we're going to be disappointed in the case $p = 4$. The number $6!$ is divisible by 4^2, but since 6 is written as 12 in base 4, $\delta_4(6) = (6 - 3)/3 = 1 \neq 2$. I'll tell you the reason right away: p must be prime.

Exercise 2. Prove the statement for any prime p.

Fortunately, prime numbers suffice to answer questions about divisibility. But what can we do with factorials—where can we apply the knowledge we've just acquired? First of all, certainly, to binomial coefficients $C_n^k = n!/k!(n - k)!$ (often denoted by $\binom{n}{k}$).

The formula they are most closely linked with is the Binomial Theorem:

$$(x + y)^n = x^n + C_n^1 x^{n-1} y + C_n^2 x^{n-2} y^2 + \ldots + C_n^{n-1} x y^{n-1} + y^n.$$

[1]This is admittedly a rather slipshod expression, but we'll allow ourselves to use it at times for the sake of brevity. To be accurate, I should have said "the exponent of the greatest power of two . . ." I hope this shortcut won't mislead you—neither here, nor below.

For greater symmetry we'll use the formula

$$C^n_{m+1} = \frac{(m+n)!}{n!m!}.$$

Thanks to what we've learned, we're able to compute the greatest exponent with which a prime p enters the factorization of this binomial coefficient. It's exactly equal to $\delta(m+n) - \delta_p(m) - \delta_p(n)$, or

$$\frac{(m+n)-\sigma_p(m+n)-(m-\sigma_p(m))-(n-\sigma_p(n))}{p-1}$$

$$= \frac{\sigma_p(m)+\sigma_p(n)-\sigma_p(m+n)}{p-1}.$$

Beautiful, isn't it? For instance, if $\sigma_p(m+n) = \sigma_p(m) + \sigma_p(n)$, then C^n_{m+n} is not divisible by p, and vice versa. And when does this happen? Well, at least when there are no carries from digit to digit as we add m and n in the p-nary notation. For instance, adding 23 and 32 in base 7, we get 55 without carries. The conclusion: since $3 + 2 \cdot 7 = 17$ and $2 + 3 \cdot 7 = 23$, then C^{17}_{40} is not divisible by 7.

What if there are carries? Suppose the ith digits of m and n are $r < p$ and $s < p$, respectively, and their sum $r + s > p$. Then we'll have to carry 1 to the next digit, and in the ith digit $r + s - p$ will replace $r + s$. Therefore, the contribution of this digit to $\sigma_p(m) + \sigma_p(n) - \sigma_p(m+n)$ will be $p - 1$. But we divide it by $p - 1$. Remarkable! How could we have failed to guess at once that the following theorem is true?

Theorem 2. *If p is a prime, then the exponent of the greatest power of p that is a divisor of C^n_{m+n} equals the number of carries when the numbers m and n are added in the p-nary number system.*

Now that's some theorem we've managed to reach! What could we derive from such a nontrivial fact? You don't know where to look first! For the sake of simplicity let's start by studying C^n_{2n}, the largest binomial coefficient among all that enter the expansion of the binomial $(x + y)^{2n}$. (By the way, can you prove that this coefficient is indeed the largest?) The greatest power of p into which it divides is equal to the number of carries when n is added to itself in the base-p number system.

Suppose $n < p < 2n$. Then the p-nary notation of n consists of one digit (the digit n), and $2n$ consists of two digits—say, $2n = r + 1 \cdot p$. So there is exactly one carry, and the factor p enters C^n_{2n} once. (You can check this simple result directly.) It follows that *the product of all prime numbers between n and $2n$ does not exceed C^n_{2n}*. Can you imagine a more attractive result? Let's do a rough estimate of C^n_{2n}.

Setting $x = y = 1$ in the binomial theorem, we get

$$1 + C^1_{2n} + C^2_{2n} + \cdots + C^{2n-1}_{2n} + 1 = 2^{2n},$$

which implies

$$C^n_{2n} < 4^n.$$

Similarly, the product of all primes between $n/2$ and n is less than $4^{n/2}$; between $n/4$ and $n/2$ it's less than $4^{n/4}$; and so on. Then the product of all primes between 1 and n is less than

$$4^{n/2} \cdot 4^{n/4} \cdot 4^{n/8} \cdot \ldots < 4^n.$$

Thus, free of charge, we get the following fact, which is not at all obvious.

Theorem 3. *The product of all prime numbers less than n does not exceed 4^n.*

Exercise 3. Prove this strictly. (We were too careless with division by two, having "forgotten" that some numbers may be odd. Perhaps the best rigorous approach is to use induction.)

Now suppose that $p \le n$. Then the notation of n contains at least two p-nary digits. If $2n$ is written with exactly two digits, then $2n < p^2$, and there surely is no more than one carry. This proves the following statement.

Lemma 1. *If $p > \sqrt{2n}$, then the greatest power of p that is a divisor of C^n_{2n} has an exponent no greater than 1.*

What if there is no divisibility at all? Since $2n < p^2$, we can write $n = a_0 + a_1 p$, where $a_0 < p$, $a_1 < p/2$. To have no carries, it's necessary that $a_0 < p/2$. In particular, for a_1 we see that C^n_{2n} is not divisible by p if $a_0 = n - p < p/2$. This leads to the following lemma.

Lemma 2. *If $n \ge p > 2n/3$, then C^n_{2n} is not divisible by n ($n > 2$).*

To prove this, it suffices to note that the condition $n - p < p/2$ is equivalent to $n < 3p/2$, or $p > 2n/3$.

Now let's reckon what happens with small values of p, $p \le \sqrt{2n}$. Here several carries may occur, but no more than k, if $2n = a_0 + a_1 p + a_2 p^2 + \ldots + a_k p^k$. Since $2n \ge p^k$, we have $\log_p 2n \ge k$, so the greatest power of p that is a divisor of C^n_{2n} has an exponent no greater than $\log_p 2n$. This means that for arbitrary prime p the following lemma is true.

Lemma 3. *If $N = p^m$ is a divisor of C^n_{2n}, then $N \le 2n$.*

Indeed,

$$p^m \le p^{\log_p 2n} = 2n.$$

Well then, now we have a more or less clear idea of the structure of the number C^n_{2n}. Its factorization into powers of primes consists of three types of factors:

1. Prime numbers greater than n (and, naturally, less than $2n$)—each entering the factorization exactly once;
2. Prime numbers less than $2n/3$, but greater than $\sqrt{2n}$—each appearing no more than once;
3. Prime numbers less than $\sqrt{2n}$. Here the divisibility by p^k with $k > 1$ is possible, but all the same the total contribution p^k of each of these primes is no greater than $2n$.

Could it be possible that the first group is absent—that is, that there are no primes between n and $2n$? In that case all prime factors must be concentrated in the second and third groups. Can we estimate their actual contribution? The product of all the numbers of the second group, by theorem 3, is no greater than $4^{2n/3}$. In the third group there are certainly less than $\sqrt{2n} - 1$ primes, so their contribution does not exceed $(2n)^{\sqrt{2n}-1}$. To sum up: if there are no primes between n and $2n$, then the inequality

$$C_{2n}^{n} < 4^{2n/3} \cdot (2n)^{\sqrt{2n}-1} \qquad (1)$$

must hold.

What a thought! If we prove that this inequality is wrong, we'll prove the famous postulate of Bertrand: *There is at least one prime number between n and 2n.*

Let's try to estimate C_{2n}^{n}. Since it's the largest of the binomial coefficients in $(1 + 1)^{2n}$, and since there are $2n + 1 < 4n$ of them in all, we can be sure that $C_{2n}^{n} > 4^n/4n$. Along with equation (1), this yields successively (for large enough n)

$$\frac{4^n}{4n} < 4^{2n/3} \cdot (2n)^{\sqrt{2n}-1},$$

$$4^{n/3} < 2 \cdot (2n)^{\sqrt{2n}},$$

$$\frac{n}{3} < \sqrt{18}\log_4 2n + \frac{1}{2},$$

and finally

$$\sqrt{n} < \sqrt{18}\log_4 2n + \frac{1}{2}. \qquad (2)$$

But the logarithm is a slow function, so that $\sqrt{n}$ overtakes it sooner or later. What we need to know is when. Let's try $n = 1{,}000$. Clearly, $\sqrt{1000} > 30$, while $\log_4 2000 < \log_4 4096 = \log_4 4^6 = 6$. Since $30 > \sqrt{18} \cdot 6 + 1/2$, our inequality is violated for $n = 1{,}000$ and (as you, I hope, will prove on your own, using derivatives[2]) for all $n > 1{,}000$. Therefore, for these values of n the following theorem is true.

CHEBYSHEV'S THEOREM (BERTRAND'S POSTULATE). *There is always at least one prime between n and 2n.*

That's very nice, but what do we do with small values of n? For them inequality (2) seems to be correct. Never mind—Bertrand's postulate is also true for them. We assure ourselves of this simply by searching through a table of prime numbers, or writing a tiny computer program. Or maybe you're naturally punctilious in making estimates and would like to obtain a more exact inequality. It's a matter of taste.

However, we've been "strolling" for quite a while now. It's high time we rested, don't you think? But if you want to take another stroll on your own, here are some problems to get you started.

Problems

1. Prove that for any prime p and integer x and y the number $(x + y)^p - x^p - y^p$ is divisible by p.

2. Extend the previous problem to the case of several summands and derive Fermat's "Little" Theorem: $x^p - x$ is divisible by p (for a prime p, of course).

3. Prove that if a binomial coefficient C_n^k is divisible by a power of a prime $N = p^m$, then $N \le n$.

4. Prove that for $n > 5$ there are two prime numbers between n and $2n$.

5. Let p^k be the kth prime number. Prove that $p_{k+2} < 2p_k$.

6. Prove that $n!$ (for $n > 1$) is never a power of an integer.

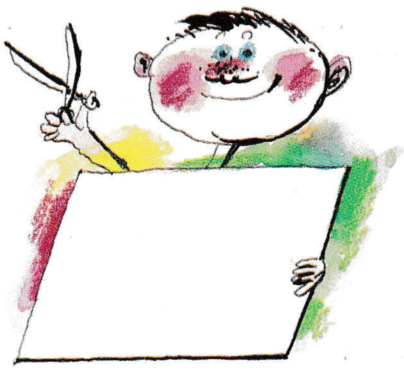
[2]This technique is explained in "Derivatives in Algebraic Problems" in the November/December 1993 issue of *Quantum*.—*Ed.*

BRAINTEASERS

Just for the fun of it!

B126

Sort it out. If the brainteaser you solved before you solved this one was harder than the brainteaser you solved after you solved the brainteaser that you solved before you solved this one, was the brainteaser you solved before you solved this one harder than this one? (N. Rozov)

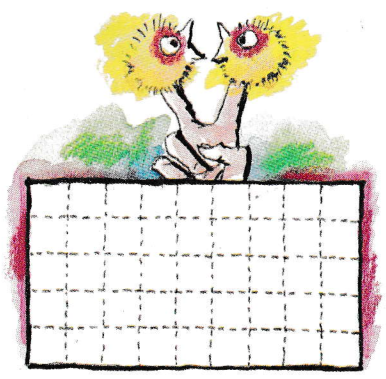


B127

Making a rhombus. Cut a parallelogram along a straight line through its center so that the two pieces can be rearranged to make a rhombus. (A. Savin)

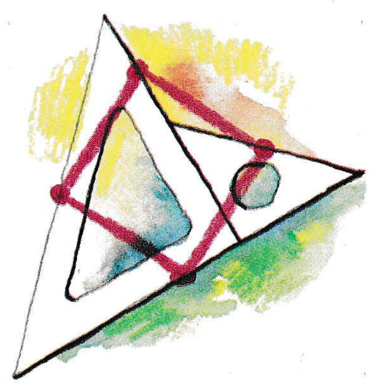
B128

Not too heavy. The mass of each weight in a set is no greater than 10 kg. If the set is arbitrarily divided into two groups, the total mass of one of the groups is also no greater than 10 kg. What is the greatest possible mass of all the weights in the set?



B129

Integer rectangles. A rectangle measuring 5×9 is cut into 10 rectangles with integer side lengths. Prove that at least two of them are congruent. (K. Kohas)



B130

Midpoint square. Two isosceles right triangles are brought together as shown in the figure. Prove that the midpoints of the sides of the nonconvex quadrilateral they form are the vertices of a square. (V. Proizvolov)

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

Art by Pavel Chernusky

Love and hate in the molecular world

"Drawn by Love, they gather in a common order,
Then, by Hostility of discord, they are driven from one another,
Until, resigned to their fate, they become one entity."
—Empedocles

by Albert Stasenko

THE GREAT RUSSIAN WRITER Anton Chekhov wrote in one of his stories: "Now the doorkeeper of the State Chamber piped up . . . He said that in St. Petersburg there was a frost of two hundred degrees . . . The people, he said, were terrified. But whether it was in St. Petersburg or Moscow—I don't exactly recall." Chekhov, a physician by education, had in mind the Celsius temperature scale, of course, so a frost of two hundred degrees corresponds to an absolute temperature of 273 – 200 = 73 K. If we look in a reference book on physics, we can see that air becomes a liquid at this temperature, so it would be quite a feat even to be "terrified" in such weather.

Only in this century have scientists managed to convert all the gases into liquids—even those that seemed to be incondensable, or "true" gases. Indeed, the primary feature of gases is that they strive to occupy all of the available space, so that molecules of gas in outer space, for instance, would be able to fly off to the ends of the universe.

What are the ties that keep molecules together in liquids? Might they actually have some sort of hooks and pegs, as the ancient atomists imagined? From the perspective of the Greek philosopher Empedocles, the condensation of gas into liquid could be described like this: attracted by Love, the molecules are drawn together, but then Discord arises among them, trying to separate them, and finally, the equilibrium of these emotions creates the condensed phase.

But what is the quantitative measure of these molecular emotions—or in the terminology of modern physics, of the forces of attraction and repulsion? Moreover, why do these forces appear between electrically neutral molecules?

First of all, neutrality of charge doesn't imply the absence of electric field. Let's consider two point charges $+q$ and $-q$ a distance l apart (see the left portion of figure 1). Such a system is known as a *dipole*. At an arbitrary point, the total electric field of these charges is the vector sum of the two fields $\mathbf{E}_+$ and $\mathbf{E}_-$. Calculating this sum at any point, one can draw continuous lines of the

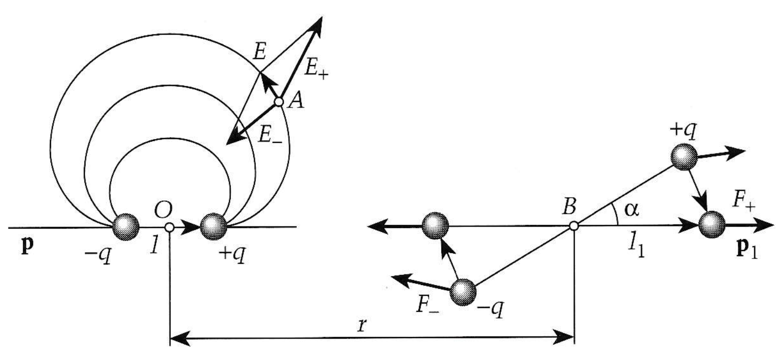
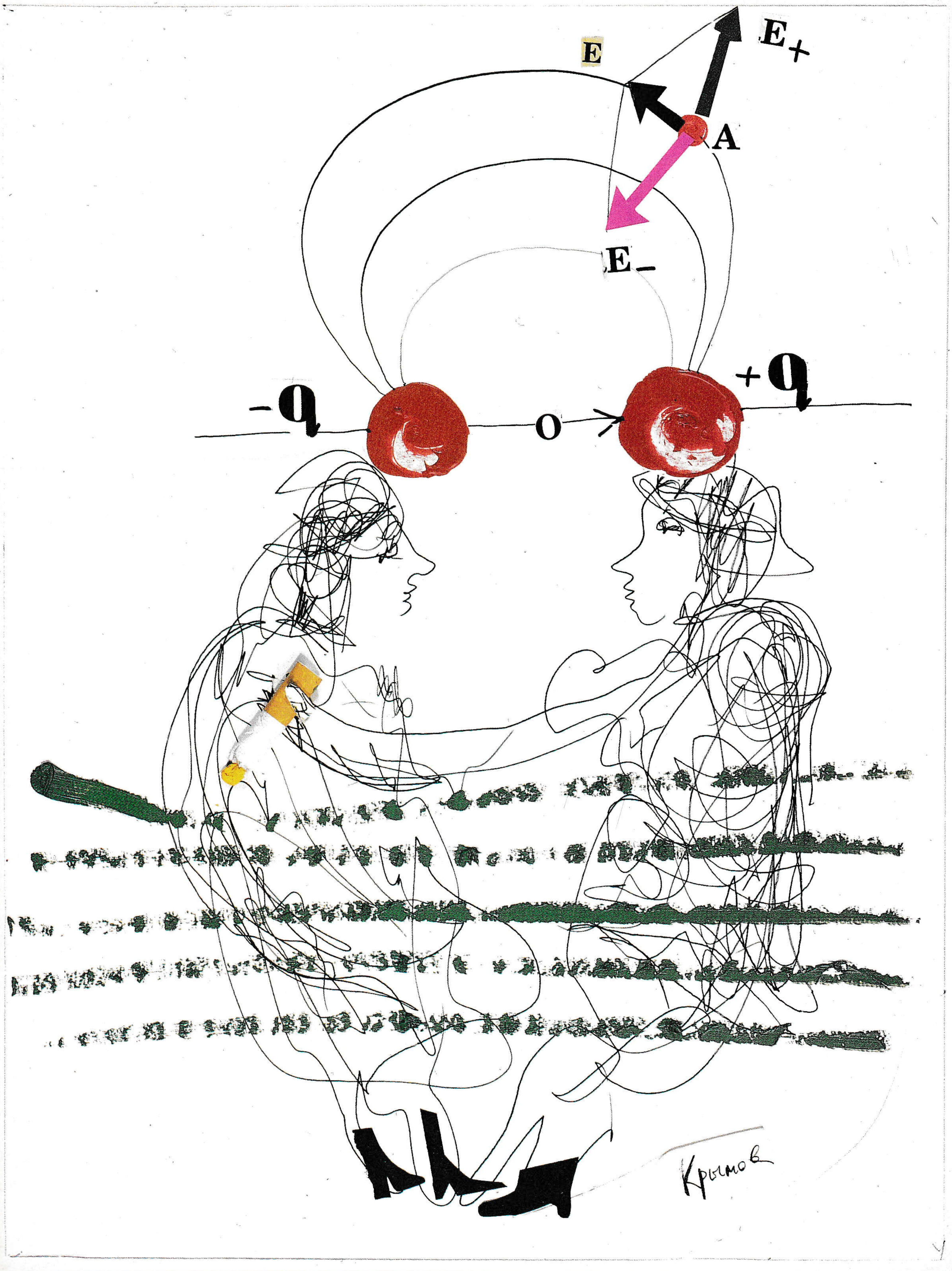


Figure 1

Art by Dmitry Krymov


E
E+
A
E−
−q
o
+q
Крымов

electric field from the positive charge to the negative charge (the figure shows the cross section of the field **E** with the axis of symmetry passing through the charges).

The product $\mathbf{p} = q\mathbf{l}$ (the vector $\mathbf{l}$ is drawn from the negative charge to the positive charge) is called the dipole moment. There are so-called *polar molecules* in which the "centers of gravity" of the positive and negative charges do not coincide (which, of course, doesn't prevent them of being electrically neutral). Such molecules have a marked dipole moment even in the absence of an external electric field, and hence they produce a dipole electrostatic field. It's convenient to define a "proper scale"—that is, a characteristic value for the molecular dipole moment. Let's take the elementary charge of a proton, $e_0 = 1.6 \cdot 10^{-10}$ C, as the characteristic charge, and a distance of one angstrom (introduced in physics as a distance scale in the atomic realm) as the length of the dipole: $l_0 = 10^{-10}$ m. Then $p_0 = e_0 l_0 = 1.6 \cdot 10^{-29}$ C · m. For example, a molecule of water has a large intrinsic dipole moment of $p = 0.62p_0$.

Now we find the electric field of a dipole at a distance r along its axis (that is, in the direction of the vector $\mathbf{l}$). To this end we must add the electric fields of both charges:

$$E = \frac{q}{4\pi\varepsilon_0(r - l/2)^2} - \frac{q}{4\pi\varepsilon_0(r + l/2)^2}$$
$$= \frac{q}{4\pi\varepsilon_0} \frac{2rl}{\left(r^2 - (l/2)^2\right)^2}. \qquad (1)$$

Let's see how this field changes at large distances from the dipole. What do we mean by "large"? Surely, the distance should be large in comparison with the size of the dipole itself : $r \gg l$. Then it's possible to neglect the value $(l/2)^2$ relative to r^2 in the denominator of equation (1), and in this dipole approximation we have

$$E \cong \frac{2}{4\pi\varepsilon_0} \frac{ql}{r^3}. \qquad (2)$$

Obviously, the dipole field decreases quicker (as the cube of distance) than that of a point charge.

What will happen if another dipole with a dipole moment $\mathbf{p}_1 = q\mathbf{l}_1$ is placed in the electric field of the original dipole (see the right portion of figure 1)? For simplicity's sake, let's assume that its center is on the axis OB. As the field produced here by the left dipole is directed to the right, a repulsive force $\mathbf{F}_+ = q\mathbf{E}_1$ acts on the positive charge, and an attractive force $\mathbf{F}_1 = -q\mathbf{E}_2$ acts on the negative charge ($\mathbf{E}_1$ and $\mathbf{E}_2$ are the electric fields produced at these points by the left dipole). Let the dipole $\mathbf{p}_1$ be rotated relative to the axis OB by an angle α. Then a torque will arise that tries to turn this dipole clockwise about point B. This rotational moment disappears only when $\alpha = 0$—that is, when $\mathbf{p}_1$ is parallel to the electric field of the left dipole.

Let's find the resultant of the forces $\mathbf{F}_+$ and $\mathbf{F}_-$ when $\alpha = 0$. To simplify our investigation of the dependence of this force upon the distance r between the centers of the dipoles, we assume, as above, that this distance is much larger than the size of either dipole: $r \gg l, l_1$. For example, at room temperature the average distance between molecules is tens of times greater than molecular diameters.

In this case the electric field produced by the left dipole can be calculated from equation (2), and the net force is

$$F_+ - F_- = q(E_1 - E_2)$$
$$= \frac{2q^2 l}{4\pi\varepsilon_0}\left(\frac{1}{(r + l_1/2)^3} - \frac{1}{(r - l_1/2)^3}\right)$$
$$= \frac{2q^2 l}{4\pi\varepsilon_0}\left(\frac{-3r^2 l_1 - l_1^3/4}{\left(r^2 - (l_1/2)^2\right)^3}\right).$$

Neglecting the small terms in the numerator and denominator yields

$$F_+ - F_- \cong -\frac{q^2}{4\pi\varepsilon_0}\frac{6ll_1}{r^4}. \qquad (3)$$

From this it's clear, first, that the force is attractive (notice the minus sign); and, second, provided the value of l_1 is fixed, the force decreases as the inverse fourth power of the distance between the dipoles.

So we have seen how polar molecules are mutually attracted. However, molecules exist that have no intrinsic dipole moment (they are called, naturally enough, *nonpolar*)—for example, the familiar molecules of oxygen and nitrogen. But under certain conditions they also condense to the liquid phase, which means that the molecules attract one another. Why?

To begin with, when a molecule is placed in an external field **E**, it can be polarized even if previously it had no dipole moment. Let's consider a simple model of a neutral particle: a positive central nucleus of charge q and a negatively charged ring of radius a concentric with the nucleus (fig. 2). The positive charge is displaced in the direction of **E**, and the negative charge is displaced in the opposite direction. As a result, a nonzero distance l_1 appears between the "centers of gravity" of the charges q and $-q$ (that is, a dipole moment arises). To come up with a value for this dipole moment, let's formulate the equilibrium condition for the Couloumb force of attraction between the nucleus and the ring and for the force due to the external field.

For an arbitrary section of the ring

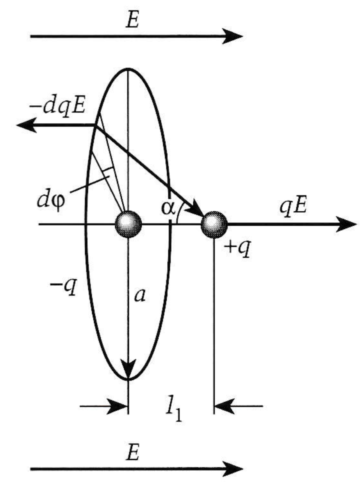


Figure 2

with a length $ds = ad\phi$ carrying a charge

$$dq = -q\frac{ds}{2\pi a},$$

we obtain

$$-dqE + \frac{1}{4\pi\varepsilon_0}\frac{qdq}{a^2 + l_1^2}\cos\alpha = 0,$$

or, taking into account that

$$\cos\alpha = \frac{l_1}{\sqrt{a^2 + l_1^2}},$$

canceling dq out of the equation, and using the condition that the deformation of our system in the external field is small ($l_1^2 \ll a^2$), we finally have

$$ql_1 = 4\pi\varepsilon_0 a^3 E.$$

It's particularly important that this *induced* dipole moment is proportional to the strength of the external field and is directed along it.

Now, returning to the situation depicted in figure 1 and using equation (2)—which quantifies the electrostatic field of the left dipole—as the external field, we see that a nonpolar molecule at a distance r acquires an induced dipole moment

$$p_1 = ql_1 = 4\pi\varepsilon_0 a^3\frac{2ql}{4\pi\varepsilon_0 r^3} = 2ql\frac{a^3}{r^3}.$$

Then, according to equation (3) it will be attracted by a force

$$-\frac{2(ql)^2}{4\pi\varepsilon_0}\frac{6a^3}{r^7} \sim -\frac{1}{r^7}.$$

How, then, does a nonpolar molecule acquire an initial dipole moment ql (fig. 1)? First of all, just because a molecule has no dipole moment on average, that doesn't mean it doesn't have one *at any particular moment*. Thus, the *average* dipole moment can be zero if calculated for a long period of time. For example, you might run quickly in the hallway from one wall to the other, and get very tired of it, but your coordinate $x(t)$, averaged for an hour, a day, or a year, would show that your displacement is zero (on average).

In the same way the dipole moment of a molecule can vary quickly with time while being zero on average (such a situation is shown qualitatively in figure 3). Imagine, for

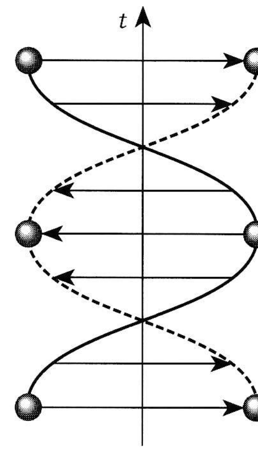


Figure 3

example, that the positive charge in figure 2 oscillates with some frequency, moving from the ring's center once to the left, then to the right. The dipole moment will change correspondingly with the same frequency, and so will the electric field quantified in equation (2)—if we neglect the time lag resulting from relativistic effects. If another molecule of the same substance were at some point in space, the alternatiing field of the first molecule would polarize this molecule at the same frequency. Consequently, at any time the molecules have dipole moments of the same direction and so are attracted with the force shown in equation (3), and their mean dipole moments are zero. (This is reminiscent of the interaction of two tuning forks: strike one of them, and the other responds at the same frequency in what is called a "sympathetic vibration.")

The attractive force described above, which increases sharply as the distance decreases, was given the name *van der Waals force* after the scientist who offered a relatively simple equation for a real gas that differed from the well-known Clapeyron equation for an ideal gas. The existence of intermolecular forces causes molecules to "feel" one another at great distances and not only when they collide, as is the case with solid balls. When the molecules are cooled and their average kinetic energy and speed decrease, they can be near one another for a longer time, and the work of the attractive forces (corresponding to potential energy) at long last will "win" over the kinetic energy, resulting in condensation.

Of course, as the molecules get closer, attraction must give way to repulsion ("Love" is replaced by "Hate," in the terminology of the ancient philosophers), because the distance between molecules can't be zero, as you know. The repulsive forces depend more strongly on the distance, as $\sim 1/r^n$, where n varies for various substances from 9 to 15. At a certain distance $r = r_0$ the attraction and repulsion become equal (fig. 4). This is simply the mean dis-

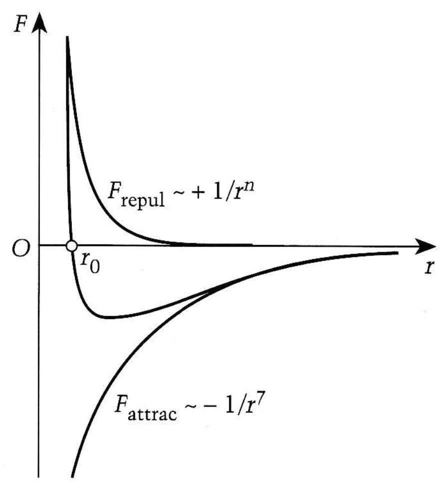
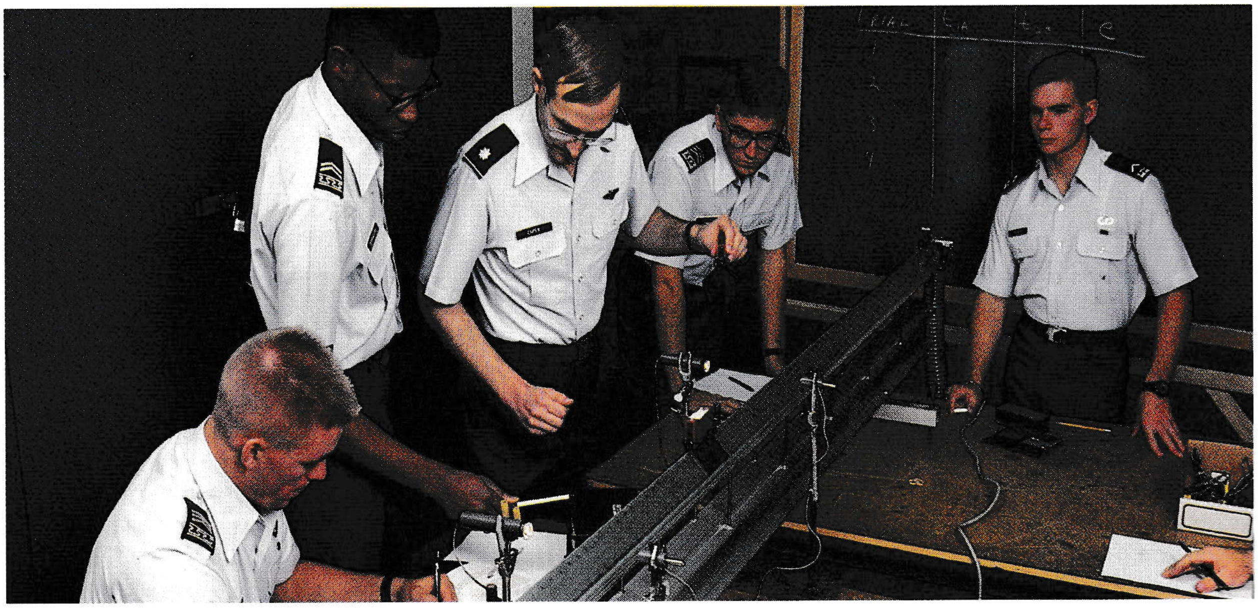


Figure 4

tance between molecules in the condensed substance—in a liquid or solid body. It doesn't mean, of course, that the molecules stopped moving completely.

Now is the time to recall that these are nevertheless microscopic objects, and that only quantum mechanics can give a correct description of their interaction—a particularly pleasant reminder when it appears in the pages of *Quantum* magazine!



TAKE IT HIGHER

As one of the brightest and best math and science students in the nation, you're capable of reaching the top. And you want to attend a college or university that can help you do it. Consider the Air Force Academy. It's a college and more. It's a special place for students who seek excellence in all that they do.

At the Air Force Academy, you can "take it higher." The Academy offers a full four-year scholarship, plus room and board. You'll graduate with a bachelor of science degree in one of 26 majors. Cadets who take honors courses build an excellent foundation for graduate studies.

"The higher the climb, the broader the view."

American proverb

Selection for the Air Force Academy is based on academic, athletic and extracurricular performance. In addition to the math and laboratory science courses you've already taken, we recommend that you complete a solid college prep program, including four years of English, three years of social studies, two years of foreign language and one year of computer science. In addition, we suggest that you develop your leadership abilities through school and community activities. You should also prepare physically by taking part in group and individual strength development and endurance programs.

The Academy's outstanding academic, athletic and leadership programs can prepare you to be an air and space leader in the 21st century.

For more details, call (719) 472-2520. Or write: HQ USAFA/RRS, 2304 Cadet Drive, Suite 200, USAF Academy, Colorado 80840-5025.

The ACADEMY

AIM HIGH
AIR FORCE

Circle No. 1 on Reader Service Card

Look, Ma—no calculus!

Using computer spreadsheets to study population dynamics

by Kurt Kreith

Art by Sergey Ivanov

ENVIRONMENTAL CONCERNS are finding their way into many aspects of our lives. Even the popular and entertaining book *Jurassic Park* features an enigmatic mathematician who, after being mauled by a *Tyrannosaurus rex*, offers some provocative views on the environmental issues before us. Noting that the Earth had existed for billions of years before humankind entered the scene, he rejects the notion that we should "save the planet." Rather, his final words are: "Let's be clear. The planet is not in jeopardy. *We* are in jeopardy. We do not have the power to destroy the planet—or to save it. But we might have the power to save ourselves."

In keeping with this point of view, population dynamics is a branch of mathematics that deals with abstract models for describing the growth, interaction, and decline of "populations" (usually species of plants and animals, but also including humans). Since calculus had its roots in efforts to explain "how things change with time," the field of population dynamics is often identified with calculus-based tools. However, in this article I'll show that it's possible to deal with very important ideas from population dynamics *without* invoking calculus.

The possibility of bypassing calculus is the result of twentieth-century developments in technology that make it practical to study "discrete models for change." Such models are based on mathematical processes that, while easy to understand, can also be very time consuming. From a historical point of view, calculus was important because it provided techniques for circumventing the time-consuming processes these models entailed. Now, however, computers enable us to deal with many such models directly, including ones that describe how populations change.

One reason for taking interest in population dynamics is that it relates to important changes likely to occur during our lifetimes. The world's (human) population stands at about 5.6 billion, and it is currently increasing by almost 100 million people every year. Within 50 years it's very likely to exceed 10 billion. This raises questions such as: What effect will future human population growth have upon other species of plants and animals? What effect will it have upon us?

While population dynamics does not provide conclusive answers, it does enable us to approach such questions on the basis of rational thought and calculation. And, given access to a modest computer, one can develop powerful insights into the underlying processes and mechanisms without invoking calculus-based tools.

How money grows

Although population dynamics usually deals with animals and plants, it will be useful to begin by thinking instead about money. Consider the familiar fact that money deposited in a bank earns interest, and that the amount of interest received is a percentage of the amount you have on deposit. Assuming (as we shall) that you choose to let the interest accrue to your account, the growth of money in a bank is an example of a "feedback loop." That is, the amount of money you have on deposit determines the amount of interest you receive, which determines the amount you have on deposit, which determines . . .

Given a deposit of 100 dollars earning interest at 10 percent a year, you can calculate its future value as follows. At the end of the first year you will have your original $100 plus 10% of $100 = $(1 + 1/10) \cdot 100$ = $110; after two years you will have $110 plus 10% of $110 = $(1 + 1/10) \cdot 110$ = $121; and so on. These considerations lead to the following table:

k	0	1	2	3	4	...
$N(k)$	100	110	121	133.10	146.41	...

In this particular process we're able to express the outcome by means of a formula: *the kth entry in this table can be obtained by multiplying 100 by (1 + 1/10) k times.* In functional terminology,

$$N(k) = 100 \cdot (1 + 1/10)^k = 100 \cdot 1.1^k.$$

Many hand calculators provide a dynamic way of representing this process of change. My old Casio fx-911 requires that one enter 1.1 and then push the multiplication button *twice*. This programs the calculator to do *repeated* multiplication by 1.1. Now entering 100 and pushing the equal sign repeatedly, one obtains the displays 110, 121, 133.1, 146.41, and so on.

A more versatile way of doing such repeated calculations is to use a computer spreadsheet. In this article I'll illustrate the underlying ideas with a Microsoft spreadsheet program called Excel. However, other spreadsheet programs will also enable you to follow along and to work on the projects suggested below, even though some of my specific instructions may have to be modified accordingly.

To make a table representing the value of $100 invested at 10% a year for k years, we'll use two columns (fig. 1). The first column is labeled k and the second $N(k)$.

	A	B
1	k	N(k)
2	0	100
3	=A2+1	
4		

Figure 1

Our starting entries are $k = 0$ (in cell A2) and $N(0) = 100$ (in cell B2). In order to have k increase by 1 at each downward step, we must enter an appropriate rule (or formula) in cell A3. An entry preceded by an equal sign signals to Excel that we are about to enter such a formula. In this case we enter "=A2+1" in cell A3.

By "copying down" (you'll have to read your software's instructions on how to do this), the spreadsheet is programmed to continue this rule downward—that is, to enter "=A3+1" in cell A4, "=A4+1" in cell A5, and so on. Similarly, entering "=(1.1)*B2" in cell B3 and copying down establishes the rules "=(1.1)*B3" in cell B4, "=(1.1)*B4" in cell B5, and so on. (The asterisk is how one gets the spreadsheet program to multiply. I'll dispense with it below, trusting that you'll change the raised dot to an asterisk if need be. Later on we'll use the slash mark to indicate division in the spreadsheet program.)

In figures 2 and 3 I've copied these two formulas downward for ten places (in the range of cells A3 to B12) and relied on the spreadsheet to calculate the resulting values of k and $N(k)$. (Usually we print out only the numerical values of the entries, but in figure 2 I show the formulas.)

	A	B
1	k	N(k)
2	0	100
3	=A2+1	=1.1*B2
4	=A3+1	=1.1*B3
5	=A4+1	=1.1*B4
6	=A5+1	=1.1*B5
7	=A6+1	=1.1*B6
8	=A7+1	=1.1*B7
9	=A8+1	=1.1*B8
10	=A9+1	=1.1*B9
11	=A10+1	=1.1*B10
12	=A11+1	=1.1*B11

Figure 2

k	N(k)
0	100
1	110.00
2	121.00
3	133.10
4	146.41
5	161.05
6	177.16
7	194.87
8	214.36
9	235.79
10	259.37

Figure 3

	A	B	C
1	f =	4	
2			
3	k	Time	N(k)
4	0	=A4/$B$1	100
5	=A4+1	=A5/$B$1	=C4*(1+0.1/$B$1)
6	=A5+1	=A6/$B$1	=C5*(1+0.1/$B$1)
7	=A6+1	=A7/$B$1	=C6*(1+0.1/$B$1)
8	=A7+1	=A8/$B$1	=C7*(1+0.1/$B$1)
9	=A8+1	=A9/$B$1	=C8*(1+0.1/$B$1)
10	=A9+1	=A10/$B$1	=C9*(1+0.1/$B$1)

Figure 4

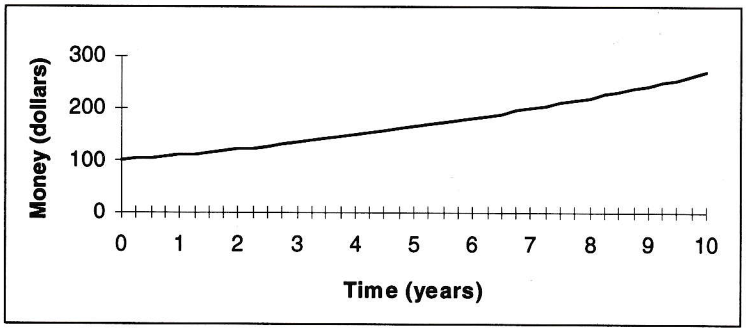


Figure 6

f =	4	
k	Time	N(k)
0	0	100
1	0.25	102.50
2	0.5	105.06
3	0.75	107.69
4	1	110.38
5	1.25	113.14
6	1.5	115.97
7	1.75	118.87
8	2	121.84
9	2.25	124.89
10	2.5	128.01
11	2.75	131.21
12	3	134.49
13	3.25	137.85
14	3.5	141.30
15	3.75	144.83
16	4	148.45
17	4.25	152.16
18	4.5	155.97
19	4.75	159.87
20	5	163.86
21	5.25	167.96
22	5.5	172.16
23	5.75	176.46
24	6	180.87
25	6.25	185.39
26	6.5	190.03
27	6.75	194.78
28	7	199.65
29	7.25	204.64
30	7.5	209.76
31	7.75	215.00
32	8	220.38
33	8.25	225.89
34	8.5	231.53
35	8.75	237.32
36	9	243.25
37	9.25	249.33
38	9.5	255.57
39	9.75	261.96
40	10	268.51

Figure 5

What do these tables have to do with population dynamics? Well, given a population of 100 rabbits that grows at 10% a year, this table would also give the number of rabbits after k years. The mathematics of money and rabbits seems to be the same.

Returning to our banking problem, suppose that a competitor to the bank paying you 10% a year offers to pay you not only 10% annually but also to *compound interest quarterly*. This means that, instead of receiving 10% interest at the end of the year, you would receive 2.5% interest four times a year. Under this plan you would have $\$100 \cdot (1 + 0.025) = \$102.50$ after three months, $\$102.50 \cdot (1 + 0.025) = \$100 \cdot (1 + 0.025)^2 = \$105.06$ after six months, . . . , and $\$100 \cdot (1 + 0.025)^4 = \$110.38$ after twelve months. The extra \$0.38 is the result of quarterly, rather than annual, compounding.

What would happen if we were to continue quarterly compounding for ten years? The answer can readily be computed, either with a hand calculator or by means of a spreadsheet model. By hand calculator the answer would be $N(40) = 100 \cdot (1+.025)^{40} = \$268.50$. As expected, this exceeds the amount one would receive with annual compounding.

Now let's see how we can include both of these examples (and more!) in a single spreadsheet model. To do this, we'll let f denote the *frequency of compounding* (we've already considered the cases $f = 1$ and $f = 4$) and set up the spreadsheet so that f can assume arbitrary positive values. In this model, the year will be divided into f equal parts at the end of each of which the bank will pay $(10/f)$% interest on your current deposit.

Figure 4 shows what the formulas for such a spreadsheet program might look like for $f = 4$. The "dollar signs" preceding the reference to cell B1 have the effect of "locking" this reference. That is, when the formulas in the table above are copied downward and A4 becomes A5, then A6, and so on, the \$ preceding the 1 in \$B\$1 will prevent the second index from changing. With this entry protected in this way, the result of continuing these formulas for 40 intervals (corresponding to 10 years) is as shown in figure 5.

If (like Excel 4.0) your spreadsheet has a graphing capability, you will also be able to generate graphs for various values of f (see figure 6).

The reason for placing the value "4" of f in a separate cell is that this way of entering data makes it easy to vary the frequency of compounding. Simply changing the entry in cell B1 from 4 to 10 leads Excel to redo its prior calculation and even redraw the corresponding graph!

Are we doing calculus?

Before turning from money to rabbits (and other populations), there is a curious fact that deserves our attention. The growth of money is usually discussed in terms of compound interest and calculated by the iterative methods described above. By contrast, the growth of rabbits tends to be discussed in terms of differential equations by people accustomed to using tools rooted in calculus rather than simple arithmetic. That is, the traditions of "population dynamics" are such that one would avoid modeling the growth of populations by the simple methods we have used for money. Instead of iterating

$$N(k+1) = (1 + 1/10)N(k);\ N(0) = 100$$

to obtain

$$N(k) = 100(1 + 0.1)^k N(0),$$

the growth of populations would tend to be modeled in terms of a differential equation

$$\frac{dN}{dt} = 0.1 \cdot N(t);\ N(0) = 100 \qquad (1)$$

whose calculus-based solution is

$$N(t) = 100e^{t/10}$$

and involves the Euler constant e = 2.71828... .

That these techniques are somewhat different is reflected by the fact that the compound interest approach yields

$$N(10) = 100(1 + 1/10)^{10} = 259.37,$$

while the calculus-based approach yields

$$N(10) = 100e^1 = 271.83.$$

The key to reconciling these two approaches is to consider a more frequent compounding of interest. For example, if the bank compounds interest quarterly, then $N(10) = 100 \cdot (1 + 0.025)^{40}$ = \$268.51, which is closer to the calculus-based answer of \$271.83 than that obtained by annual compounding. If the bank were to compound interest ten times a year for ten years, the answer would be given by $N(100) = 100 \cdot (1 + 0.01)^{100}$ = \$270.48, which is closer yet. Thus, we see that the calculus-based answer seems to correspond to a very frequent compounding of interest.

It is, in fact, easy to see the connection between compound interest and differential equations. If we agree to measure time in discrete units—

$$t = k, \text{ where } k = 0, 1, 2, 3, \ldots$$

—and replace the symbol dN/dt ("the rate of change of N with respect to t") with $N(k + 1) - N(k)$, then the differential equation (1) becomes

$$N(k+1) - N(k) = 0.1 \cdot N(k);\ N(0) = 100.$$

Transposing $N(k)$ to the right side, this *difference equation* becomes identical to the scheme used to calculate compound interest! The transition to calculus corresponds to very frequent compounding (for example, reducing the duration of time between successive compounding from a year to a quarter of a year, to a month, to a day, . . .).

In making such changes in f it's important to note that, while ten years of annual compounding requires only 10 iterations on our spreadsheet model, quarterly compounding requires 40 iterations. More generally, increasing the value of f will also increase correspondingly the length of the spreadsheet. This would be disconcerting except for one fact. At some point, increasing f will have very little effect on the outcomes. For example, increasing f from 10 to 100 or 1,000 would change the value of your deposit by a little more than a penny after 10 years. Once such a point is reached, we can be reasonably sure that our spreadsheet solution corresponds very closely to the calculus solution. Indeed, differential equations are frequently solved by the very techniques I've been describing.

How people grow

People are neither dollars nor rabbits, and to suggest that our demographic future is determined by a formula or differential equation is clearly nonsense. Differential equations provide good representations of *deterministic* processes. For example, given the initial position and velocity of a falling object, one can effectively use differential equations to predict its position at subsequent times.

By contrast, we have the capacity to influence our future numbers on Earth through a variety of individual and collective decision-making processes. That is, our demographic future involves *free will,* so using mathematics to model human population growth does not preclude our ability to achieve a future different from that predicted by a particular model. Rather, by enabling us to anticipate the outcomes of specific courses of action, mathematics may help us avoid the unpleasant future predicted by a particular model.

Having said this, the fact remains that human population growth is determined by human birth and death rates. Given specific estimates for these rates in the present and future, one can use techniques from population dynamics (the same ones we used with dollars and rabbits) to determine corresponding changes in human population.

Starting with the biblical "three score and ten" as an estimate for human longevity, it's reasonable to begin with an estimate of annual death rate of 1/70, or about 14.3 people per thousand. To estimate the annual birth rate, we note that human population is currently doubling every 40 years. This corresponds roughly to a bank account in which your initial deposit doubles every 40 years. With annual compounding, the interest rate required to achieve such returns can be determined by solving

$$200 = 100(1 + r)^{40}$$

for

$$r = 2^{1/40} - 1 = 0.0174797.$$

With very frequent (continuous) compounding, this rate would be determined by solving

$$200 = 100e^{40r}$$

for

$$r = (1/40) \ln (2) = 0.0173286.$$

Given the closeness of these answers, it seems reasonable to proceed on the basis of annual compounding and r = 0.0174, corresponding to an increase of 17.4 people per thousand per year. Recalling our prior estimate of a death rate of 14.3, this corresponds to approximately 31.7 live births per thousand people per year. Applied to the current human population of about 5.6 billion, these data correspond to an annual increase of 97.4 million people.

It's difficult to develop a sense of scale for such large numbers in and of themselves. One approach to grasping their implications is to use some 2,000-year-old mathematics to estimate the number of acres of cropland on Earth, and then to divide this number by $5.6 \cdot 10^9$. The Greek mathematician Eratosthenes established R = 4,000 miles as a very accurate estimate for the radius of the Earth (see problem 1 below). Also, Archimedes discovered the formula $S = 4\pi R^2$ for the surface area of a sphere. An order-of-magnitude calculation now yields $S = 12 \cdot 16 \cdot 10^6$, or about 200 million square miles. Noting that slightly more than 70% of the Earth is covered by water and assuming that 11% of the land is arable, we arrive at about 6 million square miles of cropland.

To complete this calculation, we need the fact that one square mile is equivalent to 640 acres, leading to a final back-of-the-envelope estimate of 640 · 6 million, or about 4 billion acres of cropland.

During the 1920s (when the world's human population stood at about 2 billion) the American humorist Will Rogers observed, "Real estate is a good investment because they aren't making it any more." The fact that there is now less than one acre of cropland per person on Earth provides one basis for thinking about the implications of human population growth for our collective future.

We can, of course, use a spreadsheet to represent these estimates and project them, say, 100 years into the future. The outcome is both stark and graphic, indicating that current trends are unlikely to be maintainable. However, such projections are not in themselves likely to encourage creative thought about actually addressing the underlying issues.

So what about free will? If we want to achieve a different future, one where human population stabilizes at some preassigned goal (say, 10 billion), this would require that the current rate of increase be reduced with time. Put mathematically, we would have to replace the constant $r = 0.0174$ with a function $r(t)$—one that satisfies $r(0) = 0.0174$ and decreases sufficiently quickly so as to achieve a leveling off of $N(t)$ at some preassigned value. One example of such a function is

$$r(t) = 0.0174/(1 + 0.001 \cdot t^2). \quad (2)$$

This would correspond to solving the differential equation

$$\frac{dN}{dt} = r(t) \cdot N(t);\ N(0) = 5.6.$$

and imposing the requirement

$$\lim_{t\to\infty} N(t) = 10.$$

Working backward from the solution to determine a function appearing in the differential equation itself is a problem from *control theory.*

In fact, however, we can use spreadsheets to approach this problem in another way. Consider depositing 5.6 units of money in a bank that pays a *variable* interest rate $r(k)$ during the kth time period in which interest is to be compounded. This isn't completely fanciful. Interest rates do change (though usually not by formula), and banks also make loans based on variable interest rates. In this context, it requires only a small change in our prior spreadsheet program to accommodate replacement of the parameter r with a function $r(k)$.

I. Population Futures Project

1. Define a function $r(k)$ for which $r(0) = 0.0174$ and $r(k)$ tends to zero as k becomes very large (for example, see equation (2)).
2. Develop a spreadsheet program for calculating the future value of $N(0)$ dollars invested at $r(k)$% per year and compounded annually.
3. Vary the function $r(k)$ until the solution corresponding to $N(0) = 5.6$ levels off near $N = 10$.
4. Refine your model to correspond to "more frequent compounding." Does $N(k)$ still approach 10? If not, redefine $r(k)$ and repeat your calculations.

The logistic equation

While reducing the growth rate provides a natural approach to stabilizing human population, there is another model that occupies a more prominent place in theoretical population dynamics. The logistic (or Verhulst) equation reflects the assumption that the growth of a particular population will be *damped* as it gets large, and that, unlike the effects of an annual death rate of 14.3 per thousand, such damping will be proportional to the *square* of the size of the population.

The mechanism for such damping is left to our imagination. Optimists may see it as reflecting restraint or economic forces, while pessimists may see famines, plague, and pestilence. The particular rule for damping (proportional to the square of N) appears to be more of a mathematical convenience than a representation of any particular mechanism.

An intuitive way to explore the resulting phenomena is again to formulate the underlying mechanism in terms of money. Consider a bank (let's call it Murky Savings & Loan) that offers you the attractive 10% rate of interest we have studied before. But Murky also imposes a "very small" service fee ("only 0.0005"), and it's applied to the square of the amount of money you have in your account. Including this fee in our model, we would now have

$$N(k+1) = N(k) + 0.1 \cdot N(k) - 0.0005 \cdot N(k)^2.$$

The spreadsheet program in figure 7 (on the next page) enables you to compute future balances based on annual deposits $N(0)$ and various frequencies of compounding. Problem 2 below will give you some insight into why you're unlikely to become very rich depositing your money in this bank.

In the more common (calculus-based) language of population dynamics, such models make use of the differential equation

$$\frac{dN}{dt} = a \cdot N(t) - b \cdot N(t)^2. \quad (3)$$

	A	B	C	D
1	f =	4	N(0) =	100
2	k	Time	N(k)	
3	0	=A3/$B$1	=$D$1	
4	=A3+1	=A4/$B$1	=C3+(0.1*C3-0.0005*C3^2)/$B$1	

Figure 7

Reading the symbol dN/dt as "rate of change," this equation corresponds to a population having a growth rate $a \cdot N(t)$ and subject to a damping factor of the form $-b \cdot N(t)^2$. As in all such problems, the constraints a and b correspond to a specific (but often unstated) unit of time. If $N(t)$ exceeds a certain critical value (see problem 2), then dN/dt becomes negative, even though a is positive.

"Putting off until tomorrow . . ."

At this point, our spreadsheet approach has some decided advantages over trying to solve the differential equation (3) directly, for one way of making the logistic model more realistic is to note that many of the factors likely to damp the growth of populations involve delays. That is, toxic substances take time to affect our health, while poor farming practices take time to deplete the soil. Thus, it is natural to ask, "What would happen if we introduce a *delay* of duration d units of time into the damping term $-b \cdot N(k)^2$?"

For the differential equation (3) this would correspond to solving

$$\frac{dN}{dt} = a \cdot N(t) - b \cdot N(t-d)^2.$$

Such *delay differential equations* have been studied only in the last 50 years, and there are no calculus-based procedures for obtaining a specific formula for a solution. Even with calculus as a tool, one has to resort to discrete methods, much like those we have used to represent the workings of Murky Savings & Loan.

However, in our spreadsheet context, delays are easy to conceive of and deal with. For instance, what would happen to an initial deposit of \$100 if Murky informed its long-standing customers that it plans to "reduce their service fee"? Specifically, if after 10 years of depositing your money with Murky, the service fee will be based not on the square of the current balance but rather on the square of the balance that existed d years ago?

II. Delayed Damping Project

1. Set up the system

$$N(k+1) = N(k) + 0.1N(k) - 0.0005N(k)^2;\ N(0) = 100$$

on a spreadsheet. Show that $N(k)$ never grows beyond 200.

2. Study the system

$$N(k+1) = N(k) + 0.1N(k) - 0.0005N(k-d)^2;\ N(0) = 100$$

for d = 4, 6, and 8.

Explain why the various outcomes in problem 2 might be expected in a population dynamics context.

More on differential equations

Before going on to another example, it seems appropriate to ask whether our banking analogies (which are getting somewhat far-fetched) are necessary for representing a differential equation on a spreadsheet. The answer is an emphatic *no*, and I'll now describe a different computational approach for dealing with differential equations.

As we've already noted, every differential equation involving dN/dt is formulated to a specific unit for t. The fact that the constants a and b in the logistic equation $dN/dt = aN(t) - bN(t)^2$ represent "change *per year*" implies that t is to be measured in *years*.

In modeling compound interest, we've already measured time in discrete units—

$$t = k, \text{ where } k = 0, 1, 2, \ldots$$

—and interpreted the symbol dN/dt as "the change of $N(t)$ in a unit of time." The same procedure applies to more general differential equations.

For example, replacing dN/dt with $N(k+1) - N(k)$, the logistic differential equation (3) becomes

$$N(k+1) - N(k) = aN(k) - bN(k)^2,$$

or

$$N(k+1) = (1+a)N(k) - bN(k)^2,$$

which corresponds to the iterative scheme underlying our spreadsheet model.

Here's another example. The differential equation

$$\frac{dN}{dt} = t^2 + \frac{1}{N(t)} \quad (4)$$

becomes

$$N(k+1) - N(k) = k^2 + 1/N(k),$$

or

$$N(k+1) = N(k) + k^2 + 1/N(k).$$

Such equations can also be iterated on a spreadsheet.

A word of caution (already noted in the case of compound interest) is called for when we set out to solve equation (4) for a specific range of values of t and then decide to increase the frequency of compounding. To solve

$$\frac{dN}{dt} = t^2 + \frac{1}{N(t)};\ N(0) = 100$$

for $0 \le t \le 10$ with annual compounding, we would program the spreadsheet to calculate

$$N(k+1) = N(k) + [k^2 + 1/N(k)];\quad N(0) = 100$$

for $k = 0, 1, \ldots, 9$. If, however, we want to compound more often (say, four times as often), the corresponding calculation would be

$$N(k+1) = N(k) + [k^2 + 1/N(k)]/4;\quad N(0) = 100$$

for $k = 0, 1, \ldots, 39$. In this second representation it's necessary to continue $N(k)$ down for 40 cells rather than 10.

In the examples considered so far, more frequent compounding didn't change the outcome very substantially, and for this reason compounding once in each unit of time was quite adequate. This may not always be the case, however, and for this reason one should always experiment with at least one refinement (corresponding to more frequent compounding) to ensure that the spreadsheet solution corresponds reasonably well to that of the differential equation. A good way of doing this is to regard the frequency of compounding as a parameter in the spreadsheet program.

Predator–prey interaction

With this brief introduction to translating differential equations into discrete (spreadsheet) form, we are ready to consider another interesting example from population dynamics—namely, the predator–prey systems introduced by Vito Volterra (see "The World according to Malthus and Volterra" in the July/August 1992 issue of *Quantum*). Here it is the interaction of *two* populations that is to be modeled.

Consider a population of rabbits and foxes on a grassy island. Left to themselves, the rabbits would increase at 10% a month. Because they do not thrive on grass, foxes left to themselves would decline at 10% a month. In fact, however, these two species interact by occasional meetings, events that are generally bad for the rabbit and good for the fox. The question addressed by Volterra is how these populations would fare together.

Ignoring for the moment the interaction of rabbits and foxes, we begin to model such populations by letting X denote the number of rabbits, Y the number of foxes, and setting

$$\frac{dX}{dt} = aX(t); \frac{dY}{dt} = -cY(t)$$

with $a = c = 1/10$. These can, of course, be translated into spreadsheet models by setting $dX/dt = X(k + 1) - X(k)$, and so on.

In order to include "interaction" in this model, Volterra needed to quantify the number of meetings between rabbits and foxes. Here he reasoned that tripling the number of foxes should triple the number of meetings. Likewise, halving the number of rabbits should halve the number of meetings. This requires that the number of meetings (in a unit of time starting at t_0) be proportional to $X(t_0) \cdot Y(t_0)$. Introducing constants of proportionality b and d relative to this unit of time, Volterra was led to the set of differential equations

$$\frac{dX}{dt} = aX(t) - bX(t) \cdot Y(t),$$

$$\frac{dY}{dt} = -cY(t) + dX(t) \cdot Y(t)$$

(see the aforementioned *Quantum* article). The sign preceding the second

	A	B	C	D
1	X(0) =	110	Y(0)=	100
2	a =	0.1	b =	0.001
3	c =	0.1	d =	0.001
4	f =	1		
5	k	Time	X(k)	Y(k)
6	0	=A6/$B$4	=$B$1	=$D$1
7	=A6+1	=A7/$B$4	=C6+($B$2*C6-$D$2*C6*D6)/$B$4	=D6+(-$B$3*D6+$D$3*C6*D6)/$B$4

Figure 8

term on the right side of each equation reflects the effect of meetings on the two populations—that is, it's a form of interaction that's "bad for the number of rabbits and good for the number of foxes."

To represent these differential equations on a spreadsheet, we write $dX/dt = X(k+1) - X(k)$, and so on, to obtain

$$X(k+1) = X(k) + [a \cdot X(k) - b \cdot X(k) \cdot Y(k)]/f,$$
$$Y(k+1) = Y(k) + [-c \cdot Y(k) + d \cdot X(k) \cdot Y(k)]/f.$$

Here f can be thought of as denoting the number of times rabbits and foxes are to be counted during each unit of time (in our case, each month). Figure 8 shows a versatile way of programming such systems on a spreadsheet.

If we now enter $X(0) = 110$, $Y(0) = 100$, $a = c = 1/10$, $b = d = 1/1000$, and $f = 1$, we obtain a graph of the numerical outcomes (fig. 9). Increasing f from 1 to 10 leads to the graph shown in figure 10. Note that the two graphs are similar, but that the growth in amplitudes has been reduced by increasing f.

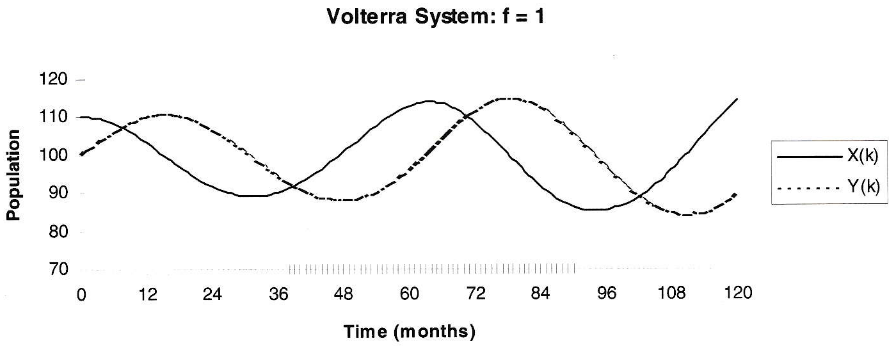


Figure 9

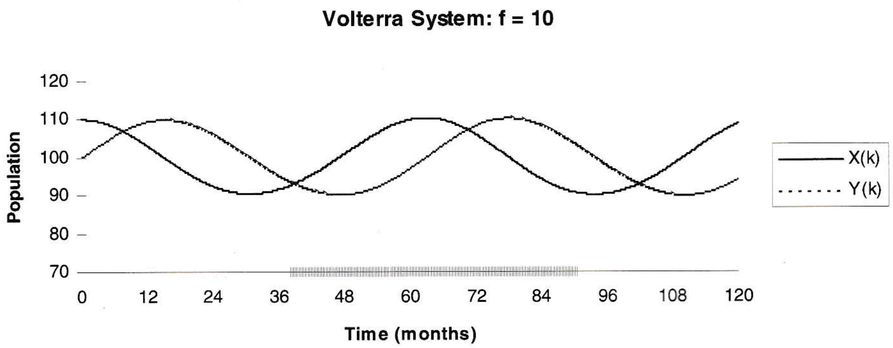


Figure 10

III. Harvesting Project

Robinson Crusoe has landed on this island and joins the foxes in the occasional hunting of rabbits. Use a spreadsheet to study the effects of harvesting an additional h rabbits per month over an extended period of time. Are there any surprising results?

Problems

1. The city of Syene was known to Eratosthenes as a place where the Sun shines directly to the bottom of a well at noon on midsummer day. In Alexandria, 500 miles to the north of Syene, the Sun makes an angle of 7.5° with a plumb line at noon on midsummer day. Eratosthenes also knew that the Earth is spherical and that light from the Sun forms parallel rays. How might he have calculated the radius of the Earth to be (about) 4,000 miles?

2. In the logistic equation, your money is increasing or decreasing at time t depending on whether $0.1N(t) - 0.0005N(t)^2$ is positive or negative. For what values of N is $N(t)$ increasing? For what values of N is $N(t)$ decreasing? Are the answers to these questions consistent with the outcomes of your model for Murky Savings & Loan?

ANSWERS IN THE NEXT ISSUE

The cover artist decided to graphically depict a binary code and chose three letters "at random" (actually, according to his refined aesthetic judgment) to color red in the ship's name: C, R, C. *Quantum* readers who use communications software may have come across references to "CRC" error detection. According to a recent software manual, "16-bit Cyclic Redundancy Check . . . provides a 99.9984% assurance of detecting any transmission errors. With the checksum method, it is possible for 1 out of 256 bad packets to have a valid checksum and thus go undetected. With the CRC method only 1 out of 700 billion bad packets will generate a valid CRC." Norbert Wiener would have been pleased with the progress being made.

HOW DO YOU FIGURE?

Challenges in physics and math

Math

M126

Equation in reciprocals. Prove that the equation $1/x - 1/y = 1/n$ in positive integers x and y has a unique solution for a given positive integer n if and only if n is a prime number. (A. Danielyan)

M127

Circular arrangement. N numbers are written around a circle such that each of them is obtained from its counterclockwise neighbor either by adding 1 or by reversing the sign. Prove that if N is odd, then all the numbers are integers and any number m is found among them as many times as $-m$. (A. Veselov)

M128

Back after further reflection. Points A, B, C, and D are chosen in the plane with $AB = BC = CD = 1$. The four points are repeatedly subjected to the following transformation that leaves points B and C fixed and preserves the lengths of AB, BC, CD, and DA. First, point A is reflected about BD; then D is reflected about AC, where A is in the new, reflected position; then the new point A is again reflected about BD (with the new D); then D is reflected; and so on. Prove that after a number of reflections points A and C will return to their starting positions (M. Kontsevich)

M129

Weird minority. Call a person with fewer than 10 acquaintances *unsociable* and a person all of whose acquaintances are unsociable a *weirdo.* Assume that the relation of "acquaintanceship" is *symmetric*—that is, if person X is acquainted with person Y, then person Y must also be acquainted with person X. Prove that the number of weirdos is smaller than the number of unsociable persons. (F. Nazarov)

M130

A circle here, a circle there. The opposite sides of a convex quadrilateral are extended to intersect at two points. A line is drawn through each of these points. These two lines divide the quadrilateral into four smaller quadrilaterals. If some pair of these quadrilaterals that don't share a common side each has an inscribed circle, show that the original quadrilateral also has an inscribed circle. (I. Sharygin)

Physics

P126

Long train. A train moving under its own momentum goes up an incline of angle α. When the train stops, one half of its length is on the incline (fig. 1).

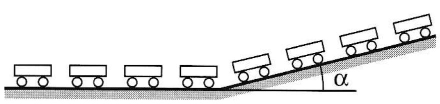


Figure 1

What time elapses from the moment the train begins to go up the incline until the moment it stops? The length of the train is L. Disregard the friction between the train's wheels and the incline. (A. Buzdin)

P127

Almost a matryoshka. A great number of thin-walled cylindrical vessels of water are submerged one inside another such that each subsequent vessel floats in the preceding one. The area S_0 of the bottom of the smallest vessel is far less than that of the largest vessel. Additional water of volume V_0 is added to the smallest vessel. How much does this vessel sink relative to the ground? (All the vessels continue to float after the water is added.) (S. Krotov)

P128

Ice on the lake. During a cold night in autumn, ice begins to form on the undisturbed surface of a very deep lake, and after 10 hours it is 10 cm thick. How thick will the ice be if the temperature doesn't change for 1,000 hours? Consider the thermal conductivity of ice to be far greater than that of water. (V. Skorovarov)

P129

Just a spoiled grid. Several conductors are removed from an infinite square conducting grid (fig. 2). The

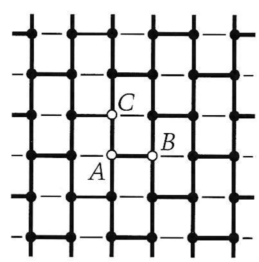


Figure 2

resistance of each rib is r. Estimate the resistance between the nodes A and B; B and C; A and C. (S. Krotov)

P130

Don't shake. A point of light is photographed from a distance of 1 m. The photographer's hand shook, and instead of a point, a small spot appeared on the film. Estimate the size of this spot if the amplitude of the displacement of any point of the camera does not exceed 1 mm. The focal length of the objective is 50 mm. The exposure time is very long. (A. Zilberman)

ANSWERS, HINTS & SOLUTIONS ON PAGE 57

Beyond the reach of Ohm's law

Where many useful phenomena are free to take place

by Sergey Murzin, Mikhail Trunin, and Dmitry Shovkun

IN MOMENTS OF BRAVADO or defiance, we say that "laws are made to be broken," but that's not how we act, now, is it? Especially when we're taking about physical laws. Take, for instance, Ohm's law, which tells us that current is proportional to voltage. Nature obeys the law, right? Well, not always. Which is just fine. If this law were observed at all times, we would be deprived of many useful electrical devices. Fortunately, Ohm's law, like many other physical laws, has a limited range of applicability. Outside this range, interesting physical phenomena arise that underlie the operation of these devices. Although these phenomena are interesting in their own right, in this article we'll address another question: why is Ohm's law violated?

Ohm's law

Let's connect a conductor in an electric circuit and measure the electric current I for different values of the applied voltage V. In this way we obtain the dependence $I = I(V)$—that is, the current–voltage relationship for the conductor. According to Ohm's law, electric current I is directly proportional to the applied voltage, which means that the current–voltage relationship is a linear function:

$$I(V) = \frac{V}{R},$$

where the resistance R doesn't depend on V. If this is not so, Ohm's law isn't obeyed and the current–voltage relationship is nonlinear.

The simplest example of a conductor in which Ohm's law doesn't hold is the filament in a light bulb. The current–voltage relationship for a 40-W bulb is shown in figure 1. The graph is linear only for $V < 5$ V; at higher voltages the current increases more slowly than a linear function. It's not hard to guess why: as the voltage is increased, the filament heats up and its resistance increases. This example illustrates the general rule: Ohm's law is valid only for sufficiently small values of I and V, and it's broken for large values.

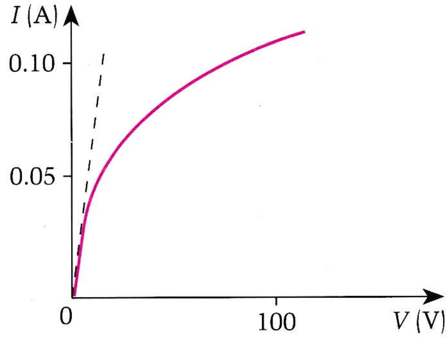


Figure 1

Let's write Ohm's law in another form. To do this, we introduce the current density $j = I/S$, where S is the cross-sectional area of the conductor. Then

$$j = \frac{I}{S} = \frac{V}{RS} = \frac{V}{(\rho L/S)S} = \frac{1}{\rho}\frac{V}{L} = \sigma E.$$

Here ρ is the resistivity of the conductor, the inverse value $\sigma = 1/\rho$ is called the conductivity, L is the length of the conductor, and $E = V/L$ is the strength of the electric field. Ohm's law implies a linear relationship between the current density j and the electric field E. If for any reason the conductivity σ varies with the electric field, the dependence of j on E becomes nonlinear and Ohm's law is broken. To find out why Ohm's law is broken, let's look at the movement of electrons in the conductors both with and without the presence of an electric field.

Art by Dmitry Krymov


Крымов

How electrons move in a conductor

Many substances that conduct electric current are crystals. Their atoms aren't located arbitrarily, but form a periodic spatial structure—a crystal lattice. In conductors some of the atoms are ionized, and the electrons that have been given up can move through them. The concentration n of such electrons (known as conduction electrons) depends on the type of conductor. In metals the concentration of conduction electrons doesn't depend on temperature. For example, in copper $n = 8.4 \cdot 10^{28}\ \text{m}^{-3}$. On the other hand, in semiconductors n varies with temperature. In germanium $n = 2.4 \cdot 10^{19}\ \text{m}^{-3}$ at 300 K.

It may seem that the electron has a hard time squeezing through the crystal—that it's continually running into atoms. But that's not the case at all. Quantum mechanics tells us that because of the strictly periodic arrangement of the atoms, the electrons move in a straight path through the ideal lattice. In this respect the conduction electrons are like free electrons in a vacuum. And just as with electrons in vacuum, the movement of the conduction electrons can be described by Newton's second law $F = m^*a$, but now m^* (the "effective mass") differs from the mass m_e of an electron in a vacuum. This distinction reflects the interaction of conduction electrons with the crystal lattice. Insofar as the lattice structure varies in different conductors, the effective masses m^* of the electrons also vary. They can be either larger or smaller than m_e.

Real conductors are never ideal crystals. They always contain disturbances in the periodic arrangement of their atoms. For example, atoms of a foreign substance can find their way into some positions in the lattice. When the conduction electrons come upon such an impurity, they scatter—that is, they change their paths. The thermal oscillations of atoms in the lattice (that is, their deviations from the equilibrium position) disturb the periodicity, which also causes scattering of the electrons. The mean time between collisions, during which an electron moves in a straight path, is called the mean free time τ. This value depends on the electron's velocity.

In the absence of an electric field the conduction electrons move in different directions, which leads to chaotic thermal motion. In semiconductors the movement of electrons is similar to the thermal motion of the molecules of an ideal gas. The mean velocity v_0 of such motion can be estimated from the condition $m^*v_0^2/2 = kT_e$, where k is the Boltzmann constant and T_e is the electron temperature. In gallium arsenide (GaAs) at $T_e \cong 300$ K, $v_0 \cong 4.5 \cdot 10^5$ m/s.

In metals, which have conduction electrons in much higher concentration than in semiconductors, conclusions drawn from the molecular-kinetic theory of gases can't be applied. Quantum theory tells us that the average velocity of the chaotic motion of electrons in metals is $v_0 \cong 10^6$ m/s and is practically independent of temperature.

Now let's see what happens when an electric field $\mathbf{E}$ is applied. The force $-e\mathbf{E}$ acting on the electron imparts an acceleration $\mathbf{a} = -e\mathbf{E}/m^*$. Let's denote the velocity of the ith electron immediately after scattering as $\mathbf{v}_i$. At an arbitrary moment the velocity of the ith electron will be $\mathbf{v}_i - e\mathbf{E}t_i/m^*$, where t_i is the time that has passed since the last collision. The mean velocity of N electrons is

$$\mathbf{u} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{v}_i - \frac{e\mathbf{E}}{m^*}t_i\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i - \frac{e\mathbf{E}}{m^*}\left(\frac{1}{N}\sum_{i=1}^{N}t_i\right).$$

The value $\frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i$ is the mean electron velocity immediately after scattering. Since the velocity immediately after a collision can be directed anywhere, we get $\frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i = 0$. The value $\frac{1}{N}\sum_{i=1}^{N}t_i = \tau$ is similar to the mean free time encountered above. Thus, under the action of an electric field, all the electrons acquire an additional velocity ("drift velocity") whose mean value is equal to $u = eE\tau/m^*$ and whose vector is directed parallel to the field $\mathbf{E}$.

Thus, in an electric field a drift is added to the chaotic motion, resulting in the appearance of a dominant direction of electron motion—that is, an electric current. If the concentration of electrons in the conductor is n, the current density is

$$j = enu = \frac{ne^2\tau}{m^*}E.$$

On the other hand, we know that $j = \sigma E$. Therefore,

$$\sigma = \frac{ne^2\tau}{m^*}.$$

This is known as Drude's formula. Clearly Ohm's law is valid if none of the physical values in this formula depends on E. If the electron concentration n, the mean free time τ, or the effective mass m^* varies with the electric field, Ohm's law is no longer valid.

When is Ohm's law valid?

First of all, let's consider the conditions when the value τ doesn't vary with the electric field E.

The time τ depends on the velocity of the electrons. The drift velocity $u = eE\tau/m^*$ that arises in an electric field increases with E. When the electric field is weak and the drift velocity u is much less than the mean velocity of chaotic motion v_0, the value u can be neglected and τ can be considered independent of the field strength E. On the other hand, when E is so high that u is comparable to v_0, the drift velocity must be taken into account. In this case it turns out that the electron velocity and hence the mean free time τ depend on the electric field.

Thus, for Ohm's law to be valid, the following condition is necessary:

$$u \ll v_0 \qquad (1)$$

—that is, the strength of the electric field in the conductor must be far less than $E = m^*v_0/e\tau$.

As we said above, in semiconductors $v_0 \sim 10^5$ m/s. To obtain a value of u comparable to v_0, an electric field $E \sim 10^6$ V/m must be applied. This is a very high value, comparable to the voltage in a lightning bolt. Nevertheless, such a field can be created in semiconductors.

There is another, even more severe limitation on the velocity u: it must be less than the speed of sound in the conductor ($v_s \sim 10^3$ m/s):

$$u < v_s. \quad (2)$$

When the velocity u reaches the value v_s, sound oscillations arise in the crystal. This may lead to a decrease in the mean free time τ and the conductivity σ, which is proportional to τ. This situation is analogous to the drastic increase in aerodynamic drag that occurs when an airplane breaks the sound barrier. Thus, when $E_s \geqslant m^*v_s/e\tau$, the conductivity begins to vary with E, resulting in a violation of Ohm's law.

The effects of the electric field are not confined to the emergence of drift motion. Current flowing in a conductor releases Joule heat, and the conductor heats up. Let's consider this process in detail.

Any conductor can be thought of as being composed of two subsystems: (1) the crystal lattice of atoms and (2) the gas of conduction electrons that fills this lattice. Both electrons and lattice can be characterized by their own temperatures T_e and T_l. In the absence of an electric field, the electron gas is in thermal equilibrium with the lattice and the surroundings (T_s): $T_e = T_l = T_s$. The electric field E acts on the conduction electrons and heats them first. Only then is the heat transmitted to the lattice and finally to the surroundings. Thus, when an electric field is present, the thermal equilibrium is disturbed such that $T_e > T_l > T_s$.

When heat transfer from a conductor to the surroundings is less than that from electrons to atoms and hence $T_e - T_l \ll T_l - T_s$, the lattice and the electrons are heated as a whole. (This takes place in the filament of an electric light bulb.) The opposite case is possible, too, where the electron temperature is far greater than that of the lattice: $T_e - T_l \gg T_l - T_s$.

As we mentioned previously, in metals the average velocity of the chaotic thermal motion of the electrons is practically independent of temperature. In semiconductors, however, the increase in T_e induced by an electric field leads to an increase in the electron thermal velocity v_0—that is, a decrease in the mean free time. If the change Δv_0 in the velocity v_0 is small—that is, $\Delta v_0 \ll v_0$—the dependence of v_0 on E and hence of τ on E can be neglected. The condition $\Delta v_0 \ll v_0$ is equivalent to the condition that the heating of the electrons ΔT_e be small relative to the equilibrium state:

$$\Delta T_e \ll T_e. \quad (3)$$

Thus, the condition that the mean free time be independent of the field strength, which is necessary for Ohm's law to be valid, leads to the following limitations on the applicability of this law:

$$u \ll v_0, \quad (1)$$
$$u < v_s, \quad (2)$$
$$\Delta T_e \ll T_e = T_s. \quad (3)$$

A violation of any of these inequalities may lead to deviations from Ohm's law. We'll see later that when the inequalities $u \ll v_0$ and $\Delta T_e \ll T_e$ are violated, the electric field E can also affect other values in Drude's formula: the effective mass m^* and the electron concentration n. The dependence of m^* and n on E can drastically change the current–voltage characteristics in semiconductors.

Semiconductors in strong electric fields

The power released in a sample where current I flows is given by

$$P = I^2R = \sigma E^2 LS.$$

Here we took into account that $I = jS = \sigma ES$, $R = \rho L/S = L/\sigma S$. The power released per unit volume is $Q = \sigma E^2$. With the same value of Q, the electric field $E = \sqrt{Q/\sigma}$ is much stronger in semiconductors than in metals, because the electron concentration in semiconductors and hence the conductivity σ are far less. Consequently, in semiconductors the conditions $u < v_s$ and $u \ll v_0$ are violated much more readily than in metals. In addition, the power per electron is higher in semiconductors than in metals. The electron gas is heated more, so the inequality $\Delta T_e \ll T_e$ is also violated more easily in them.

Which violation— of condition (1), (2), or (3)—leads to a stronger deviation from Ohm's law as the electric field increases depends on the type of semiconductor. For instance, in CdS the condition $u < v_s$ is violated first. When this takes place, a break appears in the current–voltage relationship $j(E)$ at $E_s = 1.4 \cdot 10^5$ V/m (fig. 2).

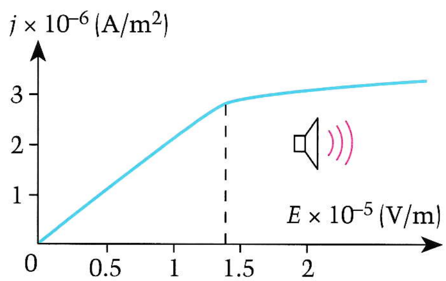


Figure 2

When $E > E_s$, the semiconductor emits an intense sound and so can be used as a sound generator. In other semiconductors, such as Ge, Si, GaAs, InP, and CdTe, the generated sound is much weaker and there is no marked break in the current–voltage relationship at E_s. In these semiconductors the deviations from Ohm's law are caused by a violation of the condition $\Delta T_e \ll T_e$. It turns out that the mean free time is inversely proportional to the field E—that is, $\tau(E) \sim 1/E$—and the dependence of current density on field strength is connected only with the changes in m^* and n. At $E > 10^6$ V/m the current–voltage curves of Ge and Si (see figure 3 on the next page) show saturation of j (m^* and n do not depend on E). There is not only a decrease with E of the mean free time in GaAs, InP, and CdTe, but also an increase in the effective mass m^*. An increase in m^* is caused by the change in the interaction of the electrons with the crystal lattice. As a result, the current density j decreases with E in these semiconductors,

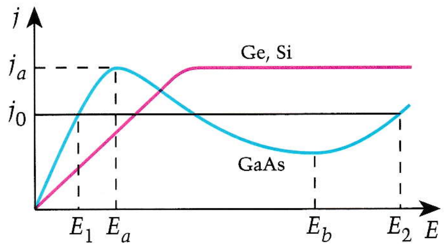


Figure 3

beginning with a certain value of the field strength E_a (the segment $E_a < E < E_b$ in figure 3). In GaAs the drop in j begins at $E_a = 3.2 \cdot 10^5$ V/m and continues up to $E_b \cong 10E_a$. The electron drift velocity for a field strength E_a is $u = j/en = 1.5 \cdot 10^5$ m/s.

When the field becomes even stronger—that is, at $E \sim 10^7$ V/m—not only is the condition $\Delta T_e \ll T_e$ violated, the condition $u \ll v_0$ is violated as well. In such a field the electrons, during the mean free time, receive enough energy to ionize atoms. When they collide with atoms, the high-speed electrons knock out extra electrons, which are also accelerated by the field and produce other charge carriers. This effect is known as ionization by collision. The total electron concentration n increases and, consequently, the conductivity increases. A further increase in the field strength ($E > 10^7$ V/m) leads to an avalanche increase in the concentration and conductivity, causing a breakdown in the semiconductor.

Thus, the current density $j = \sigma E$ increases in semiconductors in very strong electric fields more quickly than a linear function. In particular, in Ge and Si the saturation of current is replaced with a nonlinear increase, and for GaAs, InP, and CdTe the current–voltage curves become N-shaped (see figure 3): at $0 < E < E_a$, Ohm's law is valid; in the interval $E_a < E < E_b$ there is a falling portion of the curve, caused by a decrease in τ and increase in m^* in a strong electric field; and finally, in the region $E > E_b$ a rapid growth of j occurs, caused by the increase in n.

The Gunn effect

The falling portion in the current–voltage relationship underlies an interesting phenomenon found by the American engineer John Gunn. Let's apply a voltage V_0 to a GaAs sample of length L that produces the falling portion of the dependence $j(E)$. The initial electric field within the sample is assumed to be uniform and equal to V_0/L. Let the electric field E in a thin layer AB slightly exceed for some reason the field strength in other areas of the sample (fig. 4). The electron drift

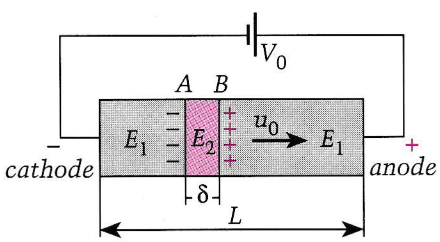


Figure 4

velocity $u = j/en$ within this layer AB will be less than that outside it. Thus, more electrons will move toward the boundary A than will move away from it, and the opposite will take place at the boundary B. An extra negative charge accumulates at boundary A, while a positive charge accumulates at boundary B. Consequently, an extra electric field arises in the layer AB that points in the same direction as the original field. This increase in the electric field results in a further decrease in the electron drift velocity, which causes a still greater increase in the electric field in this area.

Thus, it is impossible to have a uniform electric field in the falling portion of $j(E)$: any accidental disturbance of E, however small, will not disappear—on the contrary, it will grow. As a result, a narrow region (of thickness δ) with a strong electric field arises, which is called an electric domain. Since the voltage applied to the sample is constant,

$$E_2\delta + E_1(L-\delta) = V_0 = \text{const},$$

the increase in the field E_2 in the domain is accompanied by a decrease in the field E_1 outside it. A moment will come when $E_1 < E_a$ and $E_2 > E_b$ (see figure 3). The electron drift velocity outside the domain begins to decrease, while inside it increases. The increase in the field E_2 in the domain stops when these velocities are equal and the current densities in the domain and in the sample are equal:

$$j(E_1) = j(E_2) = j_0.$$

From the last two equations it follows that the steady-state current density j_0 in the sample depends on the domain's thickness δ.

Usually an electric domain arises near the cathode (there are more impurities in this region due to the soldered contacts), and then, carried off by the electron flow, it begins to move to the anode with a velocity $u_0 = j_0/en$. While it moves through the sample its size doesn't change, and so the current j_0 also remains constant. However, the domain begins to disappear near the anode—its thickness decreases and the current in the sample increases. Simultaneously, the electric field E_1 increases outside the domain. As soon as E_1 reaches the value E_a, a new domain arises near the cathode, the current begins to decrease, and the whole process starts all over again (fig. 5). The period of the current oscillations in the sample is $T_0 = L/u_0$.

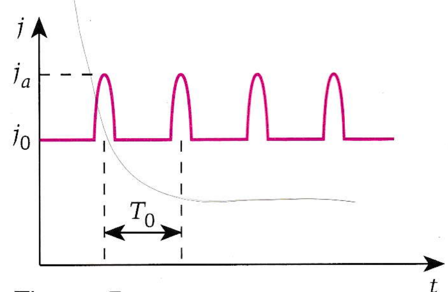


Figure 5

Thus, applying a constant voltage V_0 to the sample induces an alternating current with a frequency $f = 1/T_0 = u_0/L$. This doesn't look much like Ohm's law. In gallium arsenide (GaAs), $u_0 \cong 10^5$ m/s. Using short samples (1–100 microns), one can obtain frequencies in the range $f \sim 10^9$–10^{11} Hz. Most modern generators of ultrahigh frequencies (UHF) are based on the Gunn effect. These devices are used, for example, in television transmission via satellite and by the police to measure the speeds of cars.

Violations with large currents

Up to now we've looked at the motion of electrons under the influence of an electric field only. We know, however, that current flowing in a conductor is the source of a magnetic field. This magnetic field arises not only outside but also inside the conductor. For example, near the surface of a straight wire of diameter $d = 1$ mm with current $I = 10$ A flowing through it, a magnetic field $B = \mu_0 I/\pi d \cong 0.012$ T arises ($\mu_0 = 4\pi \cdot 10^{-7}$ N/A^2 is the permeability of free space). The current-induced magnetic field also can be responsible for violations of Ohm's law.

An electron moving in a magnetic field is affected by the Lorentz force, which bends its trajectory. If the magnetic field **B** is perpendicular to the electron velocity **v**, the electron's trajectory is a circle of radius $r = m^*v/eB$. When the angle between vectors **B** and **v** is α, the electron moves along a spiral of diameter $d = 2(m^*v/eB)$, making one turn of the spiral in a time $T = 2\pi m^*/eB$.

Spiral motion of electrons in a conductor is possible if the mean free time $\tau \gg T$ (fig. 6a). When this occurs,

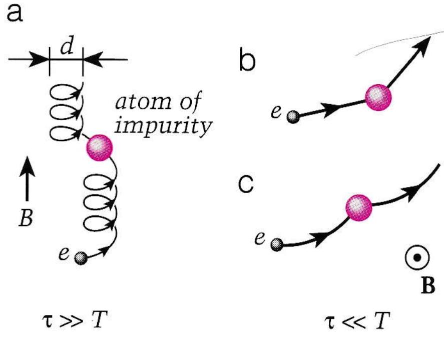


Figure 6

the spiral's diameter $d < vT$ is much smaller than the distance $l = v\tau$, which corresponds to the displacement of an electron in time τ when there is no magnetic field (fig. 6b). Thus, during time τ the electron moves as if trapped in a pipe of diameter d. As a result, the conductor's resistance in the magnetic field is greater than it is when $B = 0$. The dependence of the resistance R on the magnetic field created by its "own" current thus leads to a violation of Ohm's law.

If $\tau \ll T$, however, the electron's path between two sequential collisions doesn't deviate substantially from a straight line (fig. 6c). In this case the magnetic field doesn't change the resistance of the conductor by very much.

The magnetic field B_0, which marks the beginning of an appreciable influence of the field, can be found from the condition that the period $T = 2\pi m^*/eB_0$ of the electron's motion along the circular orbit be equal to the mean free time τ: $B_0 = 2\pi m^*/e\tau$. For GaAs, $m^* = 0.06m_e$ and $\tau \sim 10^{-13}$ s, so $B_0 \cong 3$ T. It's practically impossible to induce such a field by passing electric current through a semiconductor. The sample will be destroyed by far smaller currents. On the other hand, considerably larger currents can pass through metals. In addition, in pure metals cooled to the temperature of liquid helium (about 4 K), the mean free time τ can reach values $\sim 10^{-9}$ s, which is significantly higher than that in semiconductors. Therefore, the field B_0 is rather small in metals—about 0.01 T. Such a field arises in a wire of diameter 1 mm with a current of 10 A.

Figure 7 shows the experimental dependence of the resistance R of a metal conductor on the current I at the temperature of liquid helium. It's clear that the resistance increases with current by a factor of more than two. The electric field in this experiment was less than 10^2 V/m, which is far less than the fields in semiconductors where Ohm's law was violated. Thus, in this case the influence of the magnetic field on the resistance is the main reason that Ohm's law is violated.

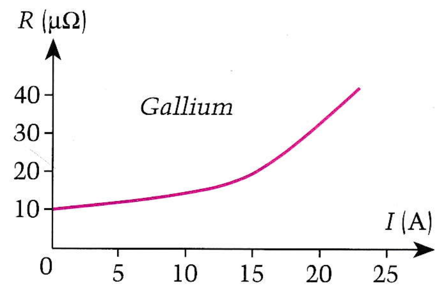


Figure 7

In this article we looked at physical phenomena that may lead to a violation of Ohm's law in semiconductors. In so doing we didn't mention the nonlinear devices—diodes and transistors—that underlie the modern electronic world. These devices are constructed in a special way to be nonhomogeneous, and Ohm's law is violated at points of contact between different conducting materials. In addition, we paid no attention to the many nonlinear effects that arise in conductors placed in alternating electric and magnetic fields. But if you try to do everything at once, nothing gets done. Perhaps others will pick up where we left off.

Panting dogs, aromatic blooms, and tea in a saucer

Tales of evaporation in the natural world

by Andrey Korzhuyev

AS YOU KNOW, EVAPORATION occurs when the fastest molecules in a liquid, after making their way to the surface, overcome the attraction of the adjacent molecules and leave the liquid. We know that the rate of evaporation depends on many factors: the nature of the liquid, its temperature, its surface area, where it's kept (in a closed or an open vessel), and so on. Also, the more intense the evaporation, the faster the liquid cools, because the most energetic ("hot") molecules are leaving it.

Evaporation is a widespread phenomenon in nature and plays an important role in the world of animals and plants. Let's look at some examples and try to explain them.

As you know, when a drought begins, the leaves of most plants start to wither. Why? Obviously, because of a lack of water in the soil. During a drought the supply of water to the roots decreases, so the plants tell their leaves (or so it seems) to roll up. This decreases their exposed surface area and consequently reduces the amount of water lost through evaporation. Likewise many desert plants are bereft of leaves and have only ugly spikes instead. That's how plants cherish the precious moisture.

What if it's necessary to increase the rate of evaporation? Why, just increase the surface area as much as possible. That's what we do when we pour hot tea from a cup into a saucer, or cut potatoes, apples, and other fruits and vegetables into small pieces when we want to dry them.

Another factor that affects the rate of evaporation is air motion over the surface of the liquid—in plain language, the wind. So, when we want to cool off our tea, we create an artificial wind—we blow on the surface. For the same reason, cut grass in a meadow dries quicker than in a forest (although, truth to say, another factor is also at play—evaporation is increased by the Sun's rays, which bathe the field but hardly penetrate the forest canopy). On the other hand, the leaves of many desert plants are covered with small, thick hairs that hinder air circulation near the leaf's surface and slow the process of evaporation.

Let's turn to another interesting phenomenon. Have you noticed how much stronger the aroma of flowers is after it rains? To explain this we need to remember that the aromas are caused by volatile oils produced in the flowers' nectaries. Water-free volatile oils evaporate more slowly than a mixture of the oils with water, which drops into the calyxes in abundance when it rains and enters the nectaries. Rapid evaporation of this mixture increases the aroma.

As I said earlier, evaporation decreases the temperature of the body from which the liquid evaporates. So if you tear off a leaf from a tree and put it against your face, you feel a pleasant coolness. Again, this is due to the intense evaporation of water from the leaf. You can notice the same effect when you go swimming—you'll cool off in hot weather if you don't immediately use a towel.

Did you ever wonder what temperature a human being can endure? Well, experiments have shown that, by gradually increasing the temperature of dry air, humans can endure temperatures up to 160°C. How can we explain it? After all, a change in temperature of only 1°C can be perceived as quite unpleasant. As it turns out, the temperature of the body itself changes only slightly.

Art by Pavel Chernusky

The organism resists being heated by perspiring profusely. Evaporation of sweat requires a great deal of heat, which is absorbed from the layer of air adjacent to the skin. After the heat is drawn off, this layer cools off. However, the air must be dry enough for this process to take place. If it contains a lot of moisture, the process of evaporation will go very slowly, and a human being wouldn't be able to bear the heat. This is why "humid heat" in St. Petersburg, for example, is much more uncomfortable than the "dry heat" in Central Asia.

Dogs, on the other hand, adapt to heat very poorly. They have sweat glands only on the pads of their paws. That's why a dog opens its mouth and lets its tongue hang out—evaporation of the saliva decreases its body temperature. Other mechanisms of heat transfer based on thermal conductivity and convection are also available to dogs: they can stretch out their legs, for instance, or lie on their backs to expose their bellies, which generally have a thinner coat of hair.

Finally, let's talk about why cold air is easier to endure when the wind isn't blowing. The feeling of cold is caused by the fact that the exposed parts of the body lose a great deal of heat in windy weather, because the air that is warmed by them is quickly replaced by more cold air, which in turn takes heat away. The stronger the wind, the quicker this replacement. However, evaporation also plays a role here. Moisture evaporates from the surface of our skin even in cold weather. If there is no wind, the evaporation is slow, because the layer of air near the skin gradually becomes saturated with vapor. If it's windy (or if you're moving quickly through still air), new packets of air come in contact with the skin, resulting in evaporation and cooling. So in windy weather, do yourself a favor and stay home!

KALEIDOSCOPE

The medians

A good tale is none the worse for being told twice

by Vladimir Dubrovsky

THE ARTICLES THAT FORM the mathematical installments of *Quantum*'s Kaleidoscope are kaleidoscopic in many different ways. There have been kaleidoscopes of problems, and kaleidoscopes of facts, and even a kaleidoscope of kaleidoscopic ornaments. This time we'll present a kaleidoscope of proofs of the same—and very well-known—theorem: the property of the medians of a triangle. We'll tell this "tale of a proof" not twice, but seven times! (Wise people say this is often more instructive than to give a number of different facts with a single proof for each.)

Each proof will be based on its own important and useful idea and will be accompanied by other applications of this idea that include several generalizations of the main theorem. But let me remind you of the statement we'll discuss—however well known it is.

THEOREM ON MEDIANS (fig. 1). *The medians AA_1, BB_1, and CC_1 of a triangle ABC meet at a certain point M and are divided by this point in the same ratio 2 : 1, counting from the vertices—$AM : MA_1 = BM : MB_1 = CM : MC_1 = 2$.*

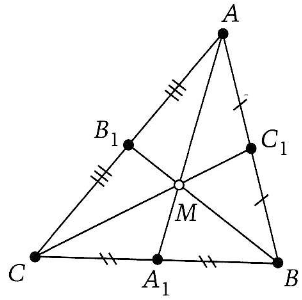


Figure 1

All our proofs, except the fourth and sixth, will follow the same scheme: we take the point M on the median AA_1 that divides it in the ratio $AM : MA_1 = 2$, and prove only that the line BM bisects AC. Replacing the line BM in any of these arguments with CM, we'll show that the third median CC_1 also passes through M. And since the medians are interchangeable, BB_1 and CC_1 are divided by M in the same ratio as AA_1. Point B', whenever it appears below, is defined as the intersection of the extended line BM and the side AC.

Proof 1: using proportional segments. Draw a line through A_1 parallel to BM and label N the point where it meets AC (fig. 2). Add lines through A and C parallel to these two. Let's apply the following well-known statement: parallel lines cut transverse lines into proportional segments. The four parallel lines are those that we added to our figure. On the transversals AC and AA_1 three of them intercept the proportional segments $AB' : B'N = AM : MA_1 = 2 : 1$; and for the transversals CA and CB three other lines give $CB' : B'N = CB : BA_1 = 2 : 1$. So $AB' : CB' = 1$, and we're done.

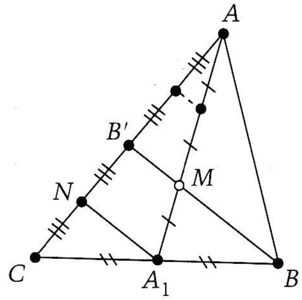


Figure 2

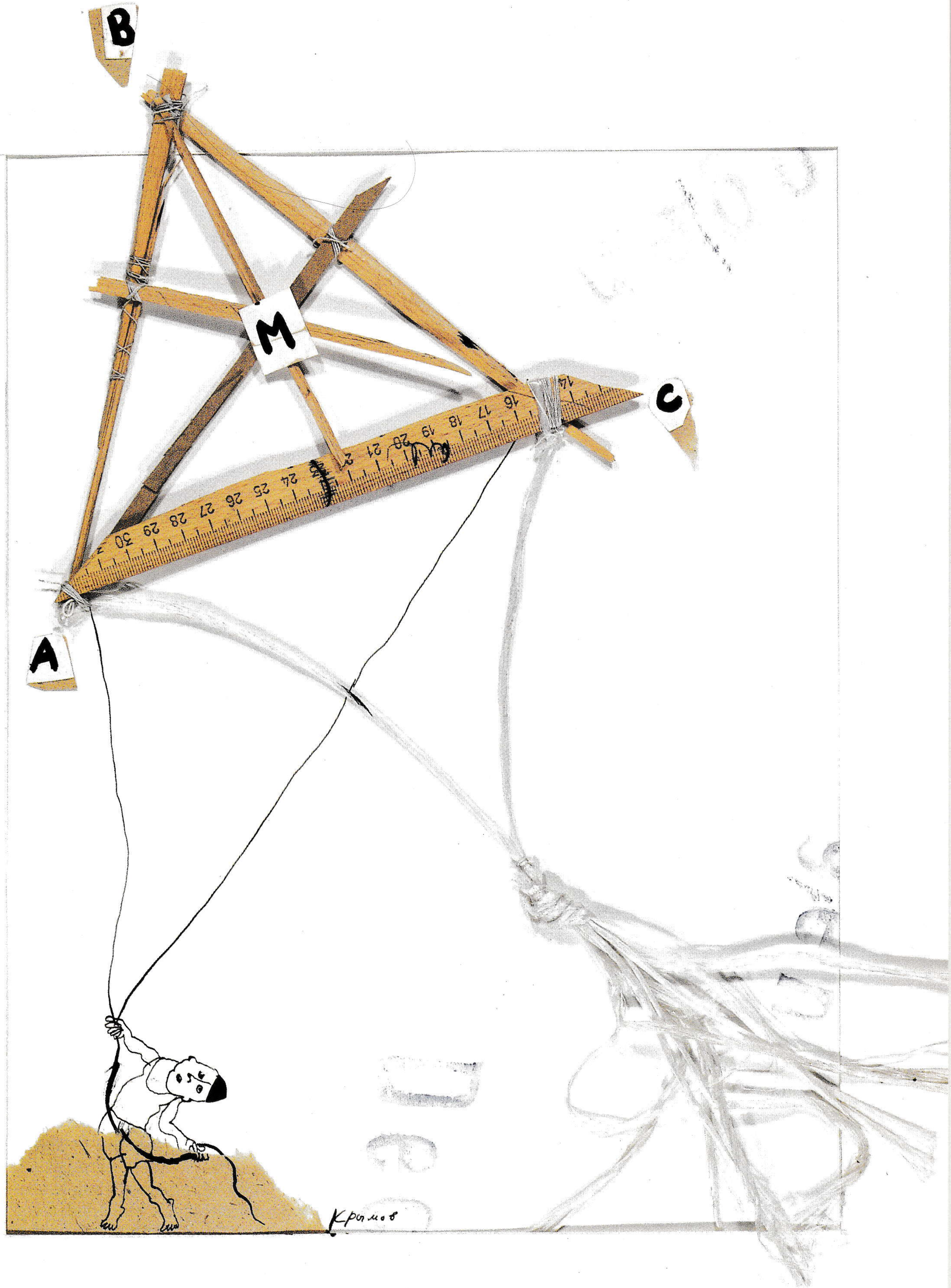
Exercise 1. In figure 2, imagine that $B'MB$ is an arbitrary line intersecting the sides CA, AA_1, and the extension of the side A_1C of triangle AA_1C, and that M is any point at all on the segment AA_1. Prove that

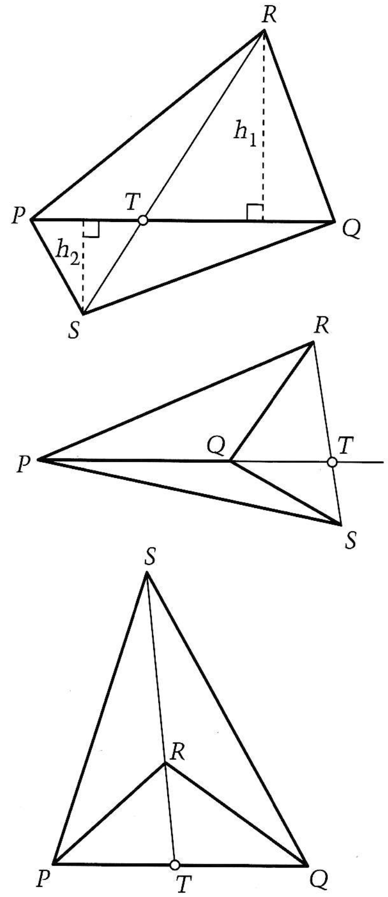
$$\frac{AM}{MA_1} \cdot \frac{A_1B}{BC} \cdot \frac{CB'}{B'A} = 1.$$

(This is Menelaus's theorem.)

The idea of the next proof is to express the ratios of segments in terms of the ratios of areas. It's based on a very simple fact (see figure 3 on page 34): *if two triangles PQR and PQS have a common base PQ, then the ratio of their areas is equal to the ratio in which the line PQ divides the segment*

Art by Dmitry Krymov


B
M
C
A



$$(PQR):(PQS) = \left(\frac{1}{2}PQ \cdot h_1\right) : \left(\frac{1}{2}PQ \cdot h_2\right)$$
$$= h_1 : h_2$$
$$= RT : TS$$

Figure 3

ST—$(PQR) : (PQS) = RT : TS$, where parentheses denote areas and T is the intersection of lines RS and PQ.

Proof 2: using areas. Let x be the area of $\triangle BA_1M$ (fig. 4). Then the lemma above yields $(BMA) = (AM/MA_1)x = 2x$ and $(BMC) = (BC/BA_1)x = 2x$, so $AB' : B'C = (BMA) : (BMC) = 1$.

Now use this method in the exercises below.

Exercises

2. Prove Ceva's theorem: if points A_1, B_1, C_1 are chosen on the sides BC, CA, AB of a triangle ABC so that the segments AA_1, BB_1, and CC_1 meet at a point, then

$$\frac{BA_1}{A_1C} \cdot \frac{CB_1}{B_1A} \cdot \frac{AC_1}{C_1B} = 1.$$

3. Each vertex of a convex pentagon is joined to the midpoint of the opposite side. Four of these segments pass through the same point. Prove that the fifth segment also passes through this point.

Proof 3: using dilation. Consider the dilation D by $-1/2$ relative to the center M—that is, the transformation that leaves point M unmoved and to any other point X assigns the point X' on the ray opposite MX half as close to M as X (formally, $\overrightarrow{MX'} = -\frac{1}{2}\overrightarrow{MX}$). Clearly, the images X' and Y' of any two points X, Y satisfy $\overrightarrow{X'Y'} = -\frac{1}{2}\overrightarrow{XY}$. By construction, the image of A is A_1; label E the image of B (fig. 5). Then $\overrightarrow{A_1E} = -\frac{1}{2}\overrightarrow{AB}$. But, by the property of a midline of a triangle, $\overrightarrow{A_1B_1} = -\frac{1}{2}\overrightarrow{AB}$ as well. Therefore, $E = B_1$—that is, the midpoint B_1 of AC lies on the line BM.

Exercises

4. Prove that the extended altitudes of a triangle meet in a point H, and that our point M divides the segment OH, where O is the circumcenter, in the ratio $OM : MH = 2 : 1$. (The line OMH is called Euler's line.)

5. Prove that the line through the intersection point of the diagonals of a trapezoid and the intersection of its extended (nonparallel) sides bisects the bases.

Proof 4: using coordinates. Denote by (x_A, y_A) the coordinates of vertex A in any fixed coordinate system; likewise, (x_B, y_B) and (x_C, y_C) are the coordinates of B and C. Find x_M, the x-coordinate of M. Since A_1 is the midpoint of BC, $x_{A_1} = (x_B + x_C)/2$. Since AM is $\frac{2}{3}$ of AA_1, $x_A - x_M = \frac{2}{3}(x_A - x_{A_1})$ (fig. 6). So

$$x_M = \frac{1}{3}\left(x_A + 2x_{A_1}\right)$$
$$= \frac{1}{3}\left(x_A + x_B + x_C\right).$$

And, of course,

$$y_M = \frac{1}{3}\left(y_A + y_B + y_C\right).$$

These expressions are symmetric relative to the coordinates of the triangle's vertices. Therefore, the points dividing the two other medians BB_1 and CC_1 in the ratio $2 : 1$ must have the same coordinates as M, so they all are the same.

Readers familiar with the notion of the center of mass must have recognized the expression for x_M and y_M as the coordinates of the centroid of points A, B, C—that is, the center of equal mass points placed at A, B, C. In the next proof we take advantage of this observation.

Proof 5: using center of mass. Put unit mass points at the vertices. The center of mass of any finite system can be found step by step: any

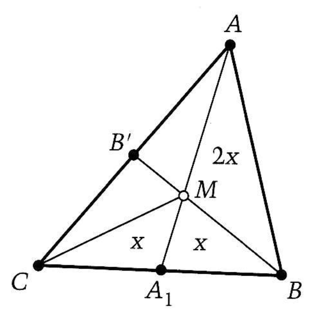


Figure 4

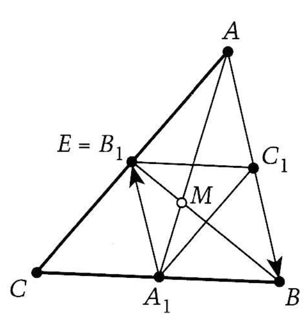


Figure 5

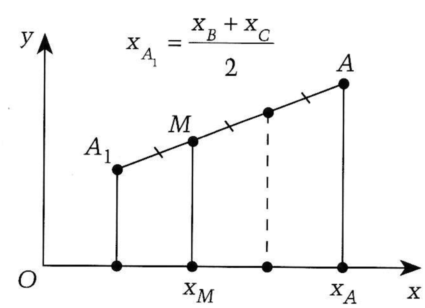


Figure 6

two masses can be replaced with their total mass placed at their center of mass, after which we get a system that has one mass point less. Then we combine another pair of masses, and so on until there is only one point left. In our case we have to make only two steps: the mass points at B and C can be replaced by a single mass point at A_1; the center of mass of this point and the unit mass point at A, according to the "lever rule," divides the segment AA_1 in inverse proportion to the masses—that is, in the ratio 2 : 1. Now, if we first unite masses at A and C, we'll see that the center of all three masses lies on another median, BB_1, which completes the proof according to our scheme.

Some readers may argue that this proof is mathematically inadequate. Indeed, the notion of the center of mass, as well as its properties, belongs to physics rather than mathematics. However, an absolutely rigorous mathematical support for the center-of-mass technique can be provided with some elementary vector algebra. In fact, this technique (in geometry) is just another "language" describing the same contents as the "language of vectors," but more convenient with a certain kind of problems.

By the way, the center of mass of a triangular *plate* coincidentally falls on the centroid M, too (see, for instance, "Halving It All" in the March/April 1992 issue of *Quantum*). However, the center of mass of a wire triangle does not coincide with the centroid unless the triangle is equilateral (see question 17 in the Kaleidoscope of the July/August 1994 issue). For other polygons, all three kinds of center of mass are, in general, distinct.

Exercise 6. Prove that the four segments joining the vertices of a tetrahedron to the centroids of the opposite faces (the medians of the tetrahedron) and the three segments joining the midpoints of its opposite edges (its bimedians) meet in a point. This point is called the *centroid* of the tetrahedron.

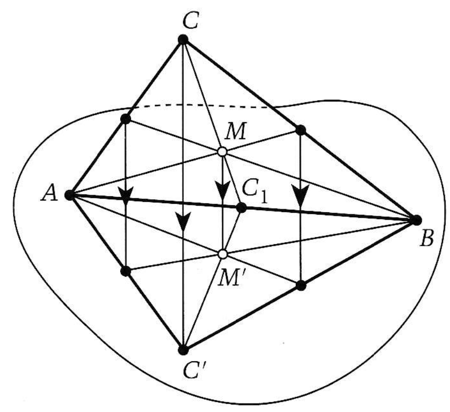


Figure 7

Proof 6: using parallel projection. Let's attach to the side AB of the given triangle an *equilateral* triangle ABC' with its vertex C' off the plane ABC (fig. 7). Now consider the parallel projection of the triangle ABC onto the plane ABC' along CC'. Clearly, triangle ABC with its medians projects on the triangle ABC' and its medians (because the midpoint of a segment is projected on the midpoint of the segment's projection). But the medians of an equilateral triangle coincide with its perpendicular bisectors (and with its bisectors and altitudes), which are known to pass through the same point—the circumcenter. Therefore, the medians of the original triangle also pass through the same point. As to the ratios, we note that the ratio in which a segment is divided by any point on it is preserved under a parallel projection (this follows from the property of proportional segments mentioned in the first proof). But for the center O (centroid, circumcenter, incenter—it's all the same) of the equilateral triangle ABC' and, say, the median CC' we have

$$C_1O : OC' = C_1O : OB = \sin 30° = \frac{1}{2}.$$

This proof brings to light and makes use of the fact that the property of medians, as well as the notion of the median itself, is preserved under parallel projection. Such properties and notions are called *affine*. For instance, "to be a parallelogram" is an affine property, whereas the properties "to be a rhombus" or "to be a rectangle" are not affine: parallel projection can make equal sides different and doesn't preserve perpendicularity. Proportional segments, ratios of areas, dilation, coordinates, center of mass, and parallel projection all are very useful and effective tools that help us solve affine problems. In fact, the best way to tackle most affine problems is to apply one of these tools. However, in our case affine proofs have a certain flaw—they are "too Euclidean." The property of medians can be extended beyond the limits of the Euclidean plane, so to speak. It can be generalized to space—to the three-dimensional analog of the triangle, the tetrahedron (see exercise 6 above)—and its affine proofs can be modified accordingly. But this theorem can be generalized to non-Euclidean planes as well. No, I'm not going to delve into hyperbolic (or Lobachevskian) geometry—the articles on this topic in the September/October 1992, November/December 1992, and March/April 1993 issues will give the interested reader plenty of opportunity to explore the property of medians. We have a much more familiar example of non-Euclidean geometry at hand—the geometry on the sphere. A spherical triangle ABC is obtained by joining three points A, B, C on the sphere with the (minor) arcs of great circles (these are the circles along which the sphere is cut by planes through its center—like meridians and the equator, but not the parallels of latitude, on a globe). The medians of triangle ABC, as well as any "straight segments" on the sphere, are also arcs of great circles. With this understanding, it turns out that *the medians of a spherical triangle always meet in a point.*

Exercise 7. Prove this statement. Hint: make use of the medians of the ordinary plane triangle with the same vertices.

Now I can explain what's wrong with affine proofs. They depend on

CONTINUED ON PAGE 39

Superconducting magnet

"Know then that this is the law: the north pole of one lodestone attracts the south pole of another."—Petrus Peregrinus (13th century A.D.)

by Arthur Eisenkraft and Larry D. Kirkpatrick

FOR THE XXV INTERNATIONAL Physics Olympiad held in July in Beijing, the Chinese hosts prepared problems that are an interesting mix of the modern and the traditional. We have used one of the theoretical problems as the basis for this month's contest problem.

For those of us who grew up with conventional electromagnets, it is very strange to see an electromagnet that is not connected to an external power source. But that is what happens with a superconducting magnet. After a current has been established in the magnet, the magnet can be disconnected from the external power and it will continue to produce a steady magnetic field for a very long time.

In a conventional electromagnet, a large current passes through a solenoid (a coil of wire) producing a magnetic field inside the solenoid. But the current produces a lot of heat due to the resistance of the wire. The production of this heat means that a source of energy is required to maintain a steady state.

In a superconducting magnet, the coil is immersed in liquid helium at a temperature of 4.2 K. At this temperature, the wire becomes superconducting—that is, its electrical resistance drops to zero. Therefore, no heat is produced and the need for an external power supply is eliminated.

The current in the magnet is controlled with a specially designed superconducting switch wired in parallel with the coil as shown in figure 1. The superconducting switch is usually a small length of superconducting wire wrapped with a heater wire and thermally insulated from the liquid helium bath. When the wire is heated, the wire reverts to the normal state and its resistance suddenly changes from $r = 0$ to $r = r_n$.

This very modern device can be analyzed using the very traditional physics that we learn in an introduction to circuits. We start out with Ohm's law, $V = IR$. We then add Kirchhoff's two laws. Kirchhoff's voltage rule tells us that the voltage drop across the superconducting coil must be the same as the voltage drop across the superconducting switch, $V_c = V_s$. This is just a statement of the conservation of energy. Kirchhoff's current rule tells us that the current flowing into a junction must equal the current flowing out of the junction—a statement of the conservation of charge. Using the directions indicated in figure 1, we have $I = I_c + I_s$.

The remaining physics that we require involves the coil. It is interesting because we have a pure inductor—the wire in the coil is superconducting and its resistance is zero. The voltage drop across an inductor

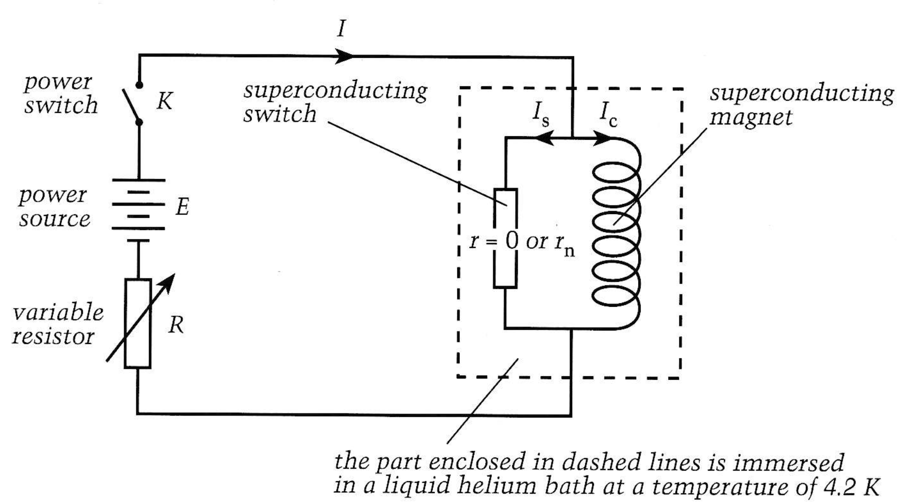


Figure 1

Art by Tomas Bunk


ELECTROMAGNETIC FIELDS FOREVER
LET ME TAKE YOU DOWN.
K
I
E = mc²
E
LIQUID HELIUM HOT
OHM
V = IR
I = Ic + Is
Tom Bunk

depends on the geometry and size of the coil (contained in the inductance L) and the change in the current. Notice that it is only the change in the current that matters, not the actual value of the current. Thus,

$$V_c = -L\frac{\Delta I_c}{\Delta t}.$$

As a consequence of this, when a solenoid is wired in series with a resistor and a battery, the current cannot instantaneously reach the final value of V/R. The voltage drop produced across the inductor as the current increases means that the current must climb to this final value exponentially. That is,

$$I = \frac{V}{R}\left(1 - e^{-t/\tau}\right),$$

where $\tau = L/R$ is the time constant characteristic of this circuit.

Let's now use these basic ideas about circuits to see how the superconducting switch can be used to control the operation of the superconducting magnet. Let's assume that $r_n = 5\ \Omega$ and $L = 10$ H. We start out with switch K closed.

A. Assume that from $t = 0$ until $t = 3$ min, we have $I = 1$ A, $I_c = I_s = 0.5$ A, and $r = 0$. We now use the variable resistor R to reduce the current I linearly to zero from $t = 3$ min to $t = 6$ min while keeping $r = 0$. Plot graphs of I_c and I_s as functions of time and explain why they behave this way.

B. Assume that from $t = 0$ until $t = 3$ min, we have $I = I_c = 0.5$ A, $I_s = 0$, and $r = 0$. At $t = 3$ min we use the heater to suddenly put the superconducting switch in the normal state with $r = r_n$. At $t = 6$ min we cool the switch so that it suddenly returns to the superconducting state with $r = 0$. Plot graphs of I, I_c, and I_s as functions of time and explain why they behave this way.

C. Let's now change the initial conditions by putting the initial current through the switch instead of the coil. Assume that the external resistor has a constant value $R = 5\ \Omega$ and that from $t = 0$ until $t = 3$ min we have $I = I_s = 0.5$ A, $I_c = 0$, and $r = 0$. At $t = 3$ min the switch changes to the normal state with $r = r_n$. At $t = 6$ min the switch returns to the superconducting state with $r = 0$. Plot graphs of I, I_c, and I_s as functions of time and explain why they behave this way.

D. When the switch is in the superconducting state, the magnet may be operated in the "persistent mode." In the persistent mode switch K is open and a current circulates through the coil and the superconducting switch indefinitely. Suppose that the magnet is operating in the persistent mode (that is, $I = 0$, $I_c = I_0$, and $I_s = -I_0$) with $I_0 = 20$ A from $t = 0$ to $t = 3$ min. We now want to shut the magnet down by reducing I_0 to zero. However, we will destroy the switch if the current through the switch exceeds 0.5 A. (Large currents cause the wire to switch to the normal state and the resulting heat melts the wire.) What steps can you use to shut the magnet down? Be sure to plot I, I_c, I_s, and r as functions of time to illustrate your method.

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*.

Laser levitation

Our contest problem in the May/June 1994 issue of *Quantum* required readers to compute the intensity of a laser beam required to suspend a triangular glass prism in the air. An excellent solution was submitted by Scott Wiley of Weslaco, Texas.

A. The preliminary problem involves suspending a box with bullets striking the bottom of a box at an angle θ. Because momentum is a vector quantity, we can break the initial and final momentum of the bullets into vertical and horizontal components. Ignoring the horizontal components as instructed, we can calculate the change in the vertical component of the momentum for each bullet:

$$\begin{aligned}\Delta p &= -mv_0 \cos\theta - mv_0 \cos\theta \\ &= -2mv_0 \cos\theta,\end{aligned}$$

where the minus sign indicates that the change is in the downward direction. Therefore, the force F_{box} exerted on the box is

$$\begin{aligned}F_{box} &= R\Delta p \\ &= 2Rmv_0 \cos\theta \\ &= 0.6\ \text{N},\end{aligned}$$

based on the data given in the problem.

B. Let's now move on to the laser beam. From the geometry of figure 2 we see that the angle of incidence for the incoming light beam is α. Using Snell's law and setting the index of refraction of air equal to 1, we have

$$\sin\alpha = n \sin\phi,$$

where ϕ is the angle of refraction at the first surface. A little more geometry shows that

$$\phi + \gamma = \alpha,$$

where γ is the angle of incidence at the second surface. Finally, using Snell's law at the second surface yields

$$n \sin\gamma = \sin\theta,$$

with θ being the angle of refraction back into the air.

Solving these equations, we find

$$\theta = \arcsin\left[n \sin(\alpha - \phi)\right],$$

with

$$\phi = \arcsin\frac{\sin\alpha}{n}.$$

C. By conservation of linear momentum, the total momentum of the system must remain unchanged. Since the wavelength of the laser

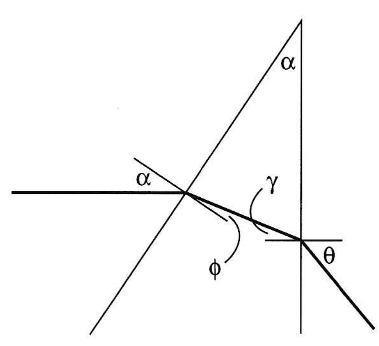


Figure 2

light is the same before and after it enters the prism, the magnitude of the momentum of each photon is the same and equal to E/c, where E is the energy of each photon. However, the angle of the beam has changed. Therefore, the horizontal and vertical components of the momentum change of the prism are

$$\Delta p_{px} = p - p\cos\theta = \frac{E}{c}(1-\cos\theta),$$

$$\Delta p_{py} = p\sin\theta = \frac{E}{c}\sin\theta.$$

Therefore, using the technique from part A, we find that the components of the force on the prism are

$$F_x = \frac{N}{t}\Delta p_{px} = \frac{NE}{ct}(1-\cos\theta),$$

$$F_y = \frac{N}{t}\Delta p_{py} = \frac{NE}{ct}\sin\theta.$$

Knowing that the power P in terms of the intensity and the number of photons N is

$$P = IA = \frac{NE}{t},$$

we have

$$F_x = \frac{IA}{c}(1-\cos\theta),$$

$$F_y = \frac{IA}{c}\sin\theta.$$

However, the intensity is not uniform over the prism. Because it falls off linearly with distance, we can avoid doing an integral and just use the average value for each face. The average intensities for the upper and lower faces for the case $h \le y_0 \le 3h$ are

$$\overline{I_u} = \frac{I(y_0) + I(y_0+h)}{2} = I_0\left(\frac{7h-2y_0}{8h}\right),$$

$$\overline{I_l} = \frac{I(y_0) + I(y_0-h)}{2} = I_0\left(\frac{9h-2y_0}{8h}\right).$$

As shown in part B, the top face provides an upward momentum to the prism. Conversely, the lower face provides a downward momentum. Therefore, the net vertical force is

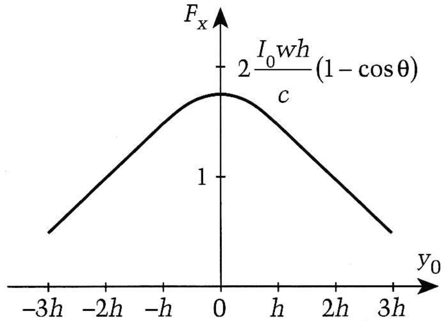


Figure 3

$$F_y = \left(\overline{I_u} - \overline{I_l}\right)\frac{wh}{c}\sin\theta$$

$$= -\frac{I_0wh}{4c}\sin\theta.$$

The minus sign indicates that the prism should be placed below the axis of the beam for it to be suspended.

Because the horizontal components act in the same direction, we have

$$F_x = \left(\overline{I_u} + \overline{I_l}\right)\frac{wh}{c}(1-\cos\theta)$$

$$= -\frac{I_0w}{2c}(4h - y_0)(1-\cos\theta).$$

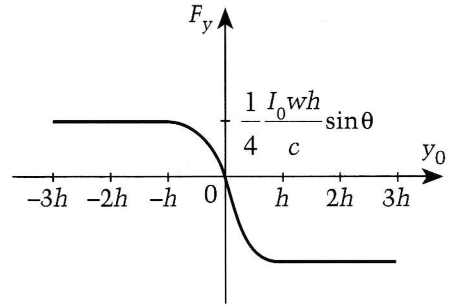


Figure 4

D. Similar calculations for the range $0 \le y_0 \le h$ yield the graphs shown in figures 3 and 4.

E. Using the dimensions and density given in the problem, the weight of the prism is $W = 1.42 \cdot 10^{-9}$ N. To levitate the prism, the upward force on the prism must equal its weight. This requires $I_0 = 6.19 \cdot 10^8$ W/m^2. The average intensity of the beam is $I_0/2$, so the power of the beam must be

$$P = IA = 24.8 \text{ W}.$$

"THE MEDIANS" CONTINUED FROM PAGE 35

parallelism, perhaps indirectly, and therefore can't be adjusted to spherical (or hyperbolic) geometry. Well, such universality may not be the first thing we want from a proof, but still I want to conclude with a proof that can be more or less applied to the sphere, too.

Proof 7: using the Law of Sines. Let's again follow the scheme we used above. Applying the Law of Sines to triangles ABB' and CBB' (fig. 4 on page 34), we get

$$\frac{AB'}{B'C} = \frac{AB\sin\angle ABB'}{\sin\angle AB'B} : \frac{BC\sin\angle CBB'}{\sin\angle CB'B}$$

$$= \frac{AB\sin\angle ABB'}{CB\sin\angle CBB'},$$

because $\sin\angle CB'B = \sin(180° - \angle AB'B) = \sin\angle AB'B$. Similarly, triangles AMB and A_1MB yield

$$\frac{AM}{A_1M} = \frac{AB\sin\angle ABB'}{A_1B\sin\angle CBB'}.$$

Since $AM = 2MA_1$ and $CB = 2A_1B$, after dividing the first relation by the second we obtain $AB'/B'C = 1$, completing the proof.

On the sphere, the Sine Law remains valid if we replace the side lengths of a triangle in it with the sines of the angular measures of the sides. So instead of the Euclidean relation $AB/\sin\angle C = BC/\sin\angle A$ in an ordinary triangle ABC, we'll have the spheric relation $s(A, B)/\sin\angle C = s(B, C)/\sin\angle A$, where $s(X, Y) = \sin\angle XOY$ and O is the center of the sphere. If we choose point M on the median AA_1 of a spheric triangle ABC such that $s(A, M)/s(M, A_1) = 2$ and replace all distances XY in proof 7 with $s(X, Y)$, it will work perfectly well for the sphere.

ANSWERS, HINTS & SOLUTIONS ON PAGE 60

AT THE BLACKBOARD

What you add is what you take

—but a profit you can make

by Andrey Yegorov

I'D LIKE TO BEGIN MY STORY WITH AN OLD chestnut. It's been around, but I found it in the old Russian book of recreational mathematics *V tsarstve smekalki* (*In the Realm of Mother Wit*) by E. I. Ignatyev, first published in 1908:

An old man who had three sons ordered that after his death they divide his herd of camels such that the oldest son took half of the herd, the middle son received one third of it, and the youngest son got one ninth. The old man died and left 17 camels. The sons started dividing the legacy, but found that the number 17 is not divisible by any of the numbers 2, 3, or 9. In a hopeless quandary about what to do, the brothers turned to a sage. The sage rode over on his own camel and divided the old man's herd in accordance with his will. How did he do it?

Art by Sergey Ivanov

Don't rack your brain too long over this baffling puzzle. It's only a joke: the sage added his camel to the herd, gave one half of the new herd (9 camels) to the oldest brother, one third (6 camels) to the middle brother, and one ninth (2 camels) to the youngest brother, and took the remaining camel ($18 - 9 - 6 - 2 = 1$), which happened to be his own. He then departed, leaving the brothers—and you, no doubt—completely befuddled.

I'm sure you'll unravel the secret of the sage's trick (although it looks very puzzling at first glance, doesn't it?). However, my goal wasn't to fool you, or even to make you smile (which would be a good enough excuse, in my book). It just so happens the sage's trick is a good illustration of one of the most frequently used techniques of recasting mathematical objects. For instance, in algebra we often add to an expression some terms and then subtract equal terms, preserving the total but making it easier to rearrange. You'll find many such algebraic transformations in this article, but before we concentrate on algebra, it should be said that this branch of mathematics is far from being the only one where the "add and subtract" trick is applied. In fact, it can be encountered almost everywhere, and my first more serious example will be from geometry. Let's derive the well-known formula for the area of a parallelogram as the product of its base and height (we already know that the area of a rectangle is the product of two adjacent sides).

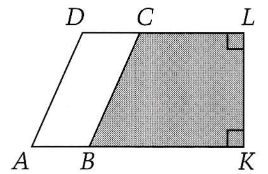


Figure 1

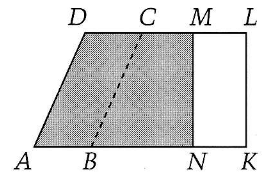


Figure 2

Consider a parallelogram $ABCD$ and add to it a trapezoid $CBKL$ (fig. 1) whose bases BK and CL are the extensions of AB and DC and whose side KL is perpendicular to the bases. Now we can draw MN parallel to LK and cutting off a trapezoid $ANMD$ that is congruent to trapezoid $BKLC$ (fig. 2). The remainder is the rectangle $KLMN$, which has the same area as $ABCD$. But this area equals $NK \cdot KL$, where NK equals the base AB of the parallelogram and KL is equal to its height.

Exercises

1. An absent-minded mathematician, instead of pouring milk into his cup of coffee, poured a spoon of coffee into a jug of milk and carefully mixed the liquid in the jug. Then he noticed his mistake and poured a spoonful of the mixture back into the cup. Is there more milk in the coffee or coffee in the milk? Does the answer depend on how carefully the drinks were mixed? And what does this problem have to do with the calculation of the area of a parallelogram above?

2. Show how a trapezoid can be transformed into a parallelogram with the same area by adding and subtracting congruent figures. Derive from this formulas for the area of a trapezoid, and also of a triangle.

3. Prove that the volume of an oblique prism equals the product of the area of its cross section perpendicular to the edges joining its bases and the length of any of these edges.

4. Find a formula for the lateral area of the prism in problem 3 in terms of the perimeter of the length of a lateral edge and the perimeter of a cross section taken perpendicular to the edge.

Completing the square

Now we turn to algebra.

One of the most frequent applications of the add-and-subtract technique is reworking an algebraic expression so as to create the expansion of the square of a sum or difference. For example, we have

$$\begin{aligned} u^2 + v^2 &= u^2 + 2uv + v^2 - 2uv \\ &= (u + v)^2 - 2uv, \end{aligned}$$

and, similarly,

$$u^2 + v^2 = (u - v)^2 + 2uv.$$

The first of these simple relations is used in the following problem.

Problem 1. For what positive integers n is the number $n^4 + 4$ a prime?

Solution. We can think of $n^4 + 4$ as $u^2 + v^2$ with $u = n^2$, $v = 2$. Then

$$\begin{aligned} n^4 + 4 &= n^4 + 4n^2 + 4 - 4n^2 \\ &= (n^2 + 2)^2 - 4n^2 \\ &= (n^2 - 2n + 2)(n^2 + 2n + 2). \end{aligned}$$

So the number $n^4 + 4$ is always the product of two integer factors the smaller of which, equal to $(n - 1)^2 + 1$, is greater than 1 unless $n = 1$. So for $n > 1$ this number is composite, and for $n = 1$ it's the prime number 5.

At the same time, we've factored the polynomial $n^4 + 4$ of n into two quadratic factors. Here's a similar factorization.

Problem 2. Factor the polynomial $x^4 + x^2 + 1$.

Solution. Add and subtract x^2:

$$\begin{aligned} x^4 + x^2 + 1 &= x^4 + 2x^2 + 1 - x^2 \\ &= (x^2 + 1)^2 - x^2 \\ &= (x^2 - x + 1)(x^2 + x + 1). \end{aligned}$$

In the same way the formula for the roots of the quadratic equation $x^2 + px + q = 0$ is derived:

$$\begin{aligned} x^2 + px + q &= x^2 + 2 \cdot \frac{p}{2} x + \frac{p^2}{4} + q - \frac{p^2}{4} \\ &= \left(x + \frac{p}{2}\right)^2 - \left(\frac{p^2}{4} - q\right) \\ &= \left(x + \frac{p}{2} - \sqrt{\frac{D}{4}}\right)\left(x + \frac{p}{2} + \sqrt{\frac{D}{4}}\right), \end{aligned}$$

where $D = p^2 - 4q$ is assumed to be nonnegative. This immediately yields the formula for the roots

$$x_{1,2} = \frac{-p \pm \sqrt{D}}{2}.$$

Next is a fourth-degree equation.

Problem 3. Solve the equation $x^4 + 4x - 1 = 0$.

Solution. We'll complete two squares at a time by adding and subtracting $2x^2 + 1$:

$$\begin{aligned} x^4 + 4x - 1 &= x^4 + 2x^2 + 1 - 2x^2 + 4x - 2 \\ &= (x^2 + 1)^2 - 2(x - 1)^2 = 0. \end{aligned}$$

It follows that $x^2 + 1 = \pm\sqrt{2}\,(x - 1)$—that is, either $x^2 + \sqrt{2}\,x + 1 - \sqrt{2} = 0$ or $x^2 - \sqrt{2}\,x + 1 + \sqrt{2} = 0$. Solving these equations, we get the answer:

$$x_{1,2} = \frac{-\sqrt{2} \pm \sqrt{4\sqrt{2} - 2}}{2},$$

$$x_{3,4} = \frac{\sqrt{2} \pm i\sqrt{4\sqrt{2} + 2}}{2}$$

(where $i = \sqrt{-1}$ is the imaginary unit).

Later we'll see how to factor an arbitrary polynomial of the fourth degree into quadratic factors by completing the square. For the time being, I'd like to mention a historical curiosity connected with the factorization of the expression $x^4 + a^4$. The great mathematician G. W. Leibniz (one of the creators of calculus) thought that this expression cannot be factored into quadratic polynomials. Yet we'll do this right now:

$$\begin{aligned} x^4 + a^4 &= x^4 + 2x^2a^2 + a^4 - 2x^2a^2 \\ &= (x^2 + a^2)^2 - (\sqrt{2}\,xa)^2 \\ &= (x^2 + \sqrt{2}\,xa + a^2)(x^2 - \sqrt{2}\,xa + a^2). \end{aligned}$$

Exercises

5. For what positive integers n is $n^4 + 4^n$ a prime?

6. Factor into quadratic polynomials (see the last example in the text above) (a) $x^4 - a^2x^2 + a^4$, (b) $x^4 + bx^2 + c$.

7. Solve the equations (a) $x^4 + 8x - 7 = 0$, (b) $(x^2 - 1)^2 = 4(2x + 1)$, (c) $x^2 + x^2/(x + 1)^2 = 1$.

I'll conclude this chain of problems with an example where the add-and-subtract trick is used to create the "incomplete square" $u^2 + uv + v^2$, which emerges in the standard factorization of $u^3 - v^3$.

Problem 4. Factor $a^5 + a + 1$ into two polynomials with integer coefficients.

Solution.

$$\begin{aligned} a^5 + a + 1 &= a^5 - a^2 + a^2 + a + 1 \\ &= a^2(a^3 - 1) + a^2 + a + 1 \\ &= (a^2 + a + 1)(a^3 - a^2 + 1). \end{aligned}$$

(Here we used the formula $u^3 - v^3 = (u - v)(u^2 + uv + v^2)$ for $u = a$, $v = 1$.)

As an additional consequence we see that the number $a^5 + a + 1$ is composite for all integer $a > 1$.

Exercises

8. Factor the polynomials (a) $a^{10} + a^5 + 1$, (b) $a^8 + a + 1$.

9. Prove that the number 1,280,000,401 is composite.

Multiply and divide

So far we performed algebraic transformations by adding and subtracting the same expression. Sometimes it's helpful to use another pair of inverse operations—multiplication and division.

Problem 5. Find the product $P = \cos x \cdot \cos 2x \cdot \cos 4x \cdot \ldots \cdot \cos 2^n x$.

Solution. Assuming $\sin x \neq 0$, multiply and divide P by $\sin x$:

$$\begin{aligned} P &= \frac{\sin x \cdot \cos x \cdot \cos 2x \cdot \ldots \cdot \cos 2^n x}{\sin x} \\ &= \frac{\sin 2x \cdot \cos 2x \cdot \ldots \cdot \cos 2^n x}{2 \sin x} \\ &= \ldots \\ &= \frac{\sin 2^{n+1} x}{2^{n+1} \sin x}. \end{aligned}$$

(If $\sin x = 0$, $P = \pm 1$.)

Thus, we have obtained a neat formula that will allow us to derive Viète's formula for the number π. To do this, take the limit of both sides of the equality

$$\cos\frac{x}{2} \cdot \cos\frac{x}{4} \cdot \ldots \cdot \cos\frac{x}{2^n} = \frac{\sin x}{2^n \sin\dfrac{x}{2^n}},$$

which is an immediate consequence of problem 5 (with $n - 1$ substituted for n and $2^{-n}x$ substituted for x), as $n \to \infty$.

Using the well-known relation $\sin\alpha/\alpha \to 1$ as $\alpha \to 0$, we see that the denominator in the right side of this equality approaches x as $n \to \infty$:

$$2^n \sin\frac{x}{2^n} = x \cdot \frac{\sin(x/2^n)}{x/2^n} \to x,$$

because $x/2^n \to 0$ as $n \to \infty$. The left side turns into an infinite product, and we arrive at

$$\cos\frac{x}{2} \cdot \cos\frac{x}{4} \cdot \ldots \cdot \cos\frac{x}{2^n} \cdot \ldots = \frac{\sin x}{x}.$$

Substituting $x = \pi/2$, we obtain

$$\cos\frac{\pi}{2} \cdot \cos\frac{\pi}{8} \cdot \ldots \cdot \cos\frac{\pi}{2^{n+1}} \cdot \ldots = \frac{2}{\pi}.$$

But for any acute angle x

$$\cos\frac{x}{2} = \sqrt{\frac{1 + \cos x}{2}}.$$

Therefore, for $n \geq 2$

$$\begin{aligned} \cos\frac{\pi}{2^n} &= \sqrt{\frac{1 + \cos\dfrac{\pi}{2^{n-1}}}{2}} \\ &= \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\cos\frac{\pi}{2^{n-1}}}} \\ &= \ldots \\ &= \underbrace{\sqrt{\frac{1}{2} + \sqrt{\frac{1}{2} + \ldots + \frac{1}{2}\sqrt{\frac{1}{2}}}}}_{n-1 \text{ roots}}, \end{aligned}$$

and so

$$\frac{\pi}{2} = \frac{1}{\sqrt{\dfrac{1}{2}}\sqrt{\dfrac{1}{2} + \dfrac{1}{2}\sqrt{\dfrac{1}{2}}}\sqrt{\dfrac{1}{2} + \dfrac{1}{2}\sqrt{\dfrac{1}{2} + \dfrac{1}{2}\sqrt{\dfrac{1}{2}}}} \cdots}.$$

Exercise 10. Compute the infinite product

$$\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}}\cdot\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}}}$$

$$\cdot\dots\cdot\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\dots+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}}}}}.$$

Certain sums can also be calculated using the multiply-and-divide device.

Problem 6. Find the sum

$$S_n = 1+11+111+\dots+\underbrace{11\dots1}_{n}.$$

Solution. Let's multiply (and later divide) by 9:

$$\begin{aligned}9S_n &= 9+99+999+\dots+\underbrace{9\dots1}_{n}\\ &= 10-1+10^2-1+10^3-1+\dots+10^n-1\\ &= 10+10^2+\dots+10^n-n\\ &= \frac{10^{n+1}-10}{9}-n.\end{aligned}$$

So the answer is

$$S_n = \frac{10^{n+1}-9n-10}{81}.$$

Problem 7. Find the sum

$$S_n = \sin x + \sin 2x + \dots + \sin nx.$$

Solution. The reader can verify (or recall) that $\sin A \sin B = ½[\cos(A-B) - \cos(A+B)]$. If $\sin(x/2) \neq 0$, then using this formula we get

$$\begin{aligned}S_n\sin\frac{x}{2} &= \sin\frac{x}{2}\cdot\sin x+\sin\frac{x}{2}\cdot\sin 2x+\dots+\sin\frac{x}{2}\cdot\sin nx\\ &= \frac{1}{2}\left(\cos\frac{x}{2}-\cos\frac{3x}{2}\right)+\frac{1}{2}\left(\cos\frac{3x}{2}-\cos\frac{5x}{2}\right)\\ &\quad+\dots+\frac{1}{2}\left(\cos\frac{2n-1}{2}x-\cos\frac{2n+1}{2}x\right)\\ &= \frac{1}{2}\left(\cos\frac{x}{2}-\cos\frac{2n+1}{2}x\right)\\ &= \sin\frac{nx}{2}\cdot\sin\frac{(n+1)x}{2}.\end{aligned}$$

Therefore,

$$S_n = \frac{\sin\frac{nx}{2}\cdot\sin\frac{(n+1)x}{2}}{\sin\frac{x}{2}}.$$

(Clearly, $S_n = 0$ for $\sin(x/2) = 0$.)

Exercise 11. Compute the sums (a) $x + 2x^2 + \dots + nx^n$, (b) $1 + \cos x + \cos 2x + \dots + \cos nx$.

Now a problem from number theory.

Problem 8. What greatest power of 2 is a divisor of the product $P_n = (n+1)\cdot(n+2)\cdot\dots\cdot 2n$?

Solution. Let's multiply and divide P_n by $n! = 1\cdot2\cdot3\cdot\dots\cdot n$ and rework the numerator:

$$\begin{aligned}P_n &= \frac{n!\cdot(n+1)\cdot\dots\cdot(2n)}{n!}\\ &= \frac{(2n)!}{n!}\\ &= \frac{2\cdot(2\cdot2)\cdot(2\cdot3)\cdot\dots\cdot(2n)\cdot(1\cdot3\cdot5\cdot\dots\cdot(2n-1))}{n!}\\ &= \frac{2^n\cdot n!(1\cdot3\cdot\dots\cdot(2n-1))}{n!}\\ &= 2^n\cdot(2n-1)!!,\end{aligned}$$

where $(2n-1)!!$ is, by definition, the product of all the odd numbers from 1 to $2n-1$. This shows that the answer is 2^n.

It's simply impossible to give in one article a more or less complete notion of the variety of problems that are solved using the add-and-subtract or multiply-and-divide tricks. (For example, I didn't even touch on the applications of this method in proving inequalities.) I hope you'll find and solve a lot of these problems yourself. Now I want to fulfil the promise I made earlier and explain how to factor polynomials of the fourth degree into quadratic polynomials.

Ferrari's method

We're going to follow in the footsteps of Lodovico Ferrari (1522–1565), who discovered a method for solving equations of the fourth degree by reducing them to quadratic equations (using an auxiliary cubic equation).

Consider the equation

$$P(x) = x^4 + ax^3 + bx^2 + cx + d = 0.$$

Applying our technique, rewrite it as

$$\begin{aligned}P(x) &= x^4+2\frac{a}{2}x^3+\frac{a^2}{4}x^2-\frac{a^2}{4}x^2+bx^2+cx+d\\ &= \left(x^2+\frac{ax}{2}\right)^2+\left(b-\frac{a^2}{4}\right)x^2+cx+d.\end{aligned}$$

Now let's try to represent the last expression as the difference of squares, which would enable us to factor it. To do this, we'll add to $P(x)$ and subtract from it the expression $2\alpha(x^2 + ax/2) + \alpha^2$, where α is an unknown number as yet. Then $P(x)$ takes the form

$$P(x) = \left(x^2+\frac{ax}{2}+\alpha\right)^2-\left(Ax^2+Bx+C\right),$$

where $A = 2\alpha + a^2/4 - b$, $B = a\alpha - c$, $C = \alpha^2 - d$. We want the trinomial $Ax^2 + Bx + C$ to be a complete square, which is true if and only if the following conditions hold: $A > 0$ and $B^2 - 4AC = 0$—that is,

$$(a\alpha - c)^2 = 4\left(2\alpha + \frac{a^2}{4} - b\right)(\alpha^2 - d).$$

This cubic equation for α is called the Ferrari resolvent for the polynomial $P(x)$. If α_0 is the resolvent's root satisfying $2\alpha_0 + a^2/4 - b > 0$ (that is, $A > 0$), then $P(x)$ is equal to the difference of squares:

$$P(x) = \left(x^2 + \frac{ax}{2} + \alpha\right)^2 - (kx + l)^2,$$

where k and l are expressed in terms of the coefficients of $P(x)$ and the number α_0, and so our original equation is reduced to two quadratic equations. (And, of course, we'll be able to factor $P(x)$ into quadratic polynomials.)

Let's make sure that the required root of the resolvent really does exist. The cubic equation for α given above can be written as

$$Q(\alpha) = 4\left(2\alpha + \frac{a^2}{4} - b\right)(\alpha^2 - d) - (a\alpha - c)^2.$$

For $\alpha = \frac{1}{2}(b - a^2/4)$ we have $Q(\alpha) = -(a\alpha - c)^2 < 0$, and for large enough α, $Q(\alpha) > 0$ (because $Q(\alpha) = 8\alpha^3$ + some quadratic polynomial in α). Therefore, there exists a number $\alpha_0 > \frac{1}{2}(b - a^2/4)$ such that $Q(\alpha_0) = 0$—which is just what we need.

To apply Ferrari's method one must know how to solve cubic equations. There is a formula, called Cardano's formula,[1] that allows us to express the roots of a cubic equation in terms of its coefficients and entails only four arithmetic operations and radicals (square and cube roots). Quadratic equations are also solved in radicals. Therefore, by Ferrari's method, we can express the roots of a fourth-degree equation in radicals—that is, there exists *a formula* that involves four arithmetic operations and square and cube roots for solving fourth-degree equations. Paolo Ruffini (1765–1822) and Niels Henrik Abel (1802–1829) proved that for equations of higher degrees there are no such formulas. Not only that, it follows from the work of Galois (see "The Short, Turbulent Life of Évariste Galois" in the November/December 1991 issue of *Quantum*) that there exists an equation of the fifth degree with integer coefficients whose roots cannot be expressed in terms of the coefficients—that is, integers—by means of a finite number of additions, subtractions, multiplications, divisions, and extraction of roots of any degrees. One such equation is, for instance, $x^5 - 25x - 5 = 0$, which has three real and two complex roots.

The reader may want to look back to problem 3 and see how this method worked there. Or, let's look at an example showing directly the application of Ferrari's method.

Problem 9. Solve the equation $x^4 - 10x^2 - 8x + 5 = 0$.

Solution. Rather than use the formula we already worked out, let's walk through the steps of Ferrari's method once again. First, rewrite the equation:

$$x^4 = 10x^2 + 8x - 5.$$

Add $2\alpha x^2 + \alpha^2$ to both sides:

$$(x^2 + \alpha)^2 = (10 + 2\alpha)x^2 + 8x + \alpha^2 - 5.$$

Equate the discriminant of the quadratic polynomial on the right side to zero:

$$16 - (10 + 2\alpha)(\alpha^2 - 5) = 0.$$

After simplifying, we arrive at the equation in α

$$\alpha^3 + 5\alpha^2 - 5\alpha - 33 = 0.$$

One of the roots of this equation can simply be guessed: $\alpha = -3$. Substituting this value of α, we get

$$(x^2 - 3)^2 = 4x^2 + 8x + 4 = 4(x + 1)^2,$$

which gives either $x^2 - 3 = 2x + 2$ or $x^2 - 3 = -2x - 2$. Solving the quadratics $x^2 - 2x - 5 = 0$ and $x^2 + 2x - 1 = 0$, we finally obtain the answer: $x_{1,2} = 1 \pm \sqrt{6}$, $x_{3,4} = -1 \pm \sqrt{2}$.

Now try Ferrari's method yourself.

Exercises

12. Solve the equations (a) $x^4 - 4x^3 + 5x^2 - 2x - 6 = 0$, (b) $x^4 + x^3 - 10x^2 - 2x + 4 = 0$.

13. Factor into quadratic polynomials (a) $x^4 + 2x^3 + 2x^2 + x - 2$, (b) $x^4 + 2x^3 - 3x^2 - 4x - 1$.

ANSWERS, HINTS & SOLUTIONS ON PAGE 61

[1]Ferrari's teacher Girolamo Cardano (1501–1576) was the first to publish this formula. Its discovery forms one of the most fascinating chapters in the history of mathematics, and we're going to publish a special article about it.—*Ed.*

Egyptian fractions

An alternate life-style without multiplication tables

by George Berzsenyi

THE ANCIENT EGYPTIANS didn't torture their children by forcing them to memorize extensive multiplication tables. Instead, they used a succession of doubling operations to perform multiplications and divisions, implicitly recognizing the fact that any number can be expressed as the sum of powers of 2. Hence their practice may be considered a forerunner of the computer age.

To deal with fractions, they avoided some of the computational difficulties by developing extensive tables of representations of fractions of the form $2/n$ as sums of distinct unit fractions $1/k$, which they denoted by an elongated ellipsis above the number k. To simplify the present exposition, we will use a bar in place of their elliptical symbol. The Ahmes (or Rhind) Papyrus, dating back to 1650 B.C., starts with a table of such expressions for $2/n$ for all odd values of n from 5 to 101. A few of the entries are reproduced below in order to provide some practice with the notation:

$$\frac{2}{5} = \overline{3} + \overline{15},$$

$$\frac{2}{7} = \overline{4} + \overline{28},$$

$$\frac{2}{11} = \overline{6} + \overline{66},$$

$$\frac{2}{13} = \overline{8} + \overline{52} + \overline{104},$$

$$\frac{2}{23} = \overline{12} + \overline{276},$$

$$\frac{2}{35} = \overline{30} + \overline{42},$$

$$\frac{2}{97} = \overline{56} + \overline{679} + \overline{776},$$

$$\frac{2}{101} = \overline{101} + \overline{202} + \overline{303} + \overline{606}.$$

Clearly, they were familiar with a number of identities, like $2/3k = \overline{2k} + \overline{6k}$, $2/n = \overline{n} + \overline{2n} + \overline{3n} + \overline{6n}$, and $1/n = \overline{n+1} + \overline{n(n+1)}$, while some of their other entries were obtained seemingly by ad hoc methods, aiming for the smallest denominators.

The world of Egyptian fractions continues to fascinate mathematicians. For example, Erdős posed the question: *Can $4/n$ always be represented as the sum of three or fewer unit fractions?* Some partial answers are provided by the identities in the box below, leaving the development of the easier identities for $4/(4k-1)$, $4/(4k-2)$, $4/(8k-3)$, and $4/(24k-7)$ as a **first challenge** to my readers. My **second challenge** is to verify that these identities prove the conjecture of Erdős for all fractions, except those of the form $4/(120k+1)$ and $4/(120k+49)$. These findings were reported in an article (in Hungarian) by János Surányi in 1981. According to Richard Guy's *Unsolved Problems in Number Theory*, published in the same year by Springer-Verlag New York, Inc., a similar conjecture by Sierpiński concerning fractions of the form $5/n$ has been verified for all $n \le 10^8$.

In conclusion, my **last challenge** to my readers consists of several easier problems: (1) For which pairs of relatively prime integers a and b will there exist positive integers x and y such that $a/b = \overline{x} + \overline{y}$? (2) Find all solutions in positive integers of the equations $\overline{x} + \overline{y} = \overline{z}$ and $\overline{x^2} + \overline{y^2} = \overline{z^2}$. (3) Are there positive integers x, y, z such that $\overline{x^n} + \overline{y^n} = \overline{z^n}$ if $n > 2$? Space limitations do not allow me to address this last question in proper detail.

$$\frac{4}{120k+25} = \overline{48k+10} + \overline{120k+25} + \overline{240k+50},$$

$$\frac{4}{120k+73} = \overline{30k+20} + \overline{(6k+4)(120k+73)} + \overline{(15k+10)(120k+73)},$$

$$\frac{4}{120k+97} = \overline{30k+25} + \overline{10(k+1)(120k+97)} + \overline{10(k+1)(6k+5)(120k+97)}$$

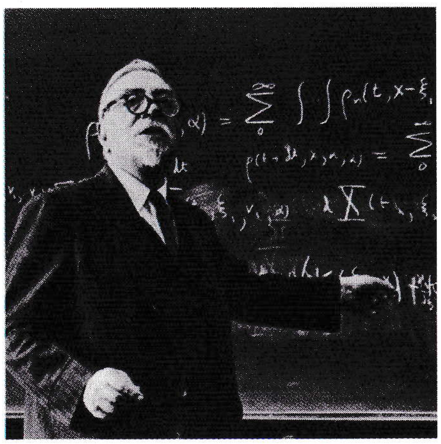
The legacy of Norbert Wiener

Part I: Childhood, boyhood, and youth

NORBERT WIENER WROTE his autobiography in two volumes, *Ex-Prodigy* and *I Am a Mathematician*. The title *Ex-Prodigy* says much: having been a prodigy was a determining fact of Wiener's life. But although it may be true that prodigies are born, they are also made. Wolfgang might not have become Mozart without Leopold, and Norbert might not have become Wiener without Leo.

For good reason, Leo Wiener (1862–1939) becomes the most compelling character in his son's autobiography. Born into a family of Jewish scholars in the Belorussian city of Bialystok (now part of Poland), Leo showed a phenomenal gift for languages and by adolescence already spoke German, Russian, French, Italian, and Polish. According to Norbert, Leo could pick up the essentials of a language in a few weeks, and later in his professional career "spoke some forty of them." He also published mathematical articles in obscure journals and passed on his knowledge to his son.

Always leaning toward Tolstoyism, Leo at 18 joined a humanitarian society and "forswore drink, tobacco and the eating of meat for the rest of his life." This last habit, at least, passed on to Norbert. The same year, Leo joined a fellow Tolstoyan in a hare-brained scheme to found a vegetarian-humanitarian-socialist community in Central America. The friend reneged, but Leo soon found himself sailing penniless for the American continent. After some useful adventures, he ended up in Kansas City, Missouri, where a sign "Gaelich Lessons Given" caught his eye. He enrolled, soon ended up teaching the class, and settled in Kansas City.



Norbert Wiener (1894–1964)

In 1893 Leo married Bertha Kahn, the daughter of a department store owner. On November 26, 1894, Bertha gave birth to Norbert. Around the same time, Leo lost his position as Professor of Modern Languages at the University of Missouri, Columbia, and the family moved east to Leonard Avenue at the border between Cambridge and Somerville, Massachusetts. Once again Leo was forced to accept odd jobs, but his remarkable talent soon landed him an instructorship at Harvard, where he remained until his retirement as Professor of Slavic Languages in 1930.

According to Leo, as quoted in the July 1911 issue of *American Magazine*, Norbert's precocity first became evident at the age of 18 months, when his nurse noticed him intently watch her draw letters in the sand at a beach. Within a few days he knew the alphabet. "Thinking that this was an indication that it would not be hard to interest him in reading, I started teaching him how to spell at the age of three. In a very few weeks he was reading quite fluently, and by six was acquainted with a number of excellent books, including works by Darwin, Ribot, and other scientists, which I had put into his hands in order to instill in him something of the scientific spirit."

Leo made no secret that he intended to mold his children into prodigies. In the same article he declared, "It is all nonsense to say, as some people do, that Norbert and Constance and Bertha [Norbert's sisters] are unusually gifted children. They are nothing of the sort. If they know more than other children of their age, it is because they have been trained differently."

The elder Wiener did indeed take almost complete charge of the younger Wiener's education. Although Norbert was enrolled in the third grade at the age of seven, not too far in advance of his years, he was soon skipped to the fourth grade. Even that move proved unsatisfactory, and Leo withdrew Norbert from school entirely, deciding to tutor him at home. This period of home instruction included large doses of algebra, Latin, and German, and lasted about

Reprinted from the program booklet for The Legacy of Norbert Wiener: A Centennial Symposium in Honor of the 100th Anniversary of Norbert Wiener's Birth, October 8–14, 1994, prepared by the MIT Department of Mathematics with the assistance of Tony Rothman.

Norbert Wiener at age 7.

two years. An important event took place when Norbert was eight. Due to already severe myopia he was forced to stop reading for six months and learn his lessons aurally. He credits the experience with a sharp improvement of his memory, which, by later accounts, was nearly photographic. According to one anecdote, he could recite a full Gilbert and Sullivan operetta having heard it once.

Undoubtedly, Leo saw himself as a well-intentioned father and fair taskmaster, but the son remembered his training otherwise. According to Norbert, whenever he made the slightest mistake, "the gentle and loving father was replaced by an avenger of the blood." Even worse, Leo's comments in *American Magazine* implied that Norbert's innate abilities were unimportant. Norbert recalls that the article "had a devastating effect on me. It declared to the public that my failures were my own but my successes were my father's." But despite what Wiener wrote in *Ex-Prodigy*, Amar Bose, Wiener's former student and colleague, and perhaps his closest associate during the last decade of his life, reports that Wiener said that "everything he had was due to his father." In sum, Norbert was ambivalent toward his father, and he displayed that ambivalence in the dedication of his book *The Human Use of Human Beings*: "To the memory of my father, Leo Wiener . . . my closest mentor and dearest antagonist."

In 1903 the family moved to Harvard, Massachusetts, where Norbert, not quite nine, was enrolled in nearby Ayer High School. He remained there three years, until his graduation in 1906. At that point his mentor and antagonist decided to enroll the eleven-year-old at Tufts, rather than risk the strain of the Harvard entrance exams. At the time Norbert's main interest was in biology, and his course work appears typical for a science major: doses of physics and mathematics along with the biology courses. Leo continued to tutor his son at home, with the result that Norbert found "the courses on calculus and differential equations quite easy." He does concede that his introduction to the theory of equations under Professor Ransom was "over my head," especially the section on Galois theory. Nonetheless, in 1909, Norbert was graduated *cum laude* in mathematics. He was fourteen.

Wiener's career at Tufts seems to have ended in a severe adolescent depression that lifted only gradually. Recurring depression would become a central feature of Wiener's life, and his own account suggests that this bout lasted fully through his graduate studies at Harvard. At Harvard Norbert had intended to pursue zoology, but this decision quickly proved to be a disaster, which he blamed on his lack of manual dexterity and his severe myopia. Earlier, at Tufts, he had had some success in philosophy, and, characteristically, it was at his father's behest that he abandoned zoology and applied to the Sage School of Philosophy at Cornell. In Ithaca, Norbert's depression continued, and his writings leave no doubt that he hated the place. Not only did he do poorly in his philosophy courses, "the theory of functions of a complex variable was beyond me."

As a result of his poor performance, Wiener's fellowship at Cornell was not renewed. The following year, he returned to Harvard's Department of Philosophy and his father. At Harvard he studied mathematical logic and, under Karl Schmidt of Tufts, wrote his thesis on theories of Schroeder, Whitehead, and Russell. Although he claims to have found the work easy, he also admits that later "under Bertrand Russell in England, I learned that I missed almost every issue of true philosophical significance." All in all, Wiener never liked Harvard much more than Cornell. In 1913 he received his Ph.D. He was not quite 19, about six years younger than the average Ph.D. recipient in that era.

While in his last year at Harvard, Wiener received a travel grant and, upon graduation, set sail for Cambridge, England, to pursue postdoctoral work in mathematical logic with Bertrand Russell. Wiener relished his new-found independence from his parents, even if his inexperience created problems: for example, he waged a quixotic battle with his landlady over the terms of his lease. But he also discovered a different breed of student who accepted his eccentricities and thrived on intellectual discussion. During that year he met another expatriate, T. S. Eliot, and they exchanged books and philosophical ideas. Wiener credits Russell with persuading him to learn some more genuine mathematics and acquainting him with the work of Einstein. But he was most inspired by G. H. Hardy, whom he calls his "master in mathematical training." Hardy introduced him properly to complex variables and to the Lebesgue integral, topics that would play a major role in his later career.

Despite the importance of Hardy's influence, Wiener came to view Hardy's renowned condescension toward applications as "pure escapism." In their later encounters, Wiener bridled at Hardy's suggestion that Wiener's beautiful work on harmonic analysis was motivated solely by the internal aesthetics of mathematics and not by applications. In keeping with his deep and abiding interest in applications, Wiener believed that mathematicians cannot ignore the outside world and must both apply mathematics and bear the moral responsibility for applications. This conviction would become even more pronounced as time passed. Indeed, Wiener has had the last laugh: even Hardy's beloved number theory has applications to telecommunications, cryptography, and computer science.

Because Russell was planning to spend the spring semester at Harvard, Wiener had decided to finish his postdoctoral year at Göttingen, home to such mathematical luminaries as David Hilbert and Edmund Husserl. After Göttingen, he returned to England, hoping to spend the academic year 1914–1915 at Cambridge again. However, he found the university effectively shut down by the war and decided to return to America. Back in America, he spent an unhappy semester at Columbia, where disagreeable encounters with dormmates appear to have overshadowed his academic activities.

As a Harvard Ph.D., Wiener had the right to give a series of docent (unpaid) lectures at Harvard. The following year, he lectured in mathematical logic and did some routine undergraduate teaching in philosophy. He hoped in this way to secure a permanent position in the Harvard Philosophy Department. He never did. Wiener was a chaotic lecturer, a trait that did not improve with time. Moreover, George David Birkhoff, America's leading mathematician at the time, disturbed Wiener's docent lectures even further by pointing out mistakes.

In *I Am a Mathematician*, Wiener accuses Birkhoff of shutting him out of Harvard and blames Birkhoff's anti-Semitism. Norbert goes on to say that Birkhoff resented Leo Wiener's extravagant boasts about Norbert and the aggressive personality of both father and son—Leo had recently confirmed his reputation as a "quarrelsome" character with vitriolic public attacks against German militarism and its defenders among the Harvard faculty. But by any objective standard, Norbert's performance that year was not good enough to have secured a position at Harvard, even for a Cabot.[1]

Leo Wiener.

Although Wiener was never reconciled to his failure to get a position at Harvard, he did ultimately win G. D. Birkhoff's respect. They ran the joint Harvard–MIT Mathematics Colloquium, and their correspondence reveals that they greatly admired each other's mathematics and developed a cordial relationship. The basis of that relationship may be guessed from the following reference to Birkhoff in *I Am a Mathematician*: "I was not alone in my competitiveness. At least one of the greatest American mathematicians, a man whose disapproval was the highest hurdle I should have to leap, was even more intensely competitive than myself."

After the Harvard debacle, Norbert, again following Leo's advice, began to look for a job in mathematics instead of philosophy. He managed, with some difficulty, to land a position at the University of Maine. But he found the place intellectually moribund, and the entire experience proved a nightmare. (For example, although he did not have to contend with the formidable Birkhoff, students dropped pennies to disrupt his lectures.) Near the end of the 1917 academic year, the United States entered the war and Norbert attempted to enlist. But he was rejected by all the services because of bad eyesight. Eventually, he did graduate from the Harvard ROTC[2] with a "document that was eminently not negotiable for a commission." There followed brief stints at General Electric and at the *Encyclopedia Americana*, where he was employed as a hack writer. Wiener actually enjoyed this work, but, during the summer of 1918, he decided to renew his job search. At this stage he was so desperate that he even applied for a position in Puerto Rico.

Around this time he received an invitation from Professor Oswald Veblen of Princeton to join Veblen's newly formed ballistics group at the Aberdeen Proving Ground in Maryland. This group's primary mission was to test new ordnance and to compute range tables that took into account the elevation angle, size of the charge, and other factors. Wiener seems to have enjoyed the direct practical application of mathematics in ballistics calculations, and his experience at Aberdeen served him well in his investigations of anti-aircraft fire during World War II.

After the war, Wiener had hoped to follow Veblen back to Princeton, where Veblen was instrumental in assembling Princeton's soon-to-be-famous department of mathematics. The invitation never came. At about the same time, the fiancé of Wiener's sister Constance died in the influenza epidemic that swept the country after World War I. Constance's fiancé had been a budding mathematician, and after his untimely death Norbert received several mathematics books from his library. Thus, by accident, Wiener became acquainted with Volterra's *Theory of Integral Equations*, Osgood's *Theory of Functions*, Lebesgue's book on the theory of integration, and Fréchet's treatise on the theory of functionals. Wiener claims that "for the first time I began to have a really good understanding of modern mathematics." As Norman Levinson, Wiener's most prominent student, remarks, this is an astounding statement from an individual who had attended Hardy's lectures five years earlier, not to mention one who had spent a semester at Hilbert's Göttingen, the fount of modern analysis. Here we confront the irony of Wiener's precocity: he received his Ph.D. at age 18, but his grasp of mathematics did not arrive until the rather advanced age of 24.

TO BE CONTINUED IN THE NEXT ISSUE

[1]A "Boston Brahmin" (that is, a member of the blue-blooded Anglo-Saxon elite of that city).—*Ed.*

[2]The Reserve Officers Training Corps.—*Ed.*

US Physics Team places third in Beijing

The host team is tops at the XXV International Physics Olympiad

THE US PHYSICS TEAM earned the third-highest number of points at the XXV International Physics Olympiad, held in Beijing, China, July 11–19, 1994. The Chinese team achieved the highest total, and Germany took second place.

The trip to China provided opportunities for team leaders and members to experience some of the differences between the US and Chinese educational systems. Chinese teachers are very demanding and do not tolerate the carelessness typical of many American high school physics students. Partial credit is not given very often, and small mistakes, such as incorrect plus and minus signs, receive large deductions. For instance, in a collision problem, students were expected to begin by writing the conditions for the conservation of linear and angular momentum, a relationship obtained from the impulse–momentum theorem and two conditions on components of the velocities. Each correct equation received a 0 or 0.8 point, and no partial credit was given. If one equation was missed, the student received no credit for solving the six simultaneous equations.

This style of grading produced a skewed distribution with almost no high scores. As a consequence, only six gold medals, five silver medals, and twenty-two bronze medals were awarded to the 229 competitors from 47 countries. Thus, only 14% of the students received medals.

China was clearly the top team, winning four gold medals and missing a fifth by only 0.05 point. Germany won one gold medal and three silver medals. The remaining gold medal went to Great Britain, and the remaining silver medal went to Poland.

The Chinese organizing committee was very gracious in awarding many special medals to students who performed particularly well on either the theoretical or experimental parts of the exam. The US Physics Team was awarded three gold and two bronze special medals.

Best finish ever

The 1994 US Physics Team achieved the highest US team finish ever. The team was led by Andrew Frey from the North Carolina School of Science and Mathematics, who placed 15th in the competition. He was awarded a bronze medal and a special gold medal for his theoretical work. Daniel Schepler of Beavercreek, Ohio, placed 16th overall and received a bronze medal. He also received a special gold medal for placing seventh on the theoretical portion of the exam. Andy Neitzke of Narberth, Pennsylvania, had a tough day on the theory problems, but came storming back to garner the second-highest score on the experimental problems. He placed 25th overall and was awarded a bronze medal and a special gold medal. Geoffrey Park of Tenafly, New Jersey, tied for 50th place and received an honorable mention and a special bronze medal for his theoretical work. Charlene Ahn from the North Carolina School of Science and Mathematics received a special bronze medal for theory and placed 82nd overall.

Something old, something new

The Chinese designed an interesting mixture of modern and traditional problems for the five-hour theoretical and experimental examinations. The first theoretical problem analyzed the one-dimensional motion of two quarks forming a meson. The quarks were assumed to be ultrarelativistic and to have an interaction that was independent of their mutual separation.

The second problem is presented as this month's Physics Contest (see page 36). In the last theoretical problem, two uniform circular discs with the same mass but different radii suffered an off-center collision. The new element in the problem was the requirement that the relative velocity along the line connecting the centers of the discs keep the same magnitude while the two final velocities of the contact points be the same along the direction perpendicular to this line. Students needed to develop six equations to find the four velocity components and the two angular velocities of the two discs.

The first experimental problem was a cleverly designed optics experiment. Each student was provided with a laser, two photodetectors, two rotatable polarizers with degree scales, a glass beam splitter, and a dielectric plate. The task was to measure the transmission axis of one polarizer, verify the linearity of the photodetector, and measure the index of refraction of the dielectric using Brewster's angle. As an added complication, the students were required to monitor the variation in the intensity of the laser and make suitable corrections.

The second experimental problem was a black-box experiment in which students used an oscilloscope and a frequency generator to determine the wiring diagram and the valves of a resistor, capacitor, and inductor inside the box.

The requirements set by the Chinese grading teams were stringent. For instance, the transmission axis had to be obtained within two degrees, and Brewster's angle to within one degree, to receive full credit. As a result, the top experimental score was only 16.9 out of a possible 20 points.

The wonders of China

Before flying to Beijing, the US team members spent three days honing their skills in the physics department at Stanford University. Then, after a 13-hour flight to Hong Kong, they spent two days trying to adjust their biological clocks. Beijing was a three-hour hop away by plane. After VIP treatment in customs, a quick bus trip brought the five US students to their hotel to make friends with high school students from 46 other nations. The academic leaders likewise made new friends and renewed old acquaintances among physics teachers from around the world.

The cultural and social programs prepared by the Olympiad hosts drew on centuries of Chinese history. Books and pictures are hardly an adequate preparation for the reality of the Great Wall. This engineering marvel snakes its way across the lush mountain terrain at Ba Da Ling pass just outside Beijing. Two hours of walking along the Great Wall gave but a brief glimpse of its military and cultural significance in the history of China.

Massive size, extraordinary art, and very long histories are traits of the Summer Palace, the Imperial Palace, the Forbidden City, the People's Hall, and the Ming Tombs. An evening of acrobats, jugglers, and magicians offered another glimpse of Chinese culture.

The Chinese people themselves, however, were the richest components of the visit. Beijing, with its 11 million people, is a thriving metropolis with hundreds of buildings under construction, thousands of cars, buses, and cars on wide boulevards, millions of bicycles, and a population with nerves of steel in the face of churning traffic patterns. The streets of Beijing seemed always to be teeming with people in a constant "street fair" atmosphere.

1994 team and sponsors

Twelve states were represented on the 1994 US Physics Team. In the list below, members who represented the team in Beijing are marked by an asterisk, and each member's physics teacher is noted in parentheses.

Matthew Ahart, Sherman Oaks, California (John Feulner, Harvard–Westlake School)

***Charlene Ahn**, Kinston, North Carolina (Hugh Haskell, North Carolina School of Science and Mathematics)

Gil Barretto, Ridley Park, Pennsylvania (Paul Pomeroy, Archmere Academy)

Rhiju Das, Norman, Oklahoma (Xifan Liu, Oklahoma School of Science and Mathematics)

Brian Doherty, Richmond, Indiana (H. Fakhruddin, Indiana School of Science, Mathematics, and the Humanities)

James Dunlop, Libertyville, Illinois (Theodore Vittitoe, Libertyville High School)

Ron Fertig, Cherry Hill, New Jersey (Hirendra Chatterjee, Cherry Hill High School West)

***Andrew Frey**, *bronze medal*, Winston-Salem, North Carolina (Hugh Haskell, North Carolina School of Science and Mathematics)

Brian Leibowitz, Manalapan, New Jersey (Jim Kovalcin, Manalapan High School)

Paul Lujan (alternate), San Francisco, California (Richard Shapiro, Lowell High School)

***Andrew Neitzke**, Narberth, Pennsylvania (Robert Schwartz, Harriton High School)

Mark Oyama, Honolulu, Hawaii (Carey Inouye, Iolani School)

***Geoffrey Park**, Tenafly, New Jersey (Zenon Ushak, Tenafly High School)

Aaron Pierce, Shaker Heights, Ohio (John Schutter, Shaker Heights High School)

***Daniel Schepler**, Beavercreek, Ohio (Margo DeBrosse, Beavercreek High School)

Mike Shubov, Lubbock, Texas (Jeff Barrows, Lubbock High School)

Ian Spielman, Albuquerque, New Mexico (David Glidden, Albuquerque Academy)

Mary Spikowski, Bay Village, Ohio (Timothy Wagner, Bay High School)

Doug Stone, Libertyville, Illinois (Theodore Vittitoe, Libertyville High School)

Aurelio Teleman, East Setauket, New York (Tania Entwistle, Ward Melville High School)

The 1994 US Physics Team was organized by the American Association of Physics Teachers (AAPT) with the financial support of the American Institute of Physics and contributions from other physics societies, industry, and individuals. The Principal Sponsor of the 1994 US Physics Team was the Physical Sciences Department of the IBM Research Division.

The XXVI International Physics Olympiad will be held in Australia, July 5–12, 1995. Teachers of students wishing to compete for positions on the 1995 US Physics Team who do not receive application materials by mid-December should contact Maria Elena Khoury at the American Association of Physics Teachers, One Physics Ellipse, College Park MD 20740-3845 (phone: 301 209-3344).

—Based on a report in the September 1994 AAPT *Announcer*

American team garners six gold medals at 35th IMO

Six perfect scores puts the US in first for the first time ever

COMPETING AGAINST TEAMS representing 69 countries, a team of six American high school students placed first in the 35th International Mathematical Olympiad (IMO), held July 8–20 in Hong Kong, with six perfect scores. The top five teams were, in order, the United States, China, Russia, Bulgaria, and Hungary.

This is the first time that a team has scored a perfect score in the IMO. Each of the six members of the US team scored the maximum number of points (42) on the nine-hour exam and each received a gold medal.

The members of the team are

Jeremy Bem, Ithaca High School, Ithaca, New York;

Aledsandr L. Khazanov, Stuyvesant High School, New York City;

Jacob A. Lurie, Montgomery Blair High School, Silver Spring, Maryland;

Noam M. Shazeer, Swampscott High School, Swampscott, Massachusetts;

Stephen S. Wang, Illinois Mathematics and Science Academy, Aurora, Illinois;

Jonathan Weinstein, Lexington High School, Lexington, Massachusetts.

The US team was led by Professor Walter E. Mientka of the University of Nebraska–Lincoln, executive director of the American Mathematics Competitions. The team was chosen on the basis of performance in the 23rd annual United States of America Mathematical Olympiad (USAMO), held earlier this year. The winners of the 1994 USAMO were honored on June 6 at the National Academy of Sciences in Washington, D.C. Prior to the competition, the US students participated in a monthlong summer program at the US Naval Academy under the direction of professors Anne Hudson, Titu Andreescu, and Paul Zeitz.

The Mathematical Olympiad is a program of the Mathematical Association of America. It is cosponsored by the American Association of Pension Actuaries, the American Mathematical Association of Two-Year Colleges, the American Mathematical Society, the American Statistical Society, the Casualty Actuarial Society, the Mathematical Association of America, Mu Alpha Theta, the National Council of Teachers of Mathematics, and the Society of Actuaries. Financial support is provided by the Army Research Office, the Office of Naval Research, Microsoft Corporation, and the Matilda R. Wilson Fund.

Problems from the 35th IMO

1. Let m and n be positive integers. Let $a_1, a_2, \ldots, a_m$ be distinct elements of $\{1, 2, \ldots, n\}$ such that whenever $a_i + a_j \le n$ for some i, j, $1 \le i \le j \le m$, there exists k, $1 \le k \le m$ with $a_i + a_j = a_k$. Prove that

$$\frac{a_1 + a_2 + \cdots + a_m}{m} \ge \frac{n+1}{2}.$$

2. ABC is an isosceles triangle with $AB = AC$. Suppose that (i) M is the midpoint of BC and O is the point on the line AM such that OB is perpendicular to AB; (ii) Q is an arbitrary point on the segment BC different from B and C; (iii) E lies on the line AB and F lies on the line AC such that E, Q, and F are distinct and collinear. Prove that OQ is perpendicular to EF if and only if $QE = QF$.

3. For any positive integer k, let f_k be the number of elements in the set $\{k + 1, k + 2, \ldots, 3k\}$ whose base 2 representation has precisely three 1's. (a) Prove that for any positive integer m, there exists at least one positive integer k such that $f(k) = m$. (b) Determine all positive integers m for which there exists exactly one k with $f(k) = m$.

4. Determine all ordered pairs (m, n) of integers such that $(n^3 + 1)/(mn - 1)$ is an integer.

5. Let S be the set of real numbers greater than -1. Find all functions $f: S \to S$ satisfying two conditions: (i) $f(x + f(y) + xf(y)) = y + f(x) + yf(x)$ for all x and y in S; (ii) $f(x)/x$ is strictly increasing for $-1 < x < 0$ and for $0 < x$.

6. Show that there exists a set A of positive integers with the following property: for any infinite set S of primes, there exist positive integers $m \in A$ and $n \notin A$, each of which is a product of k distinct elements of S for some $k \ge 2$.

—From materials submitted by Walter Mientka and Andy Liu

SOLUTIONS IN THE NEXT ISSUE

Daniel van Vliet, a Canadian high school student of Dutch heritage, accompanied Prof. Andy Liu, a member of the IMO Problem Selection Committee, to Hong Kong. Here is his account of his stay at the IMO.

AS A RESULT of a somewhat unusual arrangement, I had an opportunity to serve as a guide for the Dutch team at the 1994 IMO in Hong Kong. Guides for the IMO teams are usually local university students who accept the position of guide as a summer job. For one thing, I am not a university student. At the time of the trip, I had just finished grade 11. Moreover, I could hardly be defined as even remotely familiar with Hong Kong, let alone a local. Apart from a pit stop on July 1, I didn't arrive in Hong Kong until July 8.

The first thing one notices when one leaves the plane upon arrival is the heat. Hong Kong is in a tropical coastal area and has a sticky climate. The average daytime temperature of 35°C comes as quite a shock to someone from Edmonton. However, I learned to tolerate the heat within a day or two.

The guides were brought together on a quiet street corner on July 10 and sorted into two groups. It turned out that there was no single facility capable of housing all the competitors and deputy leaders. I was among the guides sent to a camp called Sai-Kung Outdoor Recreational Facility. Only teams with all male participants were sent there. Despite this, there were some female guides at the camp.

Hong Kong is a mosaic of tall office buildings, apartments, and shopping centers relentlessly bustling with heavy traffic from morning until night. As a result of this striking first impression, the setting of our lodging was somewhat surprising. After a seemingly lengthy drive through progressively open landscapes, we ended up in a pleasant, woodsy area. The facilities seemed out of place in a city where land was at such a premium. They included a tennis court, swimming pool, soccer field, and various other recreational facilities. We had the remainder of July 10 to become acquainted with the camp and receive last-minute instructions.

On the 11th, the teams were scheduled to arrive at various times throughout the day. We all went to the airport in the morning to wait for them. It turned out that my team (the Dutch team) would be landing quite late and, not being familiar with Hong Kong, wouldn't be able to just head into town find something to do. As all the other guides were locals, they had no problem with this. However, one of the guides showed me some of the sights I should show my team, and I passed most of the day preparing myself in this way.

When the team arrived, there was a somewhat awkward period in which I was the only source of information the team had. Not only that, we didn't know each other, so we were frequently asking each other's name, background, and so on. The students and deputy leaders went to the camp; the team leaders and observers stayed at a hotel. When we arrived at the camp, after a short orientation, we were all quite tired and retired quickly for the night.

We were allowed one day of (relative) rest. The usual breakfast time—between 7:00 and 8:00—was relaxed to allow for jet lag. Except for this anomaly, however, the day had quite a rigorous schedule. The first day's activities consisted of a museum visit, the opening ceremonies, and a welcoming dinner.

The next two mornings were taken up with the first and second contest papers. The afternoons were spent sightseeing and visiting the Hong Kong Science Museum. After the papers were written and the deputy leaders headed back to the hotel, the general atmosphere became a bit less serious as, understandably, they had a somewhat adult effect on the proceedings at the camp.

During the sightseeing and socializing, the students exchanged addresses, phone numbers, and even e-mail addresses. I do not believe that there was any student who left the IMO without making any new friends. This is the true value of this type of competition.

I was grateful to be able to share this experience with so many people from so many different countries. This, I believe, is the true value of the IMO as an educational tool. The interaction with other cultures is a valuable experience that cannot be matched by classroom instruction.

Competitive computing in Stockholm

The 1994 International Olympiad in Informatics

by Donald T. Piele

THE FOUR MEMBERS OF THIS year's US team to the International Olympiad in Informatics (IOI), had just enough time to call mom and dad and announce that they were not coming home for another week before boarding the plane for Sweden. They had just been selected as members of the 1994 US team to the 6th annual IOI to be held in Stockholm, Sweden, July 3–July 10, 1994. The winners had completed a week-long competition, the USA Computing Olympiad, June 25–July 2, hosted by the University of Wisconsin–Parkside and sponsored by The Center for Excellence in Education, USENIX, and IBM. The team—James Ayers and Mehul Patel of Houston, Texas; Brian Dean of Charlotte, North

Carolina; and Hubert Chen of Fort Washington, Pennsylvania—was headed for Sweden. They were accompanied by Greg Galperin (the deputy team leader and a graduate student at MIT) and myself.

The site of this year's competition was the Royal Institute of Technology at Riksapplet, with accommodations nearby at Hotel Najaden for the teams and in private apartments for the team leaders. Students who studied at the institute had graciously given up their apartments for the week to help meet the housing requirements. With 48 countries, each accounting for six people, housing was required for over 300 people including guests. Both the accommodations and the competition site were superb.

Competition and exploration

The first day began with opening ceremonies that included a parade of the flags from each of the participating countries. Mehul Patel, our only veteran IOI team member, was selected by the team to carry the flag for the United States. The official IOI flag was brought from Argentina and handed over to Yngve Lindberg, the president of IOI '94. Music written especially for the occasion by Johannes Dominique, 15, was played on a computer connected to a sequencer and synthesizer. Afterwards we boarded buses for a short trip to the center of Stockholm and spent the rest of the beautiful afternoon on the water touring the waterways of Stockholm—the Venice of the North.

Day two was devoted to the first round of programming challenges. The next day offered a chance to sightsee. Our touring day began with a bus ride to Ericsson, an international telecommunications company headquartered in Stockholm and the sponsor for the day's activities. We were given a scientific talk on the future of telecommunications and a demonstration of unusual robots. After lunch we boarded our buses for a short ride downtown to one of Stockholm's leading attractions—the Vasa. This seventeenth-century warship was rescued from the bottom of the harbor in 1961 after resting for 330 years in the deoxygenated water of the Baltic Sea. It survived in this water almost intact and has been restored to it original splendor. Today it stands as a unique witness to seventeenth-century shipbuilding and life at sea.

We were bused back to the Royal Institute of Technology at Riksapplet for dinner. Afterwards, a seminar on "Alias," one of the world's most advanced software tools for animation, visualization, and design, was presented.

The second round of competition on day four was a repeat of the first round, but with harder problems. On the fifth day of our visit, we had a chance to explore the Swedish archipelago. Sweden is blessed with nearly 20,000 islands, and many of them are within a short cruise from the Stockholm harbor. Some of the islands are so small they have only a single red log cabin on them—a summer vacation home. We passed many of them on our steamboat ride to the island of Uto, easily recognized by its big windmill. We stopped off for an island tour conducted by a hearty native who lives there year-round. Uto is now a vacation destination, but it was once famous for its iron mine—the first one in Sweden. The team leaders took a separate boat back and stopped off at Rosenon—a vacation resort where we met, had dinner, and made the all important cutoff decisions for the three classes of medals.

Awards ceremony

Since Stockholm is the setting for the annual Nobel prize awards, it was a wonderful idea to hold the final awards ceremony in the same location, the beautiful City Hall situated on the water in the center of the city. Most famous of all the rooms in this magnificent red brick building is the Golden Hall, where 18 million pieces of mosaic made out of ceramic, glass, and 24-carat gold leaf cover the walls. The ceremonies were held in the grand entrance hall, the Blue Room. To add to the elegance, the lovely Queen of Lake Malaren, Charlotte Mangborg, presented the medals to the very proud winners. The IFIP trophy for the highest score (195) went to Victor Bargatchev from Russia. He lead the Russian team that finished first among the 48 countries participating. Below is a listing of the total points and medals (gold, silver, bronze) for the top seven countries.

	Pts	G	S	B
Russia	617	3	1	–
China	558	3	–	1
Germany	492	2	1	1
Hungary	475	2	–	1
USA	463	1	1	2
Czech Rep	459	1	2	1
Romania	444	1	2	1

The US team finished in fifth place— up from seventh last year— with medals for everyone: Mehul Patel (gold, 156 points), Brian Dean (silver, 121 points), Hubert Chen (bronze, 95 points), and James Ayers (bronze, 91 points).

The ceremonies ended with the introduction of Ries Kock, the team leader from the Netherlands, the host of the 1995 IOI. Since the beginning of IOI in 1989, the Netherlands has brought a team consisting of two girls and two boys, and Ries has pushed for a more balanced representation of boys and girls from other countries. However, the number of women participating overall has not gone above 5%. In the Netherlands, it would be difficult to find sponsors for an event with such a low percentage of women participants. To help stimulate the participation of girls at IOI, Ries extended an invitation to each country to bring five students in 1995, as long as the team includes at least one woman. This is understandably a very controversial issue with some team leaders, but it appeared that the willingness of the Dutch to invite another student from each country in an effort to actively encourage participation by women silenced the opposition, at least for the present. When one is around the Dutch delegation for a week, it is easy to believe they know what they are doing. No one has more fun and reaches out to more people than do the Dutch. To them, the IOI is more than a competition. It is also a chance for young people from various countries to make lifelong friends.

Bouquets of flowers were handed out by Queen Charlotte to members of a very deserving Swedish organizing committee. They not only arranged for a week of perfect weather, but they also conducted an innovative Olympiad with many new time-saving improvements. IOI '94 president Yngve Lindberg's years of experience, and his strong leadership ability, shoved the bar to a new height. Just to equal this mark will be a challenge to those who follow.

I'd like to thank the Center for Excellence in Education (CEE), USENIX, IBM, and the University of Wisconsin–Parkside for sponsoring the USA Computing Olympiad at UW–Parkside. Special thanks are due to CEE, which provided funding airfare to the 3rd USACO and to the 6th IOI in Sweden. These sponsors make US participation at the IOI possible.

I would also like to thank the USACO staff, who gave freely of their time during the year to help select the 15 finalists and then traveled to UW–Parkside in June to select and train the final four. Kudos go to Rob Kolstad (chief of staff), Greg Galperin (deputy team leader), Nate Bronson, and Shawn Smith.

For more information about the International Olympiad in Informatics, write or call

Donald T. Piele
USACO Director
University of Wisconsin–Parkside
Box 2000
Kenosha WI 53141-2000
piele@cs.uwp.edu
414 595-2231 (O)
414 634-0868 (H)

PROBLEMS FROM IOI '94 WILL APPEAR IN THE NEXT ISSUE

Bulletin Board

First Step to a Nobel Prize

The Polish Academy of Sciences has announced the winners of the second annual international competition in physics for high school students. They are Can Altineller (Turkey), Anton A. Belyaev (Ukraine), Z. Cournia (Greece), Janko Isidorovic (Yugoslavia), Marcus Meuller (Switzerland), Samuel F. Schaer (Switzerland), and Michal Rewienski (Poland). These winners received a certificate and an invitation to spend one month in Poland doing research.

The deadline for the submission of research papers for the third annual competition is March 31, 1995. Interested students should contact Dr. Waldemar Gorzkowski (e-mail: gorzk@gamma1.ifpan.edu.pl; fax: 022-430926; phone: 022-435212), Institute of Physics, Polish Academy of Sciences, al. Lotnikow 32/26, Warszawa, Poland.

Peace Corps

Educational needs in the world's developing nations are immense and immediate. Thousands of teaching, teacher training, and curriculum development posts stand unfilled in secondary schools, universities, and ministries of education around the world.

Filling some of those vacancies—in Africa, Asia, central and eastern Europe, the former Soviet Union, and Latin America—is a Peace Corps job. Right now, at the request of 68 countries, the Peace Corps is looking for 1,600 men and women to become Volunteers and teach (or train others to teach) science and seven other disciplines. Men and women who serve as Peace Corps Volunteer teachers find that students in developing nations regard education as a precious gift, not a right.

A Volunteer assignment is ideal for folks looking for a mid-career break, as well as retirees who wish to continue their work in education. There is no upper age limit—in fact, the Peace Corps values maturity greatly. The average Volunteer is 32 years old, and 10% of the 6,500 Volunteers now in service are 50 or older. Three are in their eighties. Couples are eligible, too, as long as both spouses qualify as Volunteers.

The period of service is 24 months, plus three months of training. Volunteers are paid a monthly allowance that permits comfortable living at the level of their counterparts in the host country. In addition, $200 accrues for each month of service and training, which is paid upon completion of service. The Peace Corps provides all medical and dental care as well as transportation to and from the country of assignment.

For more details, you can get in touch with your local Peace Corps office by calling 800 424-8580.

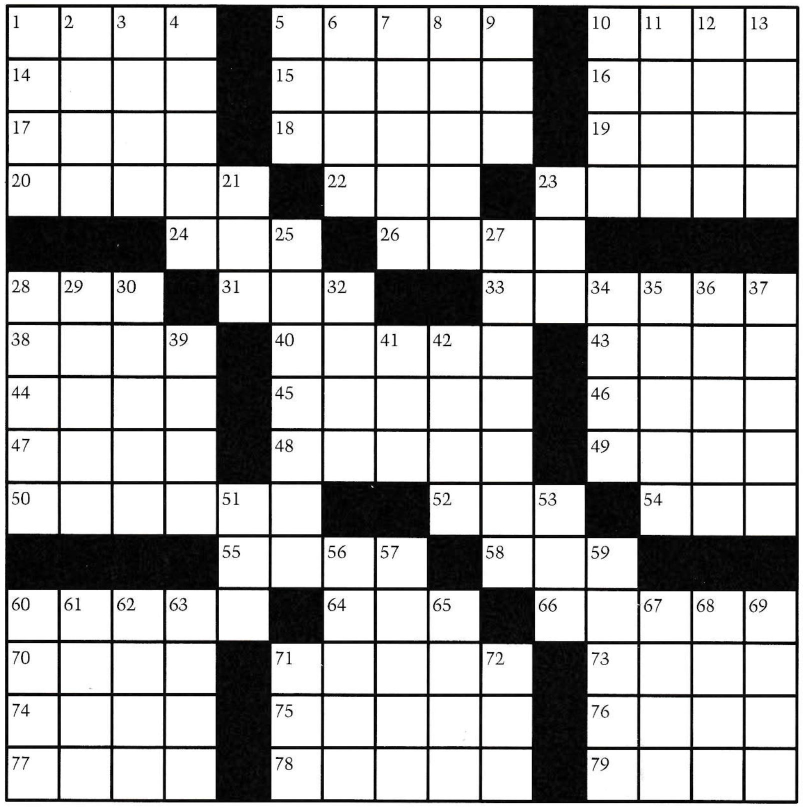
Fearless symmetry

Quantum readers who have enjoyed reading about Penrose tilings and other symmetry-related topics will want to look into *Symmetry: A Unifying Concept* by István and Magdolna Hargittai. This 221-page, profusely illustrated compendium is a pleasure to browse through yet fully repays the reader who stops to dive in more deeply. The authors range through chemistry, biology, mathematics, engineering, art, and architecture, and many of the photographs are products of their own globetrotting.

Symmetry is published by Shelter Publications and is distributed by Ten-Speed Press. Check your local bookstore, or order copies from Shelter Publications, Inc., PO Box 279, Bolinas CA 94924 ($18 per copy + $3 shipping & handling).

Criss × cross science

by David R. Martin



Across

1 Arrow poison
5 Named organism group
10 47,834 (in base 16)
14 Japanese palanquin
15 One-celled animal
16 60,845 (in base 16)
17 Type of parity
18 1964 Physiol. Nobelist
19 Fourth dimension
20 State of matter
22 ___ Angeles
23 H_2O
24 Food fish
26 Compacted snow
28 Winter month
31 English weight unit
33 C_2H_4
38 Elemental particle
40 Heathen
43 Abba of Israel
44 Rose: comb. form
45 ___ kingdom (botanist's domain)
46 Japanese island
47 44,732 (in base 16)
48 Tree dwelling primate
49 Set ___ (leave port)
50 CO_2
52 Uncle ___
54 Sounds of hesitation
55 Mine products
58 ___ Vegas
60 Units of potential difference
64 System of units: abbr.
66 Energy per unit time
70 Toward shelter
71 700,139 (in base 16)
73 A pigment
74 Equipment
75 Marathon runner ___ Bikila
76 Assert
77 Okinawan seaport
78 Radioactive gas
79 Biologist Francesco ___ (1626?–1698)

Down

1 Hawaiian instruments
2 Peacock constellation
3 ___ fiber (from the gebang palm)
4 Type of boom
5 Prescription abbr.
6 C_5H_{11}
7 Element 54
8 Fat
9 Nickname for Nancy
10 ___ particle (emitted electron)
11 Mine tunnel
12 British archaeologist ___ Kathleen Mary Kenyon (1906–1978)
13 ___ wax (ozocerite)
21 ___ product (or inner product)
23 Moist
25 Type of radar
27 Dorsal's opposite
28 Capacitance unit
29 Former hypothetical medium of space
30 ___ trap
32 1936 Physiol. Nobelist
34 1936 Physics Nobelist
35 965,294 (in base 16)
36 Point opposite zenith
37 Compounds containing -CH=C(OH)-
39 1601
41 Social visit
42 Bodily orifice
51 Trig. function
53 Function
56 981,690 (in base 16)
57 Distance divided by time
59 Type of cell or wind
60 Swedish oceanographer ___ Ekman (1874–1954)
61 Olive tree genus
62 Jacob's wife
63 Million million: pref.
65 Fat: comb. form
67 ___-particle duality
68 Supplemented
69 Soprano Grist
71 Swiss river
72 Franklin

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE SEPTEMBER/OCTOBER PUZZLE

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

ANSWERS, HINTS & SOLUTIONS

Math

M126

Since $1/x > 1/n$, we have $x < n$, so we can put $x = n - i$, where i is an integer, $1 \le i \le n - 1$. Then

$$y = \left(\frac{1}{x} - \frac{1}{n}\right)^{-1}$$

$$= \left(\frac{1}{n-i} - \frac{1}{n}\right)^{-1}$$

$$= \frac{n(n-1)}{i}.$$

We see that our equation has no positive integer solutions for $n = 1$ (since $x = n - 1 \ge 1$), and for any $n > 1$ it has at least one solution (with $i = 1$): $x = n - 1$, $y = n(n - 1)$. A necessary and sufficient condition for the existence of another solution is the divisibility of $n(n - i) = n^2 - ni$ by a certain i, $1 < i < n$, which is equivalent to the divisibility of n^2 by such i (this would ensure that y is an integer). But it's clear that such a number i exists if and only if n is composite, which completes the proof.

M127

The statement can be proved by induction over N. For $N = 1$ the only number a satisfying the condition by itself is zero (we must have $a = -a$).

Suppose the statement is true for $N - 2$ ($N \ge 3$) numbers, and consider an arrangement of N numbers a_1, $a_2, \ldots, a_N$ (numbered clockwise) satisfying the condition (fig. 1). At least one of them must be obtained from the previous one by reversing the sign; otherwise, we'd have $a_1 = a_N + 1 = a_{N-1} + 2 = \ldots a_1 + N$, which is impossible.

So we can choose any pair of two numbers that differ only in sign and

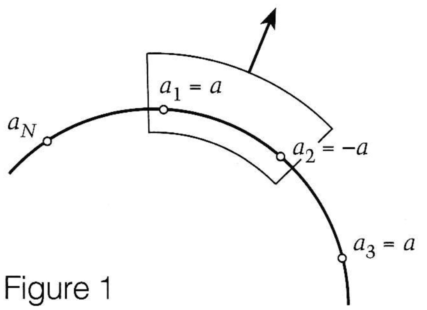


Figure 1

label them $a_1 = a = -a_2$. Now we have two possibilities for the value of a_N: it can be $-a$ or $a - 1$. We also have two possibilities for a_3: it can be a or $-a + 1 = 1 - a$. We can show that no matter what these values are, if we delete a_1 and a_2, then the deleted arrangement will satisfy the conditions of the problem. The induction hypothesis then tells us that the deleted arrangement must consist of integers, and any integer m is found as many times as $-m$. We then show that the same is true of the original arrangement, thus completing the induction.

Suppose, for instance, that $a_3 = a$. Then no matter which value of a_N we choose, it's not hard to check that the deleted arrangement satisfies the conditions of the problem: a_3 is obtained from a_N either by adding 1 (if $a_N = a - 1$) or by reversing the sign (if $a_N = -a$). Therefore, the deleted arrangement must consist of integers, with an equal number of copies of each integer and its negative. But then the original arrangement, which uses no other numbers, also consists of integers. Since the new arrangement adds one copy of a and one copy of $-a$ to the old, it follows that it also contains an equal copy of each integer and its negative.

The reader is invited to check that similar arguments hold if $a_3 = 1 - a$ (fig. 2) no matter which value of a_N we choose. This completes the induction.

While the case $N = 3$ is covered by

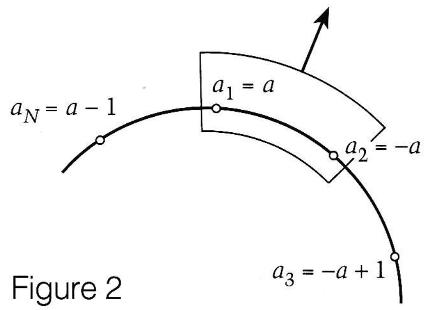


Figure 2

the induction, it may be instructive to derive the result directly for this case. For even values of N the conclusion of the problem is false; the reader may want to construct counterexamples.

M128

We can look at the diagram as "centered" around B and C, which don't move during all the transformations described. Then point A stays one unit away from B. It follows that the position of point A is uniquely determined by the signed angle $\alpha = \angle BCA$ (the sign is positive if the "shortest" rotation from the ray CB to CA is counterclockwise, minus otherwise); similarly, point D is determined by the signed angle $\delta = \angle CBD$. As the base angles of isosceles triangles, these angles take values between $\pi/2$ and $-\pi/2$. Let's see what happens to the pair (α, δ) under our transformations.

Let A' be the reflection of A about BD. For any point X on the extension of CB (fig. 3) the signed angle

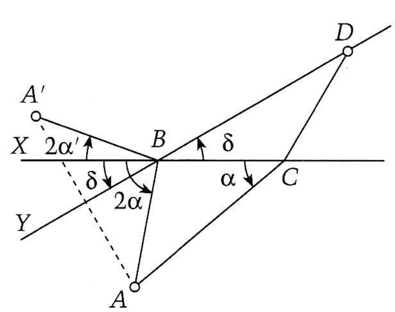


Figure 3

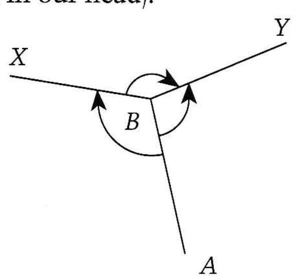
$\angle XBA = 2\alpha$ (as an exterior angle of the isosceles triangle ABC); similarly, $XBA' = 2\alpha'$, where $\alpha' = \angle BCA'$. By construction, the extension BY of DB bisects angle ABA'. Expressing equal signed angles ABY and YBA' in terms of the signed angles δ, 2α, and $2\alpha'$ (see figure 3), we have

$$\angle ABX + \angle XBY = \angle ABY = \angle YBA' = \angle YBX + \angle XBA',$$

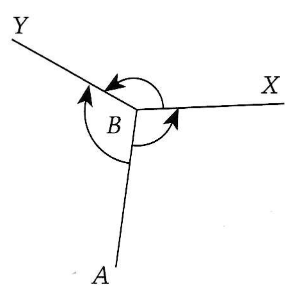
or $-2\alpha + \delta = -\delta + 2\alpha'$, or $\delta = \alpha + \alpha'$. So $\alpha' = \delta - \alpha$—that is, our first reflection replaces the pair (α, δ) that defines the entire quadrilateral with $(\delta - \alpha, \delta)$:

$$(\alpha, \delta) \to (\delta - \alpha, \delta).$$

To be honest, this argument is a bit fraudulent: it depends rather heavily on the diagram. In fact, the relation $\angle ABY = \angle ABX + \angle XBY$ used above is true in general only "modulo 2π"—that is, if the difference between the left and right sides is a multiple of 2π (see the examples in figure 4). This remark applies to $\angle YBA'$ as well. So the correct formula for α' is $\alpha' = \delta - \alpha + k\pi$ with a certain integer k ($k = 0$, 1, or -1). However, knowing $\delta - \alpha$, we can always uniquely determine k from the condition $-\pi/2 < \alpha' < \pi/2$. This allows us to omit the terms $k\pi$ in the formulas for our transformations (keeping them in our head).



$\angle ABY = \angle ABX + \angle XBY + 2\pi$



$\angle ABY = \angle ABX + \angle XBY - 2\pi$

Figure 4

The transformation by the reflection about BD is the same as the one above except that the terms in the pair exchange roles: the first remains the same, and the second is replaced by the difference between the first and the second. Alternating these transformations, we successively get

$$(\alpha, \delta) \to (\delta - \alpha, \delta) \to (\delta - \alpha, -\alpha) \to (-\delta, -\alpha) \to (-\delta, \alpha - \delta) \to (\alpha, \alpha - \delta) \to (\alpha, \delta)$$

—we returned in six steps! (To be more exact, we only know that after six steps the angles are $\alpha + n\pi$ and $\delta + m\pi$. But $n = m = 0$, because otherwise these angles wouldn't fall into $(-\pi/2, \pi/2)$.) (N. Vasilyev, M. Kontsevich)

M129

Consider three groups of persons: the group N of unsociable weirdos (let's call them *normal*), the group W of all the other weirdos, and the group U of all the other (not weird) unsociable persons. Let n, w, and u be the numbers of persons in each of these groups, respectively, and a the number of pairs of acquaintances, one from W and the other from U. We have to prove that $2 + n < u + n$, or $w < u$.

First note that a person in N cannot have an acquaintance in W, since W is made up of sociable people. Since acquaintanceship is symmetric, it follows that a person from W can have acquaintances only in U and must have more than 10 of them. So a, which counts the number of times a person from W is in an "acquaintanceship pair" with a person from U, is at least $10w$—that is, $10w \le a$. Now, a person from U can have acquaintances anywhere, but has fewer than ten of them. So the number of times a person from U is in an acquaintanceship pair with a person from W (or with anyone else, for that matter) is less than $10u$—that is, $a < 10u$. It follows that $10w < 10u$, or $w < u$.

M130

It's well known that a quadrilateral has an inscribed circle (can be circumscribed about a circle) if and only if the sum of its opposite sides are equal. This can be proved, for example, by noting pairs of equal tangent segments from each of the vertices of the quadrilateral. It turns out that there are two other conditions that are necessary and sufficient for a quadrilateral to possess an inscribed circle. We'll use these in our solution.

Let sides AB and DC of quadrilateral $ABCD$ intersect (when extended) at points E, and let sides AD and BC intersect at point F (fig. 5). Then either of the following two conditions are necessary and sufficient for the quadrilateral to be circumscribed around a circle:

$$EB + FB = ED + FD, \qquad (1)$$
$$EA - FA = EC - FC. \qquad (2)$$

First we'll show that each is necessary. Suppose $ABCD$ has an inscribed circle. Let a, b, c, d, e, f be the lengths of the tangents from points A, B, C, D, E, and F to the inscribed circle. Then $EB = e - b$, $FB = f + b$, $ED = e + d$, $FD = f - d$, and both sides of equation (1) are equal to $e - f$. An analogous proof holds for equation (2).

Now let's demonstrate the sufficiency of condition (1). Suppose the condition holds. We inscribe a circle in triangle AED (this is always possible) and draw a tangent to it from point F. Let this tangent intersect EA and ED at points B_1 and C_1, respectively. It follows from the necessity of the condition (proven above) that $EB_1 + FB_1 = ED + FD = EB + FB$; that is, $FB = FB_1 + (EB_1 - EB) = FB_1 \pm BB_1$ (whether this is a sum or a difference depends on which direction along the tangent point B is from point B_1). If B

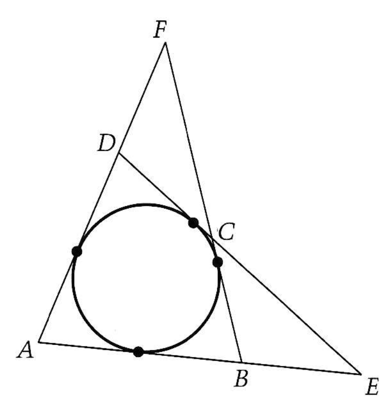


Figure 5

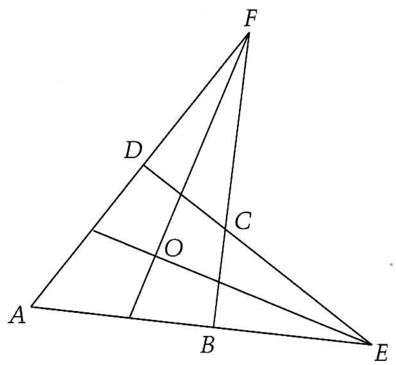


Figure 6

and B_1 are different points, this equation contradicts the triangle inequality for points F, B, and B_1. It follows that $B_1 = B$; that is, that the circle we constructed is tangent to side BC of the given quadrilateral. The sufficiency of condition (2) can be established in the same way.

Now it's not difficult to prove the assertion of the problem. Let O be the intersection of the lines dividing the quadrilateral (fig. 6). If the quadrilaterals that include points B and D have inscribed circles, then condition (1) guarantees that $EB + FB = EO + FO = ED + FD$. Condition (1) then lets us conclude that the original quadrilateral possesses an inscribed circle. For the quadrilaterals that include points A and C, the assertion follows from condition (2) and the equation $EA - FA = EO - FO = EC - FC$.

Physics

P126

Let's choose a coordinate system with the origin at the foot of the incline and the x-axis pointing up the incline. If M is the mass of the train and x is the length of the train that is on the incline, the mass of the part of the train on the incline is Mx/L. Then Newton's second law yields

$$Ma = -\frac{Mx\, g\sin\alpha}{L},$$

or

$$a = -\frac{g\sin\alpha}{L}x.$$

This equation is none other than the equation for harmonic oscillations. The period T of the oscillations is

$$T = 2\pi\sqrt{\frac{L}{g\sin\alpha}}.$$

The time t required for the train to stop is equal to one quarter of the period—that is,

$$t = \frac{T}{4} = \frac{\pi}{2}\sqrt{\frac{L}{g\sin\alpha}}.$$

Note that the answer does not depend on the momentum of the train or how much of it goes up the incline. This same behavior shows up in the independence of the period of a pendulum on its amplitude.

P127

Let's consider the equilibrium condition for any of the floating vessels: the weight of the vessel is counterbalanced by the net force due to the difference between the water pressure above and below the bottom of the vessel. Thus, both before and after water is added to any vessel, the difference in the water levels inside and outside each vessel remains the same (except for the outer vessel, which is not floating; however, the area of its bottom is so large that the water level doesn't change by an appreciable amount). This means that the position of water levels in all the vessels remains constant relative to the ground.

Thus, after the water is added, the water level in the smallest vessel will not change relative to the ground. Therefore, the bottom of this vessel will sink by

$$h = \frac{V_0}{S_0}$$

—that is, by the height of the added layer of water.

P128

Since the water is not being mixed, all the heat liberated as the water freezes is dissipated into the atmosphere only. At any moment the flow of heat is directly proportional to the temperature difference ΔT between the water and the air and inversely proportional to the thickness of the ice x. Therefore, for a change in the ice's thickness Δx during a period of time Δt we have

$$\Delta x \sim \frac{\Delta T}{x}\Delta t,$$

or, taking into account that ΔT = constant,

$$x\Delta x \sim \Delta t.$$

From here it follows that

$$x^2 \sim t,$$
$$x \sim \sqrt{t}.$$

In 1,000 hours the thickness of ice will be

$$x_{1000} = x_{10}\sqrt{\frac{1000}{10}} = 1 \text{ m}.$$

P129

Let's begin with the nodes A and B. We draw the circuit anew as shown in figure 7. Then we connect two identical batteries of emf $\mathcal{E}_1 = \mathcal{E}_2 = \mathcal{E}$ with two identical resistors $R \gg r$ to the points A and B (we neglect the internal resistance of the batteries) (fig. 8 on the next page). Let's assume that $r = 1\ \Omega$ and choose the values of E and R such that $\mathcal{E}/R = 1$ A.

First we consider the connection of one source to the node A, which results in a branching of the current flowing into node A (fig. 9). By symmetry the current along each of the three ribs from node A is $I_1/3$, where $I_1 = \mathcal{E}/(R + r_x) \cong 1$ A and r_x is the resistance between nodes A and B). Similarly, with the source $\mathcal{E}_2$

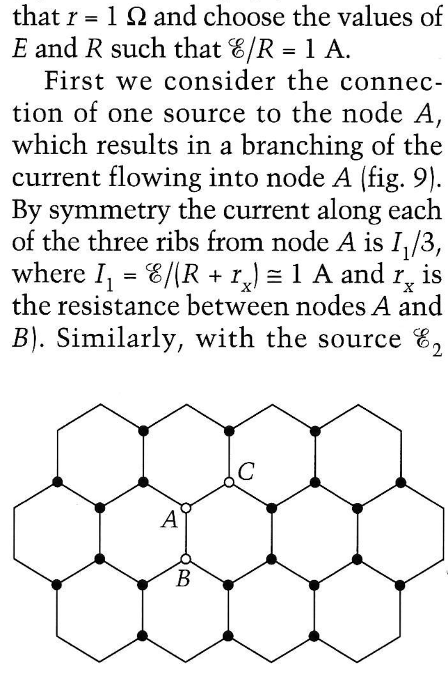


Figure 7

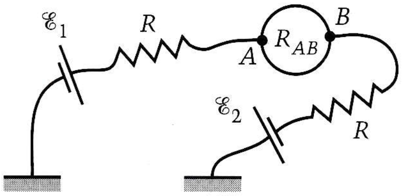


Figure 8

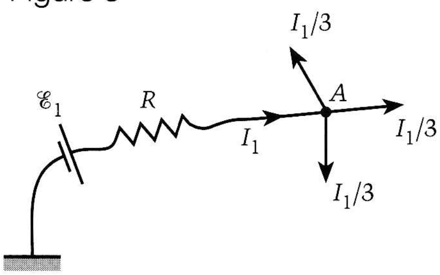


Figure 9

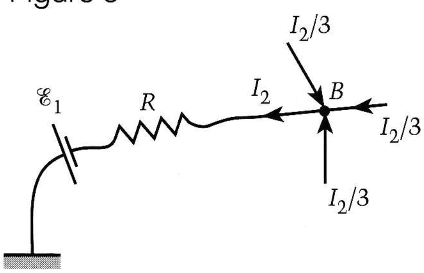


Figure 10

connected to node B, the current $I_2 = \mathscr{E}/(R + r_x) \cong 1$ A flows out of it. The current comes from the three nodes closest to B and goes off to "infinity" via the source (fig. 10).

Now let's connect both batteries to nodes A and B. Then from the principle of superposition and the condition $R \gg r$ we obtain a current $I_{AB} = (I_1 + I_2)/3 = 2/3$ A flowing in the conductor AB, which results in the voltage drop

$$V_{AB} = \tfrac{2}{3}\,\text{A} \cdot 1\,\Omega = \tfrac{2}{3}\,\text{V}.$$

On the other hand, the circuit is fed a total voltage $2\mathscr{E}$, and a current $I \cong 2\mathscr{E}/2R = 1$ A is flowing into it. Consequently,

$$R_{AB} = \frac{V_{AB}}{I} = \tfrac{2}{3}\,\Omega = \tfrac{2}{3}r.$$

The case when the voltage is applied to nodes A and C is absolutely analogous to the case considered above:

$$R_{AC} = \tfrac{2}{3}\,\Omega = \tfrac{2}{3}r.$$

All that remains is to find the resistance between nodes B and C. Connecting one source to B results in a current $I_1/3$ flowing in rib AB and a current $I_1/6$ in rib BC. Connecting the second source to node C, taking into account principles of symmetry and superposition, yields

$$V_{BC} = 1\,\Omega\left(\tfrac{1}{3}\,\text{A} + \tfrac{1}{6}\,\text{A}\right) + 1\,\Omega\left(\tfrac{1}{3}\,\text{A} + \tfrac{1}{6}\,\text{A}\right) = 1\,\text{V}$$

and

$$R_{BC} = \frac{V_{BC}}{I} = 1\,\Omega = r.$$

P130

First we need to find out what sort of camera movement leads to a spot of the largest diameter. Clearly, shifting the camera forward and backward plays virtually no role in blurring the spot (the depth of field is sufficient to keep the image sharp). Up-and-down motion isn't very harmful, because it's equivalent to a vertical movement of the object being photographed. When the distance to the object is 1 m, the angular magnitude of a point of the object is 2 mm/1m $\cong$ 0.002 radian, so the image of a point will have a diameter

$$d_1 = 0.002 \cdot 50\ \text{mm} = 0.1\ \text{mm}.$$

It's far worse if the camera "swings" relative to the line between the camera and the object—that is, when points at the edges of the camera move in opposite directions. The maximum angle is achieved when these points are closest to one another (for most cameras, this means the top and bottom of the camera).

The height of an ordinary camera is about 80 mm, so the angle of rotation is about

$$\alpha \cong \frac{2}{80} = \frac{1}{40}\ \text{rad}.$$

This results in a spot of diameter

$$d_2 = 2\alpha F \cong 2\left(\frac{1}{40}\right)50\ \text{mm} \cong 2.5\ \text{mm}.$$

So, if your hands shake, avoid using long exposure times.

Brainteasers

B126

The answer is yes. The harder brainteaser is this one.

B127

Suppose PQ is the required line (fig. 11). We can form the new rhombus by shifting $APQD$ over so that AD and BC coincide. But how do we determine PQ? For $PA'D'Q$ to be a rhombus, we must have $QP = QD' = DC$. One way to arrange this is to draw a line through M parallel to DC. It will intersect AD at its midpoint X. Then a circle with center M and radius MX will intersect DC at the required point Q.

B128

If we take three 10-kg weights, we have a total of 30 kg, and the conditions of the problem are satisfied. To prove that a greater mass is impossible, take any weight and add other weights to it, one by one, until the total mass M of the chosen weights becomes greater than 10 kg. Then the remaining mass $m \le 10$ kg. On the other hand, if a is the mass of the last weight chosen, then $a \le 10$ kg and $M - a \le 10$ kg, so the total weight is $M + m = (M - a) + a + m \le 30$ kg.

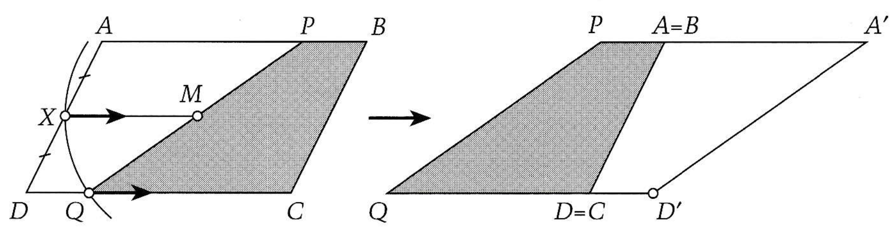


Figure 11

B129

If we have ten different rectangles, then the area of each cannot be too large. If we write the dimensions of all possible integer rectangles in order of increasing area—1×1, 1×2, 1×3, 1×4, 2×2, 1×5, 1×6, 2×3, 1×7, 1×8, 2×4, ...—the sum of the areas of the first ten rectangles in this list equals $1 + 2 + 3 + 4 + 4 + 5 + 6 + 6 + 7 + 8 = 46 > 45$. Therefore, we cannot have any larger rectangles, and at least one of these ten must be repeated.

B130

In figure 12, KN and ML are the midlines in triangles ABD and CBD, so these segments are parallel to BD and half as long. Similarly, KL and

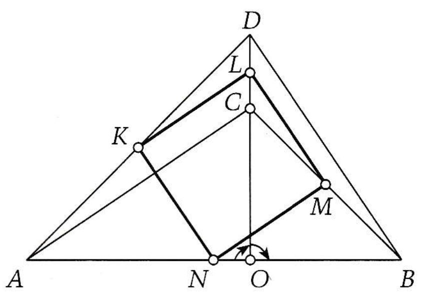


Figure 12

MN are parallel to and half as long as AC. A 90° rotation about point O takes A into D and C into B, so it takes AC into BD. But this means that the segments AC and BD, and therefore NK and NM, are perpendicular and equal in length.

Kaleidoscope

1. Inspecting figure 2 in the article, we see that the left side of the relation in question is equal to

$$\frac{AB'}{B'N} \cdot \frac{NB'}{B'C} \cdot \frac{CB'}{B'A} = 1.$$

2. Let P be the common point of the segments AA_1, BB_1, CC_1. Then the ratios in the relation in the problem can be expressed in terms of the areas of triangles ABP, BCP, CAP—say, $BA_1/A_1C = \text{area}(ABP)/\text{area}(CAP)$, and so on. After these substitutions into the statement of the theorem, all the areas cancel out.

3. Suppose points A_1, B_1, C_1, D_1 in figure 13 are the midpoints of the corresponding sides of the pentagon.

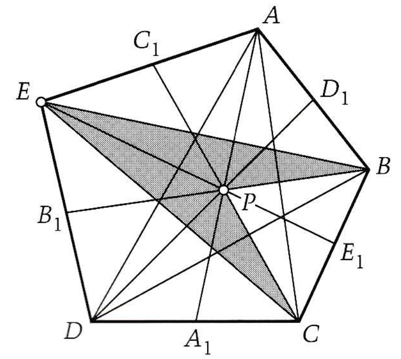


Figure 13

Then we get, successively, the equality of the areas of triangles PBE and PBD, PBD and PAD, PAD and PAC, PAC and PEC. So the areas of PBE and PEC are equal, which means that E_1 is the midpoint of BC, too.

4. The dilation considered in proof 3 takes the altitudes of a triangle ABC into the altitudes of the triangle $A_1B_1C_1$ (where A_1, B_1, C_1 are the midpoints of the sides of ABC). But the altitudes of $A_1B_1C_1$ are the perpendicular bisectors of ABC, and so they meet at O. Therefore, the original altitudes meet at a point, and this point H is taken into O under the dilation by $-1/2$ relative to M, because this dilation transforms the triangle ABC into $A_1B_1C_1$.

5. Consider two dilations that take one base of the trapezoid into the other: one dilation with the center at the intersection of extended sides of the trapezoid (by a positive factor), the other relative to the intersection point of the diagonals (with the negative of the same factor). Both dilations take the midpoint of the first base into the midpoint of the other.

6. Place unit masses at the vertices of the tetrahedron and find their center of mass by, first, uniting any three masses and then adding the fourth mass; and, second, by uniting masses in pairs.

7. If A, B, and C are the vertices of a spherical triangle and O is the center of the sphere, then the planes drawn through O and the medians of the ordinary triangle ABC have a common line OM (where M is the centroid of this triangle). These planes cut the sphere along the spherical medians. So these medians have a common point—namely, the point where the line OM meets the sphere.

Toy Store

1. See figure 14.

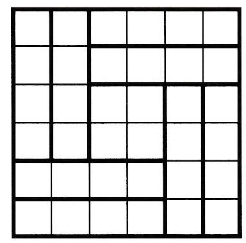


Figure 14

2. The deleted squares are the same color—say, black. Then the number of the remaining black squares is less than that of the white squares. But each domino in a tiling covers one black and one white square, so any area covered by dominoes must contain equal numbers of squares of both colors.

3. All the strips must be positioned horizontally, because they are longer than the height of the board. The greatest number of strips that can be placed on each of the n horizontal rows is $[m/k]$.

Camels and coffee

1. The amount of coffee in the milk is equal to the amount of milk in the coffee, and this doesn't depend on how well they are mixed. As to geometry, we can think of the trapezoid $ANMD$ in figure 2 in the article as the spoonful of coffee poured into the milk, and of $CBKL$ as the spoonful of the mixture poured back. Then $ABCD$ is "the coffee left in the jug" and it's equal in "amount" to $KLMN$—"the milk poured into the cup with coffee."

2. Figure 15a on the next page shows how a trapezoid $ABCD$ can

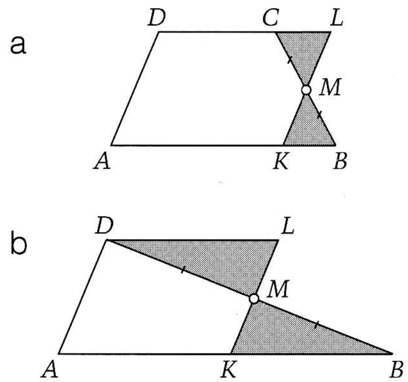


Figure 15

be transformed into a parallelogram $AKLD$ by cutting off triangle BMK and adding the congruent triangle CML (M is the midpoint of BC). Now the formula for the area of $ABCD$ can by derived from the formula for a parallelogram. An argument for a triangle, based on figure 15b, can be constructed similarly.

3. Extend the parallel edges of the prism (fig. 16) and cut them with two planes perpendicular to them at a distance equal to the edge length from

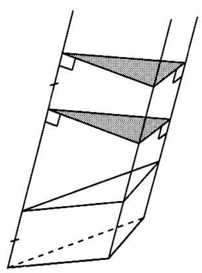


Figure 16

each other. Then repeat the argument for a parallelogram in the article, using volumes instead of areas.

4. The area in question is equal to the product of the perimeter and the edge length.

5. Only for $n = 1$. If n is even, then $n^4 + 4^n$ is also even. For $n = 2k + 1$ the number in question can be factored as follows:

$$n^4 + 2^{2n} = n^4 + 2\cdot n^2\cdot 2^n + 2^{2n} - 2^{n+1}\cdot n^2$$
$$= (n^2 + 2^n)^2 - (2^{k+1}\cdot n)^2$$
$$= (n^2 + 2^n - 2^{k+1}\cdot n)(n^2 + 2^n + 2^{k+1}\cdot n).$$

The second factor is always greater than one, and for the first one this is true whenever $n > 1$, because $n^2 + 2^n \ge 2\sqrt{n^2\cdot 2^n} > n\cdot 2^{k+1}$.

6. (a) $x^4 - a^2x^2 + a^4 = x^4 + 2a^2x^2 + a^4 - 3a^2x^2 = (x^2 + a^2)^2 - 3a^2x^2 = (x^2 - ax\sqrt{3} + a^2)(x^2 + ax\sqrt{3} + a^2)$.

(b) If $D = b^2 - 4c \ge 0$, then $x^4 + bx^2 + c = (x^2 - y_1)(x^2 - y_2)$, where y_1 and y_2 are the roots of the equation $y^2 + by + c = 0$. In the more interesting case of $D < 0$, or $4c > b^2$, we have

$$x^4 + bx^2 + c$$
$$= x^4 + 2\sqrt{c}\,x^2 + c + (b - 2\sqrt{c})x^2$$
$$= (x^2 + \sqrt{c})^2 - (2\sqrt{c} - b)x^2,$$

which is factored as the difference of squares.

7. (a) $\dfrac{-1 \pm \sqrt{4\sqrt{2} - 1}}{\sqrt{2}}$.

Hint: $x^4 + 8x - 7 = x^4 + 2x^2 + 1 - 2(x^2 - 4x + 4) = (x^2 + 1)^2 - 2(x - 1)^2$.

(b) $1 \pm \sqrt{2}$ and $1 \pm i\sqrt{2}$. Hint: the equation can be rewritten as $(x^2 + 1)^2 = 4(x + 1)^2$.

(c) $\left(\dfrac{x^2}{x+1}\right)^2 = 1 - \dfrac{2x^2}{x+1}$.

Hint: subtracting $2x^2/(x + 1)$ from both sides of the given equation and rewriting the left side as the square of the difference $x - x/(x + 1)$, we arrive at

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

This equation is solved by substituting $t = x^2/(x + 1)$.

8. (a) $a^{10} + a^5 + 1 = a^{10} - a + a^5 + a + 1 = a(a^3 - 1)(a^6 + a^3 + 1) + a^5 - a^2 + a^2 + a + 1 = (a^2 + a + 1)[(a^2 - a)(a^6 + a^3 + 1) + a^2(a - 1) + 1] = (a^2 + a + 1)(a^8 - a^7 + a^5 - a^4 + a^3 - a + 1)$.

(b) $a^8 + a + 1 = a^8 - a^5 + a^5 + a + 1 = a^5(a^3 - 1) + (a^2 + a + 1)(a^3 - a^2 + 1) = (a^2 + a + 1)(a^6 - a^5 + a^3 - a^2 + 1)$.

In both cases further factorization with integer coefficients is impossible.

9. $1{,}280{,}000{,}401 = a^7 + a^2 + 1$, where $a = 20$. But $a^7 + a^2 + 1 = a^7 - a + a^2 + a + 1 = (a^2 + a + 1)[a(a - 1)(a^3 + 1) + 1]$.

10. The answer is $3\sqrt{3}/(2\pi)$. It's obtained from the expansion of $\sin x/x$ into an infinite product given in the article after substituting $x = 2\pi/3$.

11. (a) $\dfrac{nx^{n+2} - (n+1)x^{n+1} + x}{(x-1)^2}$.

Multiply S_n by x and consider the difference $xS_n - S_n$.

(b) $\dfrac{\cos\dfrac{nx}{2}\cdot\cos\dfrac{n+1}{2}x}{\sin\dfrac{x}{2}}$.

Multiply the sum by $\sin(x/2)$ and rework it following the solution of problem 7 in the article.

12 (a) $1 \pm \sqrt{3}$ and $1 \pm i\sqrt{3}$. Rewrite the equation completing the square in the left side:

$$(x^2 - 2x)^2 = -x^2 + 2x + 6.$$

Add $2\alpha(x^2 - 2x) + \alpha^2$ to both sides:

$$(x^2 - 2x + \alpha)^2 = (2\alpha - 1)x^2 - (4\alpha - 2)x + 6 + \alpha^2.$$

Ferrari's resolvent for this equation is

$$(2\alpha - 1)^2 = (2\alpha - 1)(6 + \alpha^2),$$

and it has the root $\alpha = 1/2$. With this α our equation takes the form

$$\left(x^2 - 2x + \frac{1}{2}\right)^2 = \frac{25}{4}.$$

(b) $1 \pm \sqrt{3}$ and $-3/2 \pm \sqrt{17}/2$.

13. (a) $(x^2 + x - 1)(x^2 + x + 2)$.

(b) $x^4 + 2x^3 + x^2 - (4x^2 + 4x + 1) = (x^2 + x)^2 - (2x + 1)^2 = (x^2 - x - 1)(x^2 + 3x + 1)$.

Corrections

Vol. 4, No. 6

p. 29, item 18: a keyboarding error caused a "6" to be dropped from the third solution. Our thanks to Paul J. Blatz for pointing this out and supplying the correct answer to 12 digits: −0.766664695962… .

p. 45, col. 1, ¶3, ll. 4–5: *for* July/August 1993 *read* July/August 1992. We are proud to note that, in the meantime, Sergey has won an award from *Print* magazine for his cover to our January/February 1994 issue.

Vol. 5, No. 1

p. 13, col. 1, l. 18: the last coefficient in the display equation should be a_n (not a_0).

Strips on a board

Close packing in two dimensions (but not three)

by Boris Kotlyar

IN THE POPULAR PUZZLE "pentominoes," you have to tile a given shape with twelve different blocks, each consisting of five equal squares. This puzzle gave birth to numerous variations and mathematical problems. One such challenge is to find out whether such a tiling—for various shapes and various sets of blocks—is possible at all. Even for the simplest shapes, this problem turns out to be far from easy. Some time back it was discussed in *Quantum* along with an interesting technique in group theory (see "Getting It Together with Polyominoes" in the November/December 1991 issue). In particular, this article contained a result about tiling a rectangle measuring $m \times n$ with rectangles measuring $p \times q$ (with integer m, n, p, and q). In the even more special case of narrow tiles $1 \times k$ or $k \times 1$, called *strips* below, this theorem says that an $m \times n$ rectangle can be tiled with such strips if and only if at least one of the numbers m and n is divisible by k.

The first proofs of this rather simple fact were given independently by the Dutch mathematician N. G. de Bruijn and an American, D. F. Klarner, in 1969. Of course, the sufficiency in this statement is obvious: if, say, the horizontal dimension m of the "board" is a multiple of k, then

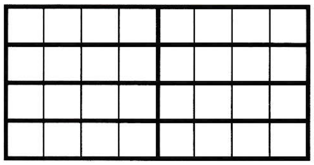


Figure 1

we can simply lay m/k strips on each of n horizontal rows (fig. 1). The whole point of the theorem is the necessity. It can be proved in a number of different ways, and one of the proofs will emerge from our subsequent investigation of this more general question: what is the greatest number $N = N(m, n, k)$ of $1 \times k$ strips that can fit on an $m \times n$ board? We imply, of course, that the strips do not overlap; not only that, we'll confine ourselves to the case where the k unit squares that constitute each strip exactly fit unit squares that constitute the board. So we are not going to consider "irregular" positions of strips like those shown in figure 2. Notice that we didn't have to stipulate this restriction in the de Bruijn–Klarner theorem, because it's clear that in any complete tiling of the board with our

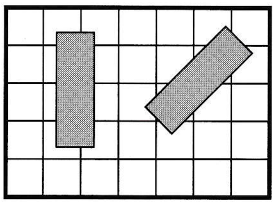


Figure 2

strips they all must be positioned "regularly."

This problems is not only interesting, it can be very useful. Questions like this arise when things are packed ("How can we put as many items as possible in this box?"—a well-known problem of packaging), or in pattern cutting ("How can we cut the greatest number of rectangles of a given size out of a rectangular piece of sheet metal?").

No more, no less

A rough estimate for N is given by the inequality $Nk \le mn$ between the areas of N strips and the board. We see that

$$N \le \left[\frac{mn}{k}\right]$$

(where $[x]$ denotes the greatest integer not exceeding x). Some additional information is provided by the de Bruijn–Klarner theorem. For instance, if we're packing 1×4 strips into a 6×6 square, the inequality above gives $N \le 6 \cdot 6/4 = 9$. But since 6, the side length of the square, is not divisible by 4, it's impossible to fit in all nine strips, so $N \le 8$.

Exercise 1. Show that in this case the estimate gives the exact value of N—that is, $N = 8$.

It's easy to derive a certain lower bound for N. Let r and s be the re-

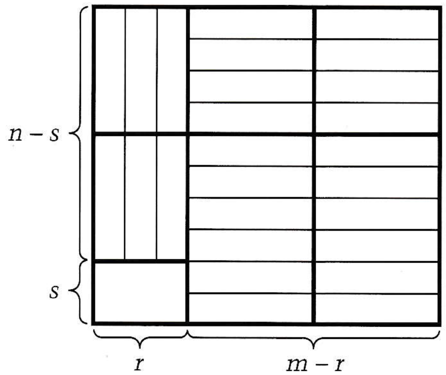


Figure 3

mainders of m and n, respectively, when divided by k:

$$m = km_1 + r,\ 0 \le r < k;$$
$$n = kn_1 + s,\ 0 \le s < k.$$

Draw the vertical line dividing the board into two rectangles measuring $(m - r) \times n$ and $r \times n$, and the horizontal line dividing the second rectangle into $r \times (n - s)$ and $r \times s$ parts. Since $m - r$ and $n - s$ are divisible by k, the rectangles $(m - r) \times n$ and $r \times (n - s)$ can be tiled with our strips $1 \times k$ (fig. 3). The tiled area will then be equal to $(m - r)n + r(n - s) = mn - rs$, so the number of tiles used is

$$N_0 = \frac{mn - rs}{k}.$$

Thus, the maximum number N of strips is no less than N_0.

Of course, the two bounds we obtained usually don't coincide, and so they don't give the exact value of N. However, we'll manage to find this value by means of a special coloring of the board—a technique often used in tiling problems.

Counting by colors

The following classic olympiad problem is a good and simple illustration of the method.

Exercise 2. Use the standard black-and-white coloring of the 8×8 chessboard to show that after cutting out its two diagonally opposite squares we'll be unable to tile it with 2×1 dominoes.

You'll surely solve this problem yourself. (Just in case, we give the answer on page 61.) And now that you've grasped the idea, let's use

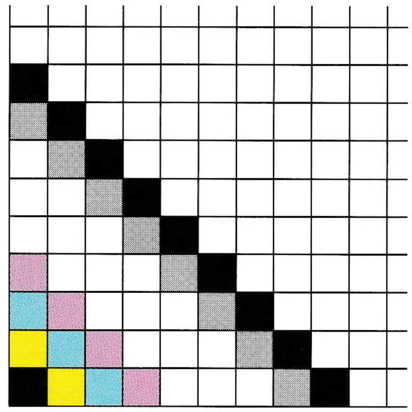


Figure 4

coloring to improve our estimates.

Imagine that the $m \times n$ board we're tiling is placed at the corner of the first quadrant of the coordinate plane divided into unit squares. Paint the squares diagonally in k colors, as shown in figure 4. The remarkable property of this coloring is that whenever a $1 \times k$ strip is placed over it (square on square), it covers exactly one square of each color. Therefore, the number of strips on the board can't be greater than the number of squares of the same color. Let's number the colors from 1 to k, split the board into three rectangles as we did in the previous section (fig. 3), and count the number of squares of the kth color on the board. In the tiled rectangles $(m - r) \times n$ and $r \times (n - s)$ it's equal to the number of strips that cover them—that is, to $N_0 = (mn - rs)/k$. As to the third rectangle $r \times s$ (fig. 5), there are two possibilities. If $r + s - 1 < k$, then the top right square of the board is colored in the color $r + s - 1$, and there are no squares of the kth color in this rectangle. In this case the total number of kth-color squares on the board is N_0, so $N \le N_0$. And in view of the inequality $N \ge N_0$ proved above, we get the exact formula

$$N = N_0 = \frac{mn - rs}{k}.$$

The other case, $r + s - 1 \ge k$ (fig. 6), is more difficult. Here the kth diagonal intersects with the $r \times s$ rectangle left untiled, and the number of squares in the intersection can be

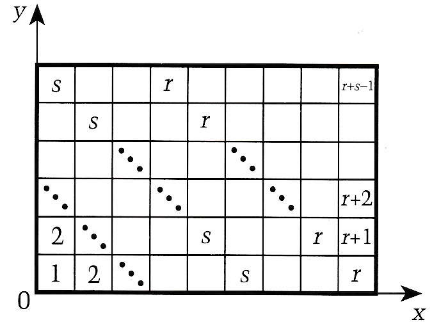


Figure 5

found by subtracting from the number k of squares on the entire kth diagonal the numbers of its squares lying to the right or above the rectangle: $k - (k - r) - (k - s) = r + s - k$. So in this case the number of kth-color squares on the entire board—which is an upper bound for N, as we know—is equal to $N_0 + (r + s - k)$. Below we'll see that, disregarding an obvious exception of $k > m$ or $k > n$, this number is the *exact* value of the greatest possible number N of strips on our board.

But at this point let's turn back to the de Bruijn–Klarner theorem to show how the necessity statement in it follows from our last estimate. Suppose the $m \times n$ board can be tiled with $1 \times k$ strips. Then it contains as many squares of the kth color as of any other—in particular, the sth color. On the other hand, suppose that neither m nor n is a multiple of k—that is, $r > 0$ and $s > 0$. Then the construction in figure 3 shows that the first two of the three rectangles we considered contain N_0 squares of

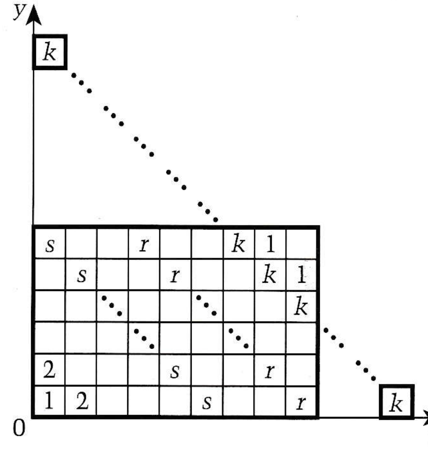


Figure 6

either color s and k, while in the third rectangle (fig. 5) there are s squares of the color s, and zero or $r + s - k < s$ squares of color k. So in this case the complete tiling is impossible.

An optimal tiling

Now I want to explain how to fit $N_1 = N_0 + (r + s - k)$ strips on the $m \times n$ board in the case $r + s - k > 0$, $k < n$, $k < m$. The construction below is due to my student L. Khariton. It's similar to the one we used (fig. 3) except that instead of the $r \times s$ rectangle in the bottom left corner of the board, we single out a $(r + k) \times (s + k)$ rectangle R (fig. 7), which will allow us to fit sufficiently many strips in it.

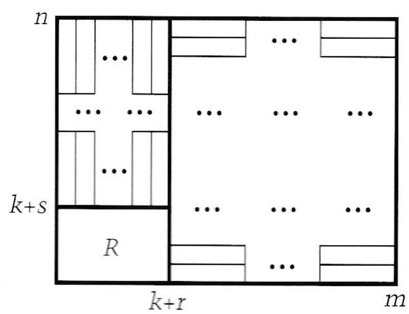


Figure 7

As to the other two rectangles, now measuring $(m - r - k) \times n$ and $(r + k) \times (n - s - k)$, they can be tiled without gaps, because each of them has a side of a length divisible by k. Figure 8 shows how to fill rectangle R.

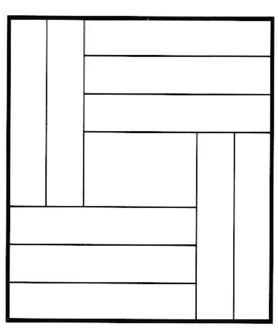


Figure 8

At two opposite corners the strips are layed horizontally (s strips at each corner), and at each of the other two corners we put r vertical strips. So there will be $2(r + s)$ strips laid on rectangle R, which leaves $(r + k)(s + k) - 2(r + s)k = (r - k)(s - k)$ squares uncovered. Thus, the strips in this tiling cover $mn - (r - k)(s - k) = (mn - rs) + (r + s - k)k$ squares of the board. Dividing this number by k, we find out that there are exactly N_1 strips.

Exercise 3. Show that in the case $k > n$ the greatest possible number of strips $N = n[m/k]$.

In the case $k > m$, the expression for N is similar: $N = m[n/k]$. (If $k > n$ and $k > m$, then both expressions give $N = 0$.)

Now let's sum up our results.

THEOREM. *The greatest number N of $1 \times k$ strips that can be ("regularly") fit into an $m \times n$ rectangle, where m and n give the remainders r and s upon division by k, can be found by the formula*

$$N = \begin{cases} N_0 + \max(r + s - k, 0), & \text{if } k \le \min(m, n); \\ \left[\dfrac{n}{k}\right] m, & \text{if } k > m; \\ \left[\dfrac{m}{k}\right] n, & \text{if } k > n, \end{cases}$$

where $N_0 = (mn - rs)/k$, $\max(x, y)$ is the greatest and $\min(x, y)$ the smallest numbers x and y. (It may seem that we skipped the cases $k = m$ and $k = n$ in the reasoning above, but they simply are special cases of the divisibility of m and n by k.)

Bricks in boxes

Our problem seems to be completely solved. But it's not time to celebrate yet. It turns out that our solution for a rectangle can't be extended even to the case of a "box" (rectangular parallelepiped) in three-dimensional space. Things are even worse for n-dimensional space with $n > 3$. I'll briefly describe some results without any proofs.

Instead of a board measuring $m \times n$, let's consider a box measuring $m \times n \times p$, and instead of strips take "bricks" measuring $1 \times 1 \times k$. Let's stack the bricks in the box in the "regular" way—that is, so that the unit cubes constituting each brick exactly fit on the unit cubes constituting the box. Then, just as on the plane, we can always fit $N_0 = (mnp - rst)/k$ bricks in the box, where r, s, and t are the remainders of m, n, and p when divided by k. The de Bruijn–Klarner theorem is true here as well (it was proved for three-dimensional and even n-dimensional space by de Bruijn). But an exact formula for the greatest possible number N of bricks in a general case has not been found yet. I only know the estimate

$$N_0 \le N \le mnp - k^3 \cdot \left\| \frac{m}{k} \right\| \cdot \left\| \frac{n}{k} \right\| \cdot \left\| \frac{p}{k} \right\|,$$

where $\|v\|$ denotes the distance from a real number v to the nearest integer ($\|v\| = \min(\{v\}, 1 - \{v\})$, where $\{v\} = v - [v]$ is the fractional part of v). If all the numbers r, s, and t are no greater than $k/2$, then this inequality turns into the exact equality $N = N_0$.

However, in general the left and right sides of this estimate do not coincide, so it doesn't give the exact value of N. Perhaps our readers will be more lucky and succeed in calculating N.

ANSWERS, HINTS & SOLUTIONS
ON PAGE 61

Do something different this summer...

Live and learn in

RUSSIA

A unique opportunity awaits high school students and teachers: the Russian-American Mathematics and Science Summer Institute will be held from July 2 to July 22, 1995, at Moscow State University. This exciting program will feature

- Two weeks of advanced classes in mathematics, physics, and biology taught in English by prominent Russian professors
- Trips to the scientific laboratories of the Russian Academy of Sciences
- Cultural and recreational activities

Following the two-week academic program in Moscow will be a one-week cultural program in St. Petersburg, the capital of Russia under the tsars.

Scholarships are available

For more information, please fill out the coupon below and send it to Dr. Edward Lozansky, President, American University in Moscow, 1800 Connecticut Ave. NW, Washington DC 20009, Phone: 202 986-6010, Fax: 202 667-4244, E-mail: lozansky@aol.com

Please send me _____ brochures to distribute among interested high school teachers and students.

Last name ______________________ First name ______________________

Address ______________________ City ______________________

State __________ Zip code __________ Phone number ______________________

I am a teacher ______ a student _______.